TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC

BÀI GIẢNG

XÁC SUẤT THỐNG KÊ

NGUYỄN THỊ THU THỦY
BỘ MÔN TOÁN ỨNG DỤNG

HÀ NỘI – 9/2020

MỤC LỤC

Lời nói đầu

Lý thuyết xác suất và thống kê toán học là một ngành khoa học đang giữ vị trí quan trọng trong các lĩnh vực ứng dụng rộng rãi và phong phú của đời sống con người. Cùng với sự phát triển mạnh mẽ của khoa học và công nghệ, nhu cầu hiểu biết và sử dụng các công cụ ngẫu nhiên trong phân tích và xử lý thông tin ngày càng trở nên đặc biệt cần thiết. Các kiến thức và phương pháp của xác suất và thống kê đã hỗ trợ hữu hiệu các nhà nghiên cứu trong nhiều lĩnh vực khoa học khác nhau như vật lý, hóa học, sinh học, nông học, kinh tế học, xã hội học, ngôn ngữ học... Do đó "Xác suất thống kê" là học phần rất cần thiết cho sinh viên bậc đại học.

Bài giảng học phần "Xác suất thống kê", mã học phần MI2020 được biên soạn theo Đề cương chi tiết với khối lượng 30 tiết lý thuyết, 30 tiết bài tập dành cho sinh viên hệ đại học chính quy (không phải chuyên ngành Toán Tin) của Trường Đại học Bách khoa Hà Nội.

Mục tiêu học phần: Cung cấp cho sinh viên những kiến thức cơ bản về xác suất là các khái niệm và quy tắc suy diễn xác suất cũng như về biến ngẫu nhiên và các phân phối xác suất thông dụng (một và hai chiều); các khái niệm cơ bản của thống kê toán học nhằm giúp sinh viên biết cách xử lý các bài toán thống kê về ước lượng, kiểm định giả thuyết.... Trên cơ sở đó sinh viên có được một phương pháp tiếp cận với mô hình thực tế và có kiến thức cần thiết để đưa ra lời giải đúng cho các bài toán đó.

Nội dung vắn tắt học phần: Sự kiện ngẫu nhiên và phép tính xác suất, đại lượng ngẫu nhiên, phân phối xác suất, véc tơ ngẫu nhiên, lý thuyết ước lượng thống kê, lý thuyết quyết định thống kê.

Chương 1

Sự kiện ngẫu nhiên và phép tính xác suất

Mục tiêu

1. Cung cấp cho sinh viên những khái niệm về phép thử, sự kiện, xác suất; mối quan hệ giữa các sự kiện; các công cụ tính toán cơ bản của lý thuyết xác suất (định lý, công thức).

2. Với các kiến thức nền tảng đó, sinh viên biết tính xác suất của các sự kiện; biết thực hiện các bài tập ứng dụng của xác suất trong các lĩnh vực kỹ thuật, kinh tế, xã hội, quản lý ra quyết định...

Nội dung

1. Sự kiện, quan hệ giữa các sự kiện

2. Nhắc lại giải tích tổ hợp, quy tắc nhân, chỉnh hợp lặp

3. Khái niệm và các định nghĩa xác suất (cổ điển, hình học, thống kê)

4. Các định lý và công thức xác suất (xác suất điều kiện; công thức nhân xác suất; công thức cộng xác suất; công thức Béc-nu-li; công thức xác suất đầy đủ, công thức Bay-ét)

Thời lượng: 8 tiết

BÀI 1 (2 tiết)

Các hiện tượng trong tự nhiên hay xã hội xảy ra một cách ngẫu nhiên (không biết trước kết quả) hoặc tất định (biết trước kết quả sẽ xảy ra). Chẳng hạn một vật nặng được thả từ trên cao chắc chắn sẽ rơi xuống đất, trong điều kiện bình thường nước sôi ở 100° C... Đó là những hiện tượng diễn ra có tính quy luật, tất nhiên. Trái lại, khi tung đồng xu ta không biết sẽ xuất hiện mặt sấp hay mặt ngửa; ta không thể biết trước có bao nhiêu cuộc gọi đến tổng đài; có bao nhiêu khách hàng đến điểm phục vụ trong khoảng thời gian nào đó; ta không thể xác định trước chỉ số chứng khoán trên thị trường chứng khoán... Đó là những hiện tượng ngẫu nhiên. Tuy nhiên, nếu tiến hành quan sát nhiều lần một hiện tượng ngẫu nhiên trong những hoàn cảnh như nhau, thì trong nhiều trường hợp ta có thể rút ra những kết luận có tính quy luật về những hiện tượng này. Lý thuyết xác suất nghiên cứu các quy luật của các hiện tượng ngẫu nhiên. Việc nắm bắt các quy luật này sẽ cho phép dự báo các hiện tượng ngẫu nhiên đó sẽ xảy ra như thế nào. Chính vì vậy các phương pháp của lý thuyết xác suất được ứng dụng rộng rãi trong việc giải quyết các bài toán thuộc nhiều lĩnh vực khác nhau của khoa học tự nhiên, kỹ thuật và kinh tế–xã hội.

1.1 Sự kiện. Quan hệ giữa các sự kiện

1.1.1 Phép thử. Sự kiện

Định nghĩa 1.1 (Phép thử. Sự kiện)**. (a)** Việc thực hiện một nhóm các điều kiện cơ bản để quan sát một hiện tượng nào đó được gọi là một phép thử (experiment).

(b) Hiện tượng, kết quả xét trong phép thử gọi là sự kiện hay biến cố (event).

(c) Sự kiện sơ cấp hay kết cục của phép thử là một kết quả mà ta không chia nhỏ hơn được, ký hiệu là ω.

(d) Sự kiện phức hợp là sự kiện có thể phân tích thành các sự kiện nhỏ hơn.

(e) Tập hợp tất cả các kết cục của một phép thử tạo thành không gian các sự kiện sơ cấp, ký hiệu là

$$\Omega = \left\{\omega_i,\ i \in I\right\}, \quad I \text{ là tập chỉ số.}$$

Ví dụ 1.1. (a) Gieo một con xúc xắc (cân đối, đồng chất, trên mặt phẳng cứng) là một phép thử. Xúc xắc xuất hiện mặt 1, 2, 3, 4, 5, 6 chấm là các sự kiện.

(b) Gieo một đồng xu (cân đối, đồng chất, trên mặt phẳng cứng) là một phép thử. Đồng xu xuất hiện mặt sấp, mặt ngửa là các sự kiện.

Ví dụ 1.2. Gieo một con xúc xắc, khi đó

(a) Sự kiện A_i "xuất hiện mặt i chấm", $i = 1, \ldots, 6$, là sự kiện sơ cấp.

(b) Sự kiện A "xuất hiện mặt chấm chẵn" là sự kiện phức hợp vì có thể phân tích nó thành các sự kiện "xuất hiện mặt 2, 4, 6 chấm".

Ví dụ 1.3. (a) Phép thử gieo một đồng xu (cân đối, đồng chất, trên mặt phẳng cứng) có không gian các sự kiện sơ cấp là $\Omega = \{S, N\}$, ở đây S là sự kiện "xuất hiện mặt sấp", N là sự kiện "xuất hiện mặt ngửa".

(b) Phép thử gieo đồng thời hai đồng xu (cân đối, đồng chất, trên mặt phẳng cứng) có không gian các sự kiện sơ cấp là $\Omega = \{SS, SN, NS, NN\}$.

Chú ý 1.1. **(a)** Chú ý rằng bản chất của các sự kiện sơ cấp không có vai trò đặc biệt gì trong lý thuyết xác suất. Chẳng hạn có thể mã hóa các kết quả và xem không gian các sự kiện sơ cấp của phép thử tung đồng xu là $\Omega = \{0, 1\}$, trong đó 0 là sự kiện sơ cấp chỉ "mặt sấp xuất hiện" và 1 để chỉ "mặt ngửa xuất hiện".

(b) Mỗi kết cục ω của phép thử $\mathcal{C}$ được gọi là kết cục thuận lợi cho sự kiện A nếu A xảy ra khi kết cục của phép thử $\mathcal{C}$ là ω.

Ví dụ 1.4. Nếu gọi sự kiện A "xuất hiện mặt chấm chẵn" trong phép thử gieo con xúc xắc thì A có các kết cục thuận lợi là 2, 4 và 6.

1.1.2 Phân loại sự kiện

Có 3 loại sự kiện.

(a) **Sự kiện chắc chắn** là sự kiện nhất định sẽ xảy ra khi thực hiện một phép thử. Ký hiệu là U hoặc S.

(b) **Sự kiện không thể có** là sự kiện nhất định không xảy ra khi thực hiện một phép thử. Ký hiệu là V hoặc $\emptyset$.

(c) **Sự kiện ngẫu nhiên** là sự kiện có thể xảy ra, cũng có thể không xảy ra khi thực hiện một phép thử. Ký hiệu là A, B, C, A_1, A_2 ...

Ví dụ 1.5. Gieo một con xúc xắc, khi đó

(a) Sự kiện "xuất hiện mặt có số chấm ≤ 6 và ≥ 1" là sự kiện chắc chắn S.

(b) Sự kiện "xuất hiện mặt 7 chấm" là sự kiện không thể $\emptyset$.

(c) Sự kiện "xuất hiện mặt chấm chẵn" là sự kiện ngẫu nhiên A.

1.1.3 Quan hệ giữa các sự kiện

Một cách tương ứng với các phép toán của tập hợp, trong lý thuyết xác suất người ta xét các quan hệ sau đây cho các sự kiện trong cùng một phép thử.

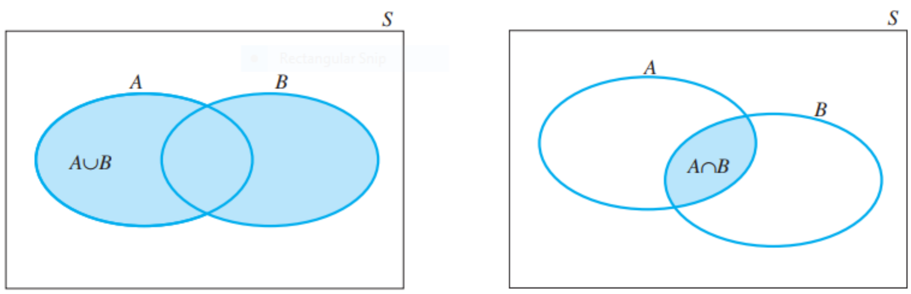
(a) **Quan hệ kéo theo**: Sự kiện A kéo theo sự kiện B, ký hiệu $A \subset B$, nếu khi A xảy ra thì B xảy ra. Nếu $A \subset B$ và $B \subset A$ thì ta nói hai sự kiện A và B trùng nhau, viết là $A = B$.

(b) **Tổng các sự kiện**: Sự kiện A được gọi là tổng của các sự kiện A_1, A_2,..., A_n nếu A xảy ra khi và chỉ khi ít nhất một trong các sự kiện A_i xảy ra, $i = 1, 2, \ldots, n$. Viết là:

$$\boxed{A = A_1 + A_2 + \cdots + A_n}$$

hoặc

$$\boxed{A = A_1 \cup A_2 \cup \cdots \cup A_n}$$



Hình 1.1: Sơ đồ Venn của $A \cup B$ và $A \cap B$

(c) **Tích các sự kiện**: Sự kiện B được gọi là tích của các sự kiện A_1, A_2,..., A_n nếu B xảy ra khi và chỉ khi tất cả các sự kiện A_i xảy ra, $i = 1, 2, \ldots, n$. Viết là:

$$\boxed{B = A_1 A_2 \ldots A_n}$$

hoặc

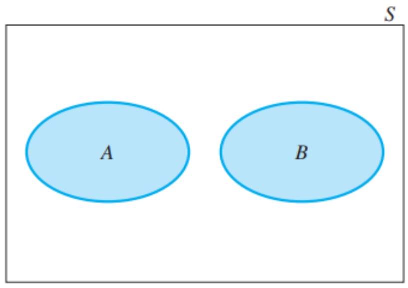
$$\boxed{B = A_1 \cap A_2 \cap \cdots \cap A_n}$$

(d) **Sự kiện xung khắc**: Hai sự kiện A và B được gọi xung khắc với nhau nếu chúng không đồng thời xảy ra trong cùng một phép thử.

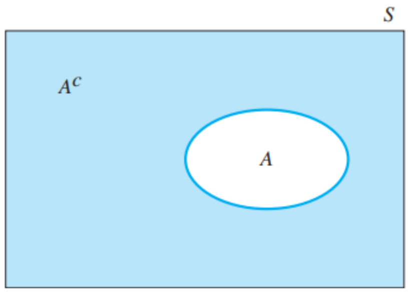
Nếu A và B xung khắc thì $A \cap B = \emptyset$.

(e) **Sự kiện đối lập**: Sự kiện không xảy ra sự kiện A được gọi là sự kiện đối lập của A, ký hiệu là $\overline{A}$ hoặc A^c.

Như vậy A và $\overline{A}$ thỏa mãn tính chất: $A \cup \overline{A} = S$ và $A \cap \overline{A} = \emptyset$.



Hình 1.2: Hai sự kiện xung khắc



Hình 1.3: Sự kiện đối lập

(f) Hiệu hai sự kiện: Hiệu của 2 sự kiện A và B, ký hiệu là $A - B$, là sự kiện xảy ra khi và chỉ khi A xảy ra nhưng B không xảy ra.

Trường hợp hay sử dụng sự kiện hiệu: $\overline{A} = S - A$, $A = S - \overline{A}$.

Trường hợp tổng quát, ta biến đổi thành sự kiện tích như sau: $A - B = A \cap \overline{B}$.

(g) Hệ (nhóm) đầy đủ các sự kiện: Hệ n sự kiện A_1, A_2,..., A_n được gọi là hệ đầy đủ các sự kiện nếu nhất định phải xảy ra một và chỉ một trong các sự kiện ấy sau phép thử.

Như vậy hệ $\{A_1, A_2, \ldots, A_n\}$ là hệ đầy đủ nếu

$$\begin{cases} A_i \cap A_j = \emptyset, & i \neq j, \\ A_1 \cup A_2 \cup \cdots \cup A_n = S. \end{cases}$$

Nhận xét 1.1. Các sự kiện trong cùng một phép thử với phép toán tổng, tích và lấy sự kiện đối tạo thành đại số Boole, do đó các phép toán này có các tính chất như các phép toán hợp, giao, lấy phần bù đối với các tập con của không gian các sự kiện sơ cấp. Chẳng hạn

1. $A \cap \emptyset = \emptyset$.
2. $A \cup \emptyset = A$.
3. $A \cap \overline{A} = \emptyset$.
4. $A \cup \overline{A} = S$.
5. $\overline{S} = \emptyset$.
6. $\overline{\emptyset} = S$.
7. $\overline{(\overline{A})} = A$.
8. $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.
9. $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$.
10. $\overline{A \cup B} = \overline{A} \cap \overline{B}$; $\overline{A \cap B} = \overline{A} \cup \overline{B}$.
11. $A = A \cap (B \cup \overline{B}) = (A \cap B) \cup (A \cap \overline{B})$.

Chú ý 1.2. **(a)** Mọi sự kiện ngẫu nhiên đều có thể biểu diễn dưới dạng tổng của một số sự kiện sơ cấp nào đó. Sự kiện chắc chắn S là tổng của mọi sự kiện sơ cấp có thể. Do đó S còn được gọi là không gian các sự kiện sơ cấp Ω.

(b) Đối với một sự kiện A thì ta có hệ đầy đủ $\left\{A, \overline{A}\right\}$. Đối với hai sự kiện A và B, một hệ đầy đủ là $\left\{A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}\right\}$.

Tính chất 1.1. **(a)** $A \cup B = B \cup A$, $A \cap B = B \cap A$ (giao hoán).

(b) $A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$, $A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$ (kết hợp).

(c) $A \cap (B \cup C) = A \cap B \cup A \cap C$ (phân phối của phép cộng và phép nhân).

Đặc biệt $A \cup A = A$; $A \cap A = A$; $A \cup S = S$; $A \cap S = A$; $A \cup \emptyset = A$; $A \cap \emptyset = \emptyset$.

Tà đây ta sẽ sử dụng dấu $" + "$ thay cho $" \cup "$ và dấu $"."$ thay cho $" \cap "$.

Ví dụ 1.6. (a) Một mạng điện gồm ba bóng đèn mắc nối tiếp. Gọi A_i là sự kiện "bóng đèn thứ i bị cháy", $i = 1, 2, 3$. Gọi A là sự kiện "mạng mất điện". Ta thấy rằng mạng bị mất điện khi ít nhất một trong ba bóng bị cháy. Vậy $A = A_1 + A_2 + A_3$.

(b) Một mạng điện gồm ba bóng đèn mắc song song. Gọi B_i là sự kiện "bóng đèn thứ i bị cháy", $i = 1, 2, 3$. Gọi B là sự kiện "mạng mất điện". Ta thấy rằng mạng bị mất điện khi cả ba bóng bị cháy. Vậy $B = B_1 B_2 B_3$.

(c) Một nhà máy có ba phân xưởng sản xuất ra cùng một loại sản phẩm. Giả sử rằng mỗi sản phẩm của nhà máy chỉ do một trong ba phân xưởng này sản xuất. Chọn ngẫu nhiên một sản phẩm, gọi C_i là sự kiện "sản phẩm được chọn do phân xưởng thứ i sản xuất", $i = 1, 2, 3$. Khi đó hệ ba sự kiện $\{C_1, C_2, C_3\}$ là hệ đầy đủ.

Ví dụ 1.7. Ba xạ thủ A, B, C mỗi người bắn một viên đạn vào mục tiêu. Gọi A_1, A_2, A_3 lần lượt là các sự kiện "A, B, C bắn trúng mục tiêu".

(a) Hãy mô tả các sự kiện $A_1 A_2 A_3$, $\overline{A}_1 \overline{A}_2 \overline{A}_3$, $A_1 + A_2 + A_3$.

(b) Biểu diễn các sự kiện sau theo A_1, A_2, A_3:

A: Có ít nhất hai xạ thủ bắn trúng;

B: Có nhiều nhất một xạ thủ bắn trúng;

C: Chỉ có xạ thủ A bắn trúng;

D: Chỉ có một xạ thủ bắn trúng.

(c) Các sự kiện A_1, A_2, A_3 có xung khắc không?

Lời giải Ví dụ 1.7

(a) $A_1A_2A_3$: "cả ba xạ thủ đều bắn trúng";

$\overline{A}_1\overline{A}_2\overline{A}_3$: "cả ba xạ thủ đều bắn trượt";

$A_1 + A_2 + A_3$: "có ít nhất một xạ thủ bắn trúng".

(b) $A = A_1A_2 + A_1A_3 + A_2A_3$;

$B = \overline{A}_1\overline{A}_2 + \overline{A}_1\overline{A}_3 + \overline{A}_2\overline{A}_3$;

$C = A_1\overline{A}_2\overline{A}_3$;

$D = A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3$.

(c) Các sự kiện A_1, A_2, A_3 không xung khắc vì có thể cả ba xạ thủ đều bắn trúng mục tiêu.

1.2 Giải tích kết hợp

1.2.1 Quy tắc cộng. Quy tắc nhân

1.2.1a Quy tắc cộng

Định nghĩa 1.2 (Quy tắc cộng)**.** Nếu một công việc được chia ra thành k trường hợp để thực hiện, trường hợp một có n_1 cách thực hiện xong công việc, trường hợp hai có n_2 cách thực hiện xong công việc,..., trường hợp k có n_k cách thực hiện xong công việc và không có một cách thực hiện nào ở trường hợp này lại trùng với một cách thực hiện ở trường hợp khác. Khi đó ta có $n = n_1 + n_2 + \cdots + n_k$ cách thực hiện công việc.

1.2.1b Quy tắc nhân

Định nghĩa 1.3 (Quy tắc nhân)**.** Giả sử một công việc nào đó được chia thành k giai đoạn. Có n_1 cách thực hiện giai đoạn thứ nhất, n_2 cách thực hiện giai đoạn thứ hai,..., n_k cách thực hiện giai đoạn thứ k. Khi đó ta có $n = n_1n_2\ldots n_k$ cách thực hiện công việc.

Ví dụ 1.8. Giả sử để đi từ A đến C có thể đi qua B, trong đó có 2 đường khác nhau đi trực tiếp từ A đến C, có 3 đường khác nhau để đi từ A đến B và có 2 đường khác nhau để đi từ B đến C. Hỏi có bao nhiêu cách đi từ A đến C?

Lời giải Ví dụ 1.8 Đi từ A đến C có 2 lựa chọn: Đi trực tiếp từ A đến C có $n_1 = 2$ cách; đi gián tiếp từ A đến C thông qua B có $n_2 = 3 \times 2 = 6$ (cách).

Tổng số cách đi từ A đến C là $n = n_1 + n_2 = 2 + 6 = 8$ (cách).

1.2.2 Chỉnh hợp

Định nghĩa 1.4 (Chỉnh hợp). Chỉnh hợp chập k của n phần tử là một nhóm có thứ tự gồm k phần tử khác nhau lấy từ n phần tử đã cho ($k \leq n$).

Ký hiệu và công thức tính:

$$\boxed{A_n^k = \frac{n!}{(n-k)!} = n(n-1)\dots(n-k+1)} \tag{1.1}$$

Ví dụ 1.9. Từ 5 chữ số $1, 2, 3, 4, 5$ lập được bao nhiêu số có 3 chữ số khác nhau?

Lời giải Ví dụ 1.9 Số các số được lập là $A_5^3 = 5 \times 4 \times 3 = 60$ (số).

1.2.3 Chỉnh hợp lặp

Định nghĩa 1.5 (Chỉnh hợp lặp). Chỉnh hợp lặp chập k của n phần tử là một nhóm có thứ tự gồm k phần tử có thể giống nhau lấy từ n phần tử đã cho.

Ký hiệu và công thức tính:

$$\boxed{\overline{A}_n^k = n^k} \tag{1.2}$$

Ví dụ 1.10. Từ 5 chữ số $1, 2, 3, 4, 5$ lập được bao nhiêu số có 3 chữ số?

Lời giải Ví dụ 1.10 Chọn 3 chữ số từ 5 chữ số có thứ tự và có thể lặp lại. Số các số được lập là $\overline{A}_5^3 = 5^3 = 125$ (số).

1.2.4 Hoán vị

Định nghĩa 1.6 (Hoán vị). Hoán vị của n phần tử là một nhóm có thứ tự gồm đủ mặt n phần tử đã cho. Nói cách khác, hoán vị là một chỉnh hợp chập n của n phần tử.

Ký hiệu và công thức tính:

$$\boxed{P_n = A_n^n = n!} \tag{1.3}$$

Ví dụ 1.11. Có 6 người khách cần xếp vào 6 ghế trên một bàn tròn 6 chỗ.

(a) Nếu có quan tâm đến khung cảnh xung quanh thì có bao nhiêu cách sắp xếp?

(b) Nếu chỉ quan tâm đến người ngồi xung quanh là ai thì sẽ có bao nhiêu cách?

Lời giải Ví dụ 1.11 (a) $P_6 = 6! = 720$ (cách); (b) $P_5 = 5! = 120$ (cách).

1.2.5 Tổ hợp

Định nghĩa 1.7 (Tổ hợp)**.** Tổ hợp chập k của n phần tử là một nhóm không phân biệt thứ tự gồm k phần tử khác nhau lấy từ n phần tử đã cho ($k \le n$).

Ký hiệu và công thức tính:

$$\boxed{C_n^k = \frac{n!}{k!(n-k)!}} \tag{1.4}$$

Ví dụ 1.12. Mỗi đề thi gồm 3 câu hỏi lấy trong 25 câu hỏi cho trước. Hỏi có thể lập được bao nhiêu đề thi có nội dung khác nhau?

Lời giải Ví dụ 1.12 Số đề thi có thể lập nên là $C_{25}^3 = \dfrac{25 \times 24 \times 23}{3!} = 2300$ (đề).

Chú ý 1.3. (a) Qui ước $0! = 1$.

(b) $C_n^k = C_n^{n-k}$.

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

(d) Khai triển nhị thức Niu–tơn ($a, b \in \mathbb{R}, n \in \mathbb{N}^*$)

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k = C_n^0 a^n + C_n^1 a^{n-1} b + \cdots + C_n^{n-1} a b^{n-1} + C_n^n b^n.$$

BÀI 2 (2 tiết)

1.3 Khái niệm và các định nghĩa xác suất

1.3.1 Khái niệm xác suất

Mọi sự kiện ngẫu nhiên đều giống nhau ở chỗ chúng không chắc chắn, nhưng khả năng xảy ra của từng sự kiện lại có thể khác nhau. Để đặc trưng cho khả năng xảy ra (xuất hiện) của các sự kiện người ta dùng các con số, sự kiện nào có khả năng xảy ra nhiều hơn được đặc trưng bởi số lớn hơn. Con số đặc trưng cho khả năng xuất hiện của một sự kiện gọi là xác suất của sự kiện đó.

Định nghĩa 1.8 (Xác suất–Probability)**.** Xác suất của một sự kiện A là một số nằm giữa 0 và 1, số này đo lường khả năng xuất hiện của sự kiện đó khi phép thử được thực hiện.

Ký hiệu là $P(A)$.

1.3.2 Định nghĩa cổ điển

Định nghĩa 1.9 (Định nghĩa cổ điển)**.** Giả sử trong một phép thử có n kết cục đồng khả năng có thể xảy ra, trong đó có m kết cục thuận lợi cho sự kiện A. Khi đó,

$$\boxed{P(A) = \frac{m}{n} = \frac{\text{số kết cục thuận lợi cho } A}{\text{tổng số kết cục có thể}}} \tag{1.5}$$

Từ định nghĩa này ta suy ra các tính chất sau đây của xác suất.

Tính chất 1.2. **(a)** $0 \le P(A) \le 1$, A là sự kiện bất kỳ.

(b) $P(S) = 1$.

(c) $P(\emptyset) = 0$.

(d) Nếu $A \subset B$ thì $P(A) \le P(B)$.

Ví dụ 1.13. Một người khi gọi điện thoại quên mất 2 số cuối cùng của số điện thoại cần gọi mà chỉ nhớ được rằng chúng khác nhau. Tìm xác suất để người đó chọn ngẫu nhiên một số để gọi thì được đúng số cần gọi.

Lời giải Ví dụ 1.13 Gọi A là sự kiện "chọn ngẫu nhiên một số để gọi thì được đúng số cần gọi".

Số kết cục đồng khả năng là $n = A_{10}^2$.

Số kết cục thuận lợi cho A là $m = 1$.

Vậy $P(A) = \dfrac{m}{n} = \dfrac{1}{90}$.

Ví dụ 1.14. Từ bộ bài tú-lơ-khơ 52 cây đã trộn kỹ rút ngẫu nhiên ra 2 cây. Tính xác suất xảy ra các sự kiện sau:

(a) Hai cây rút ra đều là Át.

(b) Hai cây rút ra có 1 cây Át, 1 cây K.

Lời giải Ví dụ 1.14 Số kết cục đồng khả năng có thể có là $n = C_{52}^2 = 1326$. Gọi A là sự kiện "hai cây rút ra đều là Át"; B là sự kiện "hai cây rút ra có 1 cây Át, 1 cây K".

(a) Số kết cục thuận lợi cho sự kiện A là $m_A = C_4^2 = 6$. Suy ra $P(A) = \dfrac{m_A}{n} = \dfrac{6}{1326} = \dfrac{1}{221}$.

(b) Số kết cục thuận lợi cho sự kiện B là $m_B = C_4^1 \times C_4^1 = 16$, suy ra $P(B) = \dfrac{m_B}{n} = \dfrac{8}{663}$.

Ví dụ 1.15. Một đoàn tàu có 4 toa được đánh số I, II, III, IV đỗ ở sân ga. Có 6 hành khách từ sân ga lên tàu. Mỗi người độc lập với nhau chọn ngẫu nhiên một toa. Tính xác suất để:

(a) toa I có 3 người, toa II có 2 người và toa III có 1 người;

(b) một toa có 3 người, một toa 2 người, một toa có 1 người;

(c) mỗi toa có ít nhất 1 người.

Lời giải Ví dụ 1.15 Số trường hợp đồng khả năng có thể có là $n = 4^6 = 4096$.

(a) Số trường hợp thuận lợi cho sự kiện A "toa I có 3 người, toa II có 2 người và toa III có 1 người" là $C_6^3 \times C_3^2 \times C_1^1 = 60$, suy ra $P(A) = \dfrac{60}{4096} \simeq 0,0146$.

(b) Số trường hợp thuận lợi cho sự kiện B "một toa có 3 người, một toa 2 người, một toa có 1 người" là $C_6^3 \times 4 \times C_3^2 \times 3 \times C_1^1 \times 2 = 1440$, suy ra $P(B) = \dfrac{1440}{4096} \simeq 0,3516$.

(b) Gọi C là sự kiện "mỗi toa có ít nhất 1 người". Số trường hợp thuận lợi cho sự kiện C là $C_6^3 \times 4 \times 3! + C_6^2 \times C_4^2 \times C_4^2 \times 2! = 480 + 1080 = 1560$. Do đó, $P(C) = \dfrac{1560}{4096} \simeq 0,3809$.

Ví dụ 1.16. Ba nữ nhân viên phục vụ A, B và C thay nhau rửa đĩa chén và giả sử ba người này đều "khéo léo" như nhau. Trong một tháng có 4 chén bị vỡ. Tìm xác suất để:

(a) chị A đánh vỡ 3 chén và chị B đánh vỡ 1 chén;

(b) một trong ba người đánh vỡ 3 chén;

(c) một trong ba người đánh vỡ cả 4 chén.

Lời giải Ví dụ 1.16 Số kết cục đồng khả năng có thể có là $n = 3^4$.

(a) Số kết cục thuận lợi cho sự kiện D "chị A đánh vỡ 3 chén và chị B đánh vỡ 1 chén" là $m_D = C_4^3 \times 1 = 4$, suy ra $P(D) = \dfrac{4}{81} \simeq 0,0494$.

(b) Số kết cục thuận lợi cho sự kiện E "một trong 3 người đánh vỡ 3 chén" là $m_E = C_3^1 \times C_4^3 \times 2 = 24$, nên $P(E) = \dfrac{24}{81} \simeq 0,2963$.

(c) Số kết cục thuận lợi cho sự kiện F "một trong 3 người đánh vỡ 4 chén" là $m_F = C_3^1 \times C_4^4 = 3$. Vậy $P(F) = \dfrac{3}{81} \simeq 0,037$.

Nhận xét 1.2. Định nghĩa cổ điển về xác suất có ưu điểm là dễ vận dụng tuy nhiên định nghĩa này chỉ áp dụng được với các phép thử có hữu hạn kết cục đồng khả năng xảy ra. Trong trường hợp có vô hạn kết cục đồng khả năng ta sử dụng định nghĩa xác suất theo quan điểm hình học.

1.3.3 Định nghĩa hình học

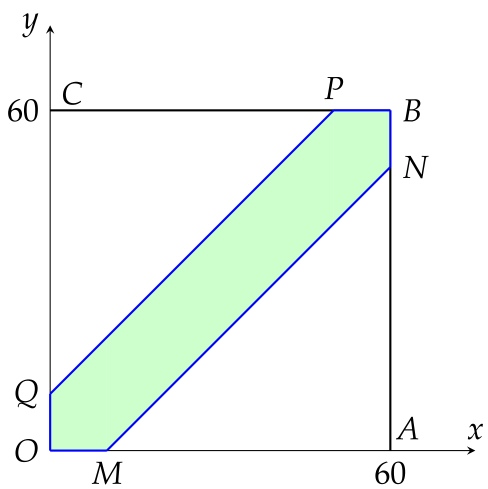
Định nghĩa 1.10 (Định nghĩa hình học)**.** Giả sử tập hợp vô hạn các kết cục đồng khả năng của một phép thử có thể biểu thị bởi một miền hình học G (đo được, hữu hạn, khác 0), còn các kết cục thuận lợi cho A bởi miền con H của G. Khi đó

$$\boxed{P(A) = \frac{\text{độ đo } H}{\text{độ đo } G}} \tag{1.6}$$

Chú ý 1.4. Tùy theo G là đoạn thẳng, miền phẳng hay khối không gian mà độ đo được hiểu là độ dài, diện tích hay thể tích.

Ví dụ 1.17. Hai người bạn hẹn gặp nhau ở một địa điểm trong khoảng thời gian từ 7h00 đến 8h00. Mỗi người có thể đến điểm hẹn một cách ngẫu nhiên tại một thời điểm trong khoảng thời gian nói trên và họ quy ước rằng ai đến trước thì chỉ đợi người kia trong vòng 10 phút. Tính xác suất để hai người gặp nhau.

Lời giải Ví dụ 1.17 Gọi x, y lần lượt là thời điểm đến điểm hẹn của hai người, $0 \leq x, y \leq 60$.



Hình 1.4: Minh họa cho Ví dụ 1.17

Vậy mỗi cặp thời điểm đến (x, y) của hai người là một điểm của miền

$$G = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 60;\ 0 \le y \le 60\} \quad \text{(hình vuông } OABC\text{)}.$$

Gọi E là sự kiện "hai người gặp nhau", khi đó E được biểu diễn bởi

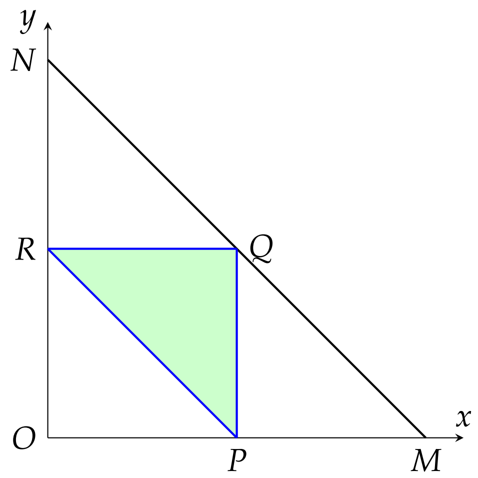
$$H = \{(x,y) \in G : |x - y| \le 10\} \quad \text{(đa giác } OMNBPQ\text{)}.$$

Sử dụng định nghĩa xác suất theo quan điểm hình học,

$$P(E) = \frac{\text{diện tích } H}{\text{diện tích } G} = \frac{\text{diện tích } (OMNBPQ)}{\text{diện tích } (OABC)} = \frac{60^2 - 50^2}{60^2} = \frac{11}{36} \simeq 0,3056.$$

Ví dụ 1.18. Cho đoạn thẳng AB có độ dài 10cm. Lấy hai điểm C, D bất kỳ trên đoạn AB (C nằm giữa A và D). Tính xác suất độ dài AC, CD, DB tạo thành 3 cạnh của một tam giác.

Lời giải Ví dụ 1.18 Gọi x là độ dài đoạn AC, y là độ dài đoạn CD thì độ dài đoạn DB là $10 - x - y$.



Hình 1.5: Minh họa cho Ví dụ 1.18

Khi đó ta có điều kiện $0 \le x \le 10$, $0 \le y \le 10$ và $0 \le x + y \le 10$. Tập hợp các giá trị (x, y) thỏa mãn điều kiện này tương ứng với miền

$$G = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 10,\ 0 \le y \le 10,\ 0 \le x + y \le 10\} \quad \text{(tam giác } OMN\text{)}.$$

Độ dài các đoạn AC, CD, DB tạo thành 3 cạnh một tam giác phải thỏa mãn tính chất "tổng hai cạnh lớn hơn một cạnh", tức là $x + y > 10 - x - y$, $x + (10 - x - y) > y$, $y + (10 - x - y) > x$ hay $x + y > 5$, $x < 5$ và $y < 5$. Tập các giá trị (x, y) thỏa mãn điều kiện này tương ứng với miền

$$H = \{(x,y) \in G : x + y > 5,\ x < 5,\ y < 5\} \quad \text{(tam giác } PQR\text{)}.$$

Theo định nghĩa hình học, xác suất cần tìm là $p = \dfrac{\text{diện tích tam giác } (PQR)}{\text{diện tích tam giác } (OMN)} = \dfrac{1}{4} = 0,25$.

Ví dụ 1.19. Trên mặt phẳng đã kẻ sẵn các đường thẳng song song cách đều nhau một khoảng có độ dài $2a$, người ta gieo ngẫu nhiên một chiếc kim dài $2b$ ($b < a$). Tính xác suất sao cho kim cắt một đường thẳng trong số những đường thẳng đó.

Lời giải Ví dụ 1.19 Gọi x là khoảng cách từ trung điểm của kim đến đường thẳng song song gần nhất và φ là góc mà kim tạo với các đường này. Ta có $0 \leq x \leq a$, $0 \leq \varphi \leq \pi$. Như vậy có thể biểu diễn miền đồng khả năng bởi hình chữ nhật $G = [a, \pi] \times [a, \pi]$. Miền thuận lợi cho sự kiện kim cắt đường thẳng song song là $H = \{(x, \varphi) \in G : 0 \leq x \leq b \sin \varphi;$ $0 \leq \varphi \leq \pi\}$. Do đó,

$$p = \frac{\text{diện tích } H}{\text{diện tích } G} = \frac{\displaystyle\int_0^{\pi} b \sin \varphi d\varphi}{a \times \pi} = \frac{2b}{a\pi}.$$

Nhận xét 1.3. Định nghĩa cổ điển và định nghĩa hình học về xác suất chỉ áp dụng được với các phép thử có kết cục đồng khả năng xảy ra. Trong nhiều bài toán thực tế, việc tính hết các kết cục của một phép thử không dễ dàng, bên cạnh đó điều kiện các kết cục đồng khả năng trên thực tế thường khó thỏa mãn.

1.3.4 Định nghĩa thống kê về xác suất

Định nghĩa 1.11 (Tần suất)**.** Giả sử trong một điều kiện nào đó ta có thể lặp lại n lần một phép thử và thấy có m lần xuất hiện sự kiện A. Khi đó, tỷ số $\dfrac{m}{n}$ gọi là tần suất xuất hiện A, ký hiệu là $f(A)$.

Như vậy

$$\boxed{f(A) = \frac{m}{n}} \tag{1.7}$$

Ví dụ 1.20. Để xác định tần suất xuất hiện mặt sấp khi tung một đồng xu nhiều lần, người ta ghi lại kết quả sau:

Người thí nghiệm	Số lần tung n	Số lần xuất hiện mặt sấp m	Tần suất f
Buýp-phông	4040	2048	0,5080
Piêc-sơn	12000	6019	0,5016
Piêc-sơn	24000	12012	0,5005

Ta thấy rằng khi số lần tung đồng xu càng lớn, tần suất xuất hiện mặt sấp càng gần với xác suất xuất hiện mặt sấp là 0,5.

Nhận xét 1.4. Tần suất của sự kiện A có tính chất ổn định, nghĩa là nó dao động rất ít xung quanh một số xác định p nào đó khi số phép thử khá lớn. Số ấy gọi là xác suất của sự kiện A theo quan điểm thống kê.

Định nghĩa 1.12 (Định nghĩa thống kê)**.** Nếu tần suất xuất hiện sự kiện A luôn luôn dao động xung quanh một số xác định p nào đó và khi số phép thử tăng lên khá lớn mà tần suất xuất hiện sự kiện A càng gần tới p thì số p được gọi là xác suất của sự kiện A theo quan điểm thống kê.

Chú ý 1.5. Bằng định nghĩa thống kê về xác suất, người ta đã tìm được xác suất để sinh con trai trong mỗi lần sinh là $p = 0,518$, con số này hầu như không thay đổi theo thời gian, địa phương và chủng tộc.

(a) Nhà toán học Láp–la–xơ trong 10 năm liền theo dõi ở thành phố Pê–tec–bua, Luân–đôn và Béc–lin thấy tỷ số đó là 22/43. Ông cũng đã theo dõi 40 năm liền ở Pa–ri thấy tỷ số đó là 25/49.

(b) Nhà toán học Crame theo dõi ở Thụy–điển năm 1935 cũng thấy tỷ số đó là 0,518.

Nhận xét 1.5. (a) Định nghĩa thống kê của xác suất khắc phục được một nhược điểm của định nghĩa cổ điển là không dùng đến khái niệm đồng khả năng.

(b) Định nghĩa này không giúp ta tính được chính xác xác suất của một sự kiện mà chỉ tìm được giá trị gần đúng; đồng thời số phép thử phải đủ lớn và chỉ dùng được cho các phép thử ngẫu nhiên có thể lặp lại nhiều lần một cách độc lập trong các điều kiện giống nhau.

Các định nghĩa trên về xác suất giúp ta một cách tích cực trong việc tính xác suất, nhưng mỗi định nghĩa đều có nhược điểm của nó. Để khắc phục các nhược điểm đó, năm 1933 nhà toán học Xô–viết Can–mơ–gơ–rôp đã đưa ra xác suất theo phương pháp tiên đề. Tuy nhiên ta không đề cập đến trong chương trình này.

1.3.5 Nguyên lý xác suất nhỏ, nguyên lý xác suất lớn

Sự kiện không thể có có xác suất bằng 0, một sự kiện có xác suất gần bằng 0 vẫn có thể xảy ra khi thực hiện một số lớn các phép thử. Tuy nhiên qua thực nghiệm và quan sát thực tế, người ta thấy rằng các sự kiện có xác suất nhỏ sẽ không xảy ra khi ta chỉ thực hiện một phép thử hay một vài phép thử. Từ đó ta thừa nhận nguyên lý sau đây, gọi là "Nguyên lý xác suất nhỏ": *Nếu một sự kiện có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử sự kiện đó sẽ không xảy ra".*

Nhận xét 1.6. (a) Mức xác suất được coi là nhỏ tùy thuộc vào từng bài toán cụ thể và gọi là mức ý nghĩa, ký hiệu là α.

(b) Nguyên lý xác suất nhỏ là cơ sở của phương pháp kiểm định giả thuyết (sẽ được đề cập ở Chương 5).

Tương tự như trên, ta có thể đưa ra nguyên lý xác suất lớn: *Nếu sự kiện A có xác suất gần bằng 1 thì trên thực tế có thể cho rằng trong một phép thử sự kiện đó sẽ xảy ra".*

Nhận xét 1.7. **(a)** Mức xác suất đủ lớn gọi là độ tin cậy, ký hiệu là $\gamma = 1 - \alpha$. Việc quy định một mức xác suất thế nào là lớn sẽ tùy thuộc vào từng bài toán cụ thể.

(b) Nguyên lý xác suất lớn là cơ sở của phương pháp ước lượng bằng khoảng tin cậy (sẽ được đề cập ở Chương 4).

BÀI 3 (2 tiết)

1.4 Công thức cộng và nhân xác suất

1.4.1 Xác suất điều kiện

Định nghĩa 1.13 (Xác suất điều kiện). Giả sử trong một phép thử ta có $P(B) > 0$. Khi đó xác suất có điều kiện của sự kiện A nào đó, biết rằng đã có B, là một số không âm ký hiệu là

$$\boxed{P(A|B) = \frac{P(AB)}{P(B)}} \tag{1.8}$$

Tương tự

$$\boxed{P(B|A) = \frac{P(AB)}{P(A)}, \quad P(A) > 0} \tag{1.9}$$

Nhận xét 1.8. **(a)** Xác suất điều kiện có mọi tính chất của một xác suất bình thường, chẳng hạn $P(A|B) \geq 0$, $P(A|A) = 1$.

(b) Ta có thể tính xác suất điều kiện bằng cách áp dụng các công thức (1.8) hoặc (1.9) hoặc tính trực tiếp.

Ví dụ 1.21. Từ một bộ bài tú-lơ-khơ 52 cây đã trộn kỹ rút ngẫu nhiên ra một cây bài. Biết đó là cây đen, tính xác suất đó là cây át.

Lời giải Ví dụ 1.21 Gọi A là sự kiện "rút được cây át" và B là sự kiện “rút được cây đen”. Xác suất cần tính là $P(A|B) = \dfrac{2}{26}$.

Ví dụ 1.22. Trong một thùng kín có N quả cầu giống nhau trong đó có M cầu trắng ($M < N$). Lấy ngẫu nhiên lần lượt không hoàn lại 2 quả cầu. Tính xác suất để cầu thứ hai lấy ra là trắng, biết rằng cầu thứ nhất lấy ra đã là trắng.

Lời giải Ví dụ 1.22 Gọi A_i là sự kiện "cầu thứ i lấy ra là trắng", $i = 1, 2$. Sử dụng công thức (1.8) ta được

$$P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)} = \frac{\frac{M\times(M-1)}{N\times(N-1)}}{\frac{M}{N}} = \frac{M-1}{N-1}.$$

1.4.2 Công thức nhân xác suất

1.4.2a Tính độc lập của các sự kiện

Định nghĩa 1.14 (Sự kiện độc lập). **(a)** Hai sự kiện A và B được gọi là độc lập với nhau nếu sự kiện này xảy ra hay không xảy ra không làm ảnh hưởng tới khả năng xảy ra của sự kiện kia, nghĩa là $P(A|B) = P(A|\overline{B}) = P(A)$, $P(B|A) = P(B|\overline{A}) = P(B)$.

(b) Các sự kiện $A_1, A_2, \ldots, A_n$ được gọi là độc lập từng đôi với nhau nếu mỗi cặp 2 trong n sự kiện đó độc lập với nhau.

(c) Các sự kiện $A_1, A_2, \ldots, A_n$ được gọi là độc lập trong tổng thể nếu mỗi sự kiện trong chúng độc lập với tích của một số bất kỳ sự kiện trong các sự kiện còn lại.

Chú ý 1.6. **(a)** Nếu A và B độc lập thì các cặp A và $\overline{B}$; $\overline{A}$ và B; $\overline{A}$ và $\overline{B}$ cũng độc lập.

(b) Thông thường tính độc lập của các sự kiện được suy ra từ ý nghĩa thực tế.

1.4.2b Công thức nhân xác suất

(a) Nếu A và B là hai sự kiện bất kỳ thì

$$\boxed{P(AB) = P(A)P(B|A) = P(B)P(A|B)} \tag{1.10}$$

Nếu A và B là hai sự kiện độc lập thì

$$\boxed{P(AB) = P(A)P(B)} \tag{1.11}$$

(b) Mở rộng cho tích n sự kiện bất kỳ $A_1, A_2, \ldots, A_n$:

$$\boxed{P(A_1A_2\ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})} \tag{1.12}$$

Nếu $A_1, A_2, \ldots, A_n$ độc lập trong tổng thể, thì:

$$\boxed{P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n)} \tag{1.13}$$

Nhận xét 1.9. Công thức nhân (1.11) cung cấp cho ta một phương pháp dễ thực hành để kiểm tra tính độc lập của hai sự kiện ngẫu nhiên.

Ví dụ 1.23. Gieo đồng thời hai đồng xu cân đối đồng chất lên mặt phẳng cứng. Gọi A_1 là sự kiện "ít nhất một đồng xu xuất hiện mặt sấp", A_2 là sự kiện "ít nhất một đồng xu xuất hiện mặt ngửa", A_3 là sự kiện "có ba đồng xu xuất hiện mặt ngửa". A_1, A_2, A_3 có độc lập không?

Lời giải Ví dụ 1.23 Mặc dù $P(A_1A_2A_3) = 0 = P(A_1)P(A_2)P(A_3)$, nhưng $P(A_1A_2) = \dfrac{1}{2} \neq \dfrac{3}{4} \times \dfrac{3}{4} = P(A_1)P(A_2)$. Do đó, A_1, A_2, A_3 không độc lập.

Ví dụ 1.24. Có 4 que thăm, trong đó có 3 que thăm dài bằng nhau và 1 que thăm ngắn hơn. Bốn người lần lượt lên rút ngẫu nhiên một que thăm. Tính xác suất người thứ i rút được thăm ngắn $(i = 1, 2, 3, 4)$.

Lời giải Ví dụ 1.24 Gọi A_i là sự kiện "lần thứ i rút được thăm ngắn", B_i là sự kiện "người thứ i rút được thăm ngắn", $i = 1, 2, 3, 4$. Khi đó,

$$
\begin{aligned}
&P\left(B_1\right) = \frac{1}{4}.\\
&P\left(B_2\right) = P(\overline{A}_1 A_2) = P\left(\overline{A}_1\right) P\left(A_2|\overline{A}_1\right) = \frac{3}{4} \times \frac{1}{3} = \frac{.1}{4}\\
&P\left(B_3\right) = P\left(\overline{A}_1\overline{A}_2 A_3\right) = P\left(\overline{A}_1\right) P\left(\overline{A}_2|\overline{A}_1\right) P\left(A_3|\overline{A}_1\overline{A}_2\right) = \frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{4}.\\
&P\left(B_4\right) = P(\overline{A}_1\overline{A}_2\overline{A}_3 A_4) = P\left(\overline{A}_1\right) P\left(\overline{A}_2|\overline{A}_1\right) P\left(\overline{A}_3|\overline{A}_1\overline{A}_2\right) P(A_4|\overline{A}_1\overline{A}_2\overline{A}_3) = \frac{1}{4}.
\end{aligned}
$$

Vậy khả năng rút được thăm ngắn của 4 người là như nhau và bằng $\frac{1}{4}$.

Ví dụ 1.25. Một tổ có 15 sinh viên trong đó có 5 sinh viên học giỏi môn Xác suất thống kê. Chia tổ này thành 5 nhóm, mỗi nhóm 3 người. Tính xác suất để nhóm nào cũng có một sinh viên học giỏi môn Xác suất thống kê.

Lời giải Ví dụ 1.25 Gọi A là sự kiện "nhóm nào cũng có một sinh viên học giỏi môn Xác suất thống kê"; A_i là sự kiện "nhóm i có một sinh viên học giỏi môn Xác suất thống kê", $i = 1, \ldots, 5$. Khi đó $A = A_1A_2A_3A_4A_5$. Sử dụng công thức nhân (1.12)

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_4|A_1A_2A_3)P(A_5|A_1A_2A_3A_4),$$

trong đó

$$
\begin{aligned}
&P(A_1) = \frac{C_5^1 \times C_{10}^2}{C_{15}^3} = \frac{45}{91}, \qquad P(A_2|A_1) = \frac{C_4^1 \times C_8^2}{C_{12}^3} = \frac{28}{55},\\
&P(A_3|A_1A_2) = \frac{C_3^1 \times C_6^2}{C_9^3} = \frac{15}{28}, \quad P(A_4|A_1A_2A_3) = \frac{C_2^1 \times C_4^2}{C_6^3} = \frac{3}{5},\\
&P(A_5|A_1A_2A_3A_4) = \frac{C_1^1 \times C_2^2}{C_3^3} = 1.
\end{aligned}
$$

Vậy $P(A) = \frac{81}{1001} \simeq 0,0809$.

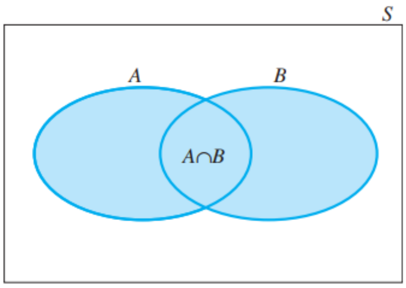
1.4.3 Công thức cộng xác suất

(a) Nếu A và B là hai sự kiện bất kỳ thì

$$\boxed{P(A+B) = P(A) + P(B) - P(AB)} \tag{1.14}$$

Nếu A và B là hai sự kiện xung khắc thì

$$\boxed{P(A+B) = P(A) + P(B)} \tag{1.15}$$



Hình 1.6: Minh họa công thức cộng

(b) Nếu A, B và C là ba sự kiện bất kỳ thì

$$\boxed{P(A+B+C) = P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)} \tag{1.16}$$

(c) Mở rộng cho tổng n sự kiện bất kỳ $A_1, A_2, \ldots, A_n$:

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \ldots + (-1)^{n-1}P(A_1A_2\ldots A_n). \tag{1.17}$$

Nếu $A_1, A_2, \ldots, A_n$ xung khắc từng đôi thì

$$\boxed{P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)} \tag{1.18}$$

Đặc biệt:

(a) Nếu $A_1, A_2, \ldots, A_n$ là hệ đầy đủ các sự kiện thì $\sum\limits_{i=1}^{n} P(A_i) = 1$.

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

(c) $P(A) = P(AB) + P(A\overline{B})$.

Ví dụ 1.26. Một lớp có 100 sinh viên, trong đó có 40 sinh viên giỏi ngoại ngữ, 30 sinh viên giỏi toán xác suất, 20 sinh viên giỏi cả ngoại ngữ lẫn toán xác suất. Chọn ngẫu nhiên một sinh viên trong lớp. Tìm xác suất để sinh viên đó giỏi ít nhất 1 trong 2 môn trên.

Lời giải Ví dụ 1.26 Gọi A là sự kiện "sinh viên đó giỏi ít nhất 1 trong 2 môn ngoại ngữ, toán xác suất"; N là sự kiện "sinh viên đó giỏi ngoại ngữ"; T là sự kiện "sinh viên đó giỏi toán xác suất". Khi đó, $A = T + N$. Suy ra

$$P(A) = P(T+N) = P(T) + P(N) - P(TN) = \frac{30}{100} + \frac{40}{100} - \frac{20}{100} = \frac{50}{100} = 0,5.$$

Ví dụ 1.27. Ba xạ thủ độc lập với nhau cùng bắn súng vào bia. Xác suất bắn trúng bia của xạ thủ thứ nhất, thứ hai, thứ ba tương ứng là 0,7, 0,8 và 0,9. Tính xác suất để:

(a) có duy nhất một xạ thủ bắn trúng bia;

(b) có đúng hai xạ thủ bắn trúng bia;

(c) có ít nhất một xạ thủ bắn trúng bia;

(d) xạ thủ thứ nhất bắn trúng bia biết rằng có hai xạ thủ bắn trúng bia.

Lời giải Ví dụ 1.27 Gọi A_i là các sự kiện "xạ thủ thứ i bắn trúng bia", $i = 1, 2, 3$.

(a) Gọi A là sự kiện "có duy nhất một xạ thủ bắn trúng bia". Khi đó,

$$A = A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3.$$

Sử dụng công thức cộng (1.18) và công thức nhân (1.13) trong trường hợp các sự kiện xung khắc và độc lập suy ra

$$\begin{aligned}P(A) &= P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3)\\ &= 0,7 \times 0,2 \times 0,1 + 0,3 \times 0,8 \times 0,1 + 0,3 \times 0,2 \times 0,9 = 0,092.\end{aligned}$$

(b) Gọi B là sự kiện "có đúng hai xạ thủ bắn trúng bia". Khi đó,

$$B = A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3.$$

Làm tương tự ý (a), $P(B) = 0,398$.

(c) Gọi C là sự kiện "ít nhất một xạ thủ bắn trúng bia", khi đó

Hoặc $C = A_1 + A_2 + A_3$,

$$\begin{aligned}P(C) &= P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3)\\ &= 0,994.\end{aligned}$$

Hoặc $\overline{C} = \overline{A}_1\overline{A}_2\overline{A}_3$,

$$P(\overline{C}) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3), \quad P(C) = 1 - P(\overline{C}) = 1 - 0,006 = 0,994.$$

Hoặc $C = A + B + A_1A_2A_3$, $P(C) = P(A) + P(B) + P(A_1A_2A_3) = 0,994$.

(d) $P(A_1|B) = \dfrac{P(A_1B)}{P(B)} = \dfrac{P(A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3)}{P(B)} = \dfrac{0,182}{0,398} = 0,46.$

BÀI 4 (2 tiết)

1.5 Công thức Béc–nu–li

1.5.1 Dãy phép thử độc lập

Định nghĩa 1.15 (Dãy phép thử độc lập)**.** Các phép thử được gọi là độc lập với nhau nếu xác suất để xảy ra một sự kiện nào đó trong từng phép thử sẽ không phụ thuộc vào việc sự kiện đó có xảy ra ở các phép thử khác hay không.

Ví dụ 1.28. Tung nhiều lần một đồng xu sẽ tạo nên các phép thử độc lập. Lấy nhiều lần sản phẩm từ một lô sản phẩm theo phương thức có hoàn lại cũng tạo nên các phép thử độc lập.

1.5.2 Lược đồ Béc–nu–li

1. Giả sử ta tiến hành n phép thử độc lập.

2. Trong mỗi phép thử chỉ có hai trường hợp: hoặc sự kiện A xảy ra, hoặc sự kiện A không xảy ra (tức là xảy ra $\overline{A}$).

3. Xác suất xảy ra A trong mỗi phép thử đều bằng p (tức là $P(A) = p$) và xác suất không xảy ra A trong mỗi phép thử đều bằng $q = 1 - p$ (tức là $P(\overline{A}) = 1 - p$).

Những bài toán thỏa mãn cả ba điều kiện trên được gọi là tuân theo lược đồ Béc-nu-li (hay dãy phép thử Béc-nu-li).

1.5.3 Công thức Béc–nu–li

Định lý 1.1. Trong lược đồ Béc–nu–li (hay dãy phép thử Béc–nu–li)

(a) Xác suất để sự kiện A xuất hiện đúng k lần, ký hiệu là $P_n(k)$, được xác định bởi

$$\boxed{P_n(k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n} \tag{1.19}$$

(b) Xác suất để sự kiện A xuất hiện từ k_1 đến k_2 lần, ký hiệu là $P_n(k_1, k_2)$:

$$\boxed{P_n(k_1, k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k}} \tag{1.20}$$

Chứng minh. (a) Gọi B là sự kiện "trong dãy n phép thử Béc–nu–li, sự kiện A xuất hiện đúng k lần". Ta thấy B có thể xảy ra nhiều phương án khác nhau miễn sao trong đó sự kiện A xuất hiện đúng k lần. Khi đó, có C_n^k phương án như vậy. Còn xác suất xảy ra một phương án sẽ là

$p^k q^{n-k}$ do các phép thử là độc lập, sự kiện A xuất hiện k lần, sự kiện $\overline{A}$ xuất hiện $n-k$ lần. Từ đó ta có công thức cần chứng minh.

(b) Suy trực tiếp từ ý (a).

Nhận xét 1.10. Nếu một bài toán thỏa mãn lược đồ Béc–nu–li thì việc sử dụng công thức (1.19) hay (1.20) sẽ đơn giản hơn rất nhiều so với việc dùng các công thức nhân xác suất và cộng xác suất. Do đó chúng có ý nghĩa thực tiễn rất lớn.

Ví dụ 1.29. Trong phân xưởng có 5 máy hoạt động độc lập. Xác suất để trong mỗi ca mỗi máy bị hỏng đều bằng 0,1.

(a) Tìm xác suất để trong ca đó có đúng 2 máy hỏng.

(b) (Đề thi giữa kỳ 20182) Biết rằng trong một ca có đúng 2 máy hỏng. Tính xác suất để máy thứ nhất không hỏng.

Lời giải Ví dụ 1.29

(a) Coi sự hoạt động của mỗi máy là một phép thử. Ta có 5 phép thử độc lập; trong mỗi phép thử chỉ có 2 trường hợp: hoặc máy hỏng, hoặc máy không hỏng; xác suất hỏng của mỗi máy đều bằng 0,1. Như vậy, bài toán thỏa mãn lược đồ Béc–nu–li với $n = 5$, $p = 0,1$ và $k = 2$. Áp dụng công thức Béc–nu–li (1.19) ta có xác suất cần tìm là:

$$P_5(2) = C_5^2 \times (0,1)^2 \times (0,9)^3 = 0,0729.$$

Nếu sử dụng công thức cộng và nhân xác suất với A là sự kiện "trong ca đó có đúng 2 máy hỏng", A_i là sự kiện "máy i bị hỏng trong ca", $i = 1, 2, \ldots, 5$, ta sẽ tính xác suất của A trên cơ sở phân tích:

$$\begin{aligned} A = {} & A_1A_2\overline{A}_3\overline{A}_4\overline{A}_5 + A_1\overline{A}_2A_3\overline{A}_4\overline{A}_5 + A_1\overline{A}_2\overline{A}_3A_4\overline{A}_5 + A_1\overline{A}_2\overline{A}_3\overline{A}_4A_5 + \\ & + \overline{A}_1A_2A_3\overline{A}_4\overline{A}_5 + \overline{A}_1A_2\overline{A}_3A_4\overline{A}_5 + \overline{A}_1A_2\overline{A}_3\overline{A}_4A_5 + \\ & + \overline{A}_1\overline{A}_2A_3A_4\overline{A}_5 + \overline{A}_1\overline{A}_2A_3\overline{A}_4A_5 + \overline{A}_1\overline{A}_2\overline{A}_3A_4A_5 \end{aligned}$$

và sử dụng tính xung khắc, tính độc lập của các sự kiện. Rõ ràng việc sử dụng công thức (1.19) cho ví dụ này đơn giản hơn rất nhiều.

(b) $P(\overline{A}_1|A) = \dfrac{P(\overline{A}_1)P(A|\overline{A}_1)}{P(A)} = \dfrac{0,9 \times C_4^2 \times (0,1)^2 \times (0,9)^2}{0,0729} = \dfrac{0,04374}{0,0729} = 0,6.$

Ví dụ 1.30. Hai vận động viên bóng bàn A và B đấu một trận gồm tối đa 5 ván (không có kết quả hòa sau mỗi ván và trận đấu sẽ dừng nếu một người nào đó thắng trước 3 ván). Xác suất để A thắng được ở một ván là 0,7.

(a) Tính các xác suất để A thắng sau x ván ($x = 3, 4, 5$).

(b) Tính xác suất để trận đấu kết thúc sau 5 ván.

Lời giải Ví dụ 1.30

(a) Việc A thắng sau x ván ($x = 3, 4, 5$) tương đương với sự kiện "ván thứ x người A thắng và trong $x - 1$ ván đầu người A thắng 2 ván". Khi đó, xác suất cần tìm là

$$p_x = 0,7 \times P_{x-1}(2) = 0,7 \times C_{x-1}^2 \times (0,7)^2 \times (0,3)^{x-3},$$

cụ thể:

$$p_3 = 0,7 \times P_2(2) = 0,7 \times C_2^2 \times (0,7)^2 \times (0,3)^0 = 0,343,$$
$$p_4 = 0,7 \times P_3(2) = 0,7 \times C_3^2 \times (0,7)^2 \times (0,3)^1 = 0,3087,$$
$$p_5 = 0,7 \times P_4(2) = 0,7 \times C_4^2 \times (0,7)^2 \times (0,3)^2 = 0,18522.$$

(b) Sự kiện "trận đấu kết thúc sau 5 ván" tương đương với sự kiện "trong 4 ván đầu mỗi người thắng 2 ván". Khi đó xác suất cần tìm là:

$$P_4(2) = C_4^2 \times (0,7)^2 \times (0,3)^2 = 0,2646.$$

Ví dụ 1.31. Tỷ lệ phế phẩm của một lô hàng là 1%. Hỏi cỡ mẫu cần chọn ra là bao nhiêu (có hoàn lại) sao cho trong mẫu có ít nhất 1 phế phẩm với xác suất lớn hơn 0,95?

Lời giải Ví dụ 1.31 Giả sử mẫu chọn ra có kích cỡ là n và việc chọn ra một sản phẩm có hoàn lại là một phép thử Béc–nu–li với $p = 0,01$. Gọi A là sự kiện "trong mẫu có ít nhất một phế phẩm" thì $\overline{A}$ sẽ là sự kiện "trong mẫu không có phế phẩm nào". Khi đó $\overline{A} = A_1 A_2 \ldots A_2$, với A_i là sự kiện "sản phẩm thứ i lấy ra không là phế phẩm", $i = 1, 2, \ldots, n$. Suy ra

$$P(A) = 1 - P(\overline{A}) = 1 - (0,99)^n.$$

Theo yêu cầu của đầu bài, $P(A) > 0,95$ tức là $1 - (0,99)^n > 0,95$ hay $0,05 > (0,99)^n$. Từ đây suy ra

$$n > \frac{\log 0,05}{\log 0,99} \simeq 298.$$

1.5.4 Số có khả năng nhất trong lược đồ Béc–nu–li

Trong lược đồ Béc–nu–li, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là số có khả năng nhất (hay số lần xuất hiện chắc chắn nhất).

Cách xác định số có khả năng nhất x_0:

(a) Nếu $np - q \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = np - q$ và $x_0 = np - q + 1$.

(b) Nếu $np = q \notin \mathbb{Z}$ thì $x_0 = [np - q] + 1$, ở đây $[np - q]$ là phần nguyên của $np - q$ (phần nguyên được xác định như sau nếu $1 \le x < 2$ thì $[x] = 1$, nếu $2 \le x < 3$ thì $[x] = 2 \ldots$).

Ví dụ 1.32. Tỷ lệ mắc một loại bệnh A ở một vùng là 10%. Trong đợt khám bệnh cho vùng đó người ta đã khám 100 người. Tìm số người bị bệnh A có khả năng nhất? Tính xác suất tương ứng.

Lời giải Ví dụ 1.32 Bài toán thỏa mãn lược đồ Béc–nu–li với $n = 100$, $p = 0,1$. Theo bài ra ta có $np - q = 100 \times 0,1 - 0,9 = 9,1 \notin \mathbb{Z}$. Vậy số người bị bệnh A có khả năng nhất khi khám 100 người là $[9,1] + 1 = 10$ người và xác suất tương ứng là $P_{100}(10) = C_{100}^{10} \times (0,1)^{10} \times (0,9)^{90} \simeq 0,1319$.

1.5.5 Công thức xấp xỉ

Khi n và k khá lớn thì việc tính toán xác suất theo (1.19) và (1.20) rất cồng kềnh và khó khăn, vì vậy người ta tìm cách tính gần đúng các xác suất đó.

(a) **Xấp xỉ Poa-xông**: Nếu n rất lớn, trong khi p rất nhỏ, xác suất theo công thức (1.19) có thể xấp xỉ bằng

$$\boxed{P_n(k) \simeq \frac{(\lambda)^k}{k!} e^{-\lambda}} \tag{1.21}$$

Xác suất này được tính sẵn trong bảng giá trị hàm khối lượng xác suất Poa–xông (Phụ lục 5) với $\lambda = np$.

(b) **Xấp xỉ chuẩn** (định lý giới hạn địa phương Moa-vrơ–Láp-la-xơ): Nếu n lớn nhưng p không quá bé và quá lớn ta có xấp xỉ

$$\boxed{P_n(k) \simeq \frac{\varphi(x_k)}{\sqrt{npq}}, \quad x_k = \frac{k - np}{\sqrt{npq}}} \tag{1.22}$$

trong đó $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ là hàm Gao–xơ với các giá trị được tính trong bảng giá trị hàm Gao–xơ (Phụ lục 1) đối với các giá trị x dương. Hàm $\varphi(x)$ là hàm chẵn, tức là $\varphi(-x) = \varphi(x)$. Khi $x > 4$ ta có thể lấy $\varphi(x) \simeq 0$.

(c) **Xấp xỉ cho công thức** (1.20) (định lý giới hạn tích phân Moa-vrơ–Láp-la-xơ): Nếu n lớn nhưng p không quá bé và quá lớn thì xác suất trong (1.20) có thể xấp xỉ bằng

$$\boxed{P_n(k_1; k_2) \simeq \phi(x_2) - \phi(x_1), \quad x_i = \frac{k_i - np}{\sqrt{npq}}, \ i = 1,2} \tag{1.23}$$

trong đó

$$\boxed{\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt} \tag{1.24}$$

là hàm Láp–la–xơ với các giá trị được tính trong bảng giá trị hàm Láp–la–xơ (Phụ lục 2) đối với các giá trị x dương. Hàm $\phi(x)$ là hàm lẻ, tức là $\phi(-x) = -\phi(x)$. Khi $x > 5$ ta có thể lấy $\phi(x) \simeq 0,5$.

Ví dụ 1.33. Xác suất để sản phẩm sau khi sản xuất không được kiểm tra chất lượng bằng 0,2. Tìm xác suất để trong 400 sản phẩm sản xuất ra có:

(a) 80 sản phẩm không được kiểm tra chất lượng;

(b) từ 70 đến 100 sản phẩm không được kiểm tra chất lượng.

Lời giải Ví dụ 1.33 Bài toán thỏa mãn lược đồ Béc–nu–li với $n = 400$, $p = 0,2$.

(a) Ta phải tính $P_{400}(80)$ theo công thức Béc–nu–li (1.19):

$$P_{400}(80) = C_{400}^{80} \times (0,2)^{80} \times (0,8)^{320}.$$

Việc tính xác suất theo công thức này khá phức tạp vì $n = 400$ khá lớn. Vì $p = 0,2$ không quá bé hoặc quá lớn, ta sẽ tính xấp xỉ theo (3.37):

$$P_{400}(80) \simeq \frac{\varphi(0)}{8} \simeq 0,04986$$

ở đây $\varphi(0) = 0,3989$ được tra từ bảng giá trị hàm Gau-xơ (Phụ lục 1).

(b) Tương tự, thay việc dùng công thức (1.20) ta sử dụng xấp xỉ (3.40):

$$P_{400}(70;100) \simeq \phi(2,5) - \phi(-1,25) \simeq 0,49379 + 0,39435 = 0,88814,$$

ở đây $\phi(-1,25) = -0,39435$, $\phi(2,5) = 0,49379$ tra từ bảng giá trị hàm Láp–la–xơ (Phụ lục 2).

Ví dụ 1.34. Vận chuyển 4000 chai rượu đến một cửa hàng. Xác suất để mỗi chai rượu bị vỡ trong quá trình vận chuyển là 0,001. Tính xác suất để có 7 chai rượu bị vỡ trong quá trình vận chuyển.

Lời giải Ví dụ 1.34 Bài toán thỏa mãn lược đồ Béc–nu–li với $n = 4000$, $p = 0,001$. Ta phải tính $P_{4000}(7)$ theo công thức Béc–nu–li (1.19):

$$P_{4000}(7) = C_{4000}^{7} \times (0,001)^{7} \times (0,999)^{3993}.$$

Vì $n = 4000$ khá lớn, $p = 0,001$ khá bé, nên ta sẽ tính xấp xỉ theo (1.21):

$$P_{4000}(7) \simeq \frac{4^7}{7!}(2,71828)^{-4} \simeq 0,05954.$$

Ta có thể tính trực tiếp hoặc tra bảng giá trị hàm khối lượng Poa-xông (Phụ lục 5).

1.6 Công thức xác suất đầy đủ. Công thức Bay–ét

1.6.1 Công thức xác suất đầy đủ

Định lý 1.2. Giả sử các sự kiện $A_1, A_2, \ldots, A_n$ lập thành một hệ đầy đủ và H là một sự kiện nào đó. Khi đó,

$$\boxed{P(H) = \sum_{i=1}^{n} P(A_i)P(H|A_i)} \tag{1.25}$$

Công thức (1.25) được gọi là công thức xác suất đầy đủ (hay công thức xác suất toàn phần). Công thức này cho phép ta tính xác suất $P(H)$ nếu biết các xác suất $P(A_i)$ và $P(H|A_i)$, $i = 1, 2, \ldots, n$.

Chứng minh. Từ giả thiết ta có $H = HS = H(A_1 + A_2 + \cdots + A_2)$. Sử dụng tính xung khắc của các sự kiện và công thức nhân suy ra

$$\begin{aligned} P(H) &= P(HA_1) + P(HA_2) + \cdots + P)HA_n \\ &= P(A_1)P(H|A_1) + P(A_2)P(H|A_2) + \cdots + P(A_n)P(H|A_n). \end{aligned}$$

1.6.2 Công thức Bay–ét

Định lý 1.3. Giả sử ta có một hệ đầy đủ $A_1, A_2, \ldots, A_n$, sau đó có thêm sự kiện H nào đó. Khi đó xác suất $P(A_k|H)$, $k = 1, 2, \ldots, n$, được xác định bởi:

$$\boxed{P(A_k|H) = \frac{P(A_k)P(H|A_k)}{\sum_{i=1}^{n} P(A_i)P(H|A_i)}, \quad k = 1, 2, \ldots, n} \tag{1.26}$$

Công thức (1.26) được gọi là công thức Bay-ét.

Chứng minh. Sử dụng công thức nhân (1.10) $P(A_kH) = P(A_k)P(H|A_k) = P(H)P(A_k|H)$. Suy ra

$$P(A_k|H) = \frac{P(A_k)P(H|A_k)}{P(H)}.$$

Từ đây sử dụng công thức xác suất đầy đủ (1.25) suy ra công thức (1.26).

Nhận xét 1.11. **(a)** Các xác suất $P(A_i)$, $i = 1, 2, \ldots, n$ đã được xác định từ trước, thường được gọi là xác suất tiên nghiệm.

(b) Các xác suất $P(A_i|H)$, $i = 1, 2, \ldots, n$ được xác định sau khi đã có kết quả thí nghiệm nào đó thể hiện qua sự xuất hiện của H, thường gọi là xác suất hậu nghiệm. Như vậy, công thức Bay–ét cho phép đánh giá lại xác suất xảy ra các sự kiện A_i sau khi đã có thêm thông tin về H.

Chú ý 1.7. **(a)** Muốn dùng công thức xác suất đầy đủ (1.25) hoặc công thức Bay–ét (1.26) nhất định phải có hệ đầy đủ.

(b) Nếu (1.25) cho ta xác suất không có điều kiện thì (1.26) cho phép tính xác suất có điều kiện, trong đó sự kiện A_i cần tính xác suất phải là một thành viên của nhóm đầy đủ đang xét. Từ đó thấy rằng việc dùng công thức Bay–ét để tính xác suất có điều kiện đã gợi ý cho ta cách chọn nhóm đầy đủ sao cho sự kiện quan tâm phải là thành viên.

(c) Trong trường hợp không có, hoặc rất khó xác định nhóm đầy đủ ta nên dùng công thức (1.26), trong trường hợp này tính $P(H)$ sẽ khó hơn là dùng công thức (1.25).

Ví dụ 1.35. Một nhà máy có ba phân xưởng sản xuất ra cùng một loại sản phẩm. Xác suất để phân xưởng 1, phân xưởng 2 và phân xưởng 3 sản xuất được sản phẩm loại một lần lượt là 0,7, 0,8 và 0,6. Từ một lô hàng gồm 20% sản phẩm của phân xưởng 1, 50% sản phẩm của phân xưởng 2 và 30% sản phẩm của phân xưởng 3 người ta lấy ra một sản phẩm để kiểm tra.

(a) Tính xác suất để sản phẩm được kiểm tra là loại một.

(b) Biết sản phẩm được kiểm tra là loại một. Tính xác suất để sản phẩm này do phân xưởng 2 sản xuất.

Lời giải Ví dụ 1.35 Gọi H là sự kiện "sản phẩm được kiểm tra là loại một"; A_i là sự kiện "sản phẩm được kiểm tra do phân xưởng i sản xuất", $i = 1, 2, 3$. Ta thấy A_1, A_2, A_3 tạo thành một hệ đầy đủ với $P(A_1) = 0,2$, $P(A_2) = 0,5$ và $P(A_3) = 0,3$.

(a) Áp dụng công thức xác suất đầy đủ (1.25) với $P(H|A_1) = 0,7$; $P(H|A_2) = 0,8$ và $P(H|A_3) = 0,6$ ta nhận được

$$\begin{aligned} P(H) &= P(A_1)P(H|A_1) + P(A_2)P(H|A_2) + P(A_3)P(H|A_3) \\ &= 0,2 \times 0,7 + 0,5 \times 0,8 + 0,3 \times 0,6 = 0,72 = 72\%. \end{aligned}$$

Ý nghĩa của xác suất này là tỷ lệ sản phẩm loại một của nhà máy.

(b) Áp dụng công thức Bay–ét (1.26) ta tính

$$P(A_2|H) = \frac{P(A_2)P(H|A_2)}{\sum_{i=1}^{3} P(A_i)P(H|A_i)} = \frac{0,5 \times 0,8}{0,72} = \frac{5}{9}.$$

Ví dụ 1.36. Có hai lô sản phẩm: lô I có 7 chính phẩm 3 phế phẩm; lô II có 6 chính phẩm 2 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm từ lô I bỏ sang lô II, sau đó từ lô II lấy ngẫu nhiên ra 2 sản phẩm.

(a) Tính xác suất để 2 sản phẩm lấy ra sau cùng là chính phẩm.

(b) Giả sử 2 sản phẩm lấy ra sau cùng là chính phẩm. Hãy tính xác suất để 2 chính phẩm này là của lô I (ban đầu).

Lời giải Ví dụ 1.36

(a) Gọi H là sự kiện "hai sản phẩm lấy ra sau cùng là chính phẩm"; A_i là sự kiện "trong 2 sản phẩm lấy từ lô I bỏ sang lô II có i chính phẩm", $i = 0, 1, 2$. Khi đó A_0, A_1, A_2 tạo thành một hệ đầy đủ với

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

và

$$P(H|A_0) = \frac{C_6^2}{C_{10}^2} = \frac{15}{45};\ P(H|A_1) = \frac{C_7^2}{C_{10}^2} = \frac{21}{45};\ P(H|A_2) = \frac{C_8^2}{C_{10}^2} = \frac{28}{45}.$$

Áp dụng công thức xác suất đầy đủ (1.25)

$$\begin{aligned} P(H) &= P(A_0)P(H|A_0) + P(A_1)P(H|A_1) + P(A_2)P(H|A_2) \\ &= \frac{1}{15} \times \frac{15}{45} + \frac{7}{15} \times \frac{21}{45} + \frac{7}{15} \times \frac{28}{45} = \frac{358}{675} \simeq 0,5304. \end{aligned}$$

(b) Ta không thể chọn nhóm đầy đủ như trong ý (a), vì sự kiện cần tính xác suất không là thành viên của nhóm này. Việc chọn nhóm đầy đủ thích hợp xem như là bài tập.

Ví dụ 1.37. Một người có ba chỗ ưa thích như nhau để câu cá. Xác suất để câu được cá ở mỗi chỗ tương ứng là 0,6; 0,7 và 0,8. Biết rằng đến một chỗ người đó thả câu 3 lần và chỉ câu được một con cá. Tính xác suất để cá câu được ở chỗ thứ nhất.

Lời giải Ví dụ 1.37 Gọi A_i là sự kiện "người đó chọn chỗ thứ i", $i = 1, 2, 3$, A là sự kiện "câu được cá". Khi đó,

$$P(A) = P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + P(A_3)P(A|A_3) = 0,191,$$

trong đó

$$\begin{aligned} &P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}, \\ &P(A|A_1) = P_3(1) = C_3^1 \times (0,6)^1 \times (0,4)^2 = 0,288, \\ &P(A|A_2) = P_3(1) = C_3^1 \times (0,7)^1 \times (0,3)^2 = 0,189, \\ &P(A|A_3) = P_3(1) = C_3^1 \times (0,8)^1 \times (0,2)^2 = 0,096. \end{aligned}$$

Từ đây suy ra

$$P(A_1|A) = \frac{P(A_1)P(A|A_1)}{P(A)} = 0,5026.$$

Ví dụ 1.38. Người ta dùng một thiết bị để kiểm tra một loại sản phẩm nhằm xác định sản phẩm có đạt yêu cầu không. Biết rằng sản phẩm có tỉ lệ phế phẩm là 0,01. Thiết bị có khả năng phát hiện đúng sản phẩm là phế phẩm với xác suất 0,85 và phát hiện đúng sản phẩm đạt chất lượng với xác suất 0,9. Kiểm tra ngẫu nhiên một sản phẩm, tìm xác suất sao cho sản phẩm này:

(a) Được kết luận là phế phẩm.

(b) Được kết luận là đạt chất lượng thì lại là phế phẩm.

(c) Được kết luận đúng với thực chất của nó.

Lời giải Ví dụ 1.38 Gọi A là sự kiện "sản phẩm được chọn là phế phẩm", $P(A) = 0,01$, $P(\overline{A}) = 0,99$.

(a) Gọi H là sự kiện "sản phẩm được kết luận là phế phẩm", khi đó $\overline{H}$ là sự kiện "sản phẩm được kết luận là đạt chất lượng". Theo đầu bài, $P(H|A) = 0,85$, $P(\overline{H}|\overline{A}) = 0,9$. Suy ra

$$P(H) = P(A)P(H|A) + P(\overline{A})P(H|\overline{A}) = 0,01 \times 0,85 + 0,99 \times 0,1 = 0,1075.$$

(b) $P(\overline{H}) = 1 - 0,1075 = 0,8925$. Suy ra

$$P(A|\overline{H}) = \frac{P(A\overline{H})}{P(\overline{H})} = \frac{P(A)P(\overline{H}|A)}{P(\overline{H})} = \frac{0,01 \times 0,15}{0,8925} = 0,0017.$$

(c) $P(AH) + P(\overline{A}\,\overline{H}) = P(A)P(H|A) + P(\overline{A})P(\overline{H}|\overline{A}) = 0,01 \times 0,85 + 0,99 \times 0,9 = 0,8995.$

Ví dụ 1.39. Một hãng hàng không cho biết rằng 5% số khách đặt trước vé cho các chuyến đã định sẽ hoãn không đi chuyến bay đó. Do đó hãng đã đưa ra một chính sách là sẽ bán 52 ghế cho một chuyến bay mà trong đó mỗi chuyến chỉ trở được 50 khách hàng. Tìm xác suất để tất cả các khách đặt chỗ trước và không hoãn chuyến bay đều có ghế. Biết rằng xác suất bán được 51 vé hoặc 52 vé là như nhau và bằng 10%.

Lời giải Ví dụ 1.39 Gọi A là sự kiện "bán được 52 vé", B là sự kiện "bán được 51 vé", C là sự kiện "bán được $\leq$ 50 vé". Khi đó A, B, C tạo thành một nhóm đầy đủ, $P(A) = P(B) = 0,1$ và $P(C) = 0,8$.

Gọi H là sự kiện "tất cả các khách hàng đặt chỗ trước và không hoãn chuyến bay đều đủ chỗ", suy ra $\overline{H}$ là sự kiện "khách hàng không đủ chỗ". Khi đó

$$P(\overline{H}) = P(A)P(\overline{H}|A) + P(B)P(\overline{H}|B) + P(C)P(\overline{H}|C),$$

trong đó

$$\begin{aligned}
P(\overline{H}|A) &= P_{52}(0) + P_{52}(1) = (0,95)^{52} + 52 \times (0,95)^{51} \times (0,05)^{1},\\
P(\overline{H}|B) &= P_{51}(0) = (0,95)^{51},\\
P(\overline{H}|C) &= 0.
\end{aligned}$$

Từ đó $P(\overline{H}) = 0,0333$, suy ra $P(H) = 0,6667$.

Ví dụ 1.40. Ba người thợ cùng may một loại áo với xác suất may được sản phẩm chất lượng cao tương ứng là 0,9; 0,9 và 0,8. Biết một người khi may 8 áo thì có 6 sản phẩm chất lượng cao. Tìm xác suất để người đó may 8 áo nữa thì có 6 áo chất lượng cao.

Lời giải Ví dụ 1.40 Gọi A là sự kiện "trong 8 áo đầu tiên có 6 áo chất lượng cao"; A_i là sự kiện "8 áo đầu tiên do người thợ thứ i may", $i = 1,2,3$ với $P(A_i) = \dfrac{1}{3}$, $i = 1,2,3$. Theo công thức xác suất đầy đủ

$$\begin{aligned}P(A) &= P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + P(A_3)P(A|A_3)\\ &= \frac{1}{3} \times \left[C_8^6 \times (0,9)^6 \times (0,1)^2 + C_8^6 \times (0,9)^6 \times (0,1)^2 + C_8^6 \times (0,8)^6 \times (0,2)^2\right] \simeq 0,2.\end{aligned}$$

Gọi B là sự kiện "trong 8 áo sau có 6 áo chất lượng cao".

$$P(B) = \sum_{i=1}^{3} P(A_i|A)P(B|A_iA) = 0,225,$$

trong đó các xác suất $P(A_1|A)$, $P(A_2|A)$, $P(A_3|A)$ được xác định theo công thức Bay-et.

1.7 Tổng hợp một số đề thi

Ví dụ 1.41 (Đề thi MI2020 kỳ 20151)**.** Ra khỏi phòng khách, 6 người cùng xỏ ngẫu nhiên vào một đôi giày trong bóng tối. Mỗi người chỉ có thể phân biệt chiếc giày phải với chiếc giày trái, còn không thể phân biệt được giày của mình với giày của người khác. Tính xác suất để

(a) Mỗi người khách xỏ vào đúng đôi giày của mình.

(b) Mỗi người khách xỏ vào đúng hai chiếc giày của cùng một đôi nào đó.

Lời giải Ví dụ 1.41

(a) Gọi A là sự kiện "cả 6 người khách đều xỏ đúng đôi giày của mình"; A_i là sự kiện "người thứ i xỏ đúng đôi giày của mình", $i = 1,2,\ldots,6$. Khi đó $A = A_1A_2A_3A_4A_5A_6$ và

$$P(A) = P(A_1)P(A_2|A_1)\ldots P(A_6|A_1A_2A_3A_4A_5) = \frac{1}{6^2} \times \frac{1}{5^2} \times \cdots \times \frac{1}{1^2} = \frac{1}{(6!)^2}.$$

(b) Gọi B là sự kiện "mỗi người khách đều xỏ đúng 2 chiếc giày của cùng một đôi"; B_i là sự kiện "người thứ i xỏ đúng 2 chiếc giày của cùng một đôi", $i = 1,2,\ldots,6$. Khi đó $B = B_1B_2B_3B_4B_5B_6$ và

$$P(B) = P(B_1)P(B_2|B_1)\ldots P(B_6|B_1B_2B_3B_4B_5) = \frac{1}{6} \times \frac{1}{5} \times \cdots \times \frac{1}{1} = \frac{1}{6!}.$$

Ví dụ 1.42 (Đề thi MI2020 kỳ 20161)**.** Biết từ vị trí A đến B có hai đường đi với xác suất bị ngập của mỗi con đường là p; từ B đến C cũng có hai đường đi với xác suất bị ngập của mỗi con đường cũng là p. Biết đường đi từ A đến C bị ngập, tính xác suất để đường đi từ A đến B không bị ngập.

Lời giải Ví dụ 1.42 Gọi E_{AB} là sự kiện "đường đi từ A đến B không ngập", khi đó $\overline{E}_{AB}$ là sự kiện "đường đi từ A đến B bị ngập". Xác suất cần tìm là

$$P(E_{AB}|\overline{E}_{AC}) = \frac{P[(E_{AB})(\overline{E}_{AC})]}{P(\overline{E}_{AC})} = \frac{P[(E_{AB})(\overline{E}_{BC})]}{P(\overline{E}_{AC})} = \frac{P(E_{AB})P(\overline{E}_{BC})}{P(\overline{E}_{AC})}.$$

Đường đi từ B đến C bị ngập nếu cả hai đường đi đều bị ngập, do đó xác suất để đường đi từ B đến C bị ngập là $P(\overline{E}_{BC}) = p^2$ và xác suất để đường đi từ A đến B không ngập là $P(E_{AB}) = 1 - p^2$.

Đường đi từ A đến C không ngập nếu đường đi từ A đến B không ngập và đường đi từ B đến C cũng không ngập, nên xác suất để đường đi từ A đến C bị ngập là $P(\overline{E}_{AC}) = 1 - (1 - p^2)^2$.

Vậy

$$P(E_{AB}|\overline{E}_{AC}) = \frac{(1 - p^2)p^2}{1 - (1 - p^2)^2}.$$

Ví dụ 1.43 (Đề thi MI2020 kỳ 20171)**.** Một phân xưởng có hai máy sản xuất cùng một loại sản phẩm với tỷ lệ phế phẩm của các máy tương ứng là 0,2% và 0,5%. Từ kho chung chứa 10 sản phẩm của máy I và 8 sản phẩm của máy II chọn ngẫu nhiên ra 2 sản phẩm.

(a) Tính xác suất để trong 2 sản phẩm được chọn có đúng 1 phế phẩm.

(b) Biết trong 2 sản phẩm được chọn có đúng 1 phế phẩm, tính xác suất để 2 sản phẩm đó do máy II sản xuất.

Lời giải Ví dụ 1.43

(a) Gọi A_1, A_2, A_3 là các sự kiện "2 sản phẩm lấy ra do máy I, máy II, 1 sản phẩm của máy I và 1 sản phẩm của máy II sản xuất". H là sự kiện "trong 2 sản phẩm được chọn có đúng 1 phế phẩm".

$$P(H) = P(A_1)P(H|A_1) + P(A_2)P(H|A_2) + P(A_3)P(H|A_3)$$

trong đó $P(A_1) = \dfrac{C_{10}^2}{C_{18}^2}, P(A_2) = \dfrac{C_8^2}{C_{18}^2}, P(A_3) = \dfrac{C_{10}^1 C_8^1}{C_{18}^2}$;

$P(H|A_1) = C_2^1(0,002)(0,998), P(H|A_2) = C_2^1(0,005)(0,995), P(H|A_3) = (0,002)(0,995) + (0,005)(0,998)$.

Từ đây suy ra $P(H)$.

(b) Cần tính $P(A_2|H) = \dfrac{P(A_2)P(H|A_2)}{P(H)} \simeq 0,274$.

Ví dụ 1.44 (Đề thi MI2020 kỳ 20173)**.** Một lô hàng có 15 sản phẩm gồm 6 sản phẩm loại A, 5 sản phẩm loại B và 4 sản phẩm loại C. Chọn ngẫu nhiên ra 4 sản phẩm.

(a) Tính xác suất để trong 4 sản phẩm được chọn có đúng 2 sản phẩm loại B.

(b) Biết trong 4 sản phẩm được chọn có đúng 2 sản phẩm loại A, tính xác suất để trong 4 sản phẩm đó có đúng 1 sản phẩm loại C.

Lời giải Ví dụ 1.44

(a) Gọi D là sự kiện"trong 4 sản phẩm được chọn có đúng 2 sản phẩm loại B". $P(D) = \dfrac{C_5^2 C_{10}^2}{C_{15}^4} \simeq 0,3297$.

(b) Gọi H: "trong 4 sản phẩm được chọn có đúng 2 sản phẩm loại A", E: "trong 4 sản phẩm đó có đúng 1 sản phẩm loại C". Cần tính $P(E|H) = \dfrac{P(EH)}{P(H)}$. Trong đó

$P(H) = \dfrac{C_6^2 C_9^2}{C_{15}^4} \simeq 0,3956$ và $P(EH) = \dfrac{C_6^2 C_4^1 C_5^1}{C_{15}^4} \simeq 0,2918$.

Vậy $P(E|H) = \dfrac{P(EH)}{P(H)} \simeq 0,5556$.

Ví dụ 1.45 (Đề thi MI2020 kỳ 20182)**.** Cho ba sự kiện A, B, C độc lập từng đôi thỏa mãn $P(A) = P(B) = P(C) = p$ và $P(ABC) = 0$.

(a) Tính $P(AB\overline{C})$; $P(A\overline{B}\,\overline{C})$; $P(\overline{A}\,\overline{B}\,\overline{C})$.

(b) Tìm giá trị p lớn nhất có thể có.

Lời giải Ví dụ 1.45

(a) Vì $AB\overline{C} + ABC = AB$; $AB\overline{C}$ và ABC xung khắc; A và B độc lập, nên

$$P(AB\overline{C}) = P(AB) - P(ABC) = P(A)P(B) - 0 = p^2.$$

Vì $A = A\overline{B}\,\overline{C} + ABC + A\overline{B}C + AB\overline{C}$, sử dụng tính xung khắc của các sự kiện,

$$P(A\overline{B}\,\overline{C}) = P(A) - P(ABC) - P(A\overline{B}C) - P(AB\overline{C}) = p - 2p^2.$$

Vì $\overline{A}\,\overline{B}\,\overline{C} + A\overline{B}\,\overline{C} = \overline{B}\,\overline{C}$ nên $P(\overline{A}\,\overline{B}\,\overline{C}) = P(\overline{B}\,\overline{C}) - P(A\overline{B}\,\overline{C}) = 1 - 3p + 3p^2$.

(b) Từ ý (a) và đầu bài ta có $P(ABC) = 0$, $P(AB\overline{C}) = P(A\overline{B}C) = P(\overline{A}BC) = p^2$, $P(A\overline{B}\,\overline{C}) = P(\overline{A}B\overline{C}) = P(\overline{A}\,\overline{B}C) = p - 2p^2$, $P(\overline{A}\,\overline{B}\,\overline{C}) = 1 - 3p + 3p^2$. Khi đó p thỏa mãn hệ

$$\begin{cases} 0 \le p^2 \le 1, \\ 0 \le p - 2p^2 \le 1, \\ 0 \le 1 - 3p + 3p^2 \le 1. \end{cases}$$

Hệ này tương đương với $0 \le p \le 0,5$. Vậy giá trị p lớn nhất là $0,5$.

Ví dụ 1.46 (Đề thi MI2020 kỳ 20183)**.** Có một nhóm 4 sinh viên, mỗi người có một chiếc mũ giống hệt nhau để trên giá. Khi ra khỏi phòng, mỗi người lấy ngẫu nhiên một chiếc mũ để đội. Tính xác suất để:

(a) Sinh viên thứ nhất và sinh viên thứ hai lấy đúng mũ của mình.

(b) Có ít nhất một sinh viên lấy đúng mũ của mình.

Lời giải Ví dụ 1.46 Gọi A là sự kiện "có ít nhất một sinh viên lấy đúng mũ của mình"; A_i là sự kiện "sinh viên thứ i lấy đúng mũ của mình", $i = 1, 2, 3, 4$.

(a) $P(A_1A_2) = P(A_1)P(A_2|A_1) = \dfrac{1}{12} \simeq 0,0833$.

(b)

$$\begin{aligned}
P(A) &= P(A_1 + A_2 + A_3 + A_4) \\
&= \sum_{i=1}^{4} P(A_i) - P(A_1A_2) - P(A_1A_3) - P(A_1A_4) - P(A_2A_3) \\
&\quad - P(A_2A_4) - P(A_3A_4) + P(A_1A_2A_3) + P(A_1A_2A_4) \\
&\quad + P(A_1A_3A_4) + P(A_2A_3A_4) - P(A_1A_2A_3A_4) \\
&= 4 \times \frac{1}{4} - 6 \times \frac{1}{12} + 4 \times \frac{1}{24} - \frac{1}{24} = 0,625.
\end{aligned}$$

Ví dụ 1.47 (Đề thi MI2020 kỳ 20191)**.** Lớp MI2020 có 80 sinh viên trong đó có 20 sinh viên thuộc tổ I, 25 sinh viên thuộc tổ II và 35 sinh viên thuộc tổ III. Chọn ngẫu nhiên 10 sinh viên trong lớp tham dự trại hè. Tính xác suất để mỗi tổ có ít nhất 1 sinh viên được chọn.

Lời giải Ví dụ 1.47 Gọi A là sự kiện "Mỗi tổ có ít nhất 1 sinh viên được chọn", A_i: "tổ i có ít nhất 1 sinh viên được chọn", $i = 1, 2, 3$. Khi đó, $A = A_1A_2A_3$ và

$$P(A) = P(A_1A_2A_3) = 1 - P(\overline{A}_1 + \overline{A}_2 + \overline{A}_3).$$

Sử dụng công thức (1.16),

$$\begin{aligned}
P(\overline{A}_1 + \overline{A}_2 + \overline{A}_3) &= P(\overline{A}_1) + P(\overline{A}_2) + P(\overline{A}_3) - P(\overline{A}_1\overline{A}_2) - P(\overline{A}_1\overline{A}_3) - P(\overline{A}_2\overline{A}_2) + P(\overline{A}_1\overline{A}_2\overline{A}_3) \\
&= \frac{1}{C_{80}^{10}}\left[(C_{60}^{10} + C_{55}^{10} + C_{45}^{10}) - (C_{35}^{10} + C_{25}^{10} + C_{20}^{10}) + 0\right] \\
&\simeq 1 - 0,06538 = 0,93462.
\end{aligned}$$

Ví dụ 1.48 (Đề thi MI2020 kỳ 20192)**.** Cho biết xác suất để một sinh viên mượn một cuốn sách Kỹ thuật ở thư viện là 0,8; còn xác suất mượn một cuốn sách Văn học là 0,2. Một ngày có 5 sinh viên đến mượn sách tại thư viện, mỗi người mượn 2 cuốn sách.

(a) Tính xác suất để trong 5 người đó có đúng 2 người, mỗi người mượn một cuốn sách Kỹ thuật và một cuốn sách Văn học.

(b) Biết trong 5 người có ít nhất 2 người, mỗi người mượn 2 cuốn sách Kỹ thuật; tính xác suất để trong 5 người đó có đúng 2 người, mỗi người mượn 1 cuốn sách Kỹ thuật và 1 cuốn sách Văn học.

Lời giải Ví dụ 1.48 (a) Xác suất để trong hai người có 1 người mượn 1 sách kỹ thuật, 1 người mượn sách văn học là $p = C_2^1(0,8)(0,2) = 0,32$.

Gọi B : "đúng 2 người, mỗi người mượn 1 sách kỹ thuật, 1 người mượn sách văn học". $P(B) = C_5^2(0,32)^2(0,68)^3 \simeq 0,3220$.

(b) Gọi H : "ít nhất 2 người, mỗi người mượn 2 sách kỹ thuật", A : "đúng 2 người, mỗi người mượn 1 sách ký thuật, 1 người mượn sách văn học". Ta có $P(H) = 1 - \sum_{k=0}^{1} C_5^k(0,64)^k(0,36)^{5-k} \simeq 0,9402$.

Xác suất cần tìm là $P(A|H) = \frac{P(AH)}{P(H)} = \frac{0,3188}{0,9402} \simeq 0,3391$.

Ví dụ 1.49 (Đề thi cuối kỳ)**.** Giả sử đặt ngẫu nhiên n bức thư vào n chiếc phong bì. Tính xác suất để không có bức thư nào đặt đúng phong bì.

Lời giải Ví dụ Ví dụ 1.49 Gọi A là sự kiện "không có bức thư nào đặt đúng phong bì", A_i là sự kiện "bức thư thứ i đặt đúng phong bì". Khi đó $P(A) = 1 - P(A_1 + A_2 + \cdots + A_n)$, trong đó

$$P(A_1 + A_2 + \cdots + A_2) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} A_iA_jA_k + \cdots + (-1)^{n-1}P(A_1A_2\ldots A_n).$$

Tính

$$\begin{aligned}
&P(A_1) = \frac{1 \times (n-1)!}{n!}, \quad \sum_{i=1}^{n} P(A_i) = 1,\\
&P(A_1A_2) = \frac{1 \times (n-2)!}{n!}, \quad \sum_{i<j} P(A_iA_j) = \frac{1}{2!},\\
&P(A_1A_2A_3) = \frac{1 \times (n-3)!}{n!}, \quad \sum_{i<j<k} A_iA_jA_k \sum_{i<j<k} A_iA_jA_k = \frac{1}{3!},\\
&\ldots\\
&P(A_1A_2\ldots A_n) = \frac{1}{n!}.
\end{aligned}$$

Vậy

$$P(A) = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-1}\frac{1}{n!}.$$

Bài tập Chương 1

Bài tập 1.1. Một hộp có 10 quả cầu cùng kích cỡ được đánh số từ 0 đến 9. Từ hộp người ta lấy ngẫu nhiên 1 quả ra và ghi lại số của quả đó, sau đó trả lại vào trong hộp. Làm như vậy 5 lần ta thu được một dãy số có 5 chữ số.

(a) Có bao nhiêu kết quả cho dãy số đó?

(b) Có bao nhiêu kết quả cho dãy số đó sao cho các chữ số trong đó là khác nhau?

Bài tập 1.2. Có 6 bạn Hoa, Trang, Vân, Anh, Thái, Trung ngồi quanh một bàn tròn để uống cà phê, trong đó bạn Trang và Vân không ngồi cạnh nhau.

(a) Có bao nhiêu cách xếp 6 bạn này trên bàn tròn nếu tất cả các ghế là không phân biệt?

(b) Có bao nhiêu cách xếp 6 bạn này trên bàn tròn nếu tất cả các ghế có phân biệt?

Bài tập 1.3. Từ một bộ bài tú lơ khơ 52 cây rút ngẫu nhiên và không quan tâm đến thứ tự 4 cây. Có bao nhiêu khả năng xảy ra trường hợp trong 4 cây đó:

(a) đều là át;

(b) có duy nhất 1 cây át;

(c) có ít nhất 1 cây át;

(d) có đủ 4 loại rô, cơ, bích, nhép.

Bài tập 1.4. Có 20 sinh viên. Có bao nhiêu cách chọn ra 4 sinh viên (không xét tới tính thứ tự) tham gia câu lạc bộ Văn và 4 sinh viên tham gia câu lạc bộ Toán trong trường hợp:

(a) một sinh viên chỉ tham gia nhiều nhất một câu lạc bộ;

(b) một sinh viên có thể tham gia cả hai câu lạc bộ.

Bài tập 1.5. Cho phương trình $x + y + z = 100$. Phương trình đã cho có bao nhiêu nghiệm:

(a) nguyên dương;

(b) nguyên không âm.

Bài tập 1.6. Thực hiện một phép thử tung 2 con xúc xắc, rồi ghi lại số chấm xuất hiện trên mỗi con. Gọi x, y là số chấm xuất hiện tương ứng trên con xúc xắc thứ nhất và thứ hai. Ký hiệu không gian mẫu $\Omega = \{(x, y) : 1 \leq x, y \leq 6\}$. Hãy liệt kê các phần tử của các sự kiện sau:

(a) A : "tổng số chấm xuất hiện lớn hơn 8";

(b) B : "có ít nhất một con xúc xắc ra mặt 2 chấm";

(c) C : "con xúc xắc thứ nhất có số chấm lớn hơn 4";

(d) $A + B$, $A + C$, $B + C$, $A + B + C$, sau đó thể hiện thông qua sơ đồ Venn;

(e) AB, AC, BC, ABC, sau đó thể hiện thông qua sơ đồ Venn.

Bài tập 1.7. Số lượng nhân viên của công ty A được phân loại theo lứa tuổi và giới tính như sau:

Tuổi \ Giới tính	Nam	Nữ
Dưới 30	120	170
Từ 30 − 40	260	420
Trên 40	400	230

Tìm xác suất để lấy ngẫu nhiên một người của công ty thì được:

(a) một nhân viên trong độ tuổi 30 – 40;

(b) một nam nhân viên trên 40 tuổi;

(c) một nữ nhân viên từ 40 tuổi trở xuống.

Bài tập 1.8. Một kiện hàng có 24 sản phẩm, trong số đó có 14 sản phẩm loại I, 8 sản phẩm loại II và 2 sản phẩm loại III. Người ta chọn ngẫu nhiên 4 sản phẩm để kiểm tra. Tính xác suất trong 4 sản phẩm đó:

(a) có 3 sản phẩm loại I và 1 sản phẩm loại II;

(b) có ít nhất 3 sản phẩm loại I;

(c) có ít nhất 1 sản phẩm loại III.

Bài tập 1.9. Có 30 tấm thẻ đánh số từ 1 tới 30. Chọn ngẫu nhiên ra 10 tấm thẻ. Tính xác suất để:

(a) tất cả tấm thẻ đều mang số chẵn;

(b) có đúng 5 số chia hết cho 3;

(c) có 5 tấm thẻ mang số lẻ, 5 tấm thẻ mang số chẵn trong đó chỉ có một số chia hết cho 10.

Bài tập 1.10. Việt Nam có 64 tỉnh thành, mỗi tỉnh thành có 2 đại biểu quốc hội. Người ta chọn ngẫu nhiên 64 đại biểu quốc hội để thành lập một ủy ban. Tính xác suất để:

(a) trong ủy ban có ít nhất một người của thành phố Hà Nội;

(b) mỗi tỉnh có đúng một đại biểu trong ủy ban.

Bài tập 1.11. Một đoàn tàu có 4 toa được đánh số I, II, III, IV đỗ ở sân ga. Có 6 hành khách từ sân ga lên tàu. Mỗi người độc lập với nhau chọn ngẫu nhiên một toa. Tính xác suất để:

(a) toa I có 3 người, toa II có 2 người và toa III có 1 người;

(b) một toa có 3 người, một toa 2 người, một toa có 1 người;

(c) mỗi toa có ít nhất 1 người.

Bài tập 1.12. Gieo hai con xúc xắc cân đối và đồng chất. Một con xúc xắc có số chấm các mặt là 1, 2, 3, 4, 5, 6, con xúc xắc còn lại có số chấm các mặt là 2, 3, 4, 5, 6, 6. Tính xác suất:

(a) có đúng 1 con xúc xắc ra mặt 6 chấm;

(b) có ít nhất 1 con xúc xắc ra mặt 6 chấm;

(c) tổng số chấm xuất hiện bằng 7.

Bài tập 1.13. Trong một thành phố có 5 khách sạn. Có 3 khách du lịch đến thành phố đó, mỗi người chọn ngẫu nhiên một khách sạn. Tìm xác suất để:

(a) mỗi người ở một khách sạn khác nhau;

(b) có đúng 2 người ở cùng một khách sạn.

Bài tập 1.14. Một lớp có 3 tổ sinh viên: tổ I có 12 người, tổ II có 10 người và tổ III có 15 người. Chọn hú họa ra một nhóm sinh viên gồm 4 người.

(a) Tính xác suất để trong nhóm có đúng một sinh viên tổ I.

(b) Biết trong nhóm có đúng một sinh viên tổ I, tính xác suất để trong nhóm đó có đúng một sinh viên tổ III.

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

(a) chị A đánh vỡ 3 chén và chị B đánh vỡ 1 chén;

(b) một trong ba người đánh vỡ 3 chén;

(c) một trong ba người đánh vỡ cả 4 chén.

Bài tập 1.16. Đội A có 3 người và đội B có 3 người tham gia vào một cuộc chạy thi, 6 người có khả năng như nhau và xuất phát cùng nhau. Tính xác suất để 3 người đội A về vị trí nhất, nhì, ba.

Bài tập 1.17. Phân phối ngẫu nhiên n viên bi vào n chiếc hộp (biết rằng mỗi hộp có thể chứa cả n viên bi). Tính xác suất để:

(a) Hộp nào cũng có bi;

(b) Có đúng một hộp không có bi.

Bài tập 1.18. Hai người hẹn gặp nhau ở công viên trong khoảng thời gian từ 5h00 đến 6h00 để cùng đi tập thể dục. Hai người quy ước ai đến không thấy người kia sẽ chỉ chờ trong vòng 10 phút. Giả sử rằng thời điểm hai người đến công viên là ngẫu nhiên trong khoảng từ 5h00 đến 6h00. Tính xác suất để hai người gặp nhau.

Bài tập 1.19. Cho đoạn thẳng AB có độ dài 10cm. Lấy một điểm C bất kỳ trên đoạn thẳng đó. Tính xác suất chênh lệch độ dài giữa hai đoạn thẳng AC và CB không vượt quá 4cm.

Bài tập 1.20. Cho đoạn thẳng AB độ dài 10cm. Lấy hai điểm C, D bất kỳ trên đoạn AB (C nằm giữa A và D). Tính xác suất độ dài AC, CD, DB tạo thành 3 cạnh một tam giác.

Bài tập 1.21. Cho các sự kiện A, B với $P(A) = P(B) = 1/2$; $P(A\overline{B}) = 1/8$. Tìm:

(a) $P(\overline{A} + \overline{B})$;

(b) $P(\overline{A}B)$, $P(A + \overline{B})$.

Bài tập 1.22. Cho ba sự kiện A, B, C độc lập từng đôi thỏa mãn $P(A) = P(B) = P(C) = p$ và $P(ABC) = 0$.

(a) Tính $P(AB\overline{C})$; $P(A\overline{B}\,\overline{C})$; $P(\overline{A}\,\overline{B}\,\overline{C})$.

(b) Tìm giá trị p lớn nhất có thể có.

Bài tập 1.23. Trong cùng một phép thử, A và B là các sự kiện thỏa mãn $P(A) = 1/4$, $P(B) = 1/2$. Tính xác suất để A không xảy ra nhưng B xảy ra trong các trường hợp sau:

(a) A và B xung khắc;

(b) A suy ra B;

(c) $P(AB) = 1/8$.

Bài tập 1.24. Cho hai sự kiện A và B trong đó $P(A) = 0,4$ và $P(B) = 0,7$. Xác định giá trị lớn nhất và nhỏ nhất của $P(AB)$ và $P(A + B)$ và điều kiện đạt được các giá trị đó.

Bài tập 1.25. Ba người A, B và C lần lượt tung một đồng xu. Giả sử rằng A tung đồng xu đầu tiên, B tung thứ hai và thứ ba C tung. Quá trình lặp đi lặp lại cho đến khi ai thắng bằng việc trở thành người đầu tiên thu được mặt ngửa. Xác định khả năng mà mỗi người sẽ giành chiến thắng.

Bài tập 1.26. Trong một thùng kín có 6 quả cầu đỏ, 5 quả cầu trắng, 4 quả cầu vàng. Lấy ngẫu nhiên lần lượt từng quả cầu cho đến khi lấy được cầu đỏ thì dừng lại. Tính xác suất để:

(a) Lấy được 2 cầu trắng, 1 cầu vàng.

(b) Không có quả cầu trắng nào được lấy ra.

Bài tập 1.27. Ba xạ thủ A, B, C độc lập với nhau cùng bắn súng vào bia. Xác suất bắn trúng bia của 3 người A, B và C tương ứng là 0,7, 0,6 và 0,9. Tính xác suất để:

(a) có duy nhất một xạ thủ bắn trúng bia;

(b) có đúng hai xạ thủ bắn trúng bia;

(c) có ít nhất một xạ thủ bắn trúng bia;

(d) xạ thủ A bắn trúng bia biết rằng có hai xạ thủ bắn trúng bia.

Bài tập 1.28. Trên một bảng quảng cáo, người ta mắc hai hệ thống bóng đèn độc lập. Hệ thống I gồm 4 bóng mắc nối tiếp, hệ thống II gồm 3 bóng mắc song song. Khả năng bị hỏng của mỗi bóng trong 18 giờ thắp sáng liên tục là 0,1. Việc hỏng của mỗi bóng của mỗi hệ thống được xem như độc lập. Tính xác suất để trong 18 giờ thắp sáng liên tục:

(a) cả hai hệ thống bị hỏng;

(b) chỉ có một hệ thống bị hỏng.

Bài tập 1.29. Có 6 khẩu súng cũ và 4 khẩu súng mới, trong đó xác suất trúng khi bắn bằng súng cũ là 0,8, còn súng mới là 0,95. Bắn hú họa bằng một khẩu súng vào một mục tiêu thì thấy trúng. Điều gì có khả năng xảy ra lớn hơn: bắn bằng khẩu súng mới hay bắn bằng khẩu súng cũ?

Bài tập 1.30. Theo thống kê xác suất để hai ngày liên tiếp có mưa ở một thành phố vào mùa hè là 0,5; còn không mưa là 0,3. Biết các sự kiện có một ngày mưa, một ngày không mưa là đồng khả năng. Tính xác suất để ngày thứ hai có mưa, biết ngày đầu không mưa.

Bài tập 1.31. Một hộp chứa a quả bóng màu đỏ và b quả bóng màu xanh. Một quả bóng được chọn ngẫu nhiên và quan sát màu sắc của nó. Sau đó bóng được trả lại cho vào hộp và k bóng cùng màu cũng được thêm vào hộp. Một quả bóng thứ hai sau đó được chọn một cách ngẫu nhiên, màu sắc của nó được quan sát, và nó được trả lại cho vào hộp với k bóng bổ sung cùng một màu. Quá trình này được lặp đi lặp lại 4 lần. Tính xác suất để ba quả bóng đầu tiên sẽ có màu đỏ và quả bóng thứ tư có màu xanh.

Bài tập 1.32. Một cửa hàng sách ước lượng rằng: trong tổng số các khách hàng đến cửa hàng có 30% khách cần hỏi nhân viên bán hàng, 20% khách mua sách và 15% khách thực hiện cả hai điều trên. Gặp ngẫu nhiên một khách trong nhà sách. Tính xác suất để người này:

(a) không thực hiện cả hai điều trên;

(b) không mua sách, biết rằng người này đã hỏi nhân viên bán hàng.

Bài tập 1.33. Một cuộc khảo sát 1000 người về hoạt động thể dục thấy có 80% số người thích đi bộ và 60% thích đạp xe vào buổi sáng và tất cả mọi người đều tham gia ít nhất một trong hai hoạt động trên. Chọn ngẫu nhiên một người hoạt động thể dục. Nếu gặp được người thích đi xe đạp thì xác suất mà người đó không thích đi bộ là bao nhiêu?

Bài tập 1.34. Để thành lập đội tuyển quốc gia về một môn học, người ta tổ chức một cuộc thi tuyển gồm 3 vòng. Vòng thứ nhất lấy 80% thí sinh; vòng thứ hai lấy 70% thí sinh đã qua vòng thứ nhất và vòng thứ ba lấy 45% thí sinh đã qua vòng thứ hai. Để vào được đội tuyển, thí sinh phải vượt qua được cả 3 vòng thi. Tính xác suất để một thí sinh bất kỳ:

(a) được vào đội tuyển;

(b) bị loại ở vòng thứ ba;

(c) bị loại ở vòng thứ hai, biết rằng thí sinh này bị loại.

Bài tập 1.35. Theo thống kê ở các gia đình có hai con thì xác suất để con thứ nhất và con thứ hai đều là trai là 0,27 và hai con đều là gái là 0,23, còn xác suất con thứ nhất và con thứ hai có một trai và một gái là đồng khả năng. Biết sự kiện khi xét một gia đình được chọn ngẫu nhiên có con thứ nhất là gái, tìm xác suất để con thứ hai là trai.

Bài tập 1.36. Một tổ có 15 sinh viên trong đó có 5 sinh viên học giỏi môn "Xác suất thống kê". Cần chia làm 5 nhóm, mỗi nhóm 3 sinh viên. Tính xác suất để nhóm nào cũng có một sinh viên học giỏi môn "Xác suất thống kê".

Bài tập 1.37. Một hộp có n áo trắng và $2n$ áo xanh. Chia ngẫu nhiên các áo trong hộp thành n nhóm mỗi nhóm 3 áo.

(a) Tính xác suất để trong mỗi nhóm đều có áo trắng;

(b) Áp dụng cho $n = 5$.

Bài tập 1.38. Hai vận động viên bóng bàn A và B đấu một trận gồm tối đa 5 ván (không có kết quả hòa sau mỗi ván và trận đấu sẽ dừng nếu một người nào đó thắng trước 3 ván). Xác suất để A thắng được ở một ván là 0,7.

(a) Tính các xác suất để A thắng sau x ván ($x = 3, 4, 5$).

(b) Tính xác suất để trận đấu kết thúc sau 5 ván.

Bài tập 1.39. Một bài thi trắc nghiệm (multiple-choice test) gồm 12 câu hỏi, mỗi câu hỏi cho 5 phương án trả lời, trong đó chỉ có 1 phương án đúng. Giả sử một câu trả lời đúng được 4 điểm và mỗi câu trả lời sai bị trừ đi 1 điểm. Một học sinh kém làm bài bằng cách chọn hú họa câu trả lời. Tìm xác suất để:

(a) Học sinh đó được 13 điểm.

(b) Học sinh đó bị điểm âm.

Bài tập 1.40. Một nhân viên bán hàng mỗi ngày đi chào hàng ở 10 nơi với xác suất bán được hàng ở mỗi nơi là 0,2. Tìm xác suất để:

(a) người đó bán được hàng ở 2 nơi;

(b) người đó bán được hàng ở ít nhất 1 nơi.

Bài tập 1.41. Xác suất trúng đích của một lần bắn là 0,4. Cần phải bắn bao nhiêu phát đạn để xác suất có ít nhất một viên bắn trúng sẽ lớn hơn 0,95?

Bài tập 1.42. Hai cầu thủ bóng rổ, mỗi người ném bóng 2 lần vào rổ. Xác suất ném trúng rổ của mỗi cầu thủ theo thứ tự lần lượt là 0,6 và 0,7. Tìm xác suất để

(a) số lần ném trúng rổ của hai người bằng nhau;

(b) số lần ném trúng rổ của cầu thủ thứ nhất nhiều hơn số lần ném trúng rổ của cầu thủ thứ hai.

Bài tập 1.43. Xác suất sản xuất ra phế phẩm của một máy là 0,005. Tìm xác suất để trong 800 sản phẩm của máy đó có đúng 3 phế phẩm.

Bài tập 1.44. Một công nhân đứng máy 1000 ống sợi. Xác suất mỗi ống bị đứt trong vòng một giờ là 0,005. Tính xác suất để trong vòng một giờ:

(a) 40 ống sợi bị đứt;

(b) không quá 40 ống sợi bị đứt.

Bài tập 1.45. Xác suất ném trúng rổ của một cầu thủ là 0,8. Tìm xác suất để trong 100 lần cầu thủ đó:

(a) ném trúng 75 lần;

(b) ném trúng không ít hơn 75 lần.

Bài tập 1.46. Một phân xưởng có 3 máy tự động: máy I sản xuất 25%, máy II sản xuất 30%, máy III sản xuất 45% số sản phẩm. Tỷ lệ phế phẩm tương ứng của các máy lần lượt là 0,1%, 0,2% và 0,3%. Chọn ngẫu nhiên ra một sản phẩm của phân xưởng.

(a) Tìm xác suất nó là phế phẩm.

(b) Biết nó là phế phẩm. Tính xác suất để sản phẩm đó do máy I sản xuất.

Bài tập 1.47. Có 3 hộp đựng bi: hộp thứ nhất có 3 bi đỏ, 2 bi trắng; hộp thứ hai có 2 bi đỏ, 2 bi trắng; hộp thứ ba không có viên nào. Lấy ngẫu nhiên 1 viên bi từ hộp thứ nhất và 1 viên bi từ hộp thứ hai bỏ vào hộp thứ ba. Sau đó từ hộp thứ ba lấy ngẫu nhiên ra 1 viên bi.

(a) Tính xác suất để viên bi đó màu đỏ.

(b) Biết rằng viên bi lấy ra từ hộp thứ ba màu đỏ, tính xác suất để lúc đầu ta lấy được viên bi đỏ từ hộp thứ nhất bỏ vào hộp thứ ba.

Bài tập 1.48. Hộp I có 4 viên bi đỏ, 2 viên bi xanh; hộp II có 3 viên bi đỏ, 3 viên bi xanh. Bỏ ngẫu nhiên một viên bi từ hộp I sang hộp II, sau đó lại bỏ ngẫu nhiên một viên bi từ hộp II sang hộp I. Cuối cùng rút ngẫu nhiên từ hộp I ra một viên bi.

(a) Tính xác suất để viên bi rút ra sau cùng màu đỏ.

(b) Nếu viên rút ra sau cùng màu đỏ, tìm xác suất lúc ban đầu rút được viên bi đỏ ở hộp I cho vào hộp II.

Bài tập 1.49. Trong một kho rượu, số lượng rượu loại A và loại B bằng nhau. Người ta chọn ngẫu nhiên một chai và đưa cho 5 người nếm thử. Biết xác suất đoán đúng của mỗi người là 0,8. Có 3 người kết luận rượu loại A, 2 người kết luận rượu loại B. Hỏi khi đó xác suất chai rượu đó thuộc loại A là bao nhiêu?

Bài tập 1.50. Có hai lô sản phẩm: lô I có 7 chính phẩm 3 phế phẩm; lô II có 6 chính phẩm 2 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm từ lô I sang lô II, sau đó từ lô II lấy ngẫu nhiên ra 2 sản phẩm được 2 chính phẩm. Tính xác suất để 2 chính phẩm lấy ra sau cùng là của lô I.

Bài tập 1.51. Có hai lô sản phẩm: lô I có 7 chính phẩm, 3 phế phẩm; lô II có 8 chính phẩm, 2 phế phẩm. Từ lô I lấy ngẫu nhiên ra 2 sản phẩm, từ lô II lấy ngẫu nhiên ra 3 sản phẩm. Sau đó từ số sản phẩm này lại lấy ngẫu nhiên 2 sản phẩm. Tính xác suất để trong 2 sản phẩm lấy ra sau cùng có ít nhất 1 chính phẩm.

Bài tập 1.52. Có ba kiện hàng (mỗi kiện hàng có 20 sản phẩm) với số sản phẩm tốt tương ứng của mỗi kiện là 18, 16, 12. Lấy ngẫu nhiên một kiện hàng, rồi từ đó lấy ngẫu nhiên một sản phẩm thì được sản phẩm tốt. Trả sản phẩm này lại kiện hàng vừa lấy, sau đó lại lấy ngẫu nhiên một sản phẩm thì được sản phẩm tốt. Tính xác suất để các sản phẩm tốt đó được lấy từ kiện hàng thứ nhất.

Bài tập 1.53. Tỷ lệ người nghiện thuốc là ở một vùng là 30%. Biết rằng tỷ lệ người bị viêm họng trong số những người nghiện thuốc là 60%, còn tỷ lệ người bị viêm họng trong số những người không nghiện là 40%.

(a) Lấy ngẫu nhiên một người thấy người ấy bị viêm họng. Tính xác suất người đó nghiện thuốc lá.

(b) Nếu người đó không bị viêm họng, tính xác suất người đó nghiện thuốc lá.

Bài tập 1.54. Một công nhân đi làm ở thành phố khi trở về nhà có 2 cách: hoặc đi theo đường ngầm hoặc đi qua cầu. Biết rằng ông ta đi lối đường ngầm trong $1/3$ các trường hợp, còn lại đi lối cầu. Nếu đi lối đường ngầm 75% trường hợp ông ta về đến nhà trước 6 giờ tối; còn nếu đi lối cầu chỉ có 70% trường hợp (nhưng đi lối cầu thích hơn). Tìm xác suất để công nhân đó đã đi lối cầu biết rằng ông ta về đến nhà sau 6 giờ tối.

Bài tập 1.55. Tại một phòng khám chuyên khoa tỷ lệ người đến khám có bệnh là 0,8. Người ta áp dụng phương pháp chẩn đoán mới thì thấy nếu khẳng định có bệnh thì đúng 9 trên 10 trường hợp; còn nếu khẳng định không bệnh thì đúng 5 trên 10 trường hợp. Tính xác suất để

(a) chẩn đoán có bệnh;

(b) chẩn đoán đúng.

Bài tập 1.56. Một hãng hàng không cho biết rằng 5% số khách đặt trước vé cho các chuyến đã định sẽ hoãn không đi chuyến bay đó. Do đó hãng đã đưa ra một chính sách là sẽ bán 52 ghế cho một chuyến bay mà trong đó mỗi chuyến chỉ trở được 50 khách hàng. Tìm xác suất để tất cả các khách đặt chỗ trước và không hoãn chuyến bay đều có ghế. Biết rằng xác suất bán được 51 vé hoặc 52 vé là như nhau và bằng 10%.

Bài tập 1.57. Một trạm chỉ phát hai loại tín hiệu A và B với xác suất tương ứng là 0,84 và 0,16. Do có nhiễu trên đường truyền nên $1/6$ tín hiệu A bị méo và được thu như là tín hiệu B, còn $1/8$ tín hiệu B bị méo thành tín hiệu A.

(a) Tìm xác suất thu được tín hiệu A;

(b) Giả sử thu được tín hiệu A, tìm xác suất để thu được đúng tín hiệu lúc phát.

Bài tập 1.58. Một người có ba chỗ ưa thích như nhau để câu cá. Xác suất để câu được cá ở mỗi chỗ tương ứng là 0,6; 0,7 và 0,8. Biết rằng đến một chỗ người đó thả câu 3 lần và chỉ câu được một con cá. Tính xác suất để cá câu được ở chỗ thứ nhất.

Bài tập 1.59. Trong học kỳ I năm học 2018-2019, sinh viên phải thi 4 học phần. Xác suất để sinh viên thi đạt một học phần trong mỗi lần thi đều là 0,8. Nếu thi không đạt học phần nào phải thi lại học phần đó. Tính xác suất để một sinh viên thi đạt cả 4 học phần trong đó không có học phần nào thi quá 2 lần.

Bài tập 1.60. Ba người thợ cùng may một loại áo với xác suất may được sản phẩm chất lượng cao tương ứng là 0,9; 0,9 và 0,8. Biết một người khi may 8 áo thì có 6 sản phẩm chất lượng cao. Tìm xác suất để người đó may 8 áo nữa thì có 6 áo chất lượng cao.

Chương 2

Biến ngẫu nhiên và quy luật phân phối xác suất

Mục tiêu

1. Thông qua các công cụ giải tích, cung cấp cho sinh viên khái niệm về biến ngẫu nhiên, phân loại các biến ngẫu nhiên, các quy luật phân phối xác suất của biến ngẫu nhiên, các tham số đặc trưng của biến ngẫu nhiên cùng một số quy luật phân phối xác suất thông dụng của biến ngẫu nhiên.

2. Với các kiến thức nền tảng đó, sinh viên biết tính các tham số đặc trưng của biến ngẫu nhiên; hiểu và vận dụng được ý nghĩa của các đặc trưng của biến ngẫu nhiên cùng các quy luật phân phối xác suất trong các bài toán xác suất thuộc các lĩnh vực kỹ thuật, kinh tế, xã hội...

Nội dung

Hai nội dung quan trọng nhất của chương là quy luật phân phối xác suất và các tham số đặc trưng của một biến ngẫu nhiên.

1. Định nghĩa và phân loại biến ngẫu nhiên

2. Quy luật phân phối xác suất của biến ngẫu nhiên (bảng phân phối xác suất, hàm phân phối xác suất, hàm mật độ xác suất)

3. Các tham số đặc trưng của biến ngẫu nhiên (kỳ vọng, phương sai, độ lệch chuẩn, mốt, trung vị...)

4. Một số phân phối xác suất thông dụng

Thời lượng: 8 tiết

BÀI 5 (2 tiết)

2.1 Định nghĩa và phân loại biến ngẫu nhiên

2.1.1 Định nghĩa biến ngẫu nhiên

Khái niệm biến ngẫu nhiên (random variable) rất thông dụng trong giải tích. Vì vậy ta tìm cách đưa vào khái niệm biến ngẫu nhiên như một đại lượng phụ thuộc vào kết cục của một phép thử ngẫu nhiên nào đó.

Ví dụ 2.1. Gieo một con xúc sắc. Nếu ta gọi biến ngẫu nhiên là "số chấm xuất hiện" thì nó phụ thuộc vào kết cục của phép thử và nhận các giá trị nguyên từ 1 đến 6.

Về mặt hình thức, có thể định nghĩa biến ngẫu nhiên như một hàm số có giá trị thực xác định trên không gian các sự kiện sơ cấp.

Ký hiệu biến ngẫu nhiên là $X, Y, Z, X_1, X_2, \ldots$. Các giá trị có thể có của chúng ký kiệu là $x, y, z, x_1, x_2, \ldots$.

Tập hợp tất cả các giá trị của X gọi là miền giá trị của X, ký hiệu là $\mathbf{S}_X$.

Nhận xét 2.1. **(a)** X được gọi là biến ngẫu nhiên vì trước khi tiến hành phép thử ta chưa có thể nói một cách chắn chắc nó sẽ nhận một giá trị bằng bao nhiêu mà chỉ dự đoán điều đó với một xác suất nhất định. Nói cách khác, việc biến ngẫu nhiên X nhận một giá trị nào đó $(X = x_1), (X = x_2), \ldots, (X = x_n)$ về thực chất là các sự kiện ngẫu nhiên.

(b) Nếu biến ngẫu nhiên X chỉ nhận các giá trị $x_1, x_2, \ldots, x_n$ thì các sự kiện $(X = x_1)$, $(X = x_2), \ldots, (X = x_n)$ tạo nên một hệ đầy đủ.

2.1.2 Phân loại biến ngẫu nhiên

Biến ngẫu nhiên được phân làm hai loại: biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

Định nghĩa 2.1 (Biến ngẫu nhiên rời rạc). X là biến ngẫu nhiên rời rạc (discrete random variable) nếu tập giá trị $\mathbf{S}_X$ của nó là tập hợp hữu hạn hoặc vô hạn đếm được phần tử. Nói cách khác, ta có thể liệt kê tất cả các giá trị của biến ngẫu nhiên đó.

Định nghĩa 2.2 (Biến ngẫu nhiên liên tục). X là biến ngẫu nhiên liên tục (continuous random variable) nếu tập giá trị $\mathbf{S}_X$ có thể có của nó lấp đầy một khoảng trên trục số.

Ví dụ 2.2. (a) Gọi X là số chấm xuất hiện khi gieo một con xúc sắc cân đối đồng chất thì X là biến ngẫu nhiên rời rạc có thể nhận một trong các giá trị 1, 2, 3, 4, 5 và 6.

(b) Một người phải tiến hành thí nghiệm cho tới khi thành công thì dừng. Gọi Y là số lần tiến hành thí nghiệm. Khi đó Y là biến ngẫu nhiên rời rạc có thể nhận các giá trị $1, 2, \ldots, n, \ldots$.

(c) Bắn một viên đạn vào bia có bán kính là 20cm và giả sử viên đạn trúng vào bia. Gọi Z là khoảng cách từ tâm bia tới điểm bia trúng đạn thì Z là biến ngẫu nhiên liên tục có thể nhận các giá trị thuộc $(0;20)$.

2.2 Quy luật phân phối xác suất của biến ngẫu nhiên

Định nghĩa 2.3 (Quy luật phân phối xác suất)**.** Bất kỳ một hình thức nào cho phép biểu diễn mối quan hệ giữa các giá trị có thể có của biến ngẫu nhiên và xác suất tương ứng để biến ngẫu nhiên nhận các giá trị đó đều được gọi là quy luật phân phối xác suất (probability distribution) của biến ngẫu nhiên.

Một số phương pháp mô tả quy luật phân phối xác suất của biến ngẫu nhiên:

1. Bảng phân phối xác suất (áp dụng cho biến ngẫu nhiên rời rạc).
2. Hàm phân phối xác suất (áp dụng cho cả biến ngẫu nhiên rời rạc và liên tục).
3. Hàm mật độ xác suất (áp dụng cho biến ngẫu nhiên liên tục).

2.2.1 Bảng phân phối xác suất của biến ngẫu nhiên rời rạc

Định nghĩa 2.4 (Hàm khối lượng xác suất)**.** Cho biến ngẫu nhiên rời rạc X. Đặt

$$\boxed{p_X(x) = P(X = x), \quad x \in \mathbb{R}} \tag{2.1}$$

Hàm $p_X(x)$ được gọi là hàm khối lượng xác suất (probability mass function) của biến ngẫu nhiên rời rạc X.

Hàm khối lượng xác suất có tính chất sau.

Định lý 2.1. **(a)** $p_X(x) > 0$ với mọi $x \in \mathbf{S}_X$;

(b) $\sum_{x \in \mathbf{S}_X} p_X(x) = 1$;

(c) $p_X(x) = 0$ với mọi $x \notin \mathbf{S}_X$.

Định nghĩa 2.5 (Bảng phân phối xác suất)**.** Bảng phân phối xác suất (probability distribution) của biến ngẫu nhiên rời rạc X là bảng ghi sự tương ứng giữa các giá trị mà biến ngẫu nhiên nhận được với giá trị của hàm khối lượng xác suất tương ứng.

1. Nếu X là biến ngẫu nhiên rời rạc có hữu hạn (n) phần tử thì bảng phân phối xác suất của biến ngẫu nhiên X là:

X	x_1	x_2	$\ldots$	x_n
$P(X = x_i)$	p_1	p_2	$\ldots$	p_n

(2.2)

trong đó $\{x_1, x_2, \ldots, x_n\}$ là tập các giá trị của X đã sắp xếp theo thứ tự tăng dần, $p_i = P(X = x_i), i = 1, 2 \ldots, n$.

2. Nếu X là biến ngẫu nhiên rời rạc có vô hạn đếm được phần tử thì bảng phân phối xác suất của biến ngẫu nhiên X là:

X	x_1	x_2	$\ldots$	x_n	$\ldots$
$P(X = x_i)$	p_1	p_2	$\ldots$	p_n	$\ldots$

(2.3)

trong đó $\{x_1, x_2, \ldots, x_n \ldots\}$ là tập các giá trị của X đã sắp xếp theo thứ tự tăng dần, $p_n = P(X = x_n), n = 1, 2 \ldots.$

Nhận xét 2.2. Trong (2.2), $\sum_{i=1}^{n} p_i = 1$ và trong (2.3), $\sum_{n=1}^{\infty} p_n = 1$.

Ví dụ 2.3. Một xạ thủ có 3 viên đạn được yêu cầu bắn lần lượt từng viên cho đến khi trúng mục tiêu hoặc hết cả 3 viên thì thôi. Tìm bảng phân phối xác suất của số đạn đã bắn, biết rằng xác suất bắn trúng đích của mỗi lần bắn là 0,8.

Lời giải Ví dụ 2.3 Gọi X là số đạn đã bắn, X là biến ngẫu nhiên rời rạc nhận các giá trị 1, 2, 3.

Gọi A_i là sự kiện "bắn trúng mục tiêu ở lần bắn thứ i", $i = 1, 2, 3$. Khi đó,

$P(X = 1) = P(A_1) = 0,8.$

$P(X = 2) = P(\overline{A}_1 A_2) = P(\overline{A}_1)P(A_2) = 0,2 \times 0,8 = 0,16.$

$P(X = 3) = P(\overline{A}_1 \overline{A}_2 (A_3 + \overline{A}_3)) = P(\overline{A}_1)P(\overline{A}_2)P(A_3 + \overline{A}_3) = 0,2 \times 0,2 \times (0,8 + 0,2) = 0,04.$

Vậy bảng phân phối xác suất của X là

X	1	2	3
$P(X = x_i)$	0,8	0,16	0,04

Ví dụ 2.4. Một người đem 10 nghìn VNĐ đi đánh một số đề. Nếu trúng thì thu được 700 nghìn VNĐ, nếu trượt thì không được gì. Gọi X (nghìn VNĐ) là số tiền thu được. Ta có bảng phân phối xác suất của X

X	0	700
$P(X = x_i)$	99/100	1/100

Ví dụ 2.5. Một chùm chìa khóa gồm 4 chiếc giống nhau, trong đó chỉ có một chiếc mở được cửa. Người ta thử ngẫu nhiên từng chiếc cho đến khi mở được cửa. Gọi X là số lần thử. Tìm phân phối xác suất của X.

Lời giải Ví dụ 2.5 X có thể nhận các giá trị 1, 2, 3, 4.

Gọi A_i là sự kiện "mở được cửa ở lần thử thứ i", $i = 1, 2, 3, 4$. Khi đó,

$$P(X = 1) = P(A_1) = \frac{1}{4}$$

$$P(X = 2) = P(\overline{A}_1 A_2) = P(\overline{A}_1)P(A_2|\overline{A}_1) = \frac{3 \times 1}{4 \times 3} = \frac{1}{4}$$

$$P(X = 3) = P(\overline{A}_1 \overline{A}_2 A_3) = P(\overline{A}_1)P(\overline{A}_2|)P(A_3|\overline{A}_1\overline{A}_2) = \frac{3 \times 2 \times 1}{4 \times 3 \times 2} = \frac{1}{4}$$

$$P(X = 4) = P(\overline{A}_1 \overline{A}_2 \overline{A}_3 A_4) = P(\overline{A}_1)P(\overline{A}_2|\overline{A}_1)P(\overline{A}_3|\overline{A}_1\overline{A}_2)P(A_4|\overline{A}_1\overline{A}_2\overline{A}_3) = \frac{1}{4}.$$

Vậy bảng phân phối xác suất của X là

X	1	2	3	4
$P(X = x_i)$	$1/4$	$1/4$	$1/4$	$1/4$

2.2.2 Hàm phân phối xác suất

Định nghĩa 2.6 (Hàm phân phối xác suất)**.** Hàm phân phối xác suất (cumulative distribution function) của biến ngẫu nhiên X, ký hiệu là $F_X(x)$, được định nghĩa như sau:

$$\boxed{F_X(x) = P(X < x), \quad x \in \mathbb{R}} \tag{2.4}$$

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì hàm phân phối (tích lũy) là:

$$F_X(x) = \begin{cases} 0, & x \le x_1, \\ p_1, & x_1 < x \le x_2, \\ p_1 + p_2, & x_2 < x \le x_3, \\ \dots & \\ 1, & x > x_n. \end{cases} \tag{2.5}$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì hàm phân phối (tích lũy) là:

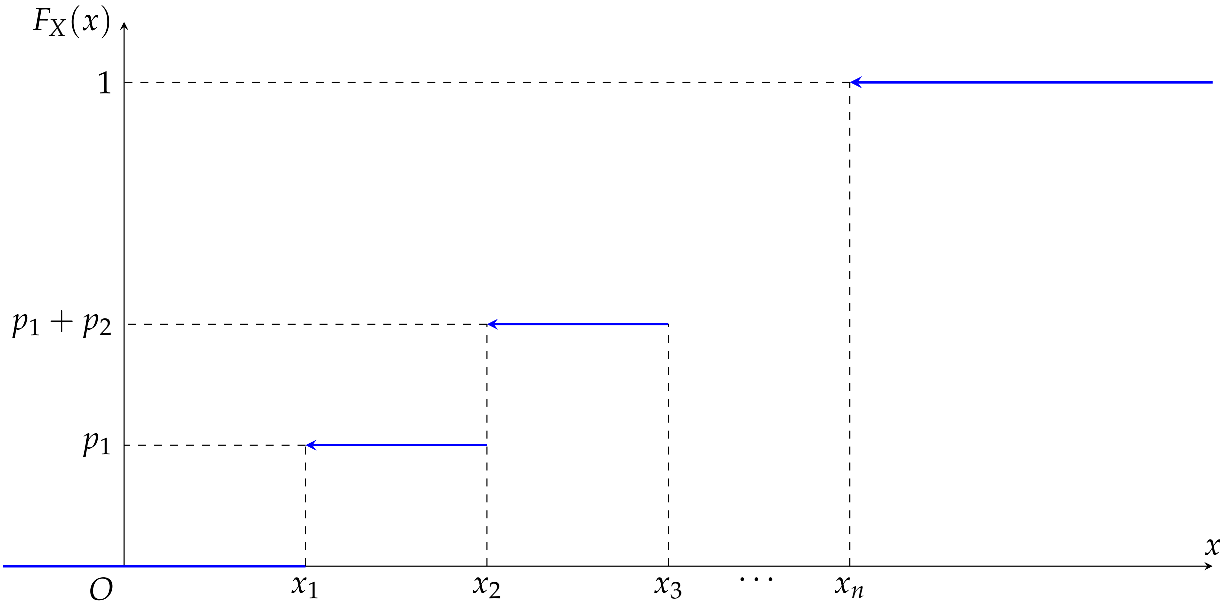
$$F_X(x) = \begin{cases} 0, & x \le x_1, \\ p_1, & x_1 < x \le x_2, \\ p_1 + p_2, & x_2 < x \le x_3, \\ \dots & \\ \sum_{i=1}^{n} p_i, & x_n < x \le x_{n+1}, \\ \dots & \end{cases} \tag{2.6}$$

Nhận xét 2.3. 1. Hàm phân phối xác suất $F_X(x)$ phản ánh mức độ tập trung xác suất ở bên trái của một số thực x nào đó.

2. Đồ thị của hàm phân phối xác suất của biến ngẫu nhiên rời rạc có dạng bậc thang (Hình 2.1)

Ví dụ 2.6. (a) Tìm hàm phân phối xác suất của biến ngẫu nhiên ở Ví dụ 2.3. (b) Vẽ đồ thị hàm phân phối.

Lời giải Ví dụ 2.6

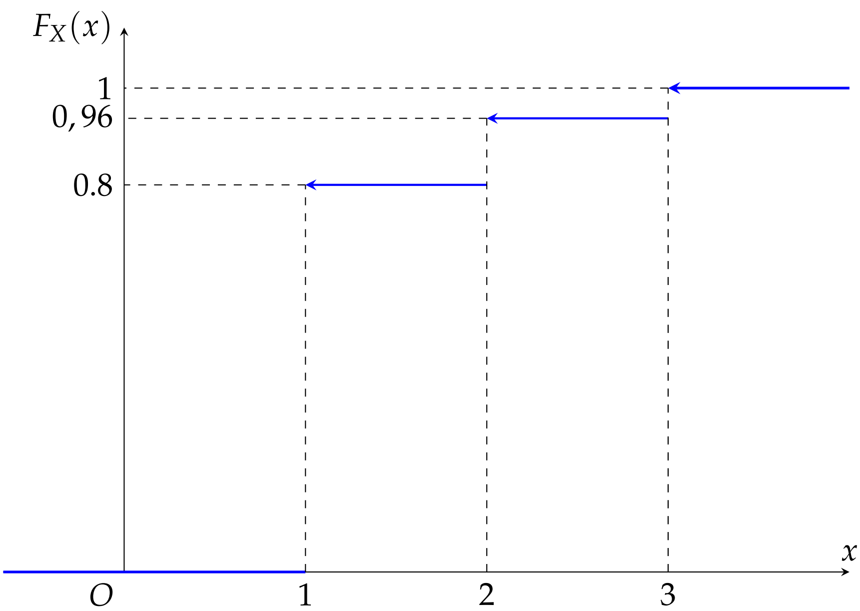


Hình 2.1: Đồ thị của hàm phân phối xác suất (2.5)

(a) Từ bảng phân phối xác suất ở Ví dụ 2.3, sử dụng (2.5) suy ra

$$F_X(x) = \begin{cases} 0, & x \le 1, \\ 0,8, & 1 < x \le 2, \\ 0,96, & 2 < x \le 3, \\ 1, & x > 3. \end{cases}$$

(b) Đồ thị của hàm $F_X(x)$ có dạng bậc thang:



Hình 2.2: Đồ thị hàm phân phối xác suất trong Ví dụ 2.6

Hàm phân phối có các tính chất sau.

Định lý 2.2. 1. $0 \leq F_X(x) \leq 1$ với mọi $x \in \mathbb{R}$.

2. $F_X(x)$ là hàm không giảm, liên tục bên trái, nghĩa là với mọi $x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$ thì $F_X(x_1) \leq F_X(x_2)$ và với mọi $a \in \mathbb{R}$, $F_X(a^-) = F_X(a)$, với $F_X(a^-) = \lim_{x\to a^-} F_X(x)$.

   Nếu X là biến ngẫu nhiên liên tục thì $F_X(x)$ là hàm liên tục.

3. $P(a \leq X < b) = F_X(b) - F_X(a)$;

   Nếu X là biến ngẫu nhiên liên tục thì $P(X = a) = 0$ và

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

4. $F_X(-\infty) = 0$, $F_X(+\infty) = 1$.

Chứng minh. 1. Suy trực tiếp từ định nghĩa (2.4) và tính chất của xác suất.

2. Giả sử $x_1 < x_2$, xét sự kiện $(X < x_2) = (X < x_1) + (x_1 \leq X < x_2)$. Do tính xung khắc của các sự kiện suy ra

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

   Từ đây kết hợp với (2.4) suy ra

$$F_X(x_2) - F_X(x_1) = P(x_1 \leq X < x_2) \geq 0.$$

3. Suy trực tiếp từ chứng minh tính chất trên.

4. $F_X(-\infty) = P(X < -\infty) = P(\emptyset) = 0$, $F_X(+\infty) = P(X < +\infty) = P(S) = 1$.

□

Ví dụ 2.7. Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq -1, \\ A + B \arcsin x, & -1 < x < 1, \\ 1, & x \geq 1. \end{cases}$$

Hãy xác định A và B?

Lời giải Ví dụ 2.7 Sử dụng Định lý 2.2(1), $0 \leq A + B \arcsin x \leq 1$ và theo Định lý 2.2(2) vì $F_X(x)$ liên tục nên $A - \dfrac{\pi}{2} \times B = 0$, $A + \dfrac{\pi}{2} \times B = 1$.

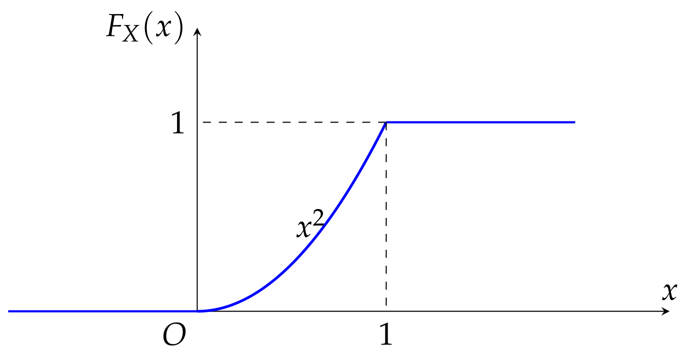
Suy ra $A = \dfrac{1}{2}$, $B = \dfrac{1}{\pi}$.

Ví dụ 2.8. Xét phép thử ném phi tiêu vào một đĩa tròn có bán kính bằng 1(m). Ký hiệu X là biến ngẫu nhiên đo khoảng cách từ điểm mũi phi tiêu cắm vào đĩa đến tâm của đĩa. Giả sử mũi phi tiêu luôn cắm vào đĩa và đồng khả năng tại mọi điểm của đĩa. (a) Tìm miền giá trị của X. (b) Tìm hàm phân phối $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Lời giải Ví dụ 2.8

(a) $\mathbf{S}_X = \{x \in \mathbb{R} : 0 \leq x < 1\}$.

(b) Sử dụng định nghĩa $F_X(x) = P(X < x) = \begin{cases} 0, & x \leq 0, \\ x^2, & 0 < x \leq 1, \\ 1, & x > 1. \end{cases}$



Hình 2.3: Đồ thị hàm phân phối xác suất của Ví dụ 2.8

BÀI 6 (2 tiết)

2.2.3 Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Định nghĩa 2.7 (Hàm mật độ xác suất)**.** Giả sử X là một biến ngẫu nhiên liên tục có hàm phân phối xác suất $F_X(x)$, $x \in \mathbb{R}$. Nếu tồn tại hàm $f_X(x)$ sao cho

$$\boxed{F_X(x) = \int\limits_{-\infty}^{x} f_X(t)dt, \quad x \in \mathbb{R}} \tag{2.7}$$

thì $f_X(x)$ được gọi là hàm mật độ xác suất (probability density function) của biến ngẫu nhiên X.

Như vậy, hàm mật độ xác suất của biến ngẫu nhiên liên tục X là đạo hàm bậc nhất của hàm phân phối xác suất của biến ngẫu nhiên đó,

$$\boxed{f_X(x) = F'_X(x), \quad x \in \mathbb{R}} \tag{2.8}$$

Nhận xét 2.4. Hàm mật độ xác suất của biến ngẫu nhiên X tại mỗi điểm x cho biết mức độ tập trung xác suất tại điểm đó.

Hàm mật độ xác suất có các tính chất sau.

Định lý 2.3. 1. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$.

2. $P(a < X < b) = \int_a^b f_X(x)dx$.

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

Chứng minh. 1. Vì $f_X(x)$ là đạo hàm của hàm không giảm.

2. Suy từ Định lý 2.2(3).

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

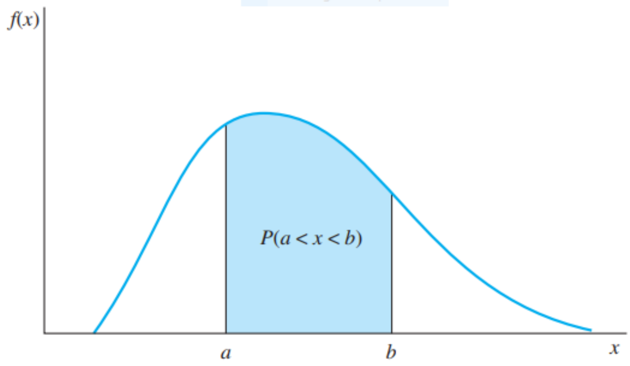
□

Ví dụ 2.9. Hàm phân phối xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = a + b \arctan x, \quad (-\infty < x < +\infty).$$

(a) Tìm a và b. (b) Tìm hàm mật độ xác suất $f_X(x)$. (c) Tìm xác suất để khi tiến hành 3 phép thử độc lập có 2 lần X nhận giá trị trong khoảng $(-1; 1)$.

Lời giải Ví dụ 2.9



Hình 2.4: $P[a < X < b]$ là diện tích miền tô màu dưới đường cong $y = f_X(x)$

(a) Sử dụng Định lý 2.2(4) ta tìm được $a = \dfrac{1}{2}$, $b = \dfrac{1}{\pi}$.

(b) Sử dụng (2.8) ta được $f_X(x) = \dfrac{1}{\pi(1+x^2)}$.

(c) Theo Định lý 2.3(2)

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

Bài toán thỏa mãn lược đồ Béc–nu–li. Áp dụng công thức (1.19) ta tính được

$$P_3(2) = C_3^2 \times p^2 \times (1-p)^1 = \frac{3}{8}.$$

Ví dụ 2.10. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất là

$$f_X(x) = \begin{cases} a\cos x, & x \in \left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right] \\ 0, & x \notin \left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right]. \end{cases}$$

(a) Tìm a. (b) Tìm hàm phân phối xác suất tương ứng. (c) Tìm xác suất để X nhận giá trị trong khoảng $\left(0, \dfrac{\pi}{4}\right)$.

Lời giải Ví dụ 2.10

(a) Sử dụng Định lý 2.3(1),(3) tính được $a = \dfrac{1}{2}$.

(b) Áp dụng (2.7).

Nếu $x \le -\dfrac{\pi}{2}$ thì $F_X(x) = \displaystyle\int_{-\infty}^{x} 0du = 0$.

Nếu $-\dfrac{\pi}{2} < x \le \dfrac{\pi}{2}$ thì $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(u)du = \int_{-\frac{\pi}{2}}^{x} \frac{1}{2}\cos u du = \frac{1}{2}(\sin x + 1)$.

Nếu $x > \dfrac{\pi}{2}$ thì $F_X(x) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}\cos x = 1$. Vậy

$$F_X(x) = \begin{cases} 0, & x \le -\dfrac{\pi}{2}, \\ \dfrac{1}{2}(\sin x + 1), & -\dfrac{\pi}{2} < x \le \dfrac{\pi}{2}, \\ 1, & x > \dfrac{\pi}{2}. \end{cases}$$

(c) $P(0 < X < \dfrac{\pi}{4}) = \int_0^{\frac{\pi}{4}} \frac{1}{2}\cos x dx = \dfrac{\sqrt{2}}{4}$.

2.3 Các tham số đặc trưng của biến ngẫu nhiên

Đặc trưng quan trọng nhất của biến ngẫu nhiên là phân phối xác suất của nó. Nhưng trong thực tế nhiều khi không xác định được hàm phân phối và không phải cứ nhất thiết phải biết hàm phân phối. Vì vậy nảy sinh vấn đề phải đặc trưng cho biến ngẫu nhiên bằng một hoặc nhiều số, mỗi số hạng đặc trưng phản ánh được các tính chất cơ bản nhất của biến ngẫu nhiên X. Trong mục này ta chỉ xét một vài tham số quan trọng nhất.

2.3.1 Kỳ vọng

Định nghĩa 2.8 (Kỳ vọng)**.** Kỳ vọng (expected value) của biến ngẫu nhiên X, ký hiệu là $E(X)$ (hoặc μ_X hoặc đơn giản là μ) được xác định như sau:

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$\boxed{\mu_X = E(X) = \sum_{i=1}^{n} x_i p_i} \tag{2.9}$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì

$$\boxed{\mu_X = E(X) = \sum_{n=1}^{\infty} x_n p_n} \tag{2.10}$$

nếu chuỗi vế phải hội tụ.

3. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$, $x \in \mathbb{R}$ thì

$$\boxed{\mu_X = E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx} \tag{2.11}$$

nếu tích phân vế phải hội tụ.

Nhận xét 2.5. 1. Kỳ vọng mang ý nghĩa là giá trị trung bình của biến ngẫu nhiên. Kỳ vọng là số xác định.

Thật vậy, giả sử đối với biến ngẫu nhiên X, tiến hành n phép thử, trong đó n_1 lần X nhận giá trị x_1, n_2 lần X nhận giá trị $x_2, \dots, n_k$ lần X nhận giá trị x_k, $n_1 + n_2 + \cdots + n_k = n$. Giá trị trung bình của biến ngẫu nhiên X trong n phép thử này là

$$\begin{aligned}\overline{X} = \frac{n_1x_2 + n_2x_2 + \cdots + n_kx_k}{n} &= x_1\frac{n_1}{n} + x_2\frac{n_2}{n} + \cdots + x_k\frac{n_k}{n} \\ &\simeq x_1p_1 + x_2p_2 + \cdots + x_kp_k = E(X).\end{aligned}$$

2. Khái niệm kỳ vọng được áp dụng rộng rãi trong nhiều lĩnh vực. Trong kinh doanh và quản lý, kỳ vọng được ứng dụng dưới dạng lợi nhuận kỳ vọng hay doanh số kỳ vọng.

Ví dụ 2.11. Theo thống kê việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000$, còn tiền đóng là 10$. Hỏi lợi nhuận trung bình của công ty bảo hiểm nhận được là bao nhiêu?

Lời giải Ví dụ 2.11 Gọi X là lợi nhuận của công ty bảo hiểm nhận được. Khi đó X là biến ngẫu nhiên rời rạc có thể nhận giá trị -990, 10. Bảng phân phối xác suất của X là

X	-990	10
$P(X = x_i)$	0,008	0,992

Suy ra $E(X) = -990 \times 0,008 + 10 \times 0,992 = 2\$$. Ta thấy lợi nhuận trung bình bằng 2$ (một số dương) vì vậy công ty bảo hiểm có thể làm ăn có lãi.

Ví dụ 2.12. Kỳ vọng của biến ngẫu nhiên X trong Ví dụ 2.4 là $\mu_X = E(X) = 0 \times {}^{99}/_{100} + 700 \times {}^{1}/_{100} = 7$ nghìn VNĐ. Như vậy bỏ ra 10 nghìn VNĐ, trung bình thu được 7 nghìn VNĐ, người chơi về lâu dài sẽ lỗ 30% tổng số tiền chơi.

Ví dụ 2.13. Xét trò chơi trả lời hai câu hỏi A và B; người chơi có quyền chọn câu hỏi nào để trả lời đầu tiên. Câu hỏi A được trả lời đúng với xác suất 0,8 và khi đó người chơi sẽ được thưởng 100 USD, câu hỏi B được trả lời đúng với xác suất 0,6 và người chơi được thưởng 200 USD. Nếu không trả lời đúng lần thứ nhất sẽ không được trả lời tiếp. Vậy người chơi nên chọn câu hỏi nào trả lời đầu tiên để tiền thưởng trung bình nhận được cao hơn.

Lời giải Ví dụ 2.13 Gọi X là số tiền thưởng nhận được khi người chơi chọn câu hỏi A trả lời đầu tiên,

X	0	100	300
$P(X = x_i)$	$0,2$	$0,32$	$0,48$

và $E(X) = 0 \times 0,2 + 100 \times 0,32 + 300 \times 0,48 = 176$ USD.

Gọi Y là số tiền thưởng nhận được khi người chơi chọn câu hỏi B trả lời đầu tiên,

Y	0	200	300
$P(Y = y_i)$	$0,4$	$0,12$	$0,48$

và $E(Y) = 0 \times 0,4 + 200 \times 0,12 + 300 \times 0,48 = 168$ USD.

Vậy nên chọn câu hỏi A để trả lời đầu tiên để có khả năng nhận thưởng cao hơn.

Ví dụ 2.14. Theo thống kê ở một cửa hàng đậu tương, người ta thấy số lượng đậu tương bán ra X là một biến ngẫu nhiên rời rạc có bảng phân phối là:

X (kg)	10	13	16	19	22
$P(X = x_i)$	$0,15$	$0,2$	$0,35$	$0,2$	$0,1$

Nếu giá nhập là 10000 VNĐ/kg thì cửa hàng sẽ lãi 5000 VNĐ/kg, nếu đến cuối ngày không bán được sẽ lỗ 8000 VNĐ/kg. (a) Tìm hàm phân phối xác suất của X. (b) Mỗi ngày cửa hàng nên nhập bao nhiêu kg để thu được lãi nhiều nhất.

Lời giải Ví dụ 2.14

(a) Từ bảng phân phối xác suất ta có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq 10, \\ 0,15, & 10 < x \leq 13, \\ 0,35, & 13 < x \leq 16, \\ 0,7, & 16 < x \leq 19, \\ 0,9, & 19 < x \leq 22, \\ 1, & x > 22. \end{cases}$$

(b) Số lượng đậu tương nhập trong ngày theo các phương án 10, 13, 16, 19, 22. Gọi T_i là "số tiền lời thu được ứng với phương án i", $i = 1, 2, \ldots, 5$, trong đó phương án 1, 2, 3, 4, 5 tương ứng là nhập 10, 13, 16, 19, 22 (kg).

(b1) Phương án nhập 10kg: chắc chắn cửa hàng sẽ bán hết vì $P(X < 10) = 0$. Do đó $E(T_1) = 1 \times 50000 = 50000$ VNĐ.

(b2) Phương án nhập 13kg: do không có thống kê số lượng bán 11, 12kg, nên xem như cửa hàng đó chỉ có 2 phương án hoặc bán 10kg, hoặc bán 13kg. Do chỉ nhập 13kg nên xem như số lượng bán trên 13kg là số lượng bán được 13kg. Suy ra

$$E(T_2) = 26000 \times 0,15 + 65000 \times 0,85 = 59150 \text{ VNĐ}.$$

(b3) Phương án nhập 16kg: số lượng bán ra có thể là 10, 13, 16 với xác suất tương ứng là 0,15; 0,2 và 0,65. Suy ra

$$E(T_3) = 2000 \times 0,15 + 41000 \times 0,2 + 80000 \times 0,65 = 60500 \text{ VNĐ}.$$

(b4) Phương án nhập 19kg: số lượng bán ra có thể là 10, 13, 16, 19 với xác suất tương ứng là 0,15; 0,2; 0,35 và 0,3. Suy ra

$$E(T_4) = (-22000) \times 0,15 + 17000 \times 0,2 + 56000 \times 0,35 + 95000 \times 0,3 = 48200 \text{ VNĐ}.$$

(b5) Phương án nhập 22kg: số lượng bán ra có thể là 10, 13, 16, 19, 22 với xác suất tương ứng là 0,15; 0,2; 0,35; 0,2 và 0,1. Suy ra

$$\begin{aligned} E(T_5) &= (-46000) \times 0,15 + (-7000) \times 0,2 + 32000 \times 0,35 + 71000 \times 0,2 \\ &\quad + 110000 \times 0,1 = 28100 \text{ VNĐ}. \end{aligned}$$

Từ các kết quả trên, ta thấy $E(T_3)$ là cao nhất nên phương án nhập hiệu quả nhất là 16kg.

Chú ý 2.1. Nếu trong bảng phân phối xác suất mà giá trị nào của biến ngẫu nhiên X không được đề cập đến thì xem như xác suất tại đó bằng 0.

Ví dụ 2.15. Một dây chuyền tự động khi hoạt động bình thường có thể sản xuất ra phế phẩm với xác suất $p = 0,001$ và được điều chỉnh ngay lập tức khi phát hiện có phế phẩm. Tính số trung bình các sản phẩm được sản xuất giữa 2 lần điều chỉnh.

Lời giải Ví dụ 2.15 Gọi X là số sản phẩm được sản xuất giữa hai lần điều chỉnh. Khi đó X là biến ngẫu nhiên rời rạc có thể nhận các giá trị 1, 2, ... với xác suất tương ứng

$$P(X=1) = 0,001, \quad P(X=2) = 0,999 \times 0,001, \quad P(X=3) = (0,999)^2 \times 0,001 \ldots$$

Vậy

$$\begin{aligned} E(X) &= 1 \times 0,001 + 2 \times 0,999 \times 0,001 + 3 \times (0,999)^2 \times 0,001 + \ldots \\ &= 0,001 \times \sum_{n=1}^{\infty} n \times (0,999)^{n-1} = 1000, \end{aligned}$$

ở đây ta sử dụng tính chất của chuỗi lũy thừa và công thức tính tổng của cấp số nhân lùi vô hạn với $x = 0,999$:

$$\sum_{n=1}^{\infty} nx^n = \sum_{n=1}^{\infty} \left(x^n\right)' = \left(\sum_{n=1}^{\infty} x^n\right)' = \left(\frac{x}{1-x}\right)' = \frac{1}{(1-x)^2}.$$

Ví dụ 2.16. Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên (đơn vị là tháng) với hàm mật độ xác suất như sau:

$$f_X(x) = \begin{cases} kx^2(4-x), & x \in [0,4], \\ 0, & x \notin [0,4]. \end{cases}$$

Tìm tuổi thọ trung bình của loại côn trùng trên.

Lời giải Ví dụ 2.16 Vì $f_X(x)$ là hàm mật độ xác suất của biến ngẫu nhiên liên tục X nên theo Định lý 2.3(1),(3), $k = \dfrac{3}{64}$.

Sử dụng công thức (2.11), tuổi thọ trung bình của loại côn trùng trên là

$$E(X) = \frac{3}{64}\int_0^4 x^3(4-x)dx = \frac{12}{5} \quad \text{(tháng)}.$$

Hàm của một biến ngẫu nhiên

Bây giờ ta xét một biến ngẫu nhiên mới $g(X)$, phụ thuộc vào X; nghĩa là, mỗi giá trị của $g(X)$ được xác định bởi giá trị của X. Chẳng hạn, $g(X)$ có thể là X^2 hoặc $3X-1$ và giả sử X nhận giá trị 2, thì $g(X)$ sẽ nhận giá trị $g(2)$.

Ví dụ 2.17. Cho X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất là

X	-1	0	1	2
$P(X=x_i)$	p_{-1}	p_0	p_1	p_2

và $g(X) = X^2$, thì $g(X)$ nhận các giá trị 0, 1, 2 với

$$\begin{aligned}
&P(g(X)=0) = P(X=0) = p_0,\\
&P(g(X)=1) = P(X=-1) + P(X=1) = p_{-1} + p_1,\\
&P(g(X)=4) = P(X=2) = p_2.
\end{aligned}$$

Bảng phân phối xác suất của $g(X)$ là

$g(X)$	0	1	4
$P\{g(X)=g(x_i)\}$	p_0	$p_{-1}+p_1$	p_2

Từ đây, theo Định nghĩa 2.8(1) suy ra

$$\begin{aligned}
E[g(X)] &= (0)(p_0) + (1)(p_{-1}+p_1) + (4)(p_2)\\
&= (-1)^2(p_{-1}) + (0)^2(p_0) + (1)^2(p_1) + (2)^2(p_2) = \sum_i g(x_i)p_i, \quad i = -1, 0, 1, 2.
\end{aligned}$$

Kết quả này được mở rộng trong định lý dưới đây cho cả biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

Định lý 2.4. Cho X là một biến ngẫu nhiên và $Y = g(X)$ là một hàm của X.

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$\boxed{E(Y) = E[g(X)] = \sum_{i=1}^{n} g(x_i)p_i} \tag{2.12}$$

2. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$E(Y) = E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx. \tag{2.13}$$

Ví dụ 2.18. Gọi X là số trang trong một bản fax. Một công ty điện thoại tính cước như sau: 10 xu cho trang thứ nhất, 9 xu cho trang thứ hai, ..., 6 xu cho trang thứ năm. Những bản fax từ 6 đến 10 trang có phí là 50 xu (công ty không nhận những bản fax quá 10 trang). Gọi Y là chi phí phải trả cho một bản fax. (a) Xác định Y. (b) Tính $E(Y)$ nếu X có phân phối là

X	1	2	3	4
$P(X = x_i)$	0,25	0,25	0,25	0,25

Lời giải Ví dụ 2.18

(a) Y là một hàm của X xác định bởi

$$Y = g(X) = \begin{cases} 10,5X - 0,5X^2, & 1 \le X \le 5, \\ 50, & 6 \le X \le 10. \end{cases}$$

(b) Theo Định lý 2.4(1),

$$\begin{aligned} E[Y] &= \sum_{i=1}^{4} g(x_i) p_i \\ &= (0,25)[(10,5)(1) - (0,5)(1)^2] + (0,25)[(10,5)(2) - (0,5)(2)^2] \\ &\quad + (0,25)[(10,5)(3) - (0,5)(3)^2] + (0,25)[(10,5)(4) - (0,5)(4)^2] \\ &= (0,25)[10 + 19 + 27 + 34] = 22,5 \text{ (xu)}. \end{aligned}$$

Ví dụ 2.19. Với biến ngẫu nhiên X trong Ví dụ 2.16, kỳ vọng của biến ngẫu nhiên $Y = X^2$ là

$$E(Y) = E(X^2) = \frac{3}{64} \int_0^4 x^4 (4 - x) dx = \frac{32}{5}.$$

Sau đây là một số tính chất hữu ích giúp đơn giản hóa trong tính toán kỳ vọng của các biến ngẫu nhiên. Các tính chất này đúng cho cả biến ngẫu nhiên rời rạc và liên tục. Các chứng minh được đưa ra cho biến ngẫu nhiên liên tục.

Định lý 2.5. Nếu a và b là các hằng số thì

$$E(aX + b) = aE(X) + b.$$

Chứng minh. Theo định nghĩa,

$$E(aX+b) = \int_{-\infty}^{+\infty} (ax+b) f_X(x)dx = a \int_{-\infty}^{+\infty} x f_X(x)dx + b \int_{-\infty}^{+\infty} f_X(x)dx = aE(X)+b$$

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

Hệ quả 2.1. 1. Nếu $a = 0$, $E(b) = b$.

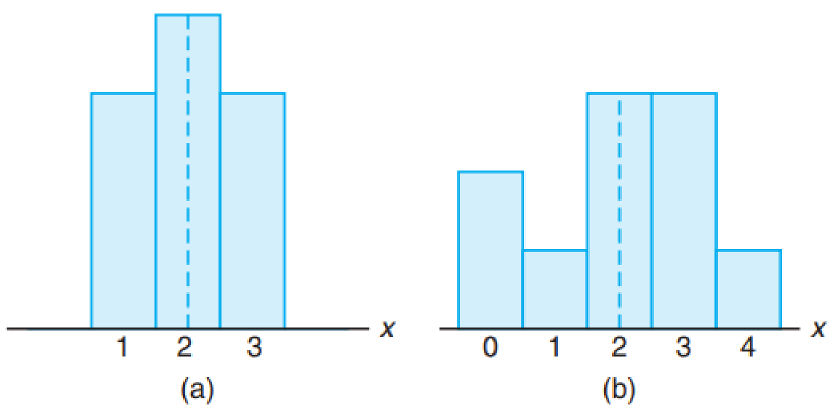
2. Nếu $b = 0$, $E(aX) = aE(X)$.

Định lý 2.6. Cho X là một biến ngẫu nhiên, $h(X)$, $g(X)$ là các hàm của X. Khi đó,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

2.3.2 Phương sai

Giá trị kỳ vọng hay giá trị trung bình của một biến ngẫu nhiên X có tầm quan trọng đặc biệt trong thống kê vì nó phản ánh giá trị trung tâm của phân phối xác suất của biến ngẫu nhiên. Tuy nhiên, giá trị trung bình không đưa ra một mô tả đầy đủ về hình dạng của phân phối. Trong Hình 2.5, ta có biểu đồ của hai phân phối xác suất rời rạc có cùng giá trị trung bình, $\mu = 2$, nhưng khác nhau đáng kể về độ biến thiên hoặc độ phân tán của các quan sát của chúng so với giá trị trung bình. Do đó cần xác định mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình của nó.



Hình 2.5: Phân phối rời rạc với kỳ vọng bằng nhau nhưng độ phân tán khác nhau

Công thức quan trọng nhất về tính biến thiên của biến ngẫu nhiên X có được bằng cách áp dụng Định lý 2.4 với $g(X) = [X - E(X)]^2$. Đại lượng này được gọi là phương sai của biến ngẫu nhiên X hoặc phương sai của phân phối xác suất của X, ký hiệu là $V(X)$ hoặc σ_X^2, hoặc đơn giản là σ^2.

Định nghĩa 2.9 (Phương sai)**.** Phương sai (variance) của biến ngẫu nhiên X được định nghĩa như sau:

$$\boxed{\sigma_X^2 = V(X) = E[X - E(X)]^2} \tag{2.14}$$

Vì $X - E(X)$ là một hàm của biến ngẫu nhiên X, nên từ Định nghĩa 2.9 và Định lý 2.4 ta nhận được các công thức sau đây:

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$\boxed{V(X) = \sum_{i=1}^{n} \Big[x_i - E(X)\Big]^2 p_i} \tag{2.15}$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì

$$\boxed{V(X) = \sum_{n=1}^{\infty} \Big[x_n - E(X)\Big]^2 p_n} \tag{2.16}$$

nếu chuỗi vế phải hội tụ.

3. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$, $x \in \mathbb{R}$ thì

$$\boxed{V(X) = \int_{-\infty}^{+\infty} \Big[x - E(X)\Big]^2 f_X(x)dx} \tag{2.17}$$

nếu tích phân vế phải hội tụ.

Ví dụ 2.20. Cho X là biến ngẫu nhiên chỉ số lượng ô tô được sử dụng cho mục đích kinh doanh chính thức trong mỗi ngày làm việc. Phân phối xác suất của công ty A, xem Hình 2.5(a), là

X_A	1	2	3
$P(X_A = x_i)$	0,3	0,4	0,3

và của công ty B, xem Hình 2.5(b), là

X_B	0	1	2	3	4
$P(X_B = y_j)$	0,2	0,1	0,3	0,3	0,1

Chỉ ra rằng phương sai của phân phối xác suất của công ty B lớn hơn so với công ty A.

Lời giải Ví dụ 2.20

Từ số liệu của công ty A ta tính

$$E(X_A) = (1)(0,3) + (2)(0,4) + (3)(0,3) = 2,0$$

và

$$V(X_A) = (1-2)^2(0,3) + (2-2)^2(0,4) + (3-2)^2(0,3) = 0,6.$$

Với công ty B ta có

$$E(X_B) = (0)(0,2) + (1)(0,1) + (2)(0,3) + (3)(0,3) + (4)(0,1) = 2,0,$$

và

$$\begin{aligned} V(X_B) &= (0-2)^2(0,2) + (1-2)^2(0,1) + (2-2)^2(0,3) + (3-2)^2(0,3) + (4-2)^2(0,1) \\ &= 1,6. \end{aligned}$$

Công thức tương đương của (2.14) được cho trong định lý dưới đây.

Định lý 2.7.

$$\boxed{V(X) = E(X^2) - [E(X)]^2} \tag{2.18}$$

Chứng minh. Giả sử X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2),

$$\begin{aligned} V(X) &= \sum_{i=1}^{n} \Big[x_i - E(X)\Big]^2 p_i = \sum_{i=1}^{n} \Big\{x_i^2 - 2E(X)x_i + [E(X)]^2\Big\} p_i \\ &= \sum_{i=1}^{n} x_i^2 p_i - 2E(X) \sum_{i=1}^{n} x_i p_i + [E(X)]^2 \sum_{i=1}^{n} p_i = E(X^2) - [E(X)]^2, \end{aligned}$$

vì theo định nghĩa $E(X) = \sum_{i=1}^n x_i p_i$, $E(X^2) = \sum_{i=1}^n x_i^2 p_i$ và $\sum_{i=1}^n p_i = 1$.

Đối với biến ngẫu nhiên liên tục ta chứng minh tương tự. □

Hệ quả 2.2.

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$\boxed{V(X) = \sum_{i=1}^{n} x_i^2 p_i - \left(\sum_{i=1}^{n} x_i p_i\right)^2} \tag{2.19}$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì

$$\boxed{V(X) = \sum_{n=1}^{\infty} x_n^2 p_n - \left(\sum_{n=1}^{\infty} x_n p_n\right)^2} \tag{2.20}$$

nếu các chuỗi vế phải hội tụ.

3. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$, $x \in \mathbb{R}$ thì

$$\boxed{V(X) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx - \left(\int_{-\infty}^{+\infty} x f_X(x)dx\right)^2} \tag{2.21}$$

nếu các tích phân vế phải hội tụ.

Ví dụ 2.21. Dùng Định lý 2.7 tính phương sai của biến ngẫu nhiên xét trong Ví dụ (2.11).

Lời giải Ví dụ 2.21 $E(X^2) = (-990)^2 \times 0,008 + (10)^2 \times 0,992 = 7940$. Suy ra

$$V(X) = E(X^2) - [E(X)]^2 = 7940 - 4 = 7936.$$

Điều này nói lên rằng mặc dù kinh doanh bảo hiểm có lãi nhưng rủi ro khá lớn.

Ví dụ 2.22. Dùng Định lý 2.7 tính phương sai của biến ngẫu nhiên xét trong Ví dụ (2.16).

Lời giải Ví dụ 2.22 Từ kết quả của Ví dụ 2.16 và 2.19 suy ra

$$V(X) = E(X^2) - [E(X)]^2 = \frac{32}{5} - \left(\frac{12}{5}\right)^2 = \frac{16}{25}.$$

Chú ý 2.2. Phương sai của biến ngẫu nhiên là một giá trị xác định không âm.

Nhận xét 2.6.

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

2. Trong kỹ thuật phương sai đặc trưng cho mức độ phân tán của các chi tiết gia công hay sai số của thiết bị. Trong quản lý và kinh doanh thì phương sai đặc trưng cho mức độ rủi ro của các quyết định.

Bây giờ ta sẽ mở rộng khái niệm về phương sai của biến ngẫu nhiên X cho biến ngẫu nhiên liên quan đến X, biến ngẫu nhiên $g(X)$.

Định lý 2.8. Cho X là một biến ngẫu nhiên và $Y = g(X)$ là một hàm của X.

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì
$$V(Y) = V[g(X)] = \sum_{i=1}^{n} \Big\{g(x_i) - E[g(X)]\Big\}^2 p_i. \tag{2.22}$$

2. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$ thì
$$V(Y) = V[g(X)] = \int_{-\infty}^{+\infty} \Big\{g(x) - E[g(X)]\Big\}^2 f_X(x)dx. \tag{2.23}$$

Chứng minh. Sử dụng Định lý 2.4 và Định nghĩa 2.9 ta được điều cần chứng minh. □

Phương sai của biến ngẫu nhiên X có tính chất sau.

Định lý 2.9. Nếu a và b là các hằng số thì

1. $V(aX) = a^2V(X)$.

2. $V(b) = 0$.

Hoặc $V(aX + b) = a^2V(X)$.

Chứng minh.

1. $V(aX) = E(a^2X^2) - [E(aX)]^2 = a^2E(X^2) - a^2[E(X)]^2 = a^2V(X)$.

2. $V(b) = E(b^2) - [E(b)]^2 = b^2 - b^2 = 0$.

□

2.3.3 Độ lệch chuẩn

Định nghĩa 2.10 (Độ lệch chuẩn). Độ lệch chuẩn (standard deviation) của biến ngẫu nhiên X, ký hiệu là $\sigma(X)$, được định nghĩa như sau:

$$\boxed{\sigma(X) = \sqrt{V(X)}} \tag{2.24}$$

Nhận xét 2.7. Khi cần đánh giá mức độ phân tán của biến ngẫu nhiên theo đơn vị đo của nó người ta dùng độ lệch chuẩn vì độ lệch chuẩn có cùng đơn vị đo với đơn vị đo của biến ngẫu nhiên.

Ví dụ 2.23. Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong Ví dụ (2.11).

Lời giải Ví dụ 2.23 Từ kết quả trong Ví dụ 2.21, $\sigma(X) = \sqrt{V(X)} = \sqrt{7936} \simeq 89,08$.

Ví dụ 2.24. Tính độ lệch chuẩn của biến ngẫu nhiên xét trong Ví dụ (2.16).

Lời giải Ví dụ 2.24 Từ kết quả của Ví dụ 2.22 suy ra $\sigma(X) = \sqrt{V(X)} = \dfrac{4}{5}$.

2.3.4 Một số đặc trưng khác

2.3.4a Mốt (mode)

Định nghĩa 2.11 (Mốt).

1. Nếu X là biến ngẫu nhiên rời rạc thì mốt là giá trị của X ứng với xác suất lớn nhất.
2. Nếu X là biến ngẫu nhiên liên tục thì mốt là giá trị làm hàm mật độ đạt max.

Ký hiệu $modX$.

2.3.4b Trung vị (median)

Định nghĩa 2.12 (Trung vị). Trung vị của biến ngẫu nhiên X, ký hiệu là $medX$, là giá trị của biến ngẫu nhiên X chia phân phối thành hai phần có xác suất giống nhau, nghĩa là

$$\boxed{P(X < medX) = P(X \geq medX) = \frac{1}{2}} \tag{2.25}$$

Nhận xét 2.8.

1. Từ định nghĩa hàm phân phối, để tìm trung vị ta cần giải phương trình $F_X(x) = \dfrac{1}{2}$.
2. Trong nhiều trường hợp ứng dụng, trung vị là đặc trưng vị trí tốt nhất, nhiều khi tốt hơn cả kỳ vọng, nhất là khi trong số liệu có những sai sót. Trung vị còn có tên là phân vị 50% của phân phối.

Ví dụ 2.25. Cho hàm mật độ xác suất của biến ngẫu nhiên liên tục X:

$$f_X(x) = \begin{cases} \dfrac{3}{4}x(2-x), & 0 \le x \le 2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

Tìm $medX$ và $modX$.

Lời giải Ví dụ 2.25 Theo (2.7), hàm phân phối xác suất của X là

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{3}{4}\left(x^2 - \dfrac{x^3}{3}\right), & 0 < x \le 2, \\ 1, & x > 2. \end{cases}$$

Khi đó $medX$ là nghiệm của phương trình $F_X(x) = \dfrac{1}{2}$. Hay $x^3 - 3x^2 + 2 = 0$ với $0 < x \le 2$. Suy ra $medX = 1$.

Hàm mật độ xác suất $f_X(x)$ có

$$f'_X(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{3}{2}(1-x), & 0 < x < 2, \\ 0, & \text{nếu trái lại,} \end{cases}$$

đổi dấu từ dương sang âm khi đi qua $x = 1$, do đó đạt cực đại tại điểm này, nên $modX = 1$.

BÀI 7 (2 tiết)

2.4 Một số phân phối xác suất thông dụng

2.4.1 Phân phối đều

2.4.1a Phân phối đều rời rạc

Định nghĩa 2.13 (Phân phối đều rời rạc). Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều rời rạc (discrete uniform distribution) với tham số n, ký hiệu là $X \sim \mathcal{U}(n)$, nếu X có bảng phân phối xác suất

X	1	2	...	n
$P(X = x_i)$	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

(2.26)

Sử dụng Định nghĩa 2.8(1) và Định nghĩa 2.9(1) ta nhận được kỳ vọng và phương sai của biến ngẫu nhiên phân phối đều.

Định lý 2.10. Kỳ vọng, phương sai của biến ngẫu nhiên có phân phối đều rời rạc:

$$\boxed{E(X) = \frac{n+1}{2}, \quad V(X) = \frac{n^2-1}{12}} \tag{2.27}$$

2.4.1b Phân phối đều liên tục

Định nghĩa 2.14 (Phân phối đều liên tục). Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều liên tục (continuous uniform distribution) trên $[a,b]$ $(a < b)$, ký hiệu là $X \sim \mathcal{U}([a,b])$, nếu X có hàm mật độ

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b], \\ 0, & x \notin [a,b]. \end{cases} \tag{2.28}$$

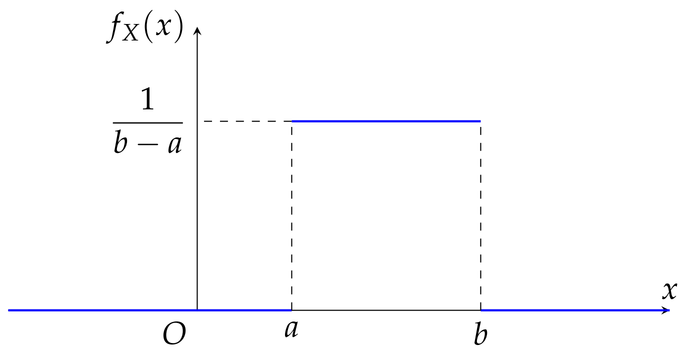
Định lý 2.11. Nếu X là biến ngẫu nhiên có phân phối đều liên tục trên $[a,b]$ thì

1. Hàm phân phối xác suất là

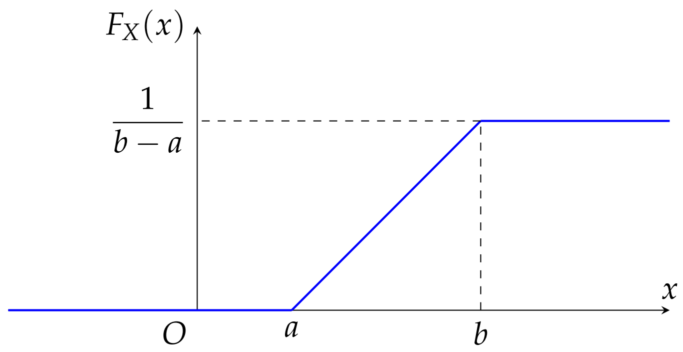
$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \begin{cases} 0, & x \le a, \\ \dfrac{x-a}{b-a}, & a < x \le b, \\ 1, & x > b. \end{cases} \tag{2.29}$$

2. Kỳ vọng $E(X) = \dfrac{a+b}{2}$.

3. Phương sai $V(X) = \dfrac{(b-a)^2}{12}$ và độ lệch chuẩn $\sigma(X) = \sqrt{\dfrac{(b-a)^2}{12}}$.



Hình 2.6: Đồ thị hàm mật độ (2.28) của biến ngẫu nhiên có phân phối đều



Hình 2.7: Đồ thị hàm phân phối (2.29) của biến ngẫu nhiên có phân phối đều

Nhận xét 2.9.
1. X có khả năng nhận giá trị trong khoảng (a, b) là "đều nhau".
2. Phân phối đều có nhiều ứng dụng trong thống kê toán như mô phỏng thống kê, đặc biệt trong phương pháp phi tham số.
3. Trong một số lý thuyết kết luận thống kê người ta thường xuất phát từ quy tắc sau đây: Nếu ta không biết gì về giá trị của tham số cần ước lượng, mỗi giá trị có thể có của tham số đó là đồng khả năng, điều đó dẫn đến việc quan niệm tham số cần ước lượng như một biến ngẫu nhiên có phân phối đều.

Ví dụ 2.26. Lịch chạy của xe bus tại một trạm xe bus như sau: chiếc xe bus đầu tiên trong ngày sẽ khởi hành từ trạm này lúc 7 giờ, cứ sau 15 phút sẽ có một xe khác đến trạm. Giả sử một hành khách đến trạm ngẫu nhiên trong khoảng thời gian từ 7 giờ đến 7 giờ 30. Tìm xác suất để hành khách này chờ (a) Ít hơn 5 phút; (b) Ít nhất 12 phút.

Lời giải Ví dụ 2.26 Gọi X là số phút từ 7 giờ đến 7 giờ 30 hành khách đến trạm, ta có $X \sim \mathcal{U}([0, 30])$.

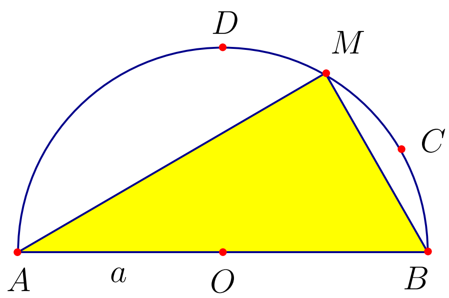
(a) Hành khách chờ ít hơn 5 phút nếu đến trạm giữa 7 giờ 10 và 7 giờ 15 hoặc giữa 7 giờ 25 và 7 giờ 30. Do đó xác suất cần tìm là:

$$P(10 < X \leq 15) + P(25 < X \leq 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

(b) Hành khách chờ ít nhất 12 phút nếu đến trạm giữa 7 giờ và 7 giờ 03 hoặc giữa 7 giờ 15 và 7 giờ 18. Xác suất cần tìm là:

$$P(0 < X \leq 3) + P(15 < X \leq 18) = \frac{3}{30} + \frac{3}{30} = 0,2.$$

Ví dụ 2.27. Lấy ngẫu nhiên một điểm M trên nửa đường tròn tâm O, đường kính $AB = 2a$. Biết rằng xác suất điểm M rơi vào cung CD bất kì của nửa đường tròn AMB chỉ phụ thuộc vào độ dài cung CD. (a) Tìm hàm phân phối xác suất của biến ngẫu nhiên Y chỉ diện tích tam giác AMB. (b) Tìm kỳ vọng của Y.



Hình 2.8: Minh họa cho Ví dụ 2.27

Lời giải Ví dụ 2.27 (a) Theo định lý hàm số sin, ta có $S_{AMB} = a^2 \sin\varphi$, ở đây φ là góc giữa trục Ox và OM. Từ giả thiết ta có φ là biến ngẫu nhiên liên tục có phân phối đều $\mathcal{U}[0,\pi]$ có hàm mật độ xác suất

$$f_\varphi(x) = \begin{cases} \frac{1}{\pi}, & x \in [0,\pi], \\ 0, & x \notin [0,\pi]. \end{cases}$$

Do đó, hàm phân phối xác suất của φ là

$$F_\varphi(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x}{\pi}, & 0 < x \leq \pi, \\ 1, & x > \pi. \end{cases}$$

Biến ngẫu nhiên $Y = a^2 \sin\varphi$, nên Y là biến ngẫu nhiên liên tục nhận giá trị trong đoạn $[0, a^2]$. Hàm phân phối xác suất của Y là

$$F_Y(x) = P(Y < x) = \begin{cases} 0, & x \leq 0, \\ \dfrac{2}{\pi} \arcsin \dfrac{x}{a^2}, & 0 < x \leq a^2, \\ 1, & x > a^2, \end{cases}$$

vì với $x \in (0, a^2]$,

$$F_Y(x) = P(Y < x) = P(a^2 \sin\varphi < x) = P(\sin\varphi < \frac{x}{a^2})$$
$$= P\left(0 < \varphi < \arcsin\frac{x}{a^2}\right) + P\left(\pi - \arcsin\frac{x}{a^2} < \varphi < \pi\right) = \frac{2}{\pi}\arcsin\frac{x}{a^2}.$$

(b) Hàm mật độ xác suất của X

$$f_X(x) = \begin{cases} \dfrac{2}{\pi\sqrt{a^4 - x^2}}, & x \in [0, a^2], \\ 0, & x \notin [0, a^2]. \end{cases}$$

Suy ra kỳ vọng của X là

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \frac{2}{\pi}\int_0^{a^2} \frac{x}{\sqrt{a^4 - x^2}}\,dx = \frac{2}{\pi}a^2.$$

2.4.2 Phân phối nhị thức

2.4.2a Phân phối Béc–nu–li

Định nghĩa 2.15 (Phân phối Béc–nu–li)**.** Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối Béc–nu–li (Bernoulli distribution) với tham số p, ký hiệu là $X \sim \mathcal{B}(1, p)$, nếu X nhận hai giá trị 0, 1 với xác suất tương ứng

$$P(X = k) = p^k q^{1-k}, \quad k = 0, 1, \tag{2.30}$$

ở đây $0 < p < 1, q = 1 - p$.

Định lý 2.12. Nếu X là biến ngẫu nhiên có phân phối Béc-nu-li $\mathcal{B}(1; p)$ thì

$$\boxed{E(X) = p, \quad V(X) = p(1 - p)} \tag{2.31}$$

Chứng minh. Theo định nghĩa kỳ vọng và phương sai của biến ngẫu nhiên rời rạc,

$$E(X) = (0)(1 - p) + (1)(p) = p \quad \text{và} \quad V(X) = (0)^2(1 - p) + (1)^2(p) - p^2 = p(1 - p).$$

□

Nhận xét 2.10. Xét phép thử Béc–nu–li với sự thành công của phép thử là sự xuất hiện của sự kiện A và giả sử xác suất xuất hiện A trong mỗi lần thử là p. Gọi X là số lần thành công trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Béc-nu-li tham số p. Biến ngẫu nhiên X còn được gọi là tuân theo phân phối không – một $\mathcal{A}(p)$.

2.4.2b Phân phối nhị thức

Định nghĩa 2.16 (Phân phối nhị thức)**.** Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối nhị thức (binomial distribution) với tham số n và p, ký hiệu là $X \sim \mathcal{B}(n,p)$, nếu X có bảng phân phối xác suất

X	0	1	...	k	...	n
$P(X=x_i)$	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	...	$C_n^k p^k q^{n-k}$	...	$C_n^n p^n q^0$

(2.32)

ở đây $q = 1-p$ và $P(X=k) = C_n^k p^k q^{n-k}$ được tính bằng công thức Béc–nu–li (1.19).

Phân phối nhị thức xuất phát từ tên thực tế của khai triển nhị thức $(p+q)^n$ có $n+1$ số hạng:

$$(p+q)^n = C_n^0(p)^0(q)^n + C_n^1(p)^1(q)^{n-1} + \cdots + C_n^n(p)^n(q)^0.$$

Nếu $p+q=1$ thì $\sum_{k=0}^n C_n^k(p)^k(q)^{n-k} = 1$, đây là điều kiện cần thiết của phân phối (2.32).

Nhận xét 2.11.

1. Thực hiện n phép thử Béc–nu–li với xác suất thành công của sự kiện A trong mỗi lần thử là p. Với mỗi $i = 1,2,\ldots,n$, nếu ở lần thử thứ i sự kiện A xuất hiện ta cho X_i nhận giá trị 1, nếu sự kiện A không xuất hiện ta cho X_i nhận giá trị 0. Như vậy $X_i \sim \mathcal{B}(1,p)$. Gọi X là số lần thành công trong n phép thử Béc–nu–li này thì
$$X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n,p). \tag{2.33}$$

2. Nếu $X \sim \mathcal{B}(n_1,p)$ và $Y \sim \mathcal{B}(n_2,p)$ và nếu X, Y độc lập thì $X+Y \sim \mathcal{B}(n_1+n_2,p)$ (xem Chương 3).

Định lý 2.13. Kỳ vọng và phương sai của biến ngẫu nhiên có phân phối nhị thức là

$$\boxed{E(X) = np, \quad V(X) = npq} \tag{2.34}$$

Chứng minh. Từ Nhận xét 2.11(1), biến ngẫu nhiên X xác định bởi (2.33) có phân phối nhị thức $\mathcal{B}(n;p)$. Theo Hệ quả 3.1, 3.3 (Chương 3) và Định lý 2.12,

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$$

và

$$V(X) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n pq = npq.$$

□

Ví dụ 2.28. Tỷ lệ phế phẩm của lô hàng là 4%. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Gọi X là số phế phẩm phát hiện được. (a) X có phân phối gì? (b) Tính xác suất có đúng 5 phế phẩm phát hiện được. (c) Lô hàng được xem là đạt tiêu chuẩn nếu số phế phẩm phát hiện được không nhiều hơn 2. Tính xác suất để lô hàng đạt tiêu chuẩn.

Lời giải Ví dụ 2.28 Có thể xem việc kiểm tra chất lượng mỗi sản phẩm là thực hiện một phép thử Béc–nu–li với sự thành công của phép thử là phát hiện ra phế phẩm. Theo giả thiết xác suất thành công của mỗi lần thử là 0,04. Kiểm tra 20 sản phẩm là thực hiện 20 phép thử.

(a) Số phế phẩm phát hiện được là số lần thành công trong 20 phép thử này. Vậy X có phân phối nhị thức $X \sim \mathcal{B}(n,p)$, với $n = 20$, $p = 0,04$.

(b) $P(X = 5) = C_{20}^5 \times (0,04)^5 \times (0,96)^{15} = 0,0008$.

(c) Xác suất để lô hàng đạt tiêu chuẩn là

$$\begin{aligned} P(X \le 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= C_{20}^0(0,04)^0(0,96)^{20} + C_{20}^1(0,04)^1(0,96)^{19} + C_{20}^2(0,04)^2(0,96)^{18} \simeq 0,956. \end{aligned}$$

2.4.3 Phân phối Poa–xông

Các phép thử mang lại các giá trị số cho biến ngẫu nhiên X, chỉ số các kết quả xảy ra trong một khoảng thời gian nhất định nào đó, được gọi là phép thử Poa-xông. Khoảng thời gian nhất định có thể là một phút, một ngày, thậm trí một năm. Chẳng hạn phép thử đếm số cuộc điện thoại gọi đến một tổng đài trong vòng 5 phút là một phép thử Poa-xông.

Một phép thử Poa-xông có nguồn gốc từ quá trình Poa-xông và có các tính chất sau.

1. Số lượng kết quả xảy ra trong một khoảng thời gian không phụ thuộc vào số lượng kết quả xảy ra trong bất kỳ khoảng thời gian nào khác. Do đó, quá trình Poa-xông có tính chất không nhớ.

2. Xác suất xảy ra một kết quả trong một khoảng thời gian ngắn tỷ lệ thuận với độ dài của khoảng thời gian và không phụ thuộc vào số lượng kết quả xảy ra bên ngoài khoảng thời gian này.

3. Xác suất có nhiều hơn một kết quả sẽ xảy ra trong một khoảng thời gian ngắn là không đáng kể.

Định nghĩa 2.17 (Phân phối Poa–xông)**.** Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối Poa-xông (Poisson distribution) với tham số λ, ký hiệu là $X \sim \mathcal{P}(\lambda)$, nếu X có bảng phân phối xác suất

X	0	1	...	k	...	n	...
$P(X=k)$	$\frac{\lambda^0}{0!}e^{-\lambda}$	$\frac{\lambda^1}{1!}e^{-\lambda}$	...	$\frac{\lambda^k}{k!}e^{-\lambda}$	...	$\frac{\lambda^n}{n!}e^{-\lambda}$	...

(2.35)

trong đó $P(X=k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$ được tính bằng công thức Poa-xông, λ là số kết quả trung bình trên mỗi đơn vị thời gian và $e = 2,71828\ldots$

Định lý 2.14. Giả sử X là biến ngẫu nhiên có phân phối Poa-xông tham số λ. Khi đó,

1. Hàm phân phối xác suất của X là $F_X(x) = e^{-\lambda} \sum_{k=0}^{n} \dfrac{\lambda^k}{k!}$, $n < x \leq n+1$.
2. Kỳ vọng $E(X) = \lambda$.
3. Phương sai $V(X) = \lambda$ và độ lệch chuẩn $\sigma(X) = \sqrt{\lambda}$.

Chứng minh. Ta sẽ chứng minh hai tính chất đầu của định lý.

1. Suy trực tiếp từ Định nghĩa (2.4) và Định nghĩa 2.17.
2. Sử dụng Định nghĩa 2.8
$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda,$$
vì $\sum_{k=1}^{\infty} \dfrac{\lambda^{k-1}}{(k-1)!} = \sum_{l=0}^{\infty} \dfrac{\lambda^l}{(l)!} = e^{\lambda}$.

□

Nhận xét 2.12. 1. Phân phối Poa–xông được ứng dụng rộng rãi trong nhiều lĩnh vực thực tế như kiểm tra chất lượng sản phẩm, lý thuyết sắp hàng, các hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài …

2. Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân phối Poa–xông tham số lần lượt λ_1, λ_2, thì $X_1 + X_2$ cũng có phân phối Poa–xông tham số $\lambda_1 + \lambda_2$ (xem Chương 3).
3. Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm sau: số cuộc gọi đến một tổng đài; số khách hàng đến một điểm phục vụ; số xe cộ qua một ngã tư; số tai nạn (xe cộ); số các sự cố xảy ra ở một địa điểm … trong một khoảng thời gian xác định nào đó sẽ có phân phối Poa–xông với tham số λ, trong đó λ là tốc độ trung bình diễn ra trong khoảng thời gian này.

Ví dụ 2.29. Ở một tổng đài bưu điện, các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau với tốc độ trung bình 2 cuộc gọi trong một phút. Tìm xác suất để (a) Có đúng 5 cuộc điện thoại trong vòng 2 phút; (b) Không có cuộc điện thoại nào trong khoảng thời gian 30 giây; (c) Có ít nhất 1 cuộc điện thoại trong khoảng thời gian 10 giây.

Lời giải Ví dụ 2.29

(a) Gọi X là số cuộc điện thoại xuất hiện trong vòng 2 phút. $X \sim \mathcal{P}(\lambda)$, λ chính là số cuộc điện thoại trung bình đến trong vòng 2 phút, $\lambda = 4$.
$P(X=5) = e^{-\lambda}\frac{\lambda^5}{5!} = e^{-4}\frac{4^5}{5!} = 0,156$.

(b) Gọi X là số cuộc điện thoại xuất hiện trong vòng 30 giây. $X \sim \mathcal{P}(\lambda)$ với $\lambda = 1$.
$P(X=0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-1} = 0,3679$.

(c) Gọi X là số cuộc điện thoại xuất hiện trong vòng 10 giây. $X \sim \mathcal{P}(\lambda)$ với $\lambda = 1/3$.

$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/3} = 0,2835$.

Ví dụ 2.30. Một ga ra cho thuê ô tô thấy rằng số người đến thuê ô tô vào thứ bảy cuối tuần là một biến ngẫu nhiên tuân theo luật phân phối Poa-xông với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ô tô. (a) Tìm xác suất để tất cả 4 ô tô đều được thuê vào thứ 7. (b) Tìm xác suất gara không đáp ứng được yêu cầu (thiếu xe cho thuê) vào thứ 7. (c) Trung bình có bao nhiêu ô tô được thuê vào ngày thứ 7?

Lời giải Ví dụ 2.30 Gọi X là biến ngẫu nhiên chỉ "số người đến thuê ô tô vào thứ bảy". Theo giả thiết X là biến ngẫu nhiên phân phối tuân theo quy luật Poa-xông $\mathcal{P}(\lambda)$ với $\lambda = 2$. Gọi Y là biến ngẫu nhiên chỉ "số xe ô tô được thuê vào thứ bảy".

(a) Áp dụng phân phối Poa-xông, $P(X = k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$,

$$\begin{aligned} P(Y = 4) = P(X \geq 4) = 1 - P(X < 4) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \\ &= 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right) = 0,1429. \end{aligned}$$

(b) $P(X > 4) = P(X \geq 4) - P(X = 4) = 0,1429 - e^{-2}\dfrac{2^4}{4!} = 0,0527$.

(c) Y có thể nhận các giá trị 0, 1, 2, 3, 4, với

$$\begin{aligned} &P(Y = 0) = P(X = 0) = 0,1353, \quad P(Y = 1) = P(X = 1) = 0,2707, \\ &P(Y = 2) = P(X = 2) = 0,2707, \quad P(Y = 3) = P(X = 3) = 0,1804, \\ &P(Y = 4) = P(X \geq 4) = 0,1429. \end{aligned}$$

Bảng phân phối xác suất của Y là:

x	0	1	2	3	4
$P(Y = x)$	0,1353	0,2707	0,2707	0,1804	0,1429

Vậy trung bình số ôtô được thuê trong ngày thứ bảy là $E(Y) = 1,9249$, tức là khoảng 2 chiếc.

Chú ý 2.3. Giá trị xác suất của phân phối Poa–xông được tính sẵn trong bảng Phụ lục 5.

2.4.4 Xấp xỉ phân phối nhị thức bằng phân phối Poa-xông

Trong Mục 1.5.5 ta đã đề cập đến việc tính xấp xỉ công thức Béc–nu–y (1.19) khi số phép thử n khá lớn bởi công thức (1.21). Ở đây ta xét mối liên hệ của hai phân phối tương ứng.

Định lý 2.15. Cho X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n, p)$. Nếu $n \to \infty$, $p \to 0$ và $np \to \lambda$ (λ là một hằng số) thì

$$\mathcal{B}(n, p) \to \mathcal{P}(\lambda) \quad \text{khi} \quad n \to \infty.$$

Trong thực tế nếu n đủ lớn và $\lambda = np$ đủ nhỏ (thỏa mãn $np < 7$) thì ta có thể xấp xỉ phân phối nhị thức $\mathcal{B}(n, p)$ bằng phân phối Poa-xông $\mathcal{P}(\lambda)$ và

$$\boxed{P_n(k) = C_n^k (p)^k (1-p)^{n-k} \simeq \frac{(np)^k}{k!} e^{-np}} \tag{2.36}$$

Ví dụ 2.31. Giả sử một công ty bảo hiểm nhân thọ bảo hiểm cho cuộc sống của 5000 người đàn ông ở độ tuổi 42. Nghiên cứu của các chuyên gia tính toán cho thấy xác suất để một người đàn ông 42 tuổi sẽ chết trong một năm (xác định) là 0,001. Hãy tìm xác suất mà công ty sẽ phải trả bảo hiểm cho 4 người trong một năm (xác định).

Lời giải Ví dụ 2.31 Gọi X là số người chết trong một năm (xác định). X là biến ngẫu nhiên có phân phối nhị thức tham số $n = 5000$ và $p = 0,001$. Khi đó,

$$P(X = 4) = P_{5000}(4) = C_{5000}^4 (0,001)^4 (1 - 0,001)^{5000-4} = \frac{5000!}{4!4996!} (0,001)^4 (0,999)^{4996}.$$

Vì $n = 5000$ đủ lớn và $\lambda = np = (5000)(0,001) = 5$ nên xác suất trên có thể được xấp xỉ bằng công thức Poa-xông:

$$P(X = 4) \simeq \frac{5^4}{4!} e^{-5} = \frac{(625)(0,006738)}{24} = 0,175.$$

BÀI 8 (2 tiết)

2.4.5 Phân phối mũ

Định nghĩa 2.18 (Phân phối mũ)**.** Biến ngẫu nhiên X được gọi là tuân theo luật phân phối mũ (exponential distribution) với tham số $\lambda > 0$ nếu nó có hàm mật độ xác suất có dạng:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{2.37}$$

Ký hiệu: $X \sim \mathcal{E}(\lambda)$.

Ví dụ 2.32. Biến ngẫu nhiên T có phân phối mũ với hàm phân phối xác suất

$$F_T(t) = \begin{cases} 1 - e^{-t/3}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

(a) Tìm hàm mật độ xác suất của T. (b) Tính $P[2 \leq T \leq 4]$. (c) Tính kỳ vọng, phương sai, độ lệch chuẩn của T.

Lời giải Ví dụ 2.32

(a)

$$f_T(t) = \frac{dF_T(t)}{dt} = \begin{cases} \dfrac{1}{3}e^{-t/3}, & t \geq 0, \\ 0, & \text{trái lại}. \end{cases}$$

Theo Định nghĩa 2.18, biến ngẫu nhiên T có phân phối mũ với tham số $\lambda = 1/3$.

(b) $P[2 \leq T \leq 4] = F_T(4) - F_T(2) = e^{-2/3} - e^{-4/3} = 0,250$.

(c) Sử phương pháp tích phân từng phần,

$$E(T) = \int_{-\infty}^{+\infty} t f_T(t)dt = \int_0^{+\infty} t\frac{1}{3}e^{-t/3}dt = -te^{-t/3}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-t/3}dt = 3.$$

Để tính phương sai của T, ta tính $E(T^2) = \int_{-\infty}^{+\infty} t^2 f_T(t)dt = \int_0^{+\infty} t^2\frac{1}{3}e^{-t/3}dt$. Tích phân từng phần ta được

$$E(T^2) = -t^2e^{-t/3}\Big|_0^{+\infty} + \int_0^{+\infty} 2te^{-t/3}dt = 2\int_0^{+\infty} te^{-t/3}dt.$$

Vì $E(T) = 3$, nên $\int_0^{+\infty} te^{-t/3}dt = 3E(T) = 9$. Do đó $E(T^2) = 6E(T) = 18$ và

$$V(T) = E(T^2) - [E(T)]^2 = 18 - 3^2 = 9.$$

Độ lệch chuẩn là $\sigma(T) = \sqrt{V(T)} = 3$.

Định lý 2.16. Nếu X là biến ngẫu nhiên có phân phối mũ $\mathcal{E}(\lambda)$ thì

1. $F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & \text{trái lại.} \end{cases}$
2. $\mu = E(X) = 1/\lambda$.
3. $\sigma^2 = V(X) = 1/\lambda^2$.

Nhận xét 2.13. 1. $P(X > x) = 1 - P(X \leq x) = 1 - P(X < x) = 1 - F_X(x) = e^{-\lambda x}$.

2. Phân phối mũ có tính chất không nhớ: $P(X > t + s|X > t) = P(X > s)$. Vì

$$P(X > t + s|X > t) = \frac{P[(X > t + s)(X > t)]}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)}.$$

Chú ý 2.4. 1. Phân phối mũ có nhiều ứng dụng trong thực tiễn.

2. Nói chung với một giả thiết nào đó, khoảng thời gian giữa hai lần xuất hiện của một sự kiện E nào đó sẽ có phân phối mũ. Vì lý do này phân phối mũ còn có tên gọi là phân phối của thời gian chờ đợi "Waiting time distribution" (khoảng thời gian giữa 2 ca cấp cứu ở một bệnh viện, khoảng thời gian giữa 2 lần hỏng hóc của một chiếc máy, khoảng thời gian giữa 2 trận lụt hay động đất...).

Ví dụ 2.33. Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân phối mũ với kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Lời giải Ví dụ 2.33 Gọi X là tuổi thọ của mạch điện tử trong máy tính. X tuân theo phân phối mũ với tham số $\lambda = \dfrac{1}{E(X)} = \dfrac{1}{6,25} = 0,16$.

$P(X \leq 5) = 1 - P(X > 5) = 1 - e^{-5\lambda} = 1 - e^{-0,8} = 0,5506$. Vậy có khoảng 55,06% mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Ví dụ 2.34. Công ty điện thoại A thu phí 0,15$ mỗi phút cho các cuộc gọi điện thoại. Với bất kỳ cuộc gọi nào trong vòng một phút, họ sẽ tính phí trong một phút. Công ty điện thoại B cũng tính phí 0,15$ mỗi phút. Tuy nhiên, công ty điện thoại B tính toán phí dựa trên thời lượng chính xác của một cuộc gọi. Cho T, thời lượng của một cuộc gọi tính bằng phút, là biến ngẫu nhiên có phân phối mũ với tham số $\lambda = 1/3$. (a) Hàm mật độ xác suất của T là gì? (b) Kỳ vọng của T là bao nhiêu? (c) Doanh thu trung bình cho mỗi cuộc gọi $E(R_A)$ và $E(R_B)$ của công ty A và B là bao nhiêu?

Lời giải Ví dụ 2.34

(a) Vì T có phân phối mũ với $\lambda = 1/3$, $f_T(t) = \begin{cases} \dfrac{1}{3}e^{-\frac{1}{3}t}, & t \geq 0, \\ 0, & \text{trái lại.} \end{cases}$

(b) Theo Định lý 2.16, $E[T] = \int_{-\infty}^{+\infty} t f_T(t)dt = \dfrac{1}{\lambda} = 3$ phút/cuộc gọi.

(c) Với công ty B, $E(R_B) = 0,15 \times E(T) = 0,45$ \$/cuộc gọi.

Với công ty A, đặt $K = [T]$ theo nghĩa nếu $0 < T \leq 1$ thì $K = 1$, nếu $1 < T \leq 2$ thì $K = 2, \ldots$. Khi đó $E(R_A) = 0,15 \times E(K)$. Để tính $E(K)$, trước hết ta tính

$$P(K = k) = P(k-1 < T \leq k) = \int\limits_{k-1}^{k} f_T(t)dt = (e^{-\lambda})^{k-1}(1 - e^{-\lambda}).$$

Suy ra $E(K) = \sum\limits_{k=1}^{\infty} kP(K = k) = \sum\limits_{k=1}^{\infty} k(1-p)^{k-1}p = \dfrac{1}{p}$ với $p = 1 - e^{-\lambda}$. Vậy

$$E(R_A) = \frac{0,15}{p} = \frac{0,15}{0,2834} = (0,15) \times (3,5285) = 0,5292 \quad \$/\text{cuộc gọi.}$$

2.4.6 Phân phối chuẩn

2.4.4a Phân phối chuẩn

Định nghĩa 2.19 (Phân phối chuẩn)**.** Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối chuẩn (normal distribution) với tham số μ, σ^2, ký hiệu là $X \sim \mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ xác suất của X có dạng

$$\boxed{f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}} \tag{2.38}$$

ở đây e và π được lấy xấp xỉ lần lượt là $2,71828$ và $3,14159$.

Định lý 2.17. Kỳ vọng và phương sai của biến ngẫu nhiên X có phân phối chuẩn

$$\boxed{E(X) = \mu, \quad V(X) = \sigma^2} \tag{2.39}$$

và độ lệch tiêu chuẩn là $\sigma(X) = \sigma$.

Chứng minh. Để xác định kỳ vọng, trước hết ta tính

$$E(X - \mu) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Đặt $z = (x-\mu)/\sigma$ và $dx = \sigma dz$, ta nhận được

$$E(X - \mu) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz = 0,$$

vì hàm số dưới dấu tích phân là hàm lẻ của z. Do đó,

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

Phương sai của biến ngẫu nhiên tuân theo luật phân phối chuẩn được cho bởi

$$E[(X-\mu)^2] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Đặt $z = (x-\mu)/\sigma$ và $dx = \sigma dz$, ta nhận được

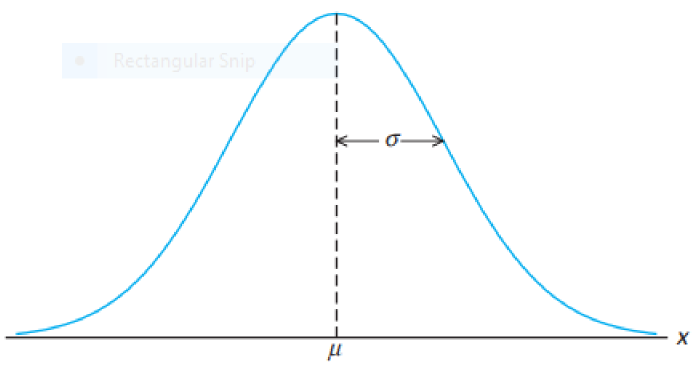
$$E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz.$$

Tích phân từng phần với $u = z$ và $dv = ze^{-z^2/2}dz$ suy ra $du = dz$ và $v = -e^{-z^2/2}$, ta tìm được

$$E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ze^{-z^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \right) = \sigma^2(0+1) = \sigma^2.$$

□

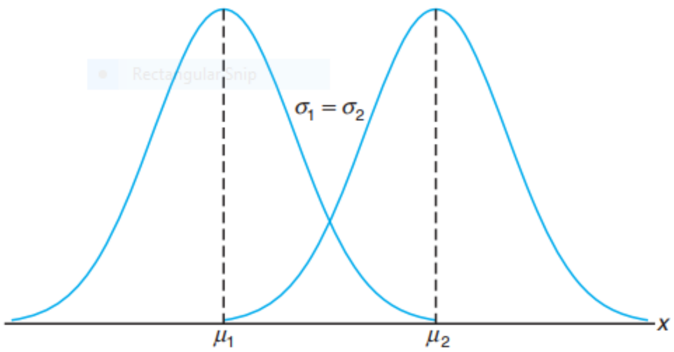
Nhận xét 2.14. Phân phối liên tục quan trọng nhất trong lĩnh vực thống kê là phân phối chuẩn. Đồ thị của hàm mật độ xác suất $f_X(x)$ của biến ngẫu nhiên tuân theo luật phân phối chuẩn, được gọi là đường cong chuẩn, có dạng hình chuông (xem Hình 2.9), mô tả gần đúng nhiều hiện tượng trong tự nhiên, công nghiệp và nghiên cứu.



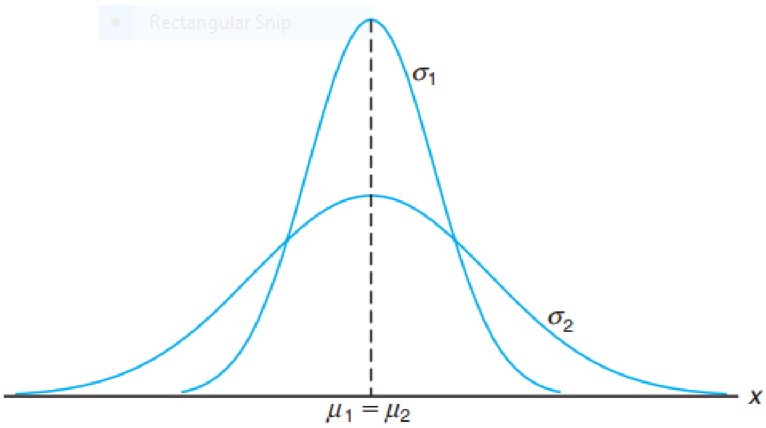
Hình 2.9: Đường cong chuẩn

Hình 2.10 mô tả hai đường cong chuẩn có cùng độ lệch chuẩn nhưng kỳ vọng khác nhau. Hai đường cong giống hệt nhau về hình thức nhưng được tập trung tại các vị trí khác nhau dọc theo trục hoành.

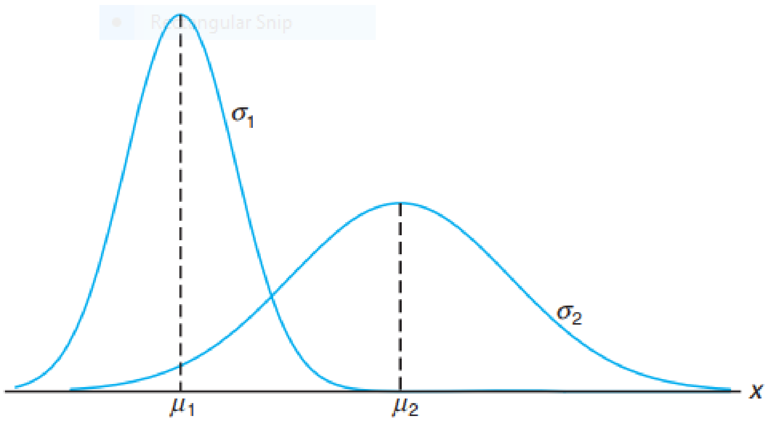
Hình 2.11 mô tả hai đường cong chuẩn có cùng kỳ vọng nhưng độ lệch chuẩn khác nhau. Hình 2.12 mô tả cho trường hợp kỳ vọng và độ lệch chuẩn khác nhau.



Hình 2.10: Đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 = \sigma_2$



Hình 2.11: Đường cong chuẩn với $\mu_1 = \mu_2$ và $\sigma_1 < \sigma_2$



Hình 2.12: Đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 < \sigma_2$

Định lý 2.18. Nếu X là biến ngẫu nhiên tuân theo luật phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$, thì biến ngẫu nhiên $Y = aX + b$ tuân theo luật phân phối chuẩn $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

Chú ý 2.5.

1. Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân phối chuẩn $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ thì $X_1 + X_2$ cũng có phân phối chuẩn $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (xem Chương 3).

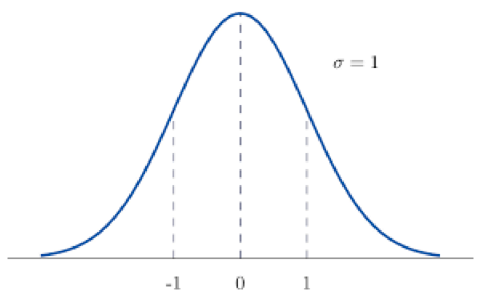
2. Nếu n biến ngẫu nhiên độc lập X_i cùng có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$, $i = 1, \dots, n$, thì (xem Chương 3).
$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

2.4.4b Phân phối chuẩn tắc

Định nghĩa 2.20 (Phân phối chuẩn tắc)**.** Phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với $\mu = 0$ và $\sigma = 1$ gọi là phân phối chuẩn tắc $\mathcal{N}(0, 1)$.

Nếu X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì
$$U = \frac{X - \mu}{\sigma} \tag{2.40}$$
là biến ngẫu nhiên có phân phối chuẩn tắc $\mathcal{N}(0, 1)$. Do đó các tính toán về X sẽ được quy về U.

Định nghĩa 2.21 (Hàm mật độ xác suất)**.** Hàm mật độ xác suất của biến ngẫu nhiên có phân phối chuẩn tắc là
$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}} \tag{2.41}$$
Đây là hàm Gau–xơ với các giá trị được tính sẵn trong Phụ lục 1.



Hình 2.13: Hàm mật độ của biến ngẫu nhiên có phân phối chuẩn tắc $\mathcal{N}(0, 1)$

Định nghĩa 2.22 (Hàm phân phối xác suất)**.** Hàm phân phối xác suất của biến ngẫu nhiên U phân phối chuẩn tắc là
$$\boxed{\Phi_U(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R}} \tag{2.42}$$

Giá trị của hàm $\Phi(x)$ được tính sẵn trong Phụ lục 3.

Hàm $\Phi(x)$ có tính chất sau.

Định lý 2.19.

$$\boxed{\Phi(-x) = 1 - \Phi(x)} \tag{2.43}$$

Chứng minh. Từ Định nghĩa 2.22,

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

□

2.4.4c Xác suất để biến ngẫu nhiên X có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ nhận giá trị trong khoảng (α, β)

Định lý 2.20. Nếu X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì

$$\boxed{P(\alpha < X < \beta) = \phi\left(\frac{\beta - \mu}{\sigma}\right) - \phi\left(\frac{\alpha - \mu}{\sigma}\right), \quad \alpha, \beta \in \mathbb{R}} \tag{2.44}$$

trong đó $\phi(x)$ là hàm số Láp–la–xơ xác định bởi (1.24).

Chứng minh. Sử dụng phép đổi biến $t = \dfrac{x - \mu}{\sigma}$ ta nhận được

$$\begin{aligned} P(\alpha < X < \beta) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-\mu}{\sigma}}^{\frac{\beta-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-\mu}{\sigma}}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\beta-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\beta-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\alpha-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt. \end{aligned}$$

Từ đây và (1.24) ta nhận được (2.44). □

Hệ quả 2.3.

1. $P(X < \beta) = P(-\infty < X < \beta) = 0,5 + \phi\left(\dfrac{\beta - \mu}{\sigma}\right)$.
2. $P(X > \alpha) = P(\alpha < X < \infty) = 0,5 - \phi\left(\dfrac{\alpha - \mu}{\sigma}\right)$.
3. $P(|X - \mu| < t\sigma) = 2\phi(t)$.

Định lý 2.21. Nếu X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì hàm phân phối xác suất của X là

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

Khi đó, xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng (α, β) là

$$\boxed{P(\alpha < X < \beta) = \Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right)} \tag{2.45}$$

Chú ý 2.6. 1. Các giá trị của hàm Láp–la–xơ (1.24) được tính trong bảng Phụ lục 2 (xem Mục 1.5.5) đối với các giá trị x dương. Hàm $\phi(x)$ là hàm lẻ, tức là $\phi(-x) = -\phi(x)$. Khi $x > 5$ ta có thể lấy $\phi(x) \simeq 0,5$.

2. Hàm phân phối xác suất của biến ngẫu nhiên phân phối chuẩn tắc $\Phi(x)$ xác định bởi (2.42) và hàm Láp–la–xơ $\phi(x)$ xác định bởi (1.24) có mối liên hệ:

$$\boxed{\Phi(x) = 0,5 + \phi(x), \quad x \geq 0} \tag{2.46}$$

Các giá trị của hàm phân phối chuẩn tắc $\Phi(x)$ được tính sẵn trong bảng Phụ lục 3 đối với các giá trị x dương.

3. Nếu $X \sim \mathcal{N}(\mu, \sigma^2)$ thì các công thức (2.44) và (2.45) là tương đương.

2.4.4d Quy tắc 3σ

Từ Hệ quả 2.3(3) suy ra xác suất để độ lệch tuyệt đối của biến ngẫu nhiên $X \sim \mathcal{N}(\mu, \sigma^2)$ khỏi trị trung bình của nó bé hơn $\varepsilon = t\sigma$ là

$$\boxed{P(|X-\mu| < \varepsilon) = 2\phi\left(\frac{\varepsilon}{\sigma}\right) \quad \text{hay} \quad P(|X-\mu| < t\sigma = \varepsilon) = 2\phi(t)} \tag{2.47}$$

Thay $t = 1, 2, 3$, tra bảng giá trị hàm số Láp–la–xơ (Phụ lục 2) ta nhận được

$$\begin{aligned} P(|X-\mu| < \sigma) &= 2\phi(1) = 0,6827, \\ P(|X-\mu| < 2\sigma) &= 2\phi(2) = 0,9545, \\ P(|X-\mu| < 3\sigma) &= 2\phi(3) = 0,9973. \end{aligned} \tag{2.48}$$

Quy tắc 3σ được phát biểu như sau: *Hầu chắc chắn rằng (với độ tin cậy 0,9973) X có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ lấy giá trị trong khoảng $(\mu - 3\sigma, \mu + 3\sigma)$.*

Trong thực tế, quy tắc 3σ được áp dụng như sau: Nếu quy luật phân phối xác suất của biến ngẫu nhiên được nghiên cứu chưa biết, song nó thỏa mãn điều kiện của Quy tắc 3σ thì có thể xem như nó là biến ngẫu nhiên có phân phối chuẩn.

Chú ý 2.7. 1. Phân phối chuẩn được Gao–xơ tìm ra năm 1809 nên nó còn được gọi là phân phối Gao–xơ.

2. Phân phối chuẩn thường được sử dụng trong các bài toán đo đạc các đại lượng vật lý, thiên văn ...

3. Trong thực tế, nhiều biến ngẫu nhiên tuân theo phân phối chuẩn hoặc tiệm cận chuẩn. Chẳng hạn, trọng lượng, chiều cao của một nhóm người nào đó; điểm thi của thí sinh; năng suất cây trồng; mức lãi suất của một công ty; nhu cầu tiêu thụ của một mặt hàng nào đó; nhiễu trắng trên các kênh thông tin ... là các biến ngẫu nhiên có phân phối chuẩn.

Ví dụ 2.35. Lãi suất (%) đầu tư vào một dự án trong năm 2018 được coi như một biến ngẫu nhiên tuân theo quy luật chuẩn. Theo đánh giá của ủy ban đầu tư thì với xác suất 0,1587 cho lãi suất lớn hơn 20% và với xác suất 0,0228 cho lãi suất lớn hơn 25%. Vậy khả năng đầu tư mà không bị lỗ là bao nhiêu?

Lời giải Ví dụ 2.35 Gọi X là lãi suất (%) của dự án trong năm 2018. Khi đó X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$. Theo đầu bài ta có

$$P(X > 20) = P(20 < X < +\infty) = 0,5 - \phi\left(\frac{20-\mu}{\sigma}\right) = 0,1587$$

và

$$P(X > 25) = P(25 < X < +\infty) = 0,5 - \phi\left(\frac{25-\mu}{\sigma}\right) = 0,0228.$$

Từ bảng giá trị hàm số Láp–la–xơ (Phụ lục 2) suy ra $\dfrac{20-\mu}{\sigma} = 1$ và $\dfrac{25-\mu}{\sigma} = 2$. Hay $\mu = 15$, $\sigma = 5$. Vậy khả năng đầu tư không bị lỗ là

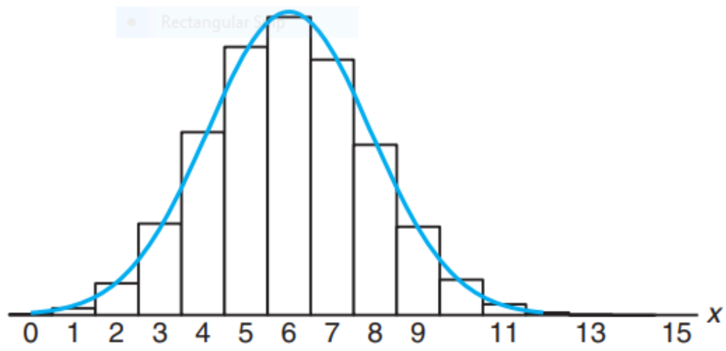
$$P(X \geq 0) = 0,5 + \phi(3) = 0,5 + 0,49865 = 0,99865.$$

2.4.4e Xấp xỉ phân phối nhị thức bởi phân phối chuẩn

Trong Mục 1.5.5 ta đã đề cập đến việc xấp xỉ công thức Béc–nu–y (1.20) bởi công thức (3.40) khi số phép thử n khá lớn. Ở đây ta xét chi tiết về mối liên hệ giữa phân phối nhị thức và phân phối chuẩn. Phân phối chuẩn có thể dùng xấp xỉ khá tốt cho một số phân phối rời rạc. Ta có định lý sau đây mang tên là Định lý Moa-vrơ–Lap-la-xơ.

Định lý 2.22. Giả sử X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n;p)$. Nếu $np > 5$ và $n(1-p) > 5$ thì X có phân phối xấp xỉ phân phối chuẩn với tham số $\mu = np$, $\sigma^2 = np(1-p)$.

Phân phối chuẩn với kỳ vọng $\mu = np$ và phương sai $\sigma^2 = np(1-p)$ không chỉ xấp xỉ khá tốt cho phân phối nhị thức khi n khá lớn và xác suất p không quá gần 0 hoặc 1 mà còn cung cấp một xấp xỉ khá tốt cho phân phối nhị thức ngay cả khi n nhỏ và p gần 1/2. Để minh họa việc xấp xỉ phân phối chuẩn cho phân phối nhị thức, ta vẽ biểu đồ của $\mathcal{B}(15;0,4)$ và vẽ đường cong chuẩn có cùng kỳ vọng $\mu = np = 15 \times 0,4 = 6$ và phương sai $\sigma^2 = np(1-p) = 15 \times 0,4 \times 0,6 = 3,6$ với biến ngẫu nhiên tuân theo luật phân phối nhị thức X (xem Hình 2.14).



Hình 2.14: Xấp xỉ phân phối chuẩn cho phân phối nhị thức $\mathcal{B}(15; 0,4)$

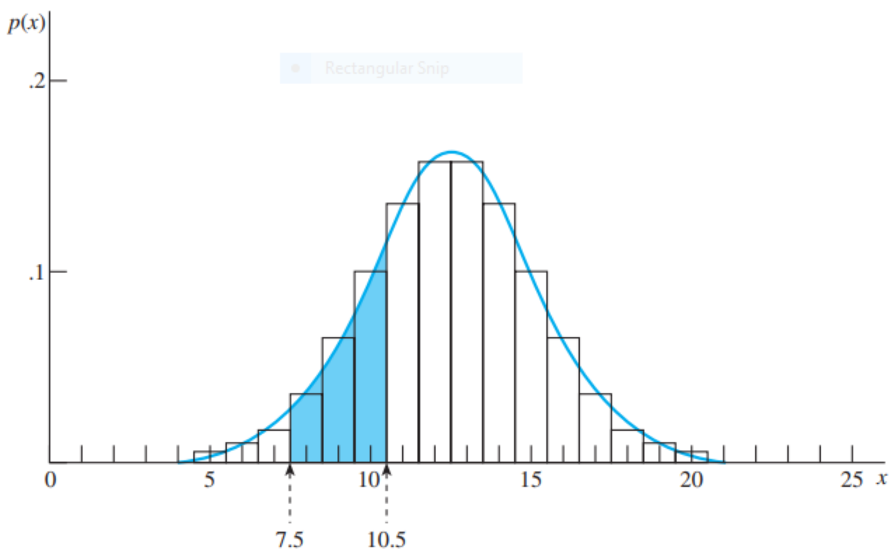
Trong hình minh họa về xấp xỉ phân phối nhị thức bởi phân phối chuẩn, vì ta xấp xỉ một phân phối rời rạc bằng một phân phối liên tục, nên cần một sự hiệu chỉnh để giảm sai số.

Định lý 2.23. Cho X là biến ngẫu nhiên tuân theo luật phân phối nhị thức $\mathcal{B}(n; p)$. Phân phối xác suất của X được xấp xỉ bởi phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với $\mu = np$ và $\sigma^2 = np(1-p)$ và

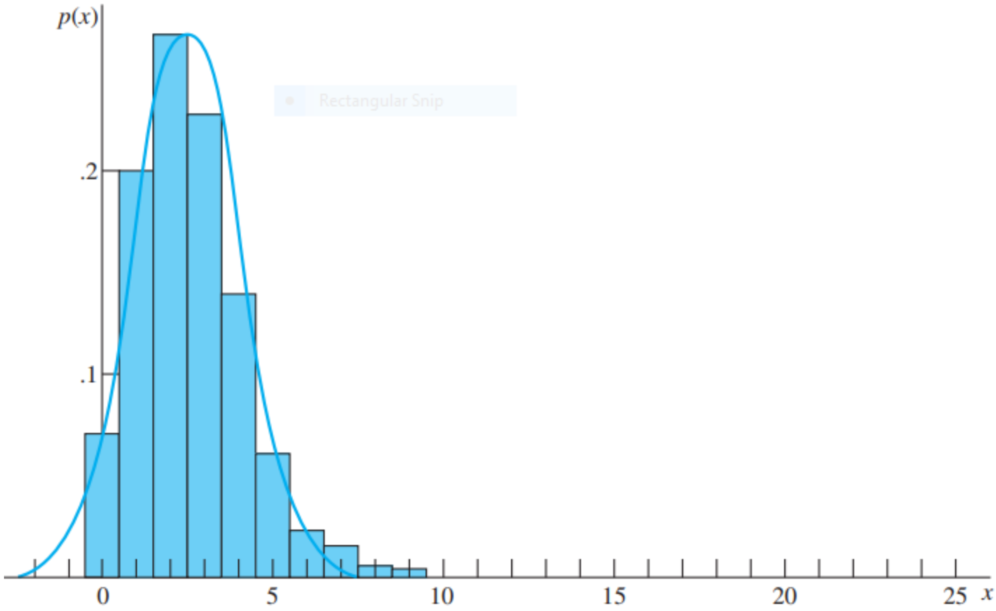
$$\boxed{P(X = k) = C_n^k p^k (1-p)^{n-k} \simeq \phi\left(\frac{k+0,5-\mu}{\sigma}\right) - \phi\left(\frac{k-0,5-\mu}{\sigma}\right)} \tag{2.49}$$

$$\boxed{P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k} \simeq \phi\left(\frac{k_2+0,5-\mu}{\sigma}\right) - \phi\left(\frac{k_1-0,5-\mu}{\sigma}\right)} \tag{2.50}$$

Nhận xét 2.15. Hình 2.15 và 2.16 biểu thị biểu đồ xác suất nhị thức với $n = 25$ và $p = 0,5$, $p = 0,1$ tương ứng. Phân phối trong Hình 2.15 là hoàn toàn đối xứng.



Hình 2.15: Phân phối nhị thức với $n = 25$ và $p = 0,5$ xấp xỉ bởi phân phối chuẩn với $\mu = 12,5$ và $\sigma = 2,5$



Hình 2.16: Phân phối nhị thức và xấp xỉ phân phối chuẩn với $n = 25$ và $p = 0,1$

Việc thêm $+0,5$ và $-0,5$ chính là yếu tố hiệu chỉnh và gọi là hiệu chỉnh liên tục.

Ví dụ 2.36. Sử dụng phân phối chuẩn xấp xỉ xác suất $X = 8, 9$, hoặc 10 cho biến ngẫu nhiên X tuân theo luật phân phối nhị thức với $n = 25$ và $p = 0,5$. So sánh với công thức tính chính xác.

Lời giải Ví dụ 2.36 Vì X là biến ngẫu nhiên tuân theo luật phân phối nhị thức với $n = 25$ và $p = 0,5$,

$$P(X = 8) + P(X = 9) + P(X = 10) = \left(C_{25}^{8} + C_{25}^{9} + C_{25}^{10}\right) \times (0,5)^{25} \simeq 0,190535.$$

Sử dụng công thức xấp xỉ (3.42) với $\mu = np = 12,5$, $\sigma = \sqrt{npq} = 2,5$ ta nhận được

$$P(8 \le X \le 10) \simeq \phi(-0.8) - \phi(-2) = 0,18911.$$

Giá trị xấp xỉ 0,18911 với giá trị thực 0,190535 là khá gần nhau.

Ví dụ 2.37. Kiểm tra chất lượng 1000 sản phẩm với tỷ lệ chính phẩm 0,95. Tìm xác suất để số chính phẩm trong lô kiểm tra từ 940 đến 960.

Lời giải Ví dụ 2.37 Gọi X là biến ngẫu nhiên chỉ số chính phẩm trong lô sản phẩm kiểm tra, ta có $X \sim \mathcal{B}(1000; 0,95)$. Với $n = 1000$, $p = 0,95$, ta có $np = 950$ và $np(1-p) = 47,5$ đủ lớn nên ta xấp xỉ bởi $X \sim \mathcal{N}(950; 47,5)$:

$$\begin{aligned} P(940 \le X \le 960) &= \phi\left(\frac{960 + 0,5 - 950}{\sqrt{47,5}}\right) - \phi\left(\frac{940 - 0,5 - 950}{\sqrt{47,5}}\right) \\ &= \phi(1,52) - \phi(-1,52) = 2\phi(1,52) = 0,8716. \end{aligned}$$

2.4.7 Phân phối khi bình phương

Định nghĩa 2.23 (Phân phối khi bình phương)**.** Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối khi bình phương với n bậc tự do, ký hiệu là $X \sim \chi^2_n$, nếu hàm mật độ xác suất của X có dạng

$$\boxed{f_X(x) = \frac{1}{2}e^{-\frac{x}{2}}\frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)}, \quad x > 0,\ n > 0} \tag{2.51}$$

ở đây

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt, \quad x > 0,$$

là hàm Gamma (đã đề cập trong Giải tích 2).

Định nghĩa sau cho cách nhận biết một biến ngẫu nhiên có phân phối khi bình phương xuất phát từ n biến ngẫu nhiên độc lập có cùng phân phối chuẩn tắc.

Định nghĩa 2.24. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối chuẩn tắc $\mathcal{N}(0,1)$ thì

$$\boxed{U_n = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2_n} \tag{2.52}$$

(U_n có phân phối khi bình phương với n bậc tự do).

Kỳ vọng và phương sai của biến ngẫu nhiên U_n:

$$\boxed{E(U_n) = n, \quad V(U_n) = 2n} \tag{2.53}$$

Tính chất 2.1.

1. Nếu X_1 và X_2 là hai biến ngẫu nhiên độc lập có phân phối khi bình phương với n_1, n_2 bậc tự do thì biến ngẫu nhiên $X_1 + X_2$ có phân phối khi bình phương với $n_1 + n_2$ bậc tự do (xem Chương 3).

2. Biến ngẫu nhiên $\dfrac{U_n - n}{\sqrt{2n}}$ có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0,1)$ khi n đủ lớn.

3. Một hệ quả quan trọng được dùng nhiều trong thống kê (xem Chương 3): Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ và

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

thì

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sim \chi^2_{(n-1)}. \tag{2.54}$$

Việc tính toán với phân phối χ^2_n đưa về việc sử dụng bảng Phụ lục 4.

2.4.8 Phân phối Student

Định nghĩa 2.25 (Phân phối Student)**.** Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối Student với n bậc tự do, ký hiệu là $X \sim t(n)$, nếu hàm mật độ xác suất của X có dạng

$$\boxed{f_X(x) = \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}, \quad -\infty < x < +\infty} \tag{2.55}$$

ở đây $\Gamma(x)$ là hàm Gamma.

Để nhận biết một biến ngẫu nhiên có phân phối Student ta sử dụng định nghĩa sau.

Định nghĩa 2.26. Nếu X và Y là hai biến ngẫu nhiên độc lập tuân theo luật $\mathcal{N}(0,1)$ và χ^2_n tương ứng thì

$$\boxed{T_n = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n)} \tag{2.56}$$

(T_n có phân phối Student với n bậc tự do).

Kỳ vọng và phương sai của biến ngẫu nhiên T_n có phân phối Student:

$$\boxed{E(T_n) = 0,\ n > 1, \quad V(T_n) = \frac{n}{n-2}, \quad n > 2} \tag{2.57}$$

Tính chất 2.2. Biến ngẫu nhiên T_n có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0,1)$ khi n đủ lớn.

Nhận xét 2.16.

1. Phân phối Student có cùng dạng và tính đối xứng như phân phối chuẩn nhưng nó phản ánh tính biến đổi của phân phối sâu sắc hơn. Phân phối chuẩn không thể dùng để xấp xỉ phân phối khi mẫu có kích thước nhỏ. Trong trường hợp này ta dùng phân phối Student.

2. Khi bậc tự do n tăng lên ($n \geq 30$) thì phân phối Student tiến nhanh về phân phối chuẩn. Do đó khi $n \geq 30$ ta có thể dùng phân phối chuẩn thay thế cho phân phối Student.

3. Một hệ quả quan trọng được dùng nhiều trong thống kê (xem Chương 3): Nếu $X_0, X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn tắc $\mathcal{N}(0,1)$ thì

$$\frac{X_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}} \sim t(n). \tag{2.58}$$

2.4.9 Phân phối Fisher

Định nghĩa 2.27. Cho $X_1, X_2, \ldots, X_n$ và $Y_1, Y_2, \ldots, Y_m$ là $n+m$ biến ngẫu nhiên độc lập, trong đó $X_i \sim \mathcal{N}(0;1)$ và $Y_j \sim \mathcal{N}(0;1)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Khi đó biến ngẫu nhiên

$$F_{n,m} = \frac{m \sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{m} Y_i^2} \tag{2.59}$$

có phân phối Fisher với (n, m) bậc tự do (xem Chương 3).

2.5 Tổng hợp một số đề thi

Ví dụ 2.38 (Đề thi MI2020 kỳ 20183)**.** Một kiện hàng có 10 sản phẩm, trong đó có 7 sản phẩm loại I và 3 sản phẩm loại II. Tiền lãi khi bán được mỗi sản phẩm loại I là 50 nghìn đồng, mỗi sản phẩm loại II là 20 nghìn đồng.

(a) Ngày thứ nhất lấy ngẫu nhiên từ kiện hàng ra 3 sản phẩm và đã bán hết cả 3 sản phẩm đó. Tìm kỳ vọng của số tiền lãi thu được.

(b) Ngày thứ hai lấy ngẫu nhiên từ kiện hàng ra 2 sản phẩm. Tính xác suất để thu được 100 nghìn đồng tiền lãi khi bán 2 sản phẩm này.

Lời giải Ví dụ 3.24

(a) Gọi X là "số tiền lãi thu được", X nhận các giá trị 60, 90, 120, 150. Khi đó,

$$E(X) = 60 \times \frac{1}{120} + 90 \times \frac{21}{120} + 120 \times \frac{63}{120} + 150 \times \frac{35}{120} = 123.$$

(b) Gọi A là sự kiện "ngày thứ hai thu được 100 nghìn đồng tiền lãi khi bán 2 sản phẩm"; A_i: "ngày thứ nhất lấy được i sản phẩm loại I", $i = 0, 1, 2, 3$; A_0, A_1, A_2, A_3 lập thành hệ đầy đủ và $P(A) = P(A_0)P(A|A_0) + P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + P(A_3)P(A|A_3)$. Khi đó,

$$P(A) = \frac{1}{120} \times \frac{21}{21} + \frac{21}{120} \times \frac{15}{21} + \frac{63}{120} \times \frac{10}{21} + \frac{35}{120} \times \frac{6}{21} = \frac{7}{15} \simeq 0,4667.$$

Ví dụ 2.39 (Đề thi MI2020 giữa kỳ 20191)**.** Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} kx^2(1-x), & \text{nếu} \quad x \in [0,1], \\ 0, & \text{nếu} \quad x \notin [0,1]. \end{cases}$$

(a) Tìm hằng số k.

(b) Tính xác suất để sau 3 lần lặp lại phép thử một cách độc lập có đúng 1 lần X nhận giá trị trong khoảng $\left(0; \frac{1}{2}\right)$.

Lời giải Ví dụ 3.25

(a) Sử dụng Tính chất 2.3(a),(c) tính được $k = 12$.

(b) $P\left(0 < X < \dfrac{1}{2}\right) = \displaystyle\int_0^{\frac{1}{2}} 12(x^2 - x^3)dx = \frac{5}{16} = 0,3125.$

Vậy, $P_3(1) = C_3^1 \times p^1 \times (1-p)^2 = C_3^1 \times (0,3125)^1 \times (0,6875)^2 = \dfrac{1815}{4096} \simeq 0,44312.$

Ví dụ 2.40 (Đề thi MI2020 kỳ 20191)**.** Cho hàm mật độ xác suất $f_X(x) = \begin{cases} 3e^{-3x}, & \text{nếu } x \geq 0 \\ 0, & \text{nếu } x < 0 \end{cases}$ của biến ngẫu nhiên liên tục X và định nghĩa $Y = [X]$ là số nguyên lớn nhất không vượt quá X (nghĩa là $[x] = 0$ nếu $0 \leq x < 1$, $[x] = 1$ nếu $1 \leq x < 2\ldots$).

(a) Tính $P(Y = 0)$.

(b) Tính $E(Y)$.

Lời giải Ví dụ 2.40 (a) $P(Y=0) = P(0 \leq X < 1) = \displaystyle\int_0^1 3e^{-3x}dx = 1 - e^{-3}$.

(b) Với $k \geq 0$, $P(Y = k) = e^{-3k}(1 - e^{-3})$ và $E(X) = \displaystyle\sum_{k=0}^{\infty} kP(Y=k) = \frac{1}{e^3 - 1}$.

Ví dụ 2.41 (Đề thi MI2020 kỳ 20191)**.** Số khách hàng đến một cửa hàng bán lẻ là một biến ngẫu nhiên có phân phối Poa-xông với trung bình 6 khách hàng đến trong vòng một giờ. Nếu có đúng 5 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:00 thì xác suất để có ít nhất 8 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:30 là bao nhiêu?

Lời giải Ví dụ 2.41 Gọi X là "số khách hàng đến cửa hàng bán lẻ trong vòng 30 phút". Khi đó X là biến ngẫu nhiên có phân phối Poa-xông, $X \sim \mathcal{P}(\lambda)$, với $\lambda = 3$. Xác suất cần tìm $P(X \geq 3)$.

$$\begin{aligned}
P(X \geq 3) = 1 - P(X < 3) &= 1 - \Big[P(X=0) + P(X=1) + P(X=2)\Big] \\
&= 1 - e^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right] \\
&= 1 - 0,42319 = 0,57681.
\end{aligned}$$

Ví dụ 2.42 (Đề thi MI2021 kỳ 20193)**.** Số máy D bán được trong ngày của một siêu thị là biến ngẫu nhiên X tuân theo phân phối Poisson tham số λ với $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, k \in \mathbb{N}$. Biết rằng xác suất bán được máy D trong một ngày là 39,35%. (a) Tính số máy D bán được trung bình trong một ngày của siêu thị đó. (b) Nếu khảo sát 30 ngày thì số ngày bán được máy D có khả năng xảy ra cao nhất là bao nhiêu?

Lời giải Ví dụ 2.42 (a) Gọi X là "số máy D bán được trong một ngày", $X \sim P(\lambda)$.

$P(X \geq 1) = 0,3935 \Rightarrow 0,6065 = P(X=0) = e^{-\lambda}$.

Trung bình số máy D bán được trong ngày là $\lambda = -\ln(0,6065) = 0,5$.
(b) Gọi Y là "số ngày bán được máy D (trong 30 ngày)"; $Y \sim B(n;p)$ với $n = 30; p = 0,3935$.
Vì $(n+1) \times p - 1 \leq \mod (Y) \leq (n+1) \times p$ nên
$11,1985 \leq \mod (Y) \leq 12,1985$ hay $\mod (Y) = 12$.

Bài tập Chương 2

Bài tập 2.1. Một chùm chìa khóa gồm 4 chiếc giống nhau, trong đó chỉ có một chiếc mở được cửa. Người ta thử ngẫu nhiên từng chiếc cho đến khi mở được cửa. Gọi X là số lần thử.

(a) Tìm phân phối xác suất của X.

(b) Tìm kỳ vọng và phương sai của X.

(b) Viết hàm phân phối xác suất của X.

Bài tập 2.2. Một xạ thủ có 5 viên đạn. Anh ta phải bắn vào bia với quy định khi nào có 2 viên trúng bia hoặc hết đạn thì dừng. Biết xác suất bắn trúng bia ở mỗi lần bắn là 0,4 và gọi X là số đạn cần bắn.

(a) Tìm phân phối xác suất của X.

(b) Tìm kỳ vọng, phương sai và viết hàm phân phối xác suất của X.

Bài tập 2.3. Tỷ lệ cử tri ủng hộ ứng cử viên A trong một cuộc bầu cử tổng thống là 40%. Người ta hỏi ý kiến 20 cử tri được chọn một cách ngẫu nhiên. Gọi X là số người bỏ phiếu cho ông A trong 20 người đó.

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

(b) Tìm $P(X = 10)$.

Bài tập 2.4. Biến ngẫu nhiên rời rạc X chỉ có 2 giá trị x_1 và x_2 ($x_1 < x_2$). Xác suất để X nhận giá trị x_1 là 0,2. Tìm luật phân phối xác suất của X, biết kỳ vọng $E(X) = 2,6$ và độ lệch tiêu chuẩn $\sigma(X) = 0,8$.

Bài tập 2.5. Mỗi khách uống cà phê tại quán cà phê mỗi ngày đều được phát ngẫu nhiên một vé bốc thăm, xác suất khách hàng trúng thăm là 0,1. Nếu khách hàng trúng thăm liên tục trong 5 ngày (từ thứ hai đến thứ sáu) sẽ nhận được 100$, nếu không sẽ không được gì. An uống cà phê liên tục tại quán này 4 tuần liên tiếp. Gọi X\$ là số tiền An được thưởng khi bốc thăm trong 4 tuần đó. Xác định kỳ vọng và phương sai của X.

Bài tập 2.6. Tung đồng xu 10 lần. Biến ngẫu nhiên X được định nghĩa như sau: $(X = 1)$ nếu sự kiện đúng 3 lần ra mặt sấp xảy ra và $(X = 0)$ trong trường hợp còn lại. Tính kỳ vọng $E(X)$ và phương sai $V(X)$.

Bài tập 2.7. Có 5 sản phẩm trong đó có 4 chính phẩm và 1 phế phẩm. Người ta lấy ra lần lượt hai sản phẩm (lấy không hoàn lại).

(a) Gọi X là "số chính phẩm gặp phải". Lập bảng phân phối xác suất của X. Tính $E(X)$ và $V(X)$.

(b) Gọi Y là "số phế phẩm gặp phải". Lập hệ thức cho mối quan hệ giữa X và Y.

Bài tập 2.8. Người ta đặt ngẫu nhiên 10 thẻ (trong đó có 5 thẻ màu đỏ và 5 thẻ màu xanh) vào 10 phong bì (5 phong bì có màu đỏ và 5 phong bì có màu xanh), mỗi phong bì một thẻ. Gọi X là số phong bì có chứa một thẻ cùng màu. Tính giá trị:

(a) $P(X = 1)$.

(b) $E(X)$.

Bài tập 2.9. Có 2 kiện hàng. Kiện I có 3 sản phẩm tốt và 2 sản phẩm xấu. Kiện II có 2 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên từ kiện I ra 2 sản phẩm và từ kiện II ra 1 sản phẩm. Lập bảng phân phối xác suất cho biến ngẫu nhiên chỉ số sản phẩm tốt trong 3 sản phẩm lấy ra.

Bài tập 2.10. Có hai kiện hàng. Kiện thứ nhất có 8 sản phẩm tốt và 2 sản phẩm xấu. Kiện thứ hai có 5 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên 2 sản phẩm từ kiện I bỏ sang kiện II. Sau đó từ kiện II lấy ngẫu nhiên ra 2 sản phẩm. Lập bảng phân phối xác suất của biến ngẫu nhiên chỉ số sản phẩm tốt có trong 2 sản phẩm lấy ra từ kiện II.

Bài tập 2.11. Gieo hai con xúc sắc đồng chất 5 lần, gọi X là số lần xuất hiện hai mặt 6.

(a) Tính xác suất của sự kiện số lần xuất hiện hai mặt 6 ít nhất là 2.

(b) Tính $E(X)$, $V(X)$.

(c) Viết hàm phân phối $F_X(x)$.

Bài tập 2.12. Một thanh niên nam vào cửa hàng thấy 5 máy thu thanh giống nhau. Anh ta đề nghị cửa hàng cho anh ta thử lần lượt các máy đến khi chọn được máy tốt thì mua, nếu cả 5 lần đều xấu thì thôi. Biết rằng xác suất để một máy xấu là 0,6 và các máy xấu tốt độc lập với nhau. Gọi X là số lần thử. Lập bảng phân phối xác suất của X.

Bài tập 2.13. Có hai hộp bi. Hộp I có 2 bi trắng, 3 bi đỏ. Hộp II có 2 bi trắng, 2 bi đỏ. Lấy ngẫu nhiên 2 bi từ hộp I bỏ sang hộp II, sau đó lại lấy ngẫu nhiên 3 bi từ hộp II bỏ vào hộp I. Lập bảng phân phối xác suất của biến ngẫu nhiên chỉ số bi trắng có mặt ở hộp I, hộp II sau khi đã chuyển xong.

Bài tập 2.14. Một người đi làm từ nhà đến cơ quan phải qua 3 ngã tư. Xác suất để người đó gặp đèn đỏ ở các ngã tư tương ứng là 0,2; 0,4 và 0,5. Gọi X là số đèn đỏ mà người đó gặp phải trong một lần đi làm (giả sử 3 đèn giao thông ở ngã tư hoạt động độc lập với nhau).

(a) Lập bảng phân phối xác suất của X. Tính kỳ vọng, phương sai của X. Tìm hàm phân phối xác suất của X.

(b) Hỏi thời gian trung bình phải ngừng trên đường là bao nhiêu biết rằng mỗi khi gặp đèn đỏ người ấy phải đợi khoảng 3 phút.

Bài tập 2.15. Một người chơi trò chơi tung con xúc sắc cân đối đồng chất ba lần. Nếu cả ba lần đều xuất hiện mặt 6 thì thu về 36$, nếu hai lần xuất hiện mặt 6 thì thu về 2,8$, nếu một lần xuất hiện mặt 6 thì thu về 0,4$. Biết rằng khi chơi người đó phải nộp $x$$.

(a) Tìm x sao cho trò chơi là vô thưởng vô phạt.

(b) x bằng bao nhiêu thì trung bình mỗi lần chơi, người chơi mất 1$?

Bài tập 2.16. Một kiện hàng có 12 sản phẩm, trong đó có 7 sản phẩm loại I và 5 sản phẩm loại II. Khi bán được một sản phẩm loại I thì được lãi 50 ngàn đồng; còn nếu bán được một sản phẩm loại II thì được lãi 20 ngàn đồng. Lấy ngẫu nhiên từ kiện hàng ra 3 sản phẩm.

(a) Tìm quy luật phân phối xác suất của số tiền lãi thu được do bán 3 sản phẩm đó; tính kỳ vọng, phương sai của số tiền lãi thu được do bán 3 sản phẩm đó.

(b) Viết hàm phân phối, vẽ đồ thị hàm phân phối của số tiền lãi thu được khi bán 3 sản phẩm đó.

Bài tập 2.17. Một hộp đựng 15 quả bóng bàn trong đó có 10 quả còn mới. Lần đầu ta lấy ra 3 quả để thi đấu, sau đó lại trả 3 quả đó vào hộp. Lần thứ hai lại lấy ra 3 quả. Gọi X là biến ngẫu nhiên chỉ số quả bóng mới trong 3 quả lấy ra. Lập bảng phân phối xác suất, tính kì vọng, phương sai của X.

Bài tập 2.18. Một cơ sở thí nghiệm có 3 phòng thí nghiệm như nhau. Xác suất thực hiện thành công một thí nghiệm của các phòng lần lượt là 0,6; 0,7 và 0,8. Một sinh viên chọn một phòng thí nghiệm bất kỳ và tiến hành 3 thí nghiệm độc lập. Gọi X là số thí nghiệm thành công.

(a) Lập bảng phân phối xác suất của X, tính kỳ vọng $E(X)$ và phương sai $V(X)$.

(b) Theo anh (chị) thì khả năng chắc chắn sẽ thành công mấy thí nghiệm?

Bài tập 2.19. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} k\sin 3x, & x \in \left(0, \dfrac{\pi}{3}\right), \\ 0, & x \notin \left(0, \dfrac{\pi}{3}\right). \end{cases}$$

(a) Xác định k và hàm phân phối $F_X(x)$.

(b) Tính $P(\pi/6 \le X < \pi/3)$.

Bài tập 2.20. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \frac{c}{e^x + e^{-x}}.$$

Xác định hằng số c và sau đó tính kỳ vọng của X.

Bài tập 2.21. Biến ngẫu nhiên liên tục X có hàm mật độ là $f_X(x) = ae^{-|x|}$, $(-\infty < x < +\infty)$.

(a) Xác định a.

(b) Tìm hàm phân phối của biến ngẫu nhiên X, biến ngẫu nhiên $Y = X^2$.

(c) Tìm $E(X)$, $V(X)$.

(d) Tính xác suất để sau ba lần lặp lại phép thử một cách độc lập có 2 lần X nhận giá trị trong khoảng $(0; \ln 3)$.

Bài tập 2.22. Nhu cầu hàng năm về loại hàng A là biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau (đơn vị: ngàn sản phẩm):

$$f_X(x) = \begin{cases} k(30 - x), & x \in (0, 30), \\ 0, & x \notin (0, 30). \end{cases}$$

(a) Tìm k.

(b) Tìm hàm phân phối $F_X(x)$.

(c) Tìm nhu cầu trung bình hàng năm về loại hàng đó.

Bài tập 2.23. Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{1}{2} - k\cos x, & 0 < x \le \pi, \\ 1, & x > \pi. \end{cases}$$

(a) Tìm k.

(b) Tìm $P\left(0 < X < \dfrac{\pi}{2}\right)$.

(c) Tìm $E(X)$.

Bài tập 2.24. Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \le -a, \\ A + B\arcsin\dfrac{x}{a}, & x \in (-a, a), \\ 1, & x \ge a. \end{cases}$$

(a) Tìm A và B.

(b) Tìm hàm mật độ xác suất $f_X(x)$.

Bài tập 2.25. Hàm phân phối xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = a + b \arctan x, \quad (-\infty < x < +\infty).$$

(a) Tìm hệ số a và b.

(b) Tìm hàm mật độ xác suất $f_X(x)$.

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

Bài tập 2.26. Biến ngẫu nhiên X liên tục trên toàn trục số và có hàm phân phối xác suất $F_X(x) = 1/2 + 1/\pi \arctan x/2$. Tìm giá trị có thể có của x_1 thỏa mãn điều kiện $P(X > x_1) = 1/4$.

Bài tập 2.27. Thu nhập của dân cư tại một vùng là biến ngẫu nhiên liên tục có hàm phân phối xác suất như sau:

$$F_X(x) = \begin{cases} 1 - \left(\dfrac{x_0}{x}\right)^{\alpha}, & x \geq x_0,\ \alpha > 0, \\ 0, & x < x_0. \end{cases}$$

Hãy xác định mức thu nhập sao cho lấy ngẫu nhiên một người ở vùng đó thì thu nhập của người này vượt quá mức trên với xác suất 0,5.

Bài tập 2.28. Thời gian phục vụ mỗi khách hàng tại một cửa hàng ăn nhanh là biến ngẫu nhiên X tuân theo quy luật lũy thừa với hàm mật độ xác suất

$$f_X(x) = \begin{cases} 5e^{-5x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

với x được tính bằng phút/khách hàng.

(a) Tìm xác suất để thời gian phục vụ một khách hàng nào đó sẽ nằm trong khoảng $(0,4;1)$(phút).

(b) Tính thời gian trung bình để phục vụ một khách hàng.

Bài tập 2.29. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} e^{-x}, & \text{khi} \quad x > 0, \\ 0, & \text{khi} \quad x \leq 0. \end{cases}$$

(a) Tính $P(X \geq 5)$.

(b) Xác định hàm phân phối xác suất của biến ngẫu nhiên $Y = -2X + 5$.

Bài tập 2.30. Cho hàm mật độ xác suất

$$f_X(x) = \begin{cases} 3e^{-3x}, & \text{nếu } x \geq 0, \\ 0, & \text{nếu } x < 0 \end{cases}$$

của biến ngẫu nhiên liên tục X và định nghĩa $Y = [X]$ là số nguyên lớn nhất không vượt quá X (nghĩa là $[x] = 0$ nếu $0 \leq x < 1$, $[x] = 1$ nếu $1 \leq x < 2 \ldots$).

(a) Tính $P(Y = 0)$.

(b) Tính $E(Y)$.

Bài tập 2.31. Bắn 5 viên đạn vào một mục tiêu. Xác suất trúng đích của mỗi lần bắn như nhau và bằng 0,2. Muốn phá hủy mục tiêu phải có ít nhất 3 viên trúng mục tiêu. Tìm xác suất mục tiêu bị phá hủy.

Bài tập 2.32. Xác suất để một sinh viên chậm giờ thi là 0,02. Tìm số sinh viên chậm giờ thi có khả năng xảy ra nhiều nhất trong 855 sinh viên dự thi.

Bài tập 2.33. Có 10 máy sản xuất sản phẩm (độc lập nhau), mỗi máy sản xuất ra 2% phế phẩm.

(a) Từ mỗi máy sản xuất lấy ngẫu nhiên ra một sản phẩm. Hỏi xác suất lấy được nhiều nhất 2 phế phẩm trong 10 sản phẩm này là bao nhiêu?

(b) Trung bình có bao nhiêu sản phẩm được sản xuất bởi máy đầu tiên trước khi nó tạo ra phế phẩm đầu tiên (giả sử các sản phẩm sản xuất ra là độc lập)?

Bài tập 2.34. Một ga ra cho thuê ôtô thấy rằng số người đến thuê ôtô vào thứ bảy cuối tuần là một biến ngẫu nhiên có phân bố Poa-xông với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ôtô.

(a) Tìm xác suất để tất cả 4 ôtô đều được thuê vào thứ 7.

(b) Tìm xác suất gara không đáp ứng được yêu cầu (thiếu xe cho thuê) vào thứ 7.

(c) Trung bình có bao nhiêu ôtô được thuê vào ngày thứ 7?

Bài tập 2.35. Số khách hàng đến một cửa hàng bán lẻ là một biến ngẫu nhiên có phân phối Poa-xông với trung bình 6 khách hàng đến trong vòng một giờ.

(a) Nếu có đúng 5 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:00 thì xác suất để có ít nhất 8 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:30 là bao nhiêu?

(b) Nếu có ít hơn 6 khách hàng đến trong khoảng thời gian từ 10:00 đến 12:00 thì cửa hàng được xem như là không có lợi nhuận. Tìm xác suất để cửa hàng có đúng 1 ngày có lãi trong một tuần (giả sử cửa hàng mở cửa 6 ngày trong tuần).

Bài tập 2.36. Gọi biến ngẫu nhiên Y là tỷ lệ người trong 1000 người Mỹ xác nhận rằng có uống nhiều hơn 5 cốc bia mỗi ngày. Giả sử rằng tỷ lệ đúng là 10% trên toàn bộ dân số Mỹ. Tính $E(Y)$, $V(Y)$.

Bài tập 2.37. Giả sử X là biến ngẫu nhiên tuân theo luật phân phối chuẩn với trung bình là 3 và phương sai là 0,16.

(a) Hãy tính $P(X > 3)$, $P(X > 3,784)$.

(b) Tìm c sao cho $P(3 - c < X < 3 + c) = 0,9$.

Bài tập 2.38. Cho biên độ dao động của một vật là biến ngẫu nhiên liên tục có hàm phân phối xác suất là

$$F_X(x) = \begin{cases} 1 - e^{\frac{-x^2}{2\sigma^2}}, & \text{nếu } x \geq 0, \\ 0, & \text{nếu } x < 0, \end{cases}$$

trong đó σ là tham số đã biết. Tính xác suất để biên độ giao động đó lớn hơn trị trung bình của nó.

Bài tập 2.39. Lãi suất (%) đầu tư vào một dự án trong năm 2019 được coi như một biến ngẫu nhiên tuân theo quy luật chuẩn. Theo đánh giá của ủy ban đầu tư thì với xác suất 0,1587 cho lãi suất lớn hơn 20% và với xác suất 0,0228 cho lãi suất lớn hơn 25%. Vậy khả năng đầu tư mà không bị lỗ là bao nhiêu?

Bài tập 2.40. Tung một đồng xu vô hạn lần, xác suất thu được mặt ngửa mỗi lần là p.

(a) Gọi X là số lần tung đến khi xuất hiện mặt ngửa lần đầu tiên (tại lần tung thứ X). Tính $E(X)$.

(b) Tính xác suất xuất hiện đúng 6 lần ngửa trong 10 lần tung.

(c) Tính xác suất để lần xuất hiện mặt ngửa thứ 6 rơi vào lần tung thứ 10.

Bài tập 2.41. Xét một phần tư hình tròn tâm O(0,0) bán kính bằng a, ký hiệu là OAB, với tọa độ tương ứng là A$(a, 0)$ và B$(0, a)$.

(a) Trên đoạn OA lấy ngẫu nhiên một điểm C. Tìm phân phối xác suất của độ dài đoạn OC.

(b) Dựng một đường thẳng đi qua C, vuông góc với OA và cắt cung tròn tại điểm D. Tính kỳ vọng và phương sai của độ dài đoạn CD.

Bài tập 2.42. Lấy ngẫu nhiên một điểm M trên nửa đường tròn tâm O, đường kính AB $= 2a$. Biết rằng xác suất điểm M rơi vào cung CD bất kì của nửa đường tròn AMB chỉ phụ thuộc vào độ dài cung CD.

(a) Tìm hàm phân phối xác suất của biến ngẫu nhiên Y chỉ diện tích tam giác AMB.

(b) Tìm giá trị trung bình của diện tích tam giác ấy.

Bài tập 2.43. Từ điểm $A(0, -a)$ $(a > 0)$ trong nửa mặt phẳng tọa độ xOy phần $x \geq 0$, người ta kẻ ngẫu nhiên một tia At hợp với tia Oy một góc φ. Biết φ là biến ngẫu nhiên có phân phối đều trong khoảng $(0, \pi/4)$. Tia At cắt Ox tại điểm M.

(a) Tìm hàm phân phối xác suất của biến ngẫu nhiên X chỉ diện tích tam giác AOM.

(b) Tìm giá trị trung bình của diện tích trên.

Bài tập 2.44. Một công ty kinh doanh mặt hàng A dự định sẽ áp dụng một trong hai phương án kinh doanh: Phương án 1: Gọi X_1 (triệu đồng/tháng) là lợi nhuận thu được. X_1 có phân phối chuẩn $\mathcal{N}(140; 2500)$. Phương án 2: Gọi X_2 (triệu đồng/tháng) là lợi nhuận thu được. X_2 có phân phối chuẩn $\mathcal{N}(200; 3600)$. Biết rằng công ty tồn tại và phát triển thì lợi nhuận thu được từ mặt hàng A phải đạt ít nhất 80 triệu đồng/tháng. Hỏi nên áp dụng phương án nào để rủi ro thấp hơn.

Bài tập 2.45. Trọng lượng của một loại trái cây tuân theo luật phân phối chuẩn với trọng lượng trung bình là 250g, độ lệch chuẩn là 5g. Trái cây loại I là trái cây có trọng lượng không nhỏ hơn 260g.

(a) Một người lấy 1 trái từ trong sọt trái cây ra. Tính xác suất người này lấy được trái cây loại I.

(b) Nếu lấy được trái loại I thì người này sẽ mua sọt đó. Người ngày kiểm tra 100 sọt. Tính xác suất người này mua được 6 sọt.

Bài tập 2.46. Một dây chuyền tự động khi hoạt động bình thường có thể sản xuất ra phế phẩm với xác suất $p = 0,001$ và được điều chỉnh ngay lập tức khi phát hiện có phế phẩm. Tính số trung bình các sản phẩm được sản xuất giữa 2 lần điều chỉnh.

Bài tập 2.47. Trong một kỳ thi điểm số trung bình của các sinh viên là 80 và độ lệch chuẩn là 10. Giả sử điểm thi của sinh viên tuân theo luật phân phối chuẩn.

(a) Nếu giáo viên muốn 25% số sinh viên đạt điểm A (nhóm điểm cao nhất) thì điểm số thấp nhất để đạt điểm A là bao nhiêu?

(b) Chọn ngẫu nhiên 50 sinh viên, tính xác suất trong đó có nhiều hơn 10 sinh viên đạt điểm A (điểm A lấy ở câu (a)).

Bài tập 2.48. Đường kính của một loại chi tiết do một máy sản xuất tuân theo luật phân phối chuẩn, với kỳ vọng là 20mm và độ lêch chuẩn là 0,2mm. Tính xác suất để lấy ngẫu nhiên một chi tiết có đường kính trong khoảng 19,9mm đến 20,3mm.

Bài tập 2.49. Chiều cao của nam giới khi trưởng thành là biến ngẫu nhiên tuân theo luật phân phối chuẩn với chiều cao trung bình là 160cm và độ lệch chuẩn là 6cm. Tìm xác suất để đo ngẫu nhiên 4 người thì có ít nhất một người có chiều cao nằm trong khoảng (158–162)cm.

Bài tập 2.50. Dùng hai phương pháp để tính sai số của một biến ngẫu nhiên. Phương pháp 1: Cho sai số đó bằng $2X$ với X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(0;25)$. Phương pháp 2: Cho sai số đó bằng tổng hai biến ngẫu nhiên độc lập $Y = Y_1 + Y_2$ trong đó $E(Y_1) = E(Y_2) = 0$ và $\sigma(Y_1) = \sigma(Y_2) = 5$. Hỏi phương pháp nào được ưa dùng hơn?

Chương 3

Biến ngẫu nhiên nhiều chiều

Thông qua công cụ giải tích, chương này giới thiệu về

1. Khái niệm và phân loại biến ngẫu nhiên nhiều chiều.
2. Phân phối xác suất (đồng thời, biên) của biến ngẫu nhiên hai chiều rời rạc và liên tục.
3. Phân phối có điều kiện.
4. Tính độc lập của các biến ngẫu nhiên.
5. Khái niệm và tính chất của kỳ vọng, phương sai, hiệp phương sai, hệ số tương quan.
6. Luật số lớn và định lý giới hạn trung tâm.

BÀI 9 (2 tiết)

3.1 Khái niệm và phân loại biến ngẫu nhiên nhiều chiều

3.1.1 Khái niệm

Trong nhiều bài toán thực tế ta thường phải xét đồng thời nhiều biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ có quan hệ với nhau.

Một biến ngẫu nhiên n-chiều (véc-tơ ngẫu nhiên n-chiều) là một bộ có thứ tự $(X_1, X_2, \ldots, X_n)$ với các thành phần $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên xác định trong cùng một phép thử.

Ký hiệu biến ngẫu nhiên hai chiều là (X, Y), trong đó X là biến ngẫu nhiên thành phần thứ nhất và Y là biến ngẫu nhiên thành phần thứ hai.

3.1.2 Phân loại

Biến ngẫu nhiên n-chiều $(X_1, X_2, \ldots, X_n)$ là liên tục hay rời rạc nếu tất cả các biến ngẫu nhiên thành phần $X_1, X_2, \ldots, X_n$ là liên tục hay rời rạc.

Để cho đơn giản, ta nghiên cứu biến ngẫu nhiên hai chiều (X, Y), trong đó X và Y là các biến ngẫu nhiên một chiều. Hầu hết các kết quả thu được đều có thể mở rộng cho trường hợp biến ngẫu nhiên n-chiều.

Trong chương này ta không xét trường hợp biến ngẫu nhiên hai chiều có một biến ngẫu nhiên rời rạc và một biến ngẫu nhiên liên tục.

Ví dụ 3.1. Một nhà máy sản xuất ra một sản phẩm.

1. Gọi X, Y, Z lần lượt là các biến ngẫu nhiên chỉ chiều dài, chiều rộng và chiều cao của sản phẩm (đơn vị tính là cm).

   (a) (X, Y, Z) là biến ngẫu nhiên ba chiều mô tả kích thước của sản phẩm.

   (b) Nếu chỉ quan tâm đến chiều dài và chiều rộng của sản phẩm ta được biến ngẫu nhiên hai chiều (X, Y).

2. Nếu quan tâm đến trọng lượng G và thể tích V của sản phẩm thì ta được biến ngẫu nhiên hai chiều (G, V).

3.2 Bảng phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

3.2.1 Bảng phân phối xác suất đồng thời

Định nghĩa 3.1 (Bảng phân phối xác suất đồng thời)**.** Bảng phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc (X, Y) là

$X \backslash Y$	y_1	$\dots$	y_j	$\dots$	y_n	$\sum_j$
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1n}	$P(X = x_1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$	p_{in}	$P(X = x_i)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	$\dots$	p_{mj}	$\dots$	p_{mn}	$P(X = x_m)$
$\sum_i$	$P(Y = y_1)$	$\dots$	$P(Y = y_j)$	$\dots$	$P(Y = y_n)$	$\sum_i \sum_j = 1$

trong đó $x_i, i = 1, \dots, m, y_j, j = 1, \dots, n$ là các giá trị có thể có của các biến thành phần X và Y tương ứng; các xác suất p_{ij} được xác định bởi

$$\boxed{p_{ij} = P(X = x_i, Y = y_j), \quad i = 1, \dots, m; \ j = 1, \dots, n} \tag{3.1}$$

Bảng này còn được gọi là bảng phân phối xác suất đồng thời của hai biến ngẫu nhiên rời rạc X và Y. Bảng này có thể ra vô hạn nếu m, n nhận giá trị ∞.

Tính chất 3.1. (a) $0 \le p_{ij} \le 1$ với mọi $i = 1, \dots, m$, $j = 1, \dots, n$.

(b) $\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1$.

Định nghĩa 3.2 (Biến ngẫu nhiên độc lập)**.** Hai biến ngẫu nhiên rời rạc X và Y độc lập với nhau nếu

$$\boxed{P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \forall i = 1, \dots, m; \ j = 1, \dots, n} \tag{3.2}$$

Ví dụ 3.2. Lấy ngẫu nhiên 3 sản phẩm từ một lô hàng gồm 5 sản phẩm loại I, 4 sản phẩm loại II và 3 sản phẩm loại III. Gọi X, Y lần lượt là số sản phẩm loại I và loại II trong 3 sản phẩm lấy ra. (a) Lập bảng phân phối xác suất đồng thời của hai biến ngẫu nhiên X và Y. (b) Tính $P(X = 0)$.

Lời giải Ví dụ 3.2 (a) X có thể nhận các giá trị $0, 1, 2, 3$, Y có thể nhận các giá trị $0, 1, 2, 3$.

$$P(X = 0, Y = 0) = \frac{C_3^3}{C_{12}^3} = \frac{1}{220}.$$

$$P(X=0,Y=1)=\frac{C_5^0C_4^1C_3^2}{C_{12}^3}=\frac{12}{220}.$$

Tương tự ta có thể tính được $P(X=i,Y=j)$ với mọi $i,j=0,1,2,3$.

Bảng phân phối xác suất cần tìm là

X \ Y	0	1	2	3
0	1/220	12/220	18/220	4/220
1	15/220	60/220	30/220	0
2	30/220	40/220	0	0
3	10/220	0	0	0

(b) $P(X=0)=\sum_{j=0}^{3}P(X=0,Y=j)=\frac{1}{220}+\frac{12}{220}+\frac{18}{220}+\frac{4}{220}=\frac{35}{220}.$

3.2.2 Bảng phân phối xác suất thành phần (biên)

Từ Định nghĩa 3.1 ta suy ra:

Định nghĩa 3.3. **(a)** Bảng phân phối xác suất của biến ngẫu nhiên thành phần X:

X	x_1	x_2	...	x_m
P	$P(X=x_1)$	$P(X=x_2)$	...	$P(X=x_m)$

trong đó $P(X=x_i)=\sum_{j=1}^{n}p_{ij}$, $i=1,\dots,m$.

(b) Bảng phân phối xác suất của biến ngẫu nhiên thành phần Y:

Y	y_1	y_2	...	y_n
P	$P(Y=y_1)$	$P(Y=y_2)$	...	$P(Y=y_n)$

trong đó $P(Y=y_j)=\sum_{i=1}^{m}p_{ij}$, $j=1,\dots,n$.

Nhận xét 3.1. Từ các bảng phân phối thành phần ta có thể dễ dàng xác định các tham số đặc trưng của các biến ngẫu nhiên thành phần X và Y.

Ví dụ 3.3. Trong Ví dụ 3.2, bảng phân phối xác suất đồng thời của hai biến ngẫu nhiên X và Y được bổ sung thêm cột $P(X=i)$, $i=0,1,2,3$ và hàng $P(Y=j)$, $j=0,1,2,3$ là

X \ Y	0	1	2	3	$P(X=i)$
0	1/220	12/220	18/220	4/220	35/220
1	15/220	60/220	30/220	0	105/220
2	30/220	40/220	0	0	70/220
3	10/220	0	0	0	10/220
$P(Y=j)$	56/220	112/220	48/220	4/220	$\sum_i\sum_j=1$

(a) Bảng phân phối xác suất của X và của Y là

X	0	1	2	3
P	35/220	105/220	70/220	10/220

Y	0	1	2	3
P	56/220	112?220	48/220	4/220

(b) $E(X) = \dfrac{(0)(35) + (1)(105) + (2)(70) + (3)(10)}{220} = 1.25.$

$E(X^2) = \dfrac{(0^2)(35) + (1^2)(105) + (2^2)(70) + (3^2)(10)}{220} \simeq 2.1591.$

$V(X) = E(X^2) - [E(X)]^2 = 2.1591 - (1.25)^2 = 0.5966.$

$\sigma(X) = \sqrt{V(X)} = \sqrt{0.5966} \simeq 0.7724.$

Ví dụ 3.4. Từ kết quả phân tích các số liệu thống kê trong tháng về doanh số bán hàng (X) và chi phí cho quảng cáo (Y) (đơn vị triệu đồng) của một công ty, thu được bảng phân phối xác suất đồng thời như sau:

$Y \backslash X$	100	200	300
1	$0,15$	$0,1$	$0,14$
$1,5$	$0,05$	$0,2$	$0,15$
2	$0,01$	$0,05$	$0,15$

(a) Tính giá trị trung bình và phương sai của doanh số bán hàng. (b) Tính giá trị trung bình và phương sai của chi phí cho quảng cáo.

Lời giải Ví dụ 3.4 Lập bảng phân phối xác suất của X và Y:

X	100	200	300
P	0,21	0,35	0,44

Y	1	1,5	2
P	0,39	0,4	0,21

(a) Trung bình và phương sai của doanh số bán hàng là $E(X)$ và $V(X)$ được tính dựa trên Định nghĩa 2.8(a) và công thức (2.18)

$$E(X) = 100 \times 0,21 + 200 \times 0,35 + 300 \times 0,44 = 223$$
$$E(X^2) = 100^2 \times 0,21 + 200^2 \times 0,35 + 300^2 \times 0,44 = 55700$$
$$V(X) = E(X^2) - [E(X)]^2 = 55700 - (233)^2 = 1411.$$

(b) Tương tự, trung bình và phương sai của chi phí cho quảng cáo là:

$$E(Y) = 1 \times 0,39 + 1,5 \times 0,4 + 2 \times 0,21 = 1,41$$
$$E(Y^2) = 1^2 \times 0,39 + 1,5^2 \times 0,4 + 2^2 \times 0,21 = 2,13$$
$$V(Y) = E(Y^2) - [E(Y)]^2 = 2,13 - (1,41)^2 = 0,2419.$$

Ví dụ 3.5. Cho X_1, X_2, X_3 là các biến ngẫu nhiên độc lập theo luật phân phối Poa-xông với tham số $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$. Tính xác suất của các sự kiện (a) Số lớn nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1. (b) Số nhỏ nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.

Lời giải Ví dụ 3.5

(a) $P(\max\{X_1, X_2, X_3\} \geq 1) = 1 - P(\max\{X_1, X_2, X_3\} < 1)$. Sử dụng tính độc lập của các biến ngẫu nhiên X_1, X_2, X_3 suy ra

$$\begin{aligned} P(\max\{X_1, X_2, X_3\} < 1) &= P(X_1 < 1, X_2 < 1, X_3 < 1) = P(X_1 < 1)P(X_2 < 1)P(X_3 < 1) \\ &= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) = \left(e^{-1}\frac{1^0}{0!}\right)\left(e^{-2}\frac{2^0}{0!}\right)\left(e^{-3}\frac{3^0}{0!}\right) \\ &\simeq 0.0025. \end{aligned}$$

Suy ra $P(\max\{X_1, X_2, X_3\} \geq 1) \simeq 0.9975$.

(b) $P(\min\{X_1, X_2, X_3\} \geq 1) = P(X_1 \geq 1, X_2 \geq 1, X_3 \geq 1)$. Sử dụng tính độc lập của các biến ngẫu nhiên X_1, X_2, X_3 ta được:

$$\begin{aligned} P(\min\{X_1, X_2, X_3\} \geq 1) &= P(X_1 \geq 1)\, P(X_2 \geq 1)\, P(X_3 \geq 1) \\ &= \big(1 - P(X_1 < 1)\big)\big(1 - P(X_2 < 1)\big)\big(1 - P(X_3 < 1)\big) \\ &= \left(1 - \frac{(e^{-1})\,(1^0)}{0!}\right) \cdot \left(1 - \frac{(e^{-2})\,(2^0)}{0!}\right) \cdot \left(1 - \frac{(e^{-3})\,(3^0)}{0!}\right) \\ &\simeq 0.5194 \end{aligned}$$

3.2.3 Phân phối có điều kiện

Từ Định nghĩa 3.1 ta suy ra:

Định nghĩa 3.4. (a) Bảng phân phối xác suất của X với điều kiện $(Y = y_j)$:

$X\|(Y = y_j)$	x_1	x_2	...	x_m
P	$p(x_1\|y_j)$	$p(x_2\|y_j)$	...	$p(x_m\|y_j)$

trong đó $p(x_i|y_j) = P[(X = x_i)|(Y = y_j)]$, $i = 1, \ldots, m; j = 1, \ldots, n$.

(b) Bảng phân phối xác suất của Y với điều kiện $(X = x_i)$:

$Y\|(X = x_i)$	y_1	y_2	...	y_n
P	$p(y_1\|x_i)$	$p(y_2\|x_i)$	...	$p(y_n\|x_i)$

trong đó $p(y_j|x_i) = P[(Y = y_j)|(X = x_i)]$, $i = 1, \ldots, m; j = 1, \ldots, n$.

Nhận xét 3.2. **(a)** Từ bảng phân phối xác suất có điều kiện ta có thể tính được kỳ vọng (có điều kiện) của từng biến ngẫu nhiên.

(b) Các xác suất có điều kiện được tính như thông thường, tức là

$$P\left(X = x_i | Y = y_j\right) = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)} \quad \text{hoặc}$$
$$P\left(X = x_i | Y \in D\right) = \frac{P(X = x_i,\ Y \in D)}{P(Y \in D)}$$

và

$$P\left(Y = y_j | X = x_i\right) = \frac{P(Y = y_j,\ X = x_i)}{P(X = x_i)} \quad \text{hoặc}$$
$$P\left(Y = y_j | X \in D\right) = \frac{P(Y = y_j,\ X \in D)}{P(X \in D)}.$$

Ví dụ 3.6. Với giả thiết của Ví dụ 3.4,

(a) Nếu chi phí cho quảng cáo 1,5 triệu đồng thì doanh số trung bình là bao nhiêu?

(b) Nếu muốn doanh số là 300 triệu đồng thì trung bình phải chi phí cho quảng cáo bao nhiêu?

Lời giải Ví dụ 3.6 Các bảng phân phối xác suất có điều kiện là:

$X\|(Y = 1,5)$	100	200	300
P	0,125	0,5	0,375

$Y\|(X = 300)$	1	1,5	2
P	$\frac{14}{44}$	$\frac{15}{44}$	$\frac{15}{44}$

(a) Nếu chỉ chi phí cho quảng cáo 1,5 triệu đồng thì doanh số trung bình là

$$E(X|Y = 1,5) = 100 \times 0,125 + 200 \times 0,5 + 300 \times 0,375 = 225 \quad \text{(triệu đồng)}.$$

(b) Nếu muốn doanh số là 300 triệu đồng thì trung bình phải chi phí cho quảng cáo là

$$E(Y|X = 300) = 1 \times \frac{14}{44} + 1,5 \times \frac{15}{44} + 2 \times \frac{15}{44} \simeq 1,5136 \quad \text{(triệu đồng)}.$$

3.3 Hàm phân phối xác suất

3.3.1 Hàm phân phối xác suất đồng thời

Định nghĩa 3.5 (Hàm phân phối đồng thời)**.** Hàm hai biến $F_{XY}(x, y)$ xác định bởi:

$$\boxed{F_{XY}(x, y) = P(X < x, Y < y), \quad x, y \in \mathbb{R}} \tag{3.3}$$

được gọi là hàm phân phối xác suất của biến ngẫu nhiên hai chiều (X, Y) và còn được gọi là hàm phân phối xác suất đồng thời của các biến ngẫu nhiên X và Y.

Từ (3.3) và Định nghĩa 3.1, hàm phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc (X,Y) được xác định bởi

$$\boxed{F_{XY}(x,y) = \sum_{x_i<x}\sum_{y_j<y} P(X=x_i, Y=y_j)} \tag{3.4}$$

Ví dụ 3.7. Từ số liệu của Ví dụ 3.16, hãy tính $F_{X,Y}(2;3)$.

Lời giải Ví dụ 3.7 Sử dụng công thức (3.4),

$$F_{X,Y}(2,3) = \sum_{x_i<2}\sum_{y_j<3} P(X=x_i, Y=y_j) = p_{11}+p_{12} = 0,12+0,15 = 0,27.$$

Sau đây là một số tính chất của hàm phân phối xác suất của biến ngẫu nhiên hai chiều (X,Y).

Tính chất 3.2. **(a)** $0 \le F_{XY}(x,y) \le 1$.

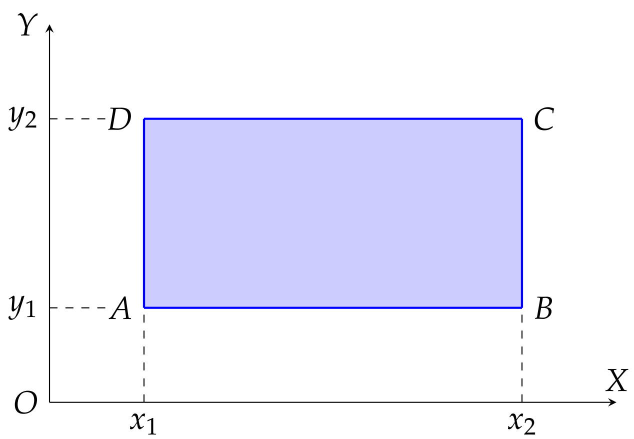
(b) Nếu $x < x_1, y < y_1$ thì $F_{XY}(x,y) \le F_{XY}(x_1,y_1)$.

(c) $F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0$; $F_{XY}(+\infty,+\infty) = 1$ (giá trị ∞ hiểu theo nghĩa lấy giới hạn).

(d) Với $x_1 < x_2, y_1 < y_2$ ta luôn có

$$P(x_1 \le X < x_2, y_1 \le Y < y_2) = F_{XY}(x_2,y_2) - F_{XY}(x_2,y_1) - F_{XY}(x_1,y_2) + F_{XY}(x_1,y_1). \tag{3.5}$$

Nhận xét 3.3. Vế phải của (3.5) chính là xác xuất để điểm ngẫu nhiên (X,Y) rơi vào hình chữ nhật $ABCD$ trong Hình 3.1.



Hình 3.1: Miền $ABCD$ trong Nhận xét 3.3

3.3.2 Hàm phân phối xác suất thành phần (biên)

Các hàm

$$F_X(x) = P(X < x) = P(X < x, Y < +\infty) = \lim_{y \to +\infty} F_{XY}(x,y) = F_{XY}(x,+\infty),$$
$$F_Y(y) = P(Y < y) = P(X < +\infty, Y < y) = \lim_{x \to +\infty} F_{XY}(x,y) = F_{XY}(+\infty, y)$$

gọi là các phân phối biên của biến ngẫu nhiên hai chiều (X, Y). Đây cũng chính là các phân phối (một chiều) thông thường của X và Y tương ứng.

Định nghĩa 3.6 (Biến ngẫu nhiên độc lập)**.** Hai biến ngẫu nhiên X và Y độc lập nếu

$$\boxed{F_{XY}(x,y) = F_X(x)F_Y(y), \quad \forall x, y \in \mathbb{R}} \tag{3.6}$$

Nhận xét 3.4. Từ Định nghĩa 3.6, nếu X và Y độc lập ta có thể nghiên cứu từng biến ngẫu nhiên theo các phương pháp đã có ở Chương 2, đồng thời từ các phân phối biên của X và Y ta có thể xác định được hàm phân phối xác suất đồng thời $F_{X,Y}(x,y)$ của X và Y theo (3.6).

3.4 Hàm mật độ xác suất của biến ngẫu nhiên hai chiều liên tục

3.4.1 Hàm mật độ xác suất đồng thời

Định nghĩa 3.7 (Hàm mật độ đồng thời)**.** Giả sử hàm phân phối xác suất của biến ngẫu nhiên hai chiều liên tục (X, Y) là $F_{XY}(x,y)$. Khi đó, hàm mật độ xác suất của biến ngẫu nhiên hai chiều liên tục (X, Y) là hàm hai biến $f_{XY}(x,y) \geq 0$ thỏa mãn:

$$\boxed{F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)dudv} \tag{3.7}$$

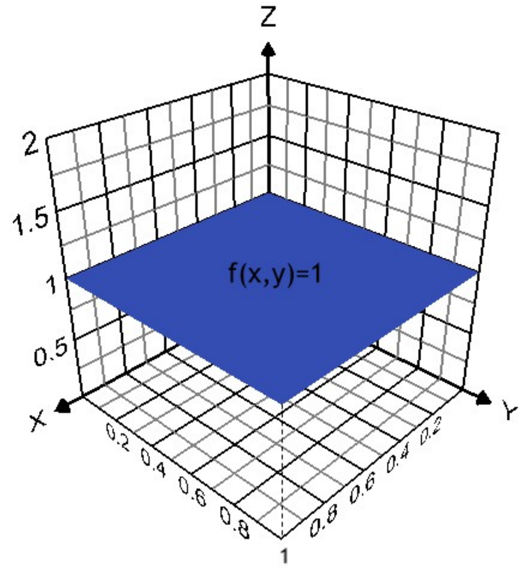
$f_{XY}(x,y)$ còn được gọi là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên liên tục X và Y.

Nhận xét 3.5. Về mặt hình học, hàm $f_{X,Y}(x,y)$ có thể xem như là một mặt cong trong $\mathbb{R}^3$ và được gọi là mặt phân phối xác suất.

Ví dụ 3.8. Hình 3.2 mô tả mặt phân phối xác suất của hàm $f_{X,Y}(x,y) = 1$ với $0 \leq x, y \leq 1$.

Hàm mật độ xác suất của biến ngẫu nhiên hai chiều (X, Y) có các tính chất sau.

Tính chất 3.3. **(a)** $f_{XY}(x,y) \geq 0$ với mọi $(x,y) \in \mathbb{R}^2$.



Hình 3.2: Mặt phân phối xác suất trong Ví dụ 3.8

(b) $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = 1.$

(c) $P\Big((X,Y)\in D\Big) = \int\int_{D\cap \mathbf{S}_{XY}} f_{XY}(x,y)dxdy$, với $D\subset \mathbb{R}^2$, $\mathbf{S}_{XY}$ là miền giá trị của biến ngẫu nhiên hai chiều (X,Y).

(d) Nếu $f_{XY}(x,y)$ liên tục theo cả hai biến thì $f_{XY}(x,y) = \dfrac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$.

Ví dụ 3.9. Biến ngẫu nhiên hai chiều liên tục (X,Y) có hàm mật độ xác suất là

$$f_{XY}(x,y) = \begin{cases} k, & 0\le x\le 5,\ 0\le y\le 3, \\ 0, & \text{trái lại.}\end{cases}$$

(a) Tìm hằng số k. (b) Tính $P(2\le X<3, 1\le Y<3)$.

Lời giải Ví dụ 3.9

(a) Sử dụng Tính chất 3.3(a),(b), ta có $k\ge 0$ và

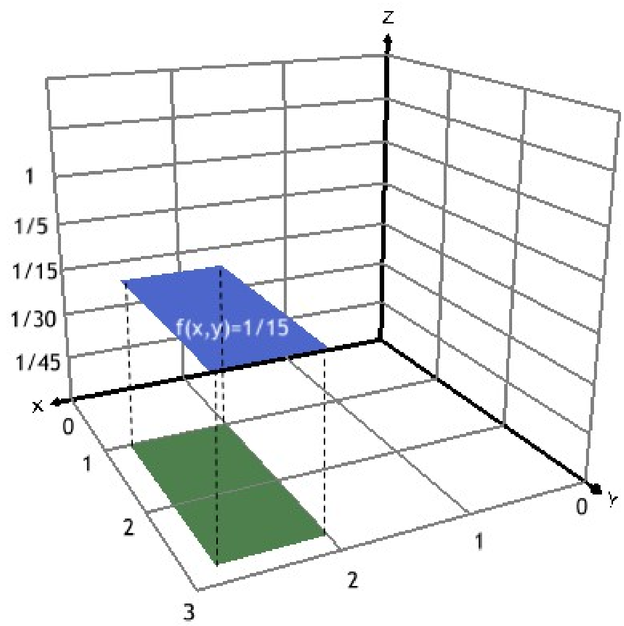
$$1 = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = \iint\limits_D kdxdy = \int\limits_0^5 dx\int\limits_0^3 kdy = 15k,$$

trong đó $D = \{(x,y)\in\mathbb{R}^2 : 0\le x\le 5,\ 0\le y\le 3\}$. Suy ra, $k = 1/15$.

(b) Sử dụng Tính chất 3.3(c),

$$P(2\le X<3, 1\le Y<3) = \int\limits_2^3 dx\int\limits_1^3 \frac{1}{15}dy = \frac{2}{15}.$$

Về mặt hình học đây là thể tích của hộp chữ nhật giới hạn bởi mặt phân phối $f_{X,Y}(x,y) = \dfrac{1}{15}$ có đáy là miền $D_1 = \{(x,y) \in \mathbb{R}^2 : 2 \le x < 3, 1 \le y < 3\}$, hình chiếu của mặt phân phối lên mặt Oxy (xem Hình 3.3).



Hình 3.3: Hình hộp chữ nhật trong Ví dụ 3.9(b)

3.4.2 Hàm mật độ xác suất biên

Định lý 3.1. Nếu (X,Y) là biến ngẫu nhiên hai chiều liên tục có hàm mật độ xác suất $f_{XY}(x,y)$ thì hàm mật độ xác suất của các biến ngẫu nhiên thành phần X và Y được xác định bởi

$$\boxed{f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx} \tag{3.8}$$

Chứng minh. Từ Tính chất 3.3(c) ta có thể viết

$$F_X(x) = P(X < x) = P(X < x, Y < \infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f_{XY}(u,y)dy \right) du.$$

Từ đây suy ra $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$. Chứng minh tương tự cho $f_Y(y)$. □

Các hàm $f_X(x)$ và $f_Y(y)$ xác định trong công thức (3.8) còn gọi là các hàm mật độ xác suất biên của biến ngẫu nhiên hai chiều liên tục (X,Y).

Định nghĩa 3.8 (Biến ngẫu nhiên độc lập)**.** Giả sử biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất $f_{XY}(x, y)$. Khi đó X và Y độc lập nếu

$$\boxed{f_{XY}(x,y) = f_X(x)f_Y(y) \quad \forall x, y \in \mathbb{R}} \tag{3.9}$$

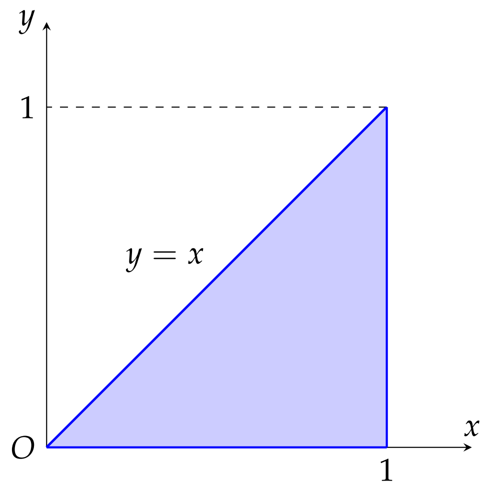
trong đó $f_X(x)$, $f_Y(y)$ lần lượt là hàm mật độ xác suất của các biến thành phần X và Y.

Ví dụ 3.10. Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} kx, & \text{nếu } 0 < y < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hằng số k. (b) Tìm các hàm mật độ xác suất của X và Y. (c) X và Y có độc lập không?

Lời giải Ví dụ 3.10 Ký hiệu $D = \left\{(x,y) \in \mathbb{R}^2 : 0 < x < 1; 0 < y < x\right\}$ (xem Hình 3.4).



Hình 3.4: Miền D của Ví dụ 3.10

(a) Theo Tính chất 3.3(a), $k \geq 0$ và theo Tính chất 3.3(b),

$$1 = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = \iint\limits_D kxdxdy = \int\limits_0^1 dx \int\limits_0^x kxdy = k\int\limits_0^1 x^2dx = \frac{k}{3}.$$

Suy ra $k = 3$ và hàm mật độ xác suất $f_{X,Y}(x,y) = \begin{cases} 3x, & \text{nếu } (x,y) \in D, \\ 0, & \text{nếu trái lại.} \end{cases}$

(b) Tìm các hàm mật độ biên từ (3.8),

$$f_X(x) = \int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \begin{cases} \int\limits_0^x 3xdy, & 0 < x < 1, \\ 0, & \text{trái lại,} \end{cases} = \begin{cases} 3x^2, & 0 < x < 1, \\ 0, & \text{trái lại,} \end{cases}$$

$$f_Y(y) = \int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dx = \begin{cases} \int\limits_y^1 3xdx, & 0 < y < 1, \\ 0, & \text{trái lại,} \end{cases} = \begin{cases} \dfrac{3}{2} - \dfrac{3}{2}y^2, & 0 < y < 1, \\ 0, & \text{trái lại.} \end{cases}$$

(c) Vì $f_{X,Y}(x,y) \neq f_X(x) \times f_Y(y)$ với $(x,y) \in D$ nên X, Y không độc lập.

3.4.3 Hàm mật độ xác suất có điều kiện

Định nghĩa 3.9 (Hàm mật độ có điều kiện)**.** Hàm mật độ có điều kiện của thành phần X biết $Y = y$ là:

$$\boxed{f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(x,y)}{\displaystyle\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx}} \tag{3.10}$$

Hàm mật độ có điều kiện của thành phần Y biết $X = x$ là:

$$\boxed{f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\displaystyle\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy}} \tag{3.11}$$

Chú ý 3.1. Các hàm mật độ có điều kiện cũng thỏa mãn các tính chất của một hàm mật độ thông thường.

Nhận xét 3.6. **(a)** Nếu $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ thì X và Y không độc lập. Trong trường hợp này, nhờ (3.10) và (3.11) ta có thể xác định được hàm mật độ đồng thời $f_{X,Y}(x,y)$ bởi

$$\boxed{f_{X,Y}(x,y) = f_Y(y)f_X(x|y) = f_X(x)f_Y(y|x)} \tag{3.12}$$

(b) Nếu $f_X(x|y) = f_X(x)$ hoặc $f_Y(y|x) = f_Y(y)$ với mọi x, y thì ta lại có điều kiện độc lập (3.9).

(c) Từ (3.10) và (3.11) và (2.7) ta có các phân phối có điều kiện sau đây

$$\boxed{F_X(x|y) = \frac{\displaystyle\int_{-\infty}^{x} f_{X,Y}(u,y)du}{f_Y(y)}, \quad F_Y(y|x) = \frac{\displaystyle\int_{-\infty}^{y} f_{X,Y}(x,v)dv}{f_X(x)}} \tag{3.13}$$

Ngoài ra

$$\boxed{F_X(x|y_1 \leq Y \leq y_2) = P(X < x|y_1 \leq Y \leq y_2) = \frac{\displaystyle\int_{-\infty}^{x} du \int_{y_1}^{y_2} f_{X,Y}(u,v)dv}{\displaystyle\int_{-\infty}^{+\infty} dx \int_{y_1}^{y_2} f_{X,Y}(x,y)dy}} \tag{3.14}$$

(d) Từ (3.14) ta nhận được hàm mật độ có điều kiện của X với điều kiện $y_1 \leq Y \leq y_2$:

$$\boxed{f_X(x|y_1 \leq Y \leq y_2) = \frac{\displaystyle\int_{y_1}^{y_2} f_{X,Y}(x,v)dv}{\displaystyle\int_{-\infty}^{+\infty} dx \int_{y_1}^{y_2} f_{X,Y}(x,y)dy}} \tag{3.15}$$

Các công thức (3.10), (3.11), (3.13)–(3.15) cần có điều kiện biểu thức ở mẫu số phải khác không.

Ví dụ 3.11. Cho biến ngẫu nhiên hai chiều (X,Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{x}, & \text{nếu } 0 < y < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm các hàm mật độ xác suất của X và Y. (b) Tìm các hàm mật độ xác suất có điều kiện $f_X(x|y)$ và $f_Y(y|x)$.

Lời giải Ví dụ 3.11

(a) Các hàm mật độ xác suất biên là:

$$f_X(x) = \int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \begin{cases} \displaystyle\int\limits_0^x \frac{1}{x}dy, & 0 < x < 1, \\ 0, & \text{trái lại,} \end{cases} = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{trái lại,} \end{cases}$$

$$f_Y(y) = \int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dx = \begin{cases} \displaystyle\int\limits_y^1 \frac{1}{x}dx, & 0 < y < 1, \\ 0, & \text{trái lại,} \end{cases} = \begin{cases} -\ln y, & 0 < y < 1, \\ 0, & \text{trái lại.} \end{cases}$$

(b) Các hàm mật độ xác suất có điều kiện là:

$$f_X(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} -\dfrac{1}{x\ln y}, & 0 < y < x < 1, \\ 0, & \text{trái lại.} \end{cases}$$

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \dfrac{1}{x}, & 0 < y < x < 1, \\ 0, & \text{trái lại.} \end{cases}$$

Ví dụ 3.12. Giả sử tại một trường đại học, một sinh viên đạt được điểm X trong bài kiểm tra năng khiếu toán học và điểm Y trong bài kiểm tra năng khiếu âm nhạc là một số trong khoảng từ 0 đến 1. Giả sử X và Y được phân phối theo hàm mật độ sau:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(2x+3y), & 0 < x,y < 1, \\ 0, & \text{nếu ngược lại.} \end{cases}$$

(a) Tính tỷ lệ sinh viên đại học đạt điểm cao hơn 0,8 trong bài kiểm tra năng khiếu toán. (b) Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu âm nhạc là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu toán học sẽ lớn hơn 0,8.

Lời giải Ví dụ 3.12

(a) $P(X > 0,8) = \int_0^1 \int_{0.8}^1 \frac{2}{5}(2x+3y)dxdy = \frac{2}{5}\int_0^1 (0,36 + 0,6y)dy = 0,264.$

(b) Ta cần tính $P(X > 0,8|Y = 0,3) = \int_{0,8}^1 f_X(x|y = 0,3)dx$, trong đó

$$f_X(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \dfrac{\frac{2}{5}(2x+3y)}{\frac{2}{5}\int_0^1 (2x+3y)dx}, & \text{nếu } 0 < x, y < 1, \\ 0, & \text{nếu trái lại} \end{cases}$$

$$= \begin{cases} \dfrac{2x+3y}{1+3y}, & \text{nếu } \quad 0 < x, y < 1, \\ 0, & \text{nếu trái lại,} \end{cases}$$

suy ra $f_X(x|y = 0,3) = \begin{cases} \dfrac{2x+0,9}{1,9}, & \text{nếu } \quad 0 < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$

Vậy,

$$P(X > 0,8|Y = 0,3) = \int_{0,8}^1 \frac{2x+0,9}{1,9}dx = \frac{0,36+0,18}{1,9} = 0,2842.$$

3.5 Tính độc lập của các biến ngẫu nhiên

Một số dấu hiệu nhận biết tính độc lập của hai biến ngẫu nhiên X và Y dựa trên tính chất của bảng phân phối xác suất đồng thời, hàm phân phối xác suất đồng thời, hàm mật độ xác suất đồng thời như trong Định nghĩa 3.2, 3.6 và 3.8 với các công thức (3.2), (3.6) và (3.9) tương ứng.

BÀI 10 (2 tiết)

3.6 Hàm của hai biến ngẫu nhiên

3.6.1 Khái niệm

Cho hai biến ngẫu nhiên X và Y và định nghĩa $Z = g(X,Y)$ là một phép cho tương ứng mỗi cặp giá trị (x,y) của (X,Y) với duy nhất một giá trị $z = g(z,y)$ của Z. Z được gọi là hàm của hai biến ngẫu nhiên X và Y. Chẳng hạn $Z = X+Y$, $Z = XY$.

3.6.2 Phân phối xác suất

Xét biến ngẫu nhiên $Z = g(X,Y)$, trong đó (X,Y) là biến ngẫu nhiên hai chiều đã biết luật phân phối. Ta sẽ xét phân phối xác suất của Z trong một số trường hợp đơn giản theo cách sau.

Định lý 3.2.

$$\boxed{F_Z(z) = P(Z < z) = P\left(g(X,Y) < z\right) = P\left((X,Y) \in D\right)} \tag{3.16}$$

trong đó $D = \left\{(x,y) \in \mathbb{R}^2 : g(x,y) < z\right\}$.

Để tìm hàm mật độ xác suất $f_Z(z)$ của biến ngẫu nhiên Z, ta sử dụng tính chất

$$f_Z(z) = \frac{dF_Z(z)}{dz}.$$

Ví dụ 3.13. Tính xác suất để hai người gặp được nhau trong Ví dụ 1.17.

Lời giải Ví dụ 3.13 Quy gốc thời gian về lúc 7h00. Gọi X và Y lần lượt là biến ngẫu nhiên chỉ thời điểm đến của hai người trong khoảng thời gian từ phút thứ 0 đến phút thứ 60. X và Y là các biến ngẫu nhiên liên tục có phân phối đều trên $[0;60]$. Do X và Y độc lập nên chúng có hàm mật độ xác suất đồng thời

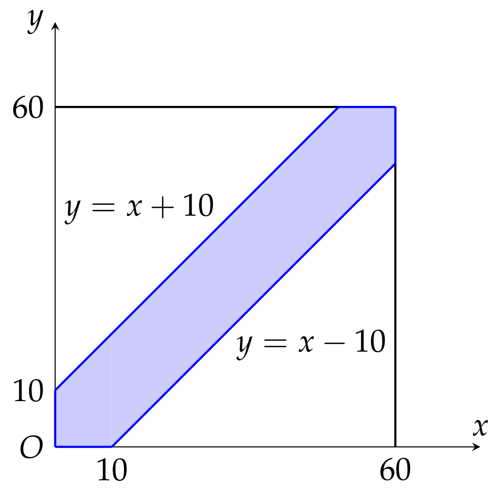
$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \begin{cases} \dfrac{1}{3600}, & (x,y) \in [0,60]^2, \\ 0, & \text{ngược lại.} \end{cases}$$

Đặt $Z = |X-Y|$. Khi đó, xác suất hai người gặp được nhau là

$$P(Z < 10) = P\left(|X-Y| < 10\right) = P\left((X,Y) \in D\right),$$

trong đó D là giao của miền $|x-y| < 10$ và hình vuông $[0;60]^2$ (xem Hình 3.5). Suy ra

$$P(Z < 10) = 1 - 2\int_{10}^{60} dx \int_0^{x-10} \frac{1}{3600}dy = 1 - \frac{2500}{3600} = \frac{11}{36}.$$



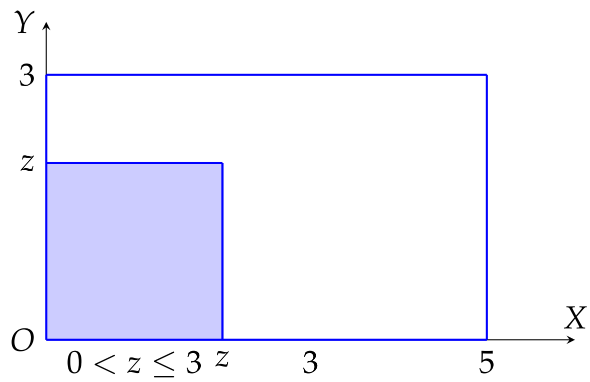
Hình 3.5: Miền D của Ví dụ 3.13

Ví dụ 3.14. Cho X và Y là hai biến ngẫu nhiên trong Ví dụ 3.9. Hãy tìm hàm mật độ của biến ngẫu nhiên $Z = \max(X, Y)$.

Lời giải Ví dụ 3.14 Trước hết ta tìm

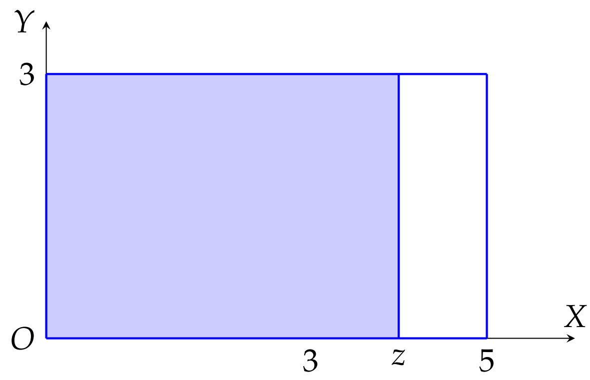
$$F_Z(z) = P(Z < z) = P(\max(X, Y) < z) = P(X < z, Y < z) = \iint\limits_D \frac{1}{15} dxdy,$$

trong đó $D = \{(x < z, y < z)\} \cap \{(0 \le x \le 3, 0 \le y \le 5)\}$.



Hình 3.6: Miền D với $0 < z \le 3$ của Ví dụ 3.14

Nếu $z \le 0$, $F_Z(z) = 0$.
Nếu $0 < z \le 3$, $F_Z(z) = \int_0^z dx \int_0^z \frac{1}{15} dy = \frac{z^2}{15}$.
Nếu $3 < z \le 5$, $F_Z(z) = \int_0^z dx \int_0^3 \frac{1}{15} dy = \frac{z}{5}$.
Nếu $z > 5$, $F_Z(z) = 1$.



Hình 3.7: Miền D với $3 < z \leq 5$ của Ví dụ 3.14

$$\text{Vậy } F_z(z) = \begin{cases} 0 & \text{nếu } z \leq 0, \\ \dfrac{z^2}{15} & \text{nếu } 0 < z \leq 3, \\ \dfrac{z}{5} & \text{nếu } 3 < z \leq 5, \\ 1 & \text{nếu } z > 5, \end{cases} \quad \text{suy ra} \quad f_z(z) = \begin{cases} \dfrac{2z}{15} & \text{nếu } 0 < z \leq 3, \\ \dfrac{1}{5} & \text{nếu } 3 < z \leq 5, \\ 0 & \text{nếu } z \leq 0,\ z > 5. \end{cases}$$

Ví dụ 3.15. Cho X, Y là 2 biến ngẫu nhiên độc lập với nhau có cùng phân phối đều trên $[0,2]$.

(a) Tìm hàm phân phối của các biến ngẫu nhiên $Z = X + Y$, $T = XY$, $U = X - Y$.

(b) Tính $P(-1 \leq Y - X \leq 1)$

Lời giải Ví dụ 3.15 Vì X và Y độc lập nên ta có hàm mật độ xác suất của (X, Y) là

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{4}, & (x,y) \in \mathcal{D}, \\ 0, & \text{trái lại}, \end{cases}$$

trong đó $D := \{0 \leq x \leq 2;\ 0 \leq y \leq 2\}$.

(a1) Hàm phân phối xác suất của $Z = X + Y$ được xác định bởi

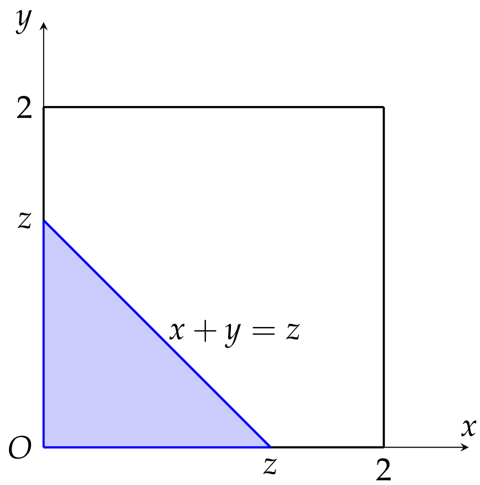
$$F_Z(z) = P\left(X + Y < z\right) = \iint\limits_{\{x+y<z\}} f_{X,Y}(x,y)dxdy = \frac{1}{4} \iint\limits_{\{x+y<z\}\cap D} dxdy.$$

Nếu $z \leq 0$ thì $F_Z(z) = 0$.

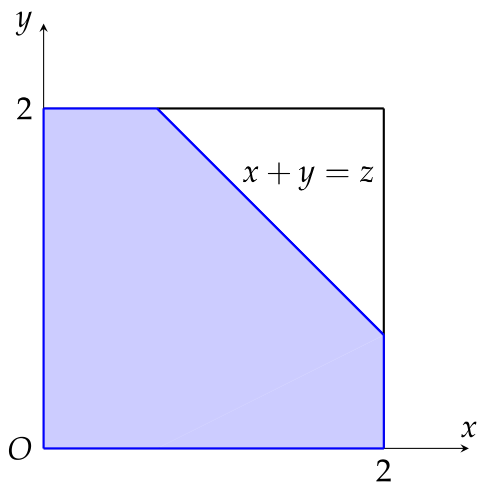
Nếu $0 < z \leq 2$ thì $F_Z(z) = \dfrac{1}{4}\displaystyle\int_0^z \left(\int_0^{z-x} dy\right)dx = \frac{1}{4} \times \frac{z^2}{2} = \frac{z^2}{8}$.

Nếu $2 < z \leq 4$ thì

$$F_Z(z) = 1 - \frac{1}{4}\int\limits_{z-2}^{2}\left(\int\limits_{z-x}^{2} dy\right)dx = 1 - \frac{1}{4}\int\limits_{z-2}^{2}(2 - z + x)dx = 1 - \frac{1}{8}\left(z^2 - 8z + 16\right)$$
$$= \frac{1}{8}(-z^2 + 8z - 8).$$



Hình 3.8: Miền $\{x + y < z\} \cap D$ với $0 < z \leq 2$ trong Ví dụ 3.15(a1)



Hình 3.9: Miền $\{x + y < z\} \cap D$ với $2 < z \leq 4$ trong Ví dụ 3.15(a1)

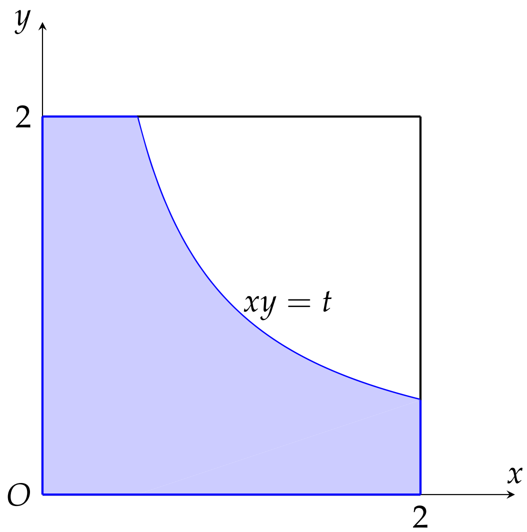
Nếu $z > 4$ thì $F_Z(z) = \dfrac{1}{4}\displaystyle\iint\limits_{\mathcal{D}} dxdy = 1$. Vậy

$$F_Z(z) = \begin{cases} 0, & \text{nếu } z \leq 0, \\ \dfrac{z^2}{8}, & \text{nếu } 0 < z \leq 2, \\ \dfrac{1}{8}(-z^2 + 8z - 8), & \text{nếu } 2 < z \leq 4, \\ 1, & \text{nếu } z > 4. \end{cases}$$

(a2) Hàm phân phối xác suất của $T = XY$ được xác định như sau:

$$F_T(t) = P(XY < t) = \iint\limits_{\{xy<t\}} f_{X,Y}(x,y)dxdy = \frac{1}{4}\iint\limits_{\{xy<t\}\cap\mathcal{D}} dxdy.$$

Nếu $t \leq 0$, $F_T(t) = 0$.



Hình 3.10: Miền $\{xy < t\} \cap D$ với $0 < t \leq 4$ trong Ví dụ 3.15(a2)

Nếu $0 < t \leq 4$, $F_T(t) = \dfrac{1}{4}\left(\displaystyle\int_0^{t/2} dx \int_0^2 dy + \int_{t/2}^2 dx \int_0^{t/x} dy\right) = \dfrac{1}{4}\left(t + t\ln 2 - t\ln\dfrac{t}{2}\right)$.

Nếu $t > 4$, $F_T(t) = 1$. Vậy

$$F_T(t) = \begin{cases} 0, & \text{nếu } t \leq 0,\\ \dfrac{1}{4}\left(t + t\ln 2 - t\ln\dfrac{t}{2}\right), & \text{nếu } 0 < t \leq 4,\\ 1, & \text{nếu } t > 4. \end{cases}$$

(a3) Hàm phân phối xác suất của $U = X - Y$ là

$$F_U(u) = P(X - Y < u) = \iint\limits_{\{x-y<u\}} f_{X,Y}(x,y)dxdy = \frac{1}{4}\iint\limits_{\{x-y<u\}\cap\mathcal{D}} dxdy.$$

Nếu $u \leq -2$, $F_U(u) = 0$.

Nếu $-2 < u \leq 0$, $F_U(u) = \dfrac{1}{4}\displaystyle\int_0^{u+2} dx \int_{x-u}^2 dy = \dfrac{1}{4} \times \dfrac{1}{2}(2+u)^2 = \dfrac{1}{8}(2+u)^2$.

Nếu $0 < u \leq 2$ thì $F_U(u) = \dfrac{1}{4}\left[4 - \dfrac{1}{2}(2-u)^2\right] = \dfrac{1}{4}\left(-\dfrac{u^2}{2} + 2u + 2\right)$.

Nếu $u > 2$, $F_U(u) = 1$. Vậy

$$F_U(u) = \begin{cases} 0, & \text{nếu } u \leq -2,\\ \dfrac{1}{8}(2+u)^2, & \text{nếu } -2 < u \leq 0,\\ \dfrac{1}{8}(-u^2 + 4u + 4), & \text{nếu } 0 < u \leq 2,\\ 1, & \text{nếu } u > 2. \end{cases}$$

(b) $P(-1 \le Y - X \le 1) = P(X - 1 \le Y \le X + 1) = \dfrac{1}{4}\left(4 - 2 \times \dfrac{1}{2}\right) = \dfrac{3}{4}.$

Định lý 3.3. Nếu X_1 và X_2 là hai biến ngẫu nhiên độc lập có phân phối Poa-xông với tham số λ_1 và λ_2 tương ứng thì $X_1 + X_2$ cũng là biến ngẫu nhiên có phân phối Poa-xông với tham số $\lambda_1 + \lambda_2$.

Chứng minh. Đặt $Y_1 = X_1 + X_2$, $Y_2 = X_2$, khi đó $X_1 = Y_1 - Y_2$ và $X_2 = Y_2$. Vì X_1 và X_2 độc lập nên với $k = 0, 1, 2, \ldots$, $l = 0, 1, 2, \ldots, k$, $l \le k$ (vì $X_1 \ge 0$),

$$\begin{aligned} P(Y_1 = k, Y_2 = l) &= P(X_1 = k - l, X_2 = l) = P(X_1 = k - l)P(X_2 = l) \\ &= \left(\frac{\lambda_1^{k-l}}{(k-l)!}e^{-\lambda_1}\right)\left(\frac{\lambda_2^{l}}{l!}e^{-\lambda_2}\right) = e^{-(\lambda_1+\lambda_2)}\frac{\lambda_1^{k-l}\lambda_2^{l}}{(k-l)!l!}. \end{aligned}$$

Từ đây suy ra

$$\begin{aligned} P(Y_1 = k) &= \sum_{l=0}^{k} P(Y_1 = k, Y_2 = l) = e^{-(\lambda_1+\lambda_2)}\sum_{l=0}^{k}\frac{\lambda_1^{k-l}\lambda_2^{l}}{(k-l)!l!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{l=0}^{k}\frac{k!}{(k-l)!l!}\lambda_1^{k-l}\lambda_2^{l} = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}, \quad k = 0, 1, 2, \ldots. \end{aligned}$$

Vậy $X_1 + X_2$ là biến ngẫu nhiên có phân phối Poa-xông với tham số $\lambda_1 + \lambda_2$. □

Định lý 3.4. If $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn với kỳ vọng $\mu_1, \mu_2, \ldots, \mu_n$ và phương sai $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ tương ứng, thì biến ngẫu nhiên

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

có phân phối chuẩn với kỳ vọng

$$\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

và phương sai

$$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2.$$

3.7 Đặc trưng của biến ngẫu nhiên hai chiều

3.7.1 Kỳ vọng, phương sai của biến ngẫu nhiên thành phần

Định lý 3.5. (a) Nếu (X, Y) là biến ngẫu nhiên hai chiều rời rạc có bảng phân phối xác suất như trong Định nghĩa 3.1 thì

$$E(X) = \sum_i x_iP(X = x_i) = \sum_i\sum_j x_ip_{ij}; \quad E(Y) = \sum_j y_jP(Y = y_j) = \sum_j\sum_i y_jp_{ij} \tag{3.17}$$

$$V(X) = \sum_i\sum_j x_i^2p_{ij} - \big(E(X)\big)^2; \qquad V(Y) = \sum_j\sum_i y_j^2p_{ij} - \big(E(Y)\big)^2. \tag{3.18}$$

(b) Nếu (X, Y) là biến ngẫu nhiên hai chiều liên tục có hàm mật độ xác suất $f_{XY}(x, y)$ thì

$$E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{XY}(x,y)dxdy; \qquad E(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f_{XY}(x,y)dxdy \tag{3.19}$$

$$V(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f_{XY}(x,y)dxdy - \big(E(X)\big)^2; \quad V(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f_{XY}(x,y)dxdy - \big(E(Y)\big)^2. \tag{3.20}$$

3.7.2 Kỳ vọng, phương sai của hàm của hai biến ngẫu nhiên

Định nghĩa 3.10. Cho biến ngẫu nhiên hai chiều (X, Y) có phân phối đã biết và ta xác định một biến mới $Z = g(X, Y)$ (g là hàm đo được).

$$E(Z) = E\left[g(X,Y)\right] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j) \quad \text{(nếu } (X,Y) \text{ rời rạc)} \tag{3.21}$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy \quad \text{(nếu } (X,Y) \text{ liên tục).} \tag{3.22}$$

Đặc biệt, khi $g(X, Y) = X$ thay vào các công thức trên ta sẽ có $E(X)$.

Ví dụ 3.16. Cho biến ngẫu nhiên hai chiều rời rạc (X, Y) có bảng phân phối xác như sau:

X \ Y	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

(a) Chứng minh rằng X và Y độc lập. (b) Tìm quy luật phân phối xác suất của biến ngẫu nhiên $Z = XY$. (c) Tính $E(Z)$ bằng hai cách và kiểm tra $E(Z) = E(X)E(Y)$.

Lời giải Ví dụ 3.16

(a) Bảng phân phối xác suất của X và Y là:

X	1	2
P	0,3	0,7

Y	1	2	3
P	0,4	0,5	0,1

Suy ra

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j), \quad \forall i = 1,2;\ j = 1,2,3$$

nên X, Y độc lập.

(b) Z nhận các giá trị $1, 2, 3, 4, 5, 6$ với

$$\begin{aligned}
&P(Z=1) = P(X=1, Y=1) = 0,12\\
&P(X=2) = P(X=1, Y=2) + P(X=2, Y=1) = 0,15 + 0,28 = 0,43\\
&P(X=3) = P(X=1, Y=3) = 0,03\\
&P(X=4) = P(X=2, Y=2) = 0,35\\
&P(Z=6) = P(X=2, Y=3) = 0,07
\end{aligned}$$

Quy luật phân phối xác suất của $Z = XY$ là:

Z	1	2	3	4	6
P	0,12	0,43	0,03	0,35	0,07

(c) Cách 1: sử dụng Định nghĩa 2.8(a),

$$E(Z) = 1 \times 0,12 + 2 \times 0,43 + 3 \times 0,03 + 4 \times 0,35 + 6 \times 0,07 = 2,89.$$

Cách 2: sử dụng Định nghĩa 3.2(a),

$$\begin{aligned}
E(Z) = E(XY) &= \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j P(X = x_i, Y = y_j)\\
&= (1)(1)(0,12) + (1)(2)(0,15) + (1)(3)(0,03) + (2)(1)(0,28)\\
&\quad + (2)(2)(0,35) + (2)(3)(0,07) = 2,89.
\end{aligned}$$

Ví dụ 3.17. Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & \text{nếu} \quad 0 < x < 2,\ 0 < y < 1,\\ 0, & \text{trong trường hợp trái lại.} \end{cases}$$

Tính $E(Y/X)$.

Lời giải Ví dụ 3.17 Sử dụng công thức (3.22),

$$E\left(\frac{Y}{X}\right) = \int_0^1\int_0^2 \frac{y(1+3y^2)}{4}dxdy = \int_0^1 \frac{y+3y^3}{2}dy = \frac{5}{8}.$$

Sau đây là một mở rộng của Định lý 2.6.

Định lý 3.6. Cho $h(X, Y)$, $g(X, Y)$ là các hàm của biến ngẫu nhiên hai chiều (X, Y). Khi đó

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)].$$

Hệ quả 3.1. (a) Nếu $g(X, Y) = g(X)$ và $h(X, Y) = h(Y)$ thì

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)].$$

(b) Nếu $g(X,Y) = X$ và $h(X,Y) = Y$ thì

$$E(X \pm Y) = E(X) \pm E(Y).$$

Định lý 3.7. Nếu X và Y là hai biến ngẫu nhiên độc lập thì

$$E(XY) = E(X)E(Y).$$

Chứng minh. Giả sử X và Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất lần lượt là $f_X(x)$ và $f_Y(y)$. Vì X và Y độc lập nên $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, ở đây $f_{X,Y}(x,y)$ là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên X và Y. Theo Định nghĩa 3.10,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_{X,Y}(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{+\infty} xf_X(x)dx \int_{-\infty}^{+\infty} yf_Y(y)dy = E(X)E(Y). \end{aligned}$$

□

3.7.3 Hiệp phương sai

Nếu $g(X,Y) = (X - E(X))(Y - E(Y))$ thì Định nghĩa 3.10 cho ta một giá trị kỳ vọng, gọi là hiệp phương sai của hai biến ngẫu nhiên X và Y.

Định nghĩa 3.11 (Hiệp phương sai)**.** Cho biến ngẫu nhiên hai chiều (X,Y), hiệp phương sai của hai thành phần X và Y, ký hiệu là $cov(X,Y)$ được xác định bởi

$$\boxed{cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right]} \tag{3.23}$$

Một công thức khác để tính hiệp phương sai, tương đương với công thức (3.23) được nêu trong định lý sau đây.

Định lý 3.8.

$$\boxed{cov(X,Y) = E(XY) - E(X).E(Y)} \tag{3.24}$$

Chứng minh. Ta chứng minh trong trường hợp biến ngẫu nhiên rời rạc. Giả sử biến ngẫu nhiên hai chiều (X,Y) có bảng phân phối xác suất đồng thời cho trong Định nghĩa 3.1. Theo Định nghĩa 3.2 và 3.11,

$$\begin{aligned} cov(X,Y) &= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_i - E(X))(y_j - E(Y))p_{ij} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_iy_jp_{ij} - E(X)\sum_{i=1}^{m}\sum_{j=1}^{n} y_jp_{ij} - E(Y)\sum_{i=1}^{m}\sum_{j=1}^{n} x_ip_{ij} + E(X)E(Y)\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) = E(XY) - E(X)E(Y), \end{aligned}$$

vì $\sum_{i=1}^{m}\sum_{j=1}^{n} y_jp_{ij} = E(Y)$, $\sum_{i=1}^{m}\sum_{j=1}^{n} x_ip_{ij} = E(X)$, $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1$.

Đối với biến ngẫu nhiên liên tục được chứng minh tương tự. □

Từ Định lý 3.7 và 3.8 ta thu được các tính chất sau đây của hiệp phương sai.

Định lý 3.9. (a) $cov(X,Y) = cov(Y,X)$.

(b) $V(X) = cov(X,X)$, $V(Y) = cov(Y,Y)$.

(c) Nếu X, Y độc lập thì $cov(Y,X) = 0$, điều ngược lại chưa chắc đã đúng.

(d) $cov(aX,Y) = acov(X,Y)$ với a là hằng số.

(e) $cov(X+Z,Y) = cov(X,Y) + cov(Z,Y)$.

(f) $cov\Big(\sum_{i=1}^n X_i, Y\Big) = \sum_{i=1}^n cov(X_i,Y)$.

Ví dụ 3.18. Cho biến ngẫu nhiên hai chiều (X,Y) có bảng phân phối xác suất là

X \ Y	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

(a) Tìm $E(X)$, $E(Y)$, $cov(X,Y)$. (b) X và Y có độc lập không?

Lời giải Ví dụ 3.18

(a) Ta có

$$E(X) = (-1)\times\frac{9}{15} + 0\times\frac{4}{15} + 1\times\frac{2}{15} = -\frac{7}{15}.$$
$$E(Y) = (-1)\times\frac{5}{15} + 0\times\frac{5}{15} + 1\times\frac{5}{15} = 0.$$
$$E(XY) = (-1)\times(-1)\times\frac{4}{15} + (-1)\times(1)\times\frac{4}{15} + 1\times(-1)\times 0 + 1\times 1\times 0 = 0.$$

Suy ra $cov(X,Y) = E(XY) - E(X)\times E(Y) = 0$.

(b) Dễ kiểm tra được $P(X=-1,Y=-1) \neq P(X=-1)\times P(Y=-1)$ nên X, Y không độc lập.

Trong Ví dụ 3.18 ta có $cov(X,Y) = 0$ nhưng hai biến ngẫu nhiên X và Y không độc lập.

Nhận xét 3.7. Hiệp phương sai được dùng làm độ đo quan hệ giữa hai biến X và Y:

(a) $cov(X,Y) > 0$ cho thấy xu thế Y tăng khi X tăng.

(b) $cov(X,Y) < 0$ cho thấy xu thế Y giảm khi X tăng.

Sau đây là một số tính chất nhằm cung cấp thêm công cụ để tính phương sai và độ lệch chuẩn.

Định lý 3.10. Cho (X, Y) là biến ngẫu nhiên hai chiều, a, b, c là các hằng số. Khi đó

$$V(aX + bY + c) = a^2V(X) + b^2V(Y) + 2ab\, cov(X, Y).$$

Chứng minh. Sử dụng Định nghĩa 2.9,

$$V(aX + bY + c) = E[(aX + bY + c) - E(aX + bY + c)]^2.$$

Theo Hệ quả 3.1 và Hệ quả 2.1,

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

Suy ra

$$\begin{aligned} V(aX + bY + c) &= E\{a[X - E(X)] + b[Y - E(Y)]\}^2 \\ &= a^2E[X - E(X)]^2 + b^2E[Y - E(Y)]^2 + 2abE\{[X - E(X)][Y - E(Y)]\} \\ &= a^2V(X) + b^2V(Y) + 2ab\, cov(X, Y). \end{aligned}$$

□

Từ Định lý 3.10, nếu cho $b = 0$, $c = 0$ ta nhận được Định lý 2.9(a), nếu cho $a = 0$, $b = 0$ ta nhận được Định lý 2.9(b).

Hệ quả 3.2. (a) Nếu $a = 1$, $b = 1$ và $c = 0$ ta có

$$V(X + Y) = V(X) + V(Y) + 2cov(X, Y).$$

(b) Nếu $a = 1$, $b = -1$ và $c = 0$ ta có

$$V(X - Y) = V(X) + V(Y) - 2cov(X, Y).$$

Từ Định lý 3.10 và Định lý 3.9(c) ta có hệ quả sau.

Hệ quả 3.3. Nếu X và Y là hai biến ngẫu nhiên độc lập thì

$$V(aX + bY) = a^2V(X) + b^2V(Y).$$
$$V(aX - bY) = a^2V(X) + b^2V(Y).$$

Định nghĩa 3.12 (Ma trận hiệp phương sai)**.** Ma trận hiệp phương sai của biến ngẫu nhiên hai chiều (X, Y) được xác định bởi

$$\Gamma = \begin{bmatrix} cov(X, X) & cov(X, Y) \\ cov(Y, X) & cov(Y, Y) \end{bmatrix} = \begin{bmatrix} V(X) & cov(X, Y) \\ cov(X, Y) & V(Y) \end{bmatrix}$$

Tính chất 3.4. (a) Ma trận hiệp phương sai là ma trận đối xứng.

(b) Ma trận hiệp phương sai là ma trận của dạng toàn phương không âm.

Nhận xét 3.8. Hiệp phương sai có hạn chế cơ bản là khó xác định được miền biến thiên, nó thay đổi từ cặp biến thiên này sang cặp biến thiên khác. Chưa kể về mặt vật lý nó có đơn vị đo bằng bình phương đơn vị đo của biến ngẫu nhiên X, Y (nếu chúng cùng đơn vị đo). Vì thế cần đưa ra một số đặc trưng khác để khắc phục hạn chế này, đó là "hệ số tương quan".

3.7.4 Hệ số tương quan

Định nghĩa 3.13 (Hệ số tương quan)**.** Hệ số tương quan của hai biến ngẫu nhiên X và Y, ký hiệu là ρ_{XY}, được xác định như sau:

$$\boxed{\rho_{XY} = \frac{cov(X,Y)}{\sqrt{V(X).V(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}} \tag{3.25}$$

Tính chất 3.5. (a) $|\rho_{XY}| \le 1$.

(b) Nếu $\rho_{XY} = \pm 1$ ta nói hai biến ngẫu nhiên X và Y có quan hệ tuyến tính (tức là tồn tại a và b sao cho $Y = aX + b$).

(c) Nếu $\rho_{XY} = 0$ ta nói hai biến ngẫu nhiên X và Y là không tương quan.

Nói chung $0 < |\rho_{XY}| < 1$, trong trường hợp này ta nói hai biến X và Y tương quan với nhau. Chú ý rằng, hai biến tương quan thì phụ thuộc (không độc lập), nhưng không tương quan thì chưa chắc độc lập.

Nhận xét 3.9. Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho_{XY}|$ càng gần 1 thì tính chất tương quan tuyến tính càng chặt. Khi $|\rho_{XY}|$ càng gần 0 thì sự phụ thuộc tuyến tính càng ít, càng lỏng lẻo. Khi $\rho_{XY} = 0$ ta nói X và Y không tương quan. Như vậy hai biến ngẫu nhiên độc lập thì không tương quan, nhưng ngược lại chưa chắc đúng.

Ví dụ 3.19. Cho biến ngẫu nhiên hai chiều rời rạc (X,Y) có bảng phân phối xác suất là

X \ Y	1	2	3
1	0,17	0,13	0,25
2	0,10	0,30	0,05

(a) Tìm ma trận hiệp phương sai của (X,Y). (b) Tìm hệ số tương quan $\rho_{X,Y}$.

Lời giải Ví dụ 3.19

(a) Tính

$$E(X) = 1 \times 0,55 + 2 \times 0,45 = 1,45;\ V(X) = 1 \times 0,55 + 4 \times 0,45 - (1,45)^2 = 0,2475.$$
$$E(Y) = 2,03;\ V(Y) = 0,5691.$$
$$E(XY) = 2,88 \Longrightarrow cov(X,Y) = E(XY) - E(X) \times E(Y) = -0,0635.$$

Vậy ma trận hiệp phương sai

$$\Gamma = \begin{pmatrix} V(X) & cov(X,Y) \\ cov(Y,X) & V(Y) \end{pmatrix} = \begin{pmatrix} 0,2475 & -0,0635 \\ -0,0635 & 0,5691 \end{pmatrix}$$

(b) Hệ số tương quan $\rho_{XY} = \dfrac{cov(X,Y)}{\sqrt{V(X) \times V(Y)}} = -0,1692.$

Ví dụ 3.20. Trọng lượng của những người chồng tuân theo luật phân phối chuẩn với trọng lượng trung bình là 70kg và độ lệch chuẩn 9kg, còn trọng lượng của những người vợ tuân theo luật phân phối chuẩn với trọng lượng trung bình là 55kg và độ lệch chuẩn 4kg. Hệ số tương quan trọng lượng giữa vợ và chồng là $\frac{2}{3}$. Tính xác suất vợ nặng hơn chồng.

Lời giải Ví dụ 3.20 Gọi X và Y lần lượt là các biến ngẫu nhiên chỉ "trọng lượng của chồng" và "trọng lượng của vợ". Ta có $X \sim \mathcal{N}(70; 9^2)$, $Y \sim \mathcal{N}(55; 4^2)$. Ta cần tính $P(X < Y)$.

Vì $X - Y$ là biến ngẫu nhiên có phân phối chuẩn với

$$E(X-Y) = E(X) - E(Y) = 70 - 55 = 15$$

và

$$V(X-Y) = V(X) + V(Y) - 2\text{cov}(X,Y) = 9^2 + 4^2 - (2)(24) = 49,$$

trong đó $\text{cov}(X,Y) = \rho_{XY}\sigma(X)\sigma(Y) = \dfrac{2}{3}(9)(4) = 24$. Vậy $X - Y \sim \mathcal{N}(15; 49)$. Suy ra

$$P(X<Y) = P(X-Y<0) = 0,5 + \phi\left(\frac{0-15}{\sqrt{49}}\right) = 0,5 - \phi(2,14) \simeq 0,5 - 0,48382 = 0.01618,$$

trong đó $\phi(2,14) = 0,48382$ được tra từ bảng hàm số Láp-la-xơ (Phụ lục 2).

3.8 Luật số lớn và định lý giới hạn trung tâm

Trong mục này ta nghiên cứu luật số lớn đó là sự hội tụ theo xác suất của dãy các biến ngẫu nhiên và định lý giới hạn trung tâm: khảo sát sự hội tụ theo phân phối xác suất của dãy các biến ngẫu nhiên.

3.8.1 Sự hội tụ của dãy các biến ngẫu nhiên

Định nghĩa 3.14 (Hội tụ theo xác suất)**.** Xét dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên X trong cùng một phép thử. Ta nói rằng dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X nếu với mọi $\varepsilon > 0$

$$\boxed{\lim_{n\to\infty} P\Big(|X_n - X| > \varepsilon\Big) = 0} \tag{3.26}$$

Như vậy dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế gần như chắc chắn ta có thể coi rằng X_n không khác mấy so với X.

Định nghĩa 3.15 (Hội tụ theo phân phối)**.** Dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo phân phối về biến ngẫu nhiên X nếu dãy các hàm phân phối xác suất $\{F_{X_n}(x)\}_{n=1}^{\infty}$ hội tụ về hàm phân phối $F_X(x)$. Tức là với mọi $x \in \mathbb{R}$

$$\boxed{\lim_{n\to\infty} F_{X_n}(x) = F_X(x)} \tag{3.27}$$

3.8.2 Luật số lớn Trê-bư-sep

Định lý 3.11 (Bất đẳng thức Markov)**.** Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{P(Y \geq \epsilon) < \frac{E(Y^2)}{\epsilon^2}} \tag{3.28}$$

Chứng minh. Ta chứng minh cho trường hợp Y là biến ngẫu nhiên liên tục có hàm mật độ xác suất là $f_Y(y)$.

$$P(Y \geq \varepsilon) = \int_{\varepsilon}^{+\infty} f_Y(y)dy = \frac{1}{\varepsilon^2}\int_{\varepsilon}^{+\infty} \varepsilon^2 f_Y(y)dy \leq \frac{1}{\varepsilon^2}\int_{\varepsilon}^{+\infty} y^2 f_Y(y)dy \leq \frac{1}{\varepsilon^2}\int_{0}^{+\infty} y^2 f_Y(y)dy = \frac{E(Y^2)}{\varepsilon^2}.$$

Dấu bằng không thể đồng thời xảy ra ở cả 2 dấu "=" và "≤" trong biểu thức trên. □

Định lý 3.12 (Bất đẳng thức Trê-bư-sep)**.** Cho X là biến ngẫu nhiên có kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$ hữu hạn. Khi đó với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{P(|X-\mu| \geq \varepsilon) < \frac{\sigma^2}{\varepsilon^2}} \tag{3.29}$$

hay tương đương

$$\boxed{P(|X-\mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}} \tag{3.30}$$

Chứng minh. Ta chứng minh cho trường hợp X là biến ngẫu nhiên liên tục. Đặt $Y = |X - \mu| \geq 0$ và áp dụng Định lý 3.11. □

Các bất đẳng thức (3.29) và (3.30) được gọi là bất đẳng thức Trê-bư-sep.

Nhận xét 3.10. Bất đẳng thức Trê-bư-sep có nhiều ứng dụng. Trước hết nó cho phép ta đánh giá cận trên hoặc cận dưới của xác suất để biến ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng $E(X)$ không quá ϵ. Bất đẳng thức Trê-bư-sep có ý nghĩa to lớn về mặt lý thuyết, nó được sử dụng để chứng minh các định lý của luật số lớn.

Ví dụ 3.21. Để ước lượng nhanh chóng sai số của số vải bán ra trong một tháng của một cửa hàng, ta tiến hành như sau:

1. Giả sử có n khách hàng trong một tháng và số vải của mỗi khách hàng được làm tròn bởi số nguyên gần nhất (ví dụ trong sổ ghi 195,6m thì được làm tròn là 196m).

2. Ký hiệu X_i là sai số giữa số mét vải thực bán và số mét vải đã làm tròn của khách hàng thứ i trong tháng, $i = 1, 2, \ldots, n$. Các sai số $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập cùng có phân bố đều trên đoạn $[-0,5; 0,5]$. Khi đó $E(X_i) = 0$ và $V(X_i) = \dfrac{1}{12}$,

$i = 1, 2, \ldots, n$. Sai số tổng cộng trong cả tháng là $X = X_1 + X_2 + \cdots + X_n$. Ta tính được $E(X) = 0$ và $V(X) = \dfrac{n}{12}$. Theo bất đẳng thức Trê-bư-sep, xác suất để sai số vượt quá ε mét sẽ được đánh giá bởi:

$$P(|X - 0| \geq \varepsilon) = P(|X| \geq \varepsilon) < \frac{n}{12\varepsilon^2}.$$

3. Bây giờ giả sử $n = 10000$. Để xác suất $P(|X| \geq \varepsilon)$ bé hơn 0,01 ta phải có $\dfrac{n}{12\varepsilon^2} \leq 0,01$ hay $\varepsilon \geq 288,6746$.

4. Vậy ta có thể kết luận: Với xác suất 0,99 sai số giữa số vải thực bán với số vải đã tính tròn không vượt quá 289m, nếu số khách hàng là 100000.

Định lý 3.13 (Định lý Trê-bư-sep). Nếu dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập, có kỳ vọng hữu hạn và phương sai bị chặn đều ($V(X_i) \leq C$, $\forall i = 1, 2, \ldots$, C là hằng số dương), thì với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) = 1} \tag{3.31}$$

Chứng minh. Áp dụng Định lý (3.12) với $X = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ ta nhận được kết quả của Định lý Trê-bư-sep. □

Hệ quả 3.4. Nếu dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập, có cùng kỳ vọng hữu hạn ($E(X_i) = \mu$, $i = 1, 2, \ldots$) và phương sai bị chặn đều ($V(X_i) \leq C$ $\forall i = 1, 2, \ldots$, C là hằng số dương), thì với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1} \tag{3.32}$$

Nhận xét 3.11. (a) Định lý Trê-bư-sep chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó. Nói cách khác có sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng với xác suất rất lớn. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên. Chẳng hạn, gieo một con xúc xắc cân đối đồng chất. Gọi X là số chấm xuất hiện ở mặt trên con xúc xắc thì $E(X) = 3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối đồng chất 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số chấm xuất hiện ở mặt trên con xúc xắc. Số chấm trung bình của 1 triệu lần gieo được tìm thấy là $\dfrac{x_1 + x_2 + \cdots + x_{10^6}}{10^6} \simeq 3,500867$

(b) Định lý Trê-bư-sep có ứng dụng rộng rãi trong nhiều lĩnh vực, chẳng hạn nó chính là cơ sở cho phương pháp đo lường trong vật lý. Để xác định giá trị của một đại lượng vật lý nào đó người ta thường tiến hành đo n lần độc lập và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Thật vậy, giả sử xem kết quả của n lần đo là các biến ngẫu nhiên $X_1, X_2, \dots, X_n$. Ta thấy rằng các biến ngẫu nhiên này độc lập, có cùng kỳ vọng bằng chính giá trị thực của đại lượng vật lý (giả sử không có sai số hệ thống), các phương sai của chúng đều bị chặn trên bởi bình phương của độ chính xác của thiết bị đo. Do đó theo định lý Trê-bư-sep ta có thể cho rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng vật lý với xác suất gần như bằng một.

(c) Định lý Trê-bư-sep còn là cơ sở cho lý thuyết mẫu, ứng dụng trong thống kê.

3.8.3 Luật số lớn Béc-nu-li

Áp dụng luật số lớn Trê-bư-sep với trường hợp $X_i \sim \mathcal{B}(1, p)$ chính là số lần xảy ra A trong phép thử thứ i, $i = 1, 2, \dots, n$ ta có luật số lớn Béc-nu-li.

Định lý 3.14 (Định lý Béc-nu-li)**.** Giả sử ta có n phép thử Béc–nu–li với $P(A) = p$ và m là số lần xảy ra A trong n phép thử đó. Khi đó với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{\lim_{n \to +\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1} \tag{3.33}$$

Nhận xét 3.12. Định lý Béc-nu-li chỉ ra rằng tần suất xuất hiện của sự kiện A trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của sự kiện đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Béc-nu-li là cơ sở lý thuyết của định nghĩa thống kê về xác suất ở Chương 1: Khi $n \to +\infty$ thì $\dfrac{m}{n} \to p$.

Ví dụ 3.22. Giả sử p là tỉ lệ cử tri sẽ bầu cho ứng cử viên A. Để ước lượng trước tỉ lệ này người ta phỏng vấn ngẫu nhiên n cử tri. Có thể coi kết quả bầu của cử tri thứ i là biến ngẫu nhiên X_i có phân phối Béc-nu-li tham số p (X_i nhận giá trị 1 nếu cử tri thứ i bầu cho ứng cử viên A và nhận giá trị 0 trong trường hợp ngược lại). Các biến ngẫu nhiên X_i, $i = 1, 2, \dots, n$ độc lập và có cùng phân phối Béc-nu-li tham số p với kỳ vọng $E(X_i) = p$ và phương sai $V(X_i) = p(1-p)$. Đặt $K_n = X_1 + X_2 + \cdots + X_n$, áp dụng bất đẳng thức Trê-bư-sep ta có

$$P(|K_n - p| \geq \varepsilon) < \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \quad \text{hay} \quad P(|K_n - p| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

Chẳng hạn, nếu $\varepsilon = 0,1$ và $n = 100$ thì $P(|K_{100} - p| \geq 0,1) < \dfrac{1}{4 \times 100 \times (0,1)^2} = 0,25$. Nói cách khác, nếu phỏng vấn 100 cử tri và lấy kết quả này để ước lượng cho tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A thì sai số vượt quá 0,1 có xác suất nhỏ hơn 0,25.

Nếu muốn ước lượng tin cậy hơn (chẳng hạn xác suất lớn hơn 95%) và chính xác hơn (với sai số 0,01) thì $P(|K_n - p| < 0,01) \geq 1 - \dfrac{1}{4n(0,01)^2}$. Vậy số cử tri phải phỏng vấn thỏa mãn

$$1 - \frac{1}{4n(0,01)^2} \geq 0,95 \quad \text{suy ra} \quad n \geq 50000.$$

Như vậy để ước lượng với độ tin cậy cao và độ chính xác cao thì cần phải lấy mẫu với số lượng lớn. Tuy nhiên ở đây ta chỉ dựa vào bất đẳng thức Trê-bư-sep để giải quyết bài toán. Trong Chương 4 ta sẽ nghiên cứu về bài toán ước lượng này và bằng phương pháp khác ta sẽ chỉ ra số cử tri phải phỏng vấn nhỏ hơn kết quả trên.

3.8.4 Định lý giới hạn trung tâm

Định lý 3.15 (Định lý giới hạn trung tâm). Giả sử $\{X_n\}$ là dãy biến ngẫu nhiên độc lập cùng phân phối với $E(X_n) = \mu$, $V(X_n) = \sigma^2$ với mọi n. Đặt $\overline{X}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$. Khi đó dãy biến ngẫu nhiên $U_n = \dfrac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$ hội tụ theo phân phối về phân phối chuẩn tắc $\mathcal{N}(0;1)$, tức là

$$\lim_{n\to\infty} F_{U_n}(x) = \Phi(x) \quad x \in \mathbb{R}, \tag{3.34}$$

ở đây $\Phi(x)$ là hàm phân phối xác suất của biến ngẫu nhiên phân phối chuẩn tắc được xác định bởi (2.42).

Nhận xét 3.13. (a) Ý nghĩa của Định lý giới hạn trung tâm là khi có nhiều nhân tố ngẫu nhiên tác động (sao cho không có nhân tố nào vượt trội lấn át các nhân tố khác) thì kết quả của chúng có dạng phân phối tiệm cận chuẩn.

(b) Trong thực hành, nếu n đủ lớn ta có:

$$\boxed{U_n = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0;1)} \tag{3.35}$$

3.8.5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Ta sử dụng định lý giới hạn để trình bày lại việc xấp xỉ phân phối nhị thức bằng phân phối Poa-xông và phân phối chuẩn đã đề cập ở Chương 2.

Giả sử $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối Béc-nu-li tham số p. Theo công thức (2.30) và (2.33) ta có $X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n,p)$. Công thức (1.19) cho phép tính xác suất $P(X = k) = P_n(k) = C_n^k p^k (1-p)^{n-k}$. Tuy nhiên khi n khá lớn ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ (vì sẽ tràn bộ nhớ khi tính toán có sử dụng máy tính).

Định lý 3.16 (Định lý giới hạn địa phương)**.** Giả sử X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n,p)$. Đặt $x_k = \dfrac{k-np}{\sqrt{np(1-p)}}$, khi đó

$$\boxed{P(X=k) = \frac{\varphi(x_k)}{\sqrt{np(1-p)}}(1+\varepsilon_{n,k})} \tag{3.36}$$

trong đó $|\varepsilon_{n,k}| < \dfrac{C}{\sqrt{n}}$ với C là hằng số.

Khi n đủ lớn ta có thể xấp xỉ xác suất $P_n(k)$ trong (1.19) bởi

$$\boxed{P_n(k) = P(X=k) \simeq \frac{\varphi(x_k)}{\sqrt{np(1-p)}}} \tag{3.37}$$

trong đó $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ là hàm Gao–xơ với các giá trị được tính trong bảng giá trị hàm Gao–xơ (Phụ lục 1) đối với các giá trị x dương. Hàm $\varphi(x)$ là hàm chẵn, tức là $\varphi(-x) = \varphi(x)$. Khi $x > 4$ ta có thể lấy $\varphi(x) \simeq 0$.

Để tính xấp xỉ giá trị của hàm phân phối xác suất nhị thức ta có thể áp dụng Định lý Moa-vrơ–Láp-la-xơ.

Định lý 3.17 (Moa-vrơ–Láp-la-xơ)**.** Đối với các các biến ngẫu nhiên độc lập $X_1, X_2, \ldots, X_n$ có cùng phân phối Béc-nu-li tham số p thì

$$\lim_{n\to\infty} P\left(\frac{X_1+X_2+\cdots+X_n-np}{\sqrt{npq}} < x\right) = \Phi(x) \quad x \in \mathbb{R}. \tag{3.38}$$

Chứng minh. Áp dụng định lý giới hạn trung tâm (Định lý 3.15) cho dãy các biến ngẫu nhiên độc lập $X_1, X_2, \ldots, X_n$ có cùng phân bố Béc-nu-li tham số p ta được kết quả của Định lý 3.17. □

Nếu X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n,p)$, thì khi n khá lớn ta có thể sử dụng công thức xấp xỉ:

$$\boxed{P(X<x) = P\left(\frac{X-np}{\sqrt{np(1-p)}} < \frac{x-np}{\sqrt{np(1-p)}}\right) \simeq \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)} \tag{3.39}$$

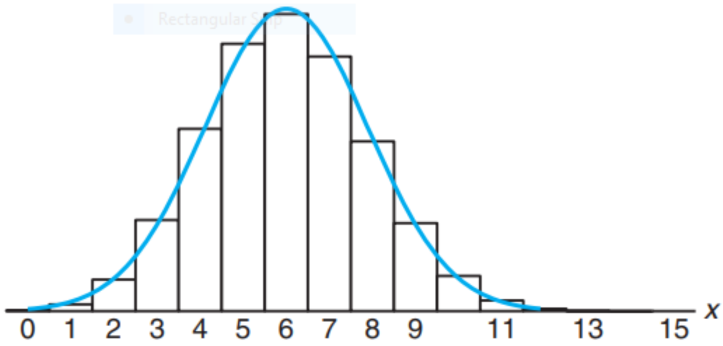
và

$$\boxed{P(k_1 \le X \le k_2) \simeq \Phi\left(\frac{k_2-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1-np}{\sqrt{np(1-p)}}\right)} \tag{3.40}$$

Công thức xấp xỉ trên là tốt nếu $np \ge 5$ và $n(1-p) \ge 5$.

Nhận xét 3.14. Khi $k_1 = k_2 = k$, $0 \le k \le n$, vế trái của công thức (3.40) sẽ là $P(X=k) \neq 0$, trong khi đó vế phải bằng 0. Điều này xảy ra vì ta đã dùng hàm phân phối liên tục để xấp xỉ phân phối rời rạc.

Ví dụ 3.23. Để minh họa việc xấp xỉ phân phối chuẩn cho phân phối nhị thức, ta vẽ biểu đồ của $\mathcal{B}(15;0,4)$ và vẽ đường cong chuẩn có cùng kỳ vọng $\mu = np = 15 \times 0,4 = 6$ và phương sai $\sigma^2 = npq = 15 \times 0,4 \times 0,6 = 3,6$ với biến ngẫu nhiên tuân theo luật phân phối nhị thức X (xem Hình 3.11).



Hình 3.11: Xấp xỉ phân phối chuẩn cho phân phối nhị thức $\mathcal{B}(15;0,4)$

Trong hình minh họa về xấp xỉ phân phối nhị thức bởi phân phối chuẩn, vì ta xấp xỉ một phân phối rời rạc bằng một phân phối liên tục, nên cần một sự hiệu chỉnh để giảm sai số.

Định nghĩa 3.16. Cho X là biến ngẫu nhiên tuân theo luật phân phối nhị thức $\mathcal{B}(n;p)$. Phân phối xác suất của X được xấp xỉ bởi phân phối chuẩn $\mathcal{N}(\mu,\sigma^2)$ với $\mu = np$ và $\sigma^2 = np(1-p)$ và

$$\boxed{P(X=k) = C_n^k p^k (1-p)^{n-k} \simeq \Phi\left(\frac{k+0,5-\mu}{\sigma}\right) - \Phi\left(\frac{k-0,5-\mu}{\sigma}\right)} \tag{3.41}$$

$$\boxed{P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k} \simeq \Phi\left(\frac{k_2+0,5-\mu}{\sigma}\right) - \Phi\left(\frac{k_1-0,5-\mu}{\sigma}\right)} \tag{3.42}$$

Xấp xỉ là khá tốt nếu $np \ge 5$ và $n(1-p) \ge 5$.

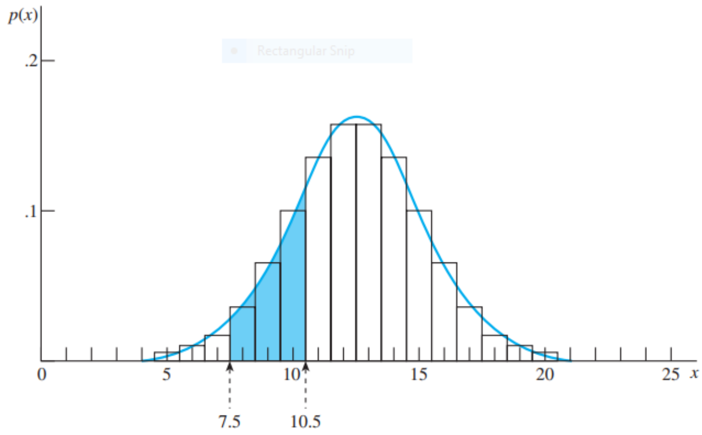
Nhận xét 3.15. Hình 3.12 và 3.13 biểu thị biểu đồ xác suất nhị thức với $n = 25$ và $p = 0,5$, $p = 0,1$ tương ứng. Phân phối trong Hình 3.12 là hoàn toàn đối xứng.

Việc thêm $+0,5$ và $-0,5$ chính là yếu tố hiệu chỉnh và gọi là hiệu chỉnh liên tục.

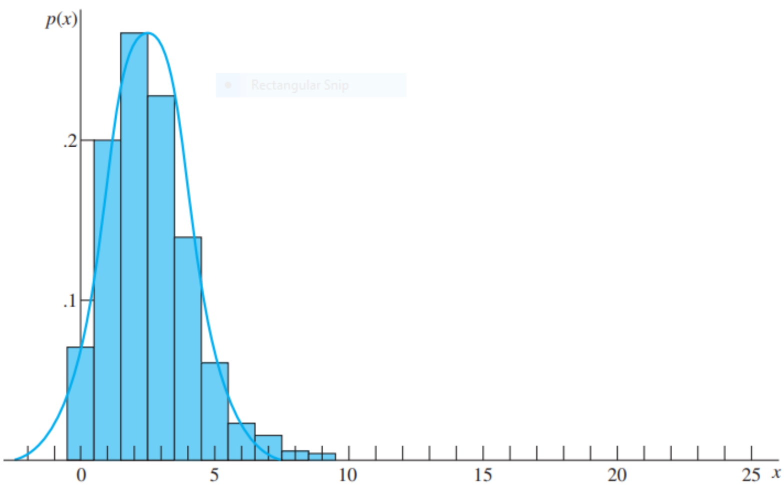
3.8.6 Xấp xỉ phân phối nhị thức bằng phân phối Poa-xông

Khi điều kiện xấp xỉ phân phối nhị thức bằng phân phối chuẩn không thỏa mãn (tức là điều kiện $np > 5$ và $n(1-p) > 5$ không thỏa mãn), ta có thể xấp xỉ phân phối nhị thức bằng phân phối Poa-xông.

Định lý 3.18. Cho dãy $\{X_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên có phân bố nhị thức, với mỗi X_n có phân phối nhị thức $X_n \sim \mathcal{B}(n, p_n)$. Nếu tồn tại giới hạn $\lim_{n\to\infty} np_n = \lambda > 0$ thì X_n hội tụ theo phân phối về biến ngẫu nhiên X có phân phối Poa-xông tham số λ.



Hình 3.12: Phân phối nhị thức với $n = 25$ và $p = 0,5$ xấp xỉ bởi phân phối chuẩn với $\mu = 12,5$ và $\sigma = 2,5$



Hình 3.13: Phân phối nhị thức và xấp xỉ phân phối chuẩn với $n = 25$ và $p = 0,1$

3.9 Tổng hợp một số đề thi

Ví dụ 3.24 (Đề thi cuối kỳ 20183)**.** Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} kx^2, & \text{nếu} \quad -1 \le x \le 1,\ 0 \le y \le x^2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

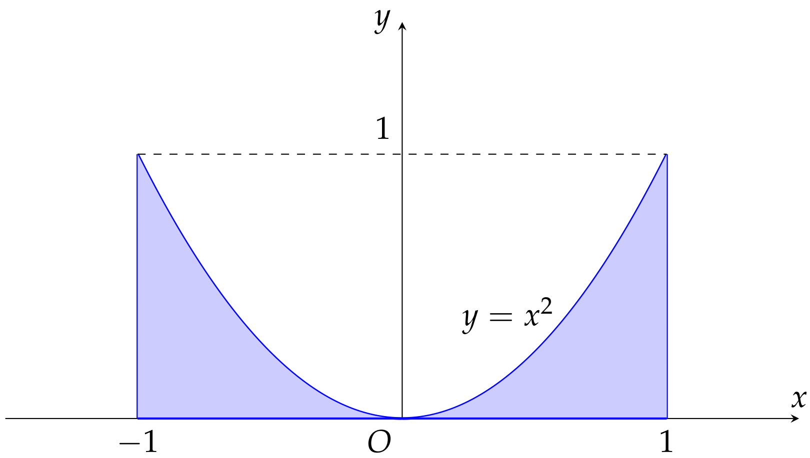
(a) Tìm k.

(b) Tính $P\left(Y \le \dfrac{1}{4}\right)$.

Lời giải Ví dụ 3.24 (a) Sử dụng Tính chất 3.3(a), $k \ge 0$. Sử dụng Tính chất 3.3(b),

$$1 = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \iint\limits_{D} kx^2 dxdy = \int\limits_{-1}^{1} \int\limits_{0}^{x^2} kx^2 dxdy = k\frac{x^5}{5}\bigg|_{-1}^{1} = \frac{2k}{5},$$

trong đó $D = \{(x,y) \in \mathbb{R}^2 : -1 \le x \le 1,\ 0 \le y \le x^2\}$ (xem Hình 3.14). Suy ra $k = 5/2$.



Hình 3.14: Miền D của Ví dụ 3.24(a)

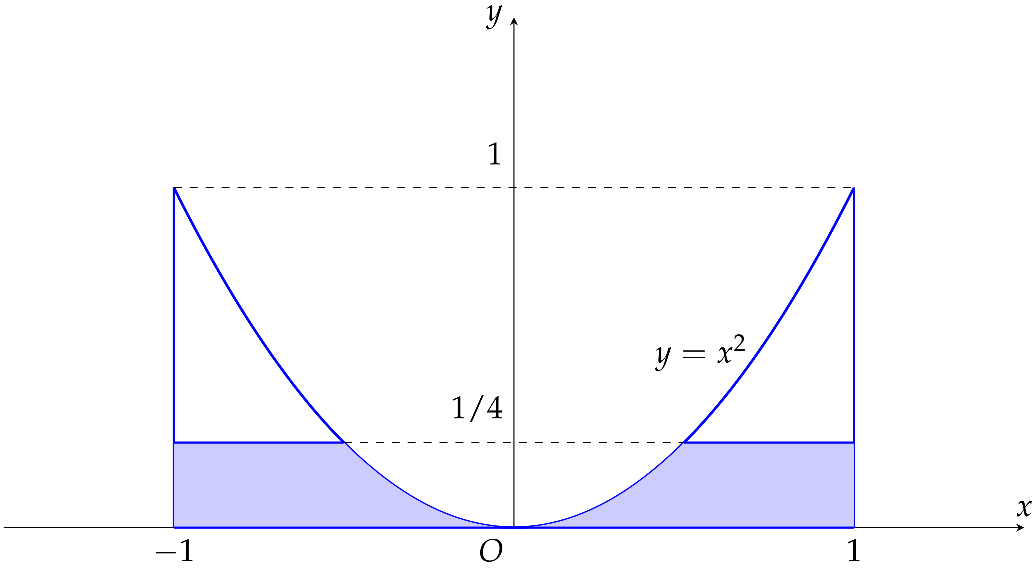
(b) Sử dụng Tính chất 3.3(c) ta tính

$$P\left(Y \le \frac{1}{4}\right) = \iint\limits_{D_1} \frac{5}{2} dxdy = 2\left[\int\limits_{0}^{1/2} dx \int\limits_{0}^{x^2} \frac{5}{2}x^2 dy + \int\limits_{1/2}^{1} dx \int\limits_{0}^{1/4} \frac{5}{2}x^2 dy\right] = \frac{19}{48} \simeq 0,3958.$$

Hoặc

$$P\left(Y \le \frac{1}{4}\right) = \iint\limits_{D_1} \frac{5}{2} dxdy = 2\int\limits_{0}^{1/4} dy \int\limits_{\sqrt{y}}^{1} \frac{5}{2}x^2 dx = \frac{19}{48} \simeq 0,3958.$$

Ở đây $D_1 = \{(x,y) \in \mathbb{R}^2 : -1 \le x \le 1,\ 0 \le y \le x^2,\ y \le 1/4\}$ (xem Hình 3.15)



Hình 3.15: Miền D_1 của Ví dụ 3.24(b)

Ví dụ 3.25 (Đề thi cuối kỳ 20191)**.** Cho U và V là hai biến ngẫu nhiên liên tục, độc lập với nhau và có cùng phân phối đều trên $[10;30]$.

(a) Tìm hàm mật độ xác suất $f_{U,V}(u,v)$ của biến ngẫu nhiên hai chiều (U,V).

(b) Tính $P(|U-V|<10)$.

Lời giải Ví dụ 3.25

(a) Vì U,V là hai biến ngẫu nhiên liên tục, có phân phối đều trên $[10;30]$ nên

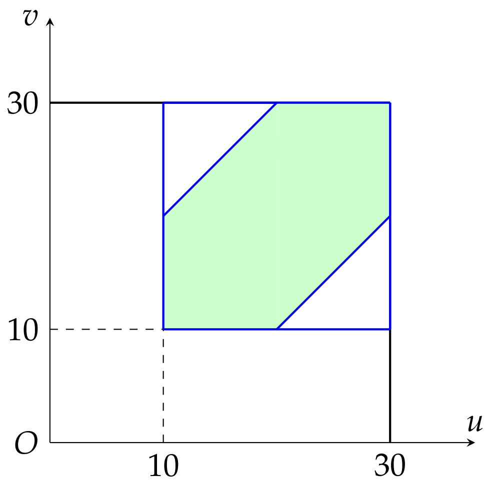
$$f_U(u)=\begin{cases}\dfrac{1}{20}, & u\in[10;30],\\ 0, & u\notin[10;30],\end{cases}\qquad f_V(v)=\begin{cases}\dfrac{1}{20}, & v\in[10;30],\\ 0, & v\notin[10;30].\end{cases}$$

Mặt khác vì U và V độc lập nên $f_{U,V}(u,v)=\begin{cases}\dfrac{1}{400}, & (u,v)\in[10;30]^2,\\ 0, & (u,v)\notin[10;30]^2.\end{cases}$

(b) $P(|U-V|<10)=\displaystyle\int\int_{D\cap S_{U,V}} f_{U,V}(u,v)dudv$ với $D=\{(u,v)\in\mathbb{R}^2:|u-v|<10\}$. Sử dụng tính chất của tích phân hai lớp suy ra

$$P(|U-V|<10)=\frac{1}{400}(20^2-10^2)=\frac{3}{4}=0,75.$$

Ví dụ 3.26 (Đề thi cuối kỳ 20192)**.** Thời gian hoạt động X_i, $i=1,2,3$, của linh kiện điện tử I, II, III là các biến ngẫu nhiên độc lập, tuân theo luật phân phối mũ với hàm mật độ xác suất tương ứng là $f_{X_i}(x)=\lambda_i e^{-\lambda_i x}$, $x>0$, $\lambda_i>0$, $i=1,2,3$. (a) Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ thời gian hoạt động của một hệ thống gồm 3 linh kiện trên mắc nối tiếp. (b) Tìm kỳ vọng và phương sai của biến ngẫu nhiên chỉ thời gian hoạt động của hệ thống đó.



Hình 3.16: Miền $D \cap S_{U,V}$ trong Ví dụ 3.25(b)

Lời giải Ví dụ 3.26 (a) Gọi X : "thời gian hoạt động của hệ thống", $X = \min\{X_1, X_2, X_3\}$.

$$F_X(x) = P(X < x) = 1 - P(\min\{X_1, X_2, X_3\} \geq x) = 1 - \Pi_{i=1}^{3} P(X_i \geq x)$$
$$= 1 - \Pi_{i=1}^{3}\Big[1 - P(X_i < x)\Big] = 1 - e^{-(\lambda_1+\lambda_2+\lambda_3)x}, \; x > 0.$$

(b) Sử dụng công thức tính kỳ vọng, phương sai của biến ngẫu nhiên phân phối mũ,

$$E(X) = \frac{1}{\lambda_1 + \lambda_2 + \lambda_3}, \quad V(X) = \frac{1}{(\lambda_1 + \lambda_2 + \lambda_3)^2}.$$

Bài tập Chương 3

Biến ngẫu nhiên rời rạc

Bài tập 3.1. Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời như sau

X \ Y	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

(a) Chứng minh rằng X và Y độc lập.

(b) Lập bảng phân phối xác suất của X và Y.

(c) Tìm quy luật phân phối của biến ngẫu nhiên $Z = XY$.

(d) Tính $E(Z)$ bằng 2 cách và kiểm tra $E(Z) = E(X).E(Y)$.

Bài tập 3.2. Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời là

X \ Y	-1	0	1
-1	$4/15$	$1/15$	$4/15$
0	$1/15$	$2/15$	$1/15$
1	0	$2/15$	0

(a) Tìm $E(X)$, $E(Y)$, $cov(X,Y)$.

(b) X và Y có độc lập không?

(c) Tìm bảng phân phối xác suất của X, của Y.

Bài tập 3.3. Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời là

X \ Y	1	2	3
1	0,17	0,13	0,25
2	0,10	0,30	0,05

(a) Lập bảng phân phối xác suất của X và của Y.

(b) Lập ma trận Covarian của (X,Y).

(c) Tìm hệ số tương quan.

(d) X và Y có độc lập không?

Bài tập 3.4. Thống kê về giá thành sản phẩm Y(triệu đồng) và sản lượng X(tấn) của một ngành sản xuất thu được bảng phân phối xác suất sau:

X \ Y	30	50	80	100
6	0,05	0,06	0,08	0,11
7	0,06	0,15	0,04	0,08
8	0,07	0,09	0,10	0,11

(a) Tìm giá thành sản phẩm trung bình và mức độ phân tán của nó.

(b) Tìm sản lượng trung bình khi giá thành bằng 8.

(c) X và Y có độc lập không?

(d) X và Y có tương quan không?

Bài tập 3.5. Cho X_1, X_2, X_3 là các biến ngẫu nhiên độc lập theo luật phân phối Poa-xông với tham số $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$. Tính xác suất của các sự kiện sau:

(a) Số lớn nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.

(b) Số lớn nhất trong các số X_1, X_2, X_3 bằng 1.

(c) Số nhỏ nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.

(d) Số nhỏ nhất trong các số X_1, X_2, X_3 bằng 1.

Bài tập 3.6. Cho X và Y là hai biến ngẫu nhiên có bảng phân phối xác suất là:

X	0	1	2	3	4	5
P	0,15	0,3	0,25	0,2	0,08	0,02

Y	0	1	2	3	4	5
P	0,3	0,2	0,2	0,15	0,1	0,05

(a) Tính $E(X), E(Y), V(X), V(Y)$.

(b) Nếu X và Y độc lập, tính $P(X + Y \leq 2)$ và lập bảng phân phối xác suất của $X + Y$.

Bài tập 3.7. Lấy ngẫu nhiên 3 viên bi từ một hộp gồm 3 bi đỏ, 5 bi xanh và 4 bi vàng. Gọi X, Y lần lượt là số bi xanh, bi vàng trong 3 bi lấy ra. Lập bảng phân phối xác suất đồng thời cho biến ngẫu nhiên hai chiều (X, Y).

Biến ngẫu nhiên liên tục

Bài tập 3.8. Cho X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} kx, & \text{nếu } 0 < y < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hằng số k

(b) X và Y có độc lập không?

Bài tập 3.9. Cho X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} k\left(x^2 + \dfrac{xy}{2}\right), & \text{nếu } 0 < x < 1,\ 0 < y < 2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hằng số k.

(b) Tìm hàm phân phối đồng thời của X và Y.

Bài tập 3.10. Cho X, Y là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{6\pi}, & \text{nếu } \dfrac{x^2}{9} + \dfrac{y^2}{4} < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hàm mật độ xác suất biên của X, của Y.

(b) Tìm xác suất để (X,Y) nằm trong hình chữ nhật $O(0,0); A(0,1); B(1,2); D(2,0)$.

Bài tập 3.11. Cho biến ngẫu nhiên hai chiều liên tục (X,Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} kx^2, & \text{nếu } \quad -1 \le x \le 1,\ 0 \le y \le x^2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm k.

(b) Tìm hàm mật độ xác suất biên $f_X(x)$, $f_Y(y)$.

(c) Tính $P\Big(Y \le \dfrac{1}{4}\Big)$.

Bài tập 3.12. Cho X và Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{x}, & \text{nếu } 0 < y < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hàm mật độ xác suất biên của X, của Y.

(b) Tìm hàm mật độ xác suất có điều kiện $f_1(x|y)$, $f_2(y|x)$.

Bài tập 3.13. Một linh kiện điện tử có thời gian hoạt động X là biến ngẫu nhiên có phân phối mũ với hàm mật độ xác suất là $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$.

(a) Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ thời gian hoạt động của một mạng gồm 2 linh kiện loại trên được mắc song song/mắc nối tiếp.

(b) Tính kỳ vọng, phương sai của thời gian hoạt động của mạng đó.

Bài tập 3.14. Cho X và Y là hai biến ngẫu nhiên độc lập với nhau có cùng phân phối đều trên $[0,2]$.

(a) Tìm hàm phân phối của các biến ngẫu nhiên $Z = X + Y$; $T = XY$; $U = X - Y$.

(b) Tính $P(-1 \le Y - X \le 1)$.

Bài tập 3.15. Hai người A và B hẹn gặp nhau tại cổng trường trong khoảng từ 7h00 đến 8h00. Gọi X và Y lần lượt là thời gian đến điểm hẹn của người A và B trong khoảng thời gian trên. Giả sử X và Y độc lập và có cùng phân phối đều trên [7;8].

(a) Tìm hàm phân phối xác suất đồng thời của X và Y.

(b) Với quy ước chỉ đợi nhau trong vòng 10 phút, tìm xác suất để 2 người được gặp nhau.

Bài tập 3.16. Cho X và Y là hai biên ngẫu nhiên độc lập, $X \sim \mathcal{N}(5;1^2)$, $Y \sim \mathcal{N}(3;(0,2)^2)$.

(a) Tìm $P(X+Y<5,5)$.

(b) Tìm $P(X<Y)$; $P(X>2Y)$.

(c) Tìm $P(X<1;Y<1)$.

Bài tập 3.17. Trọng lượng của những người chồng tuân theo luật phân phối chuẩn với kỳ vọng 70kg và độ lệch chuẩn 9kg, còn trọng lượng của những người vợ tuân theo luật phân phối chuẩn với kỳ vọng 55kg và độ lệch chuẩn 4kg. Hệ số tương quan trọng lượng giữa vợ và chồng là $2/3$. Tính xác suất vợ nặng hơn chồng.

Bài tập 3.18. Biến ngẫu nhiên liện tục X có hàm mật độ xác suất $f_X(x)$. Tìm hàm mật độ xác suất $g_Y(y)$ của biến ngẫu nhiên Y nếu:

(a) $Y = X+1, -\infty < x < +\infty$.

(b) $Y = 2X, -a < x < a$.

Bài tập 3.19. Giả sử tại một trường đại học, một sinh viên đạt được điểm X trong bài kiểm tra năng khiếu toán học và điểm Y trong bài kiểm tra năng khiếu âm nhạc là một số trong khoảng từ 0 đến 1. Giả sử X và Y được phân phối theo hàm mật độ sau

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(2x+3y) & 0<x,y<1, \\ 0 & \text{nếu ngược lại.} \end{cases}$$

(a) Tính tỷ lệ sinh viên đại học đạt điểm cao hơn 0,8 trong bài kiểm tra năng khiếu toán.

(b) Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu âm nhạc là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu toán học sẽ lớn hơn 0,8.

(c) Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu toán là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu âm nhạc sẽ lớn hơn 0,8.

Bài tập 3.20. Một mảnh đất bằng phẳng có hình tam giác vuông với một bờ phía nam dài 200m, bờ phía đông dài 100m. Ta quan tâm đến điểm mà một hạt giống rơi từ trên cao xuống tiếp đất. Giả sử rằng hạt giống nằm trong ranh giới của mảnh đất với tọa độ X và Y của nó được phân bố đều trên bề mặt của tam giác vuông.

(a) Tìm c với c là giá trị của hàm mật độ xác suất của điểm nằm trong ranh giới mảnh đất.

(b) Tìm các hàm mật độ xác suất biên của X và Y.

(c) Tìm hàm mật độ xác suất của Y biết $X = x$ và tính $P(0,1 \leq Y \leq 0,7 \mid X = 0,5)$.

Chương 4

Thống kê. Ước lượng tham số

BÀI 11 (2 tiết)

4.1 Lý thuyết mẫu

"Thống kê là một khoa học đồng thời là một công nghệ cung cấp cho ta những phương pháp, công cụ để thu thập và tạo dữ liệu, trình bày và phân tích dữ liệu để hiểu nội dung ẩn chứa trong dữ liệu. Từ đó rút ra những thông tin, tri thức hữu ích và đưa ra những quyết định, chính sách thích hợp".[1]

Thống kê toán là bộ môn toán học nghiên cứu quy luật của các hiện tượng ngẫu nhiên có tính chất số lớn trên cơ sở thu thập và xử lý số liệu thống kê các kết quả quan sát về những hiện tượng ngẫu nhiên này. Nếu ta thu thập được các số liệu liên quan đến tất cả đối tượng cần nghiên cứu thì ta có thể biết được đối tượng này (phương pháp toàn bộ). Tuy nhiên trong thực tế điều đó không thể thực hiện được vì quy mô của các đối tượng cần nghiên cứu quá lớn hoặc trong quá trình nghiên cứu đối tượng nghiên cứu bị phá hủy. Vì vậy cần lấy mẫu để nghiên cứu.

Mục này giới thiệu về phương pháp lấy mẫu ngẫu nhiên và các thống kê thường gặp của mẫu ngẫu nhiên.

4.1.1 Tổng thể và mẫu

Khái niệm tổng thể

Khi nghiên cứu các vấn đề về kinh tế - xã hội, cũng như nhiều vấn đề thuộc các lĩnh vực vật lý, sinh vật, quân sự ... thường dẫn đến khảo sát một hay nhiều dấu hiệu (định tính hoặc định lượng) thể hiện bằng số lượng trên nhiều phần tử. Tập hợp tất cả các phần tử này gọi là tổng thể hay đám đông (population). Số phần tử trong tổng thể có thể là hữu hạn hoặc vô hạn. Cần

[1] Đặng Hùng Thắng, Trần Mạnh Cường (2019), *Thống kê cho Khoa học xã hội và Khoa học sự sống*, NXB Đại học Quốc gia Hà Nội.

nhấn mạnh rằng ta không nghiên cứu trực tiếp bản thân tổng thể mà chỉ nghiên cứu dấu hiệu nào đó của nó.

Ký hiệu N là số phần tử của tổng thể; $\mathcal{X}$ là dấu hiệu cần khảo sát.

Ví dụ 4.1. (a) Muốn điều tra thu nhập bình quân của các hộ gia đình ở Hà Nội thì tập hợp cần nghiên cứu là các hộ gia đình ở Hà Nội, dấu hiệu nghiên cứu là thu nhập của từng hộ gia đình (dấu hiệu định lượng).

(b) Một doanh nghiệp muốn nghiên cứu các khách hàng của mình về dấu hiệu định tính có thể là mức độ hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, còn dấu hiệu định lượng là số lượng sản phẩm của doanh nghiệp mà khách hàng có nhu cầu được đáp ứng.

Một số lý do không thể khảo sát toàn bộ tổng thể

(a) Do quy mô của tập hợp cần nghiên cứu quá lớn nên việc nghiên cứu toàn bộ sẽ đòi hỏi nhiều chi phí về vật chất và thời gian, có thể không kiểm soát được dẫn đến bị chồng chéo hoặc bỏ sót.

(b) Trong nhiều trường hợp không thể nắm được toàn bộ các phần tử của tập hợp cần nghiên cứu, do đó không thể tiến hành toàn bộ được.

(c) Có thể trong quá trình điều tra sẽ phá hủy đối tượng nghiên cứu. . .

Do đó thay vì khảo sát tổng thể, ta chỉ cần chọn ra một tập nhỏ để khảo sát và đưa ra quyết định.

Khái niệm tập mẫu

Thông thường quy mô của một tổng thể là rất lớn. Vì thế người ta thường chọn ra một tập hợp con các cá thể để nghiên cứu. Việc chọn ra từ tổng thể một tập hợp con nào đó được gọi là phép lấy mẫu. Tập hợp con được chọn được gọi là mẫu (sample). Số cá thể trong mẫu được gọi là kích thước mẫu, ký hiệu là n.

Ví dụ 4.2. Ta muốn đánh giá số giờ trong một ngày mà một sinh viên đại học sử dụng Facebook. Vì số sinh viên đại học rất lớn, nên ta không thể điều tra trên tất cả các sinh viên được. Ta chọn ngẫu nhiên một mẫu gồm 50 sinh viên để khảo sát và tìm được số giờ trung bình dùng Facebook của 50 sinh viên này là 4,7 giờ. Con số này cho ta một hình ảnh về việc sử dụng Facebook của các sinh viên đại học.

Ví dụ 4.3. Ta muốn đánh giá tỷ lệ phế phẩm trong các sản phẩm của nhà máy A. Giả sử nhà máy chế tạo được 400000 sản phẩm. Ta không đủ thời gian và tiền bạc để xem xét được toàn bộ 400000 sản phẩm. Ta chọn ra một mẫu gồm 300 sản phẩm để kiểm tra và phát hiện ra có 18 sản phẩm mắc lỗi. Tỷ lệ phế phẩm trong mẫu kiểm tra là $18/300 = 6\%$. Từ đó ta nhận định tỷ lệ phế phẩm của nhà máy A khoảng 6%.

Chương 4 và Chương 5 sẽ nghiên cứu tổng thể thông qua mẫu. Nói nghiên cứu tổng thể có nghĩa là nghiên cứu một hoặc một số đặc trưng nào đó của tổng thể. Khi đó, ta không thể đem tất cả các phần tử trong tổng thể ra nghiên cứu mà chỉ lấy một số phần tử trong tổng thể ra nghiên cứu và làm sao qua việc nghiên cứu này có thể kết luận được về một hoặc một số đặc trưng của tổng thể mà ta quan tâm ban đầu.

Một số cách chọn mẫu cơ bản

Các kết luận suy diễn từ mẫu có đáng tin cậy không? Câu nói nổi tiếng của Mark Twain, nhà văn Anh "Có ba kiểu nói dối: Nói dối, nói dối trắng trợn và thống kê" ("Thera are three kinds of lies: Lies, Damned lies and Statistics"). Tuy nhiên, thống kê không nói dối. Kết quả sai (mà ta gọi là dối trá) do thống kê đưa ra là do phương pháp lấy mẫu không đúng:

1. Việc lấy mẫu đã được tiến hành không khách quan, theo hướng có lợi cho người nghiên cứu.
2. Mẫu được chọn không đại diện.

Ví dụ 4.4. Để điều tra mức thu nhập trung bình của sinh viên tốt nghiệp đại học mới ra trường, nếu mẫu được chọn trong số các sinh viên tốt nghiệp ngành Công nghệ thông tin thì rõ ràng mức lương trung bình trong mẫu không phản ánh trung thực mức lương trung bình của sinh viên mới ra trường nói chung.

Các kết luận suy diễn từ mẫu có đáng tin cậy chỉ đạt được nếu mẫu được chọn phản ánh trung thực, thực sự đại diện cho tổng thể. Do đó vấn đề chọn mẫu là một vấn đề rất quan trọng và phong phú của thống kê. Các kỹ thuật chọn mẫu đúng đắn sẽ giúp ta đảm bảo được tính đại diện trung thực cho tổng thể. Để trả lời cho câu hỏi đặt ra là làm sao chọn được tập mẫu có tính chất tương tự như tổng thể để các kết luận của tập mẫu có thể dùng cho tổng thể, ta sử dụng một trong những cách chọn mẫu sau:

(a) Lấy mẫu ngẫu nhiên: mỗi cá thể của tổng thể được chọn một cách độc lập với xác suất như nhau.

(b) Lấy mẫu theo khối: Tổng thể được chia làm N khối, mỗi khối xem là một tổng thể con. Chọn ngẫu nhiên ra m khối trong N khối đó. Tập hợp tất cả các cá thể của m khối được chọn sẽ được lập thành một mẫu để khảo sát.

Phương pháp này được áp dụng khi ta không liệt kê danh sách tất cả các cá thể trong tổng thể.

(c) Lấy mẫu phân tầng: Chia tổng thể ra một số tầng, sao cho các cá thể trong mỗi tầng khác nhau càng ít càng tốt. Mỗi tầng được coi là một tổng thể con. Trong mỗi tầng ta sẽ thực hiện việc lấy mẫu ngẫu nhiên.

Phương pháp này được sử dụng khi các cá thể quá khác nhau về vấn đề mà nhà nghiên cứu đang quan tâm khảo sát.

Ví dụ 4.5. Tại một trường Đại học có 20000 sinh viên với 5 hệ đào tạo khác nhau: 10000 sinh viên hệ chính quy; 2000 sinh viên hệ liên thông; 2000 sinh viên hệ văn bằng hai; 5000 sinh viên hệ tại chức và 1000 học viên hệ sau đại học. Bộ phận đảm bảo chất lượng tiến hành cuộc khảo sát về chất lượng và mức độ hài lòng của người học. Chọn ngẫu nhiên 1000 sinh viên để khảo sát. Xem mỗi hệ đào tạo là một tầng, số sinh viên ở mỗi tầng được chọn như sau:

Hệ đào tạo	**Số SV**	**% SV**	**Số SV được chọn**
Chính quy	10000	50	500
Liên thông	2000	10	200
Văn bằng hai	2000	10	200
Tại chức	2000	10	200
Sau đại học	1000	5	50
Tổng	**20000**	**100**	**1000**

4.1.2 Mẫu ngẫu nhiên

Biến ngẫu nhiên và quy luật phân phối gốc

Giả sử ta cần nghiên cứu dấu hiệu $\mathcal{X}$ của tổng thể có $E(\mathcal{X}) = \mu$ và $V(\mathcal{X}) = \sigma^2$ (μ và σ chưa biết). Ta có thể mô hình hóa dấu hiệu $\mathcal{X}$ bằng một biến ngẫu nhiên. Thật vậy, nếu lấy ngẫu nhiên từ tổng thể ra một phần tử và gọi X là giá trị của dấu hiệu $\mathcal{X}$ đo được trên phần tử lấy ra thì X là biến ngẫu nhiên có bảng phân phối xác suất là

X	x_1	x_2	$\ldots$	x_n
P	$P(X = x_1)$	$P(X = x_2)$	$\ldots$	$P(X = x_n)$

Như vậy dấu hiệu $\mathcal{X}$ mà ta nghiên cứu được mô hình hóa bởi biến ngẫu nhiên X, còn cơ cấu của tổng thể theo dấu hiệu $\mathcal{X}$ (tập hợp các xác suất) chính là quy luật phân phối xác suất của X.

Biến ngẫu nhiên X được gọi là biến ngẫu nhiên gốc. Quy luật phân phối xác suất của X là quy luật phân phối gốc, đồng thời $E(X) = \mu$, $V(X) = \sigma^2$.

Các đặc trưng của tổng thể

(a) Xét tổng thể về mặt định lượng : tổng thể được đặc trưng bởi dấu hiệu $\mathcal{X}$ được mô hình hóa bởi biến ngẫu nhiên X. Ta có các tham số đặc trưng sau đây:

1. Trung bình tổng thể: $E(X) = \mu$.
2. Phương sai tổng thể: $V(X) = \sigma^2$.
3. Độ lệch chuẩn của tổng thể: $\sigma(X) = \sigma$.

(b) Xét tổng thể về mặt định tính : tổng thể có kích thước N, trong đó có M phần tử có tính chất A. Khi đó $p = \dfrac{M}{N}$ gọi là tỷ lệ tính chất A của tổng thể.

Khái niệm mẫu ngẫu nhiên

Giả sử tiến hành n phép thử độc lập. Gọi X_i là "giá trị của dấu hiệu $\mathcal{X}$ đo lường được trên phần tử thứ i của mẫu" $i = 1, 2, \ldots, n$. Khi đó, $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên độc lập có cùng quy luật phân phối xác suất với X.

Định nghĩa 4.1 (Mẫu ngẫu nhiên)**.** Cho biến ngẫu nhiên gốc X có quy luật phân phối xác suất $F_X(x)$ nào đó. Một mẫu ngẫu nhiên kích thước n được thành lập từ biến ngẫu nhiên X là n biến ngẫu nhiên độc lập có cùng quy luật phân phối xác suất $F_X(x)$ với biến ngẫu nhiên X.

Ký hiệu mẫu ngẫu nhiên: $W_X = (X_1, X_2, \ldots, X_n)$.

Thực hiện một phép thử đối với mẫu ngẫu nhiên W_X tức là thực hiện một phép thử đối với mỗi thành phần X_i của mẫu. Giả sử X_1 nhận giá trị x_1, X_2 nhận giá trị $x_2, \ldots$, X_n nhận giá trị x_n ta thu được một mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$.

Ví dụ 4.6. Gọi X là "số chấm xuất hiện khi gieo một con xúc xắc". X là biến ngẫu nhiên có bảng phân phối xác suất

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

Nếu gieo con xúc xắc 3 lần và gọi X_i là "số chấm xuất hiện ở lần gieo thứ i", $i = 1, 2, 3$ thì ta có 3 biến ngẫu nhiên độc lập có cùng quy luật phân phối xác suất với X. Vậy ta có một mẫu ngẫu nhiên $W_X = (X_1, X_2, X_3)$ cỡ $n = 3$ được xây dựng từ biến ngẫu nhiên gốc X. Thực hiện một phép thử đối với mẫu ngẫu nhiên này (tức là gieo 3 lần một con xúc xắc). Giả sử lần thứ nhất xuất hiện mặt 6, lần thứ hai xuất hiện mặt 2, lần thứ ba xuất hiện mặt 1 thì ta có một giá trị của mẫu ngẫu nhiên $W_x = (6, 3, 1)$.

Mẫu ngẫu nhiên có thể phản ánh được kết quả điều tra, thực nghiệm bởi vì những kết quả này được coi là một giá trị của nó. Mặt khác mẫu ngẫu nhiên là tập hợp các biến ngẫu nhiên. Do vậy ta có thể nghiên cứu quy luật phân phối xác suất của nó, tức là khái quát được thực nghiệm. Quan hệ giữa mẫu ngẫu nhiên và mẫu cụ thể (hay một giá trị của nó) tương tự quan hệ giữa biến ngẫu nhiên và một giá trị có thể có của nó.

4.1.3 Mô tả giá trị của mẫu ngẫu nhiên

Phân loại dữ liệu

Từ tổng thể ta trích ra tập mẫu có n phần tử. Ta có n số liệu.

(a) Dạng liệt kê: Các số liệu thu được được ghi lại thành dãy $x_1, x_2, \ldots, x_n$.

(b) Dạng rút gọn: Số liệu thu được có sự lặp đi lặp lại một số giá trị thì ta có dạng rút gọn sau:

(b1) Dạng tần số: $(n_1 + n_2 + \ldots + n_k = n)$

Giá trị	x_1	x_2	$\ldots$	x_k
Tần số	n_1	n_2	$\ldots$	n_k

(b2) Dạng tần suất: $(f_k = n_k/n)$

Giá trị	x_1	x_2	$\ldots$	x_k
Tần suất	f_1	f_2	$\ldots$	f_k

(c) Dạng khoảng: Dữ liệu thu được nhận giá trị trong (a, b). Ta chia (a, b) thành k miền con bởi các điểm chia: $a_0 = a < a_1 < a_2 < \cdots < a_{k-1} < a_k = b$.

(c1) Dạng tần số: $(n_1 + n_2 + \ldots + n_k = n)$

Giá trị	$(a_0 - a_1]$	$(a_1 - a_2]$	$\ldots$	$(a_{k-1} - a_k]$
Tần số	n_1	n_2	$\ldots$	n_k

(c2) Dạng tần suất: $(f_k = n_k/n)$

Giá trị	$(a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$
Tần suất	f_1	f_2	$\ldots$	f_k

Chú ý, thông thường, độ dài các khoảng chia bằng nhau. Khi đó ta có thể chuyển về dạng rút gọn:

Giá trị	x_1	x_2	$\ldots$	x_k
Tần số	n_1	n_2	$\ldots$	n_k

trong đó x_i là điểm đại diện cho $(a_{i-1}, a_i]$ thường được xác định là trung điểm của đoạn đó: $x_i = \dfrac{1}{2}(a_{i-1} + a_i)$.

Đặt w_i là tần số tích lũy của x_i và $F_n(x_i)$ là tần suất tích lũy của x_i, ta sẽ có

$$w_i = \sum_{x_j < x_i} n_j; \quad F_n(x_i) = \frac{w_i}{n} = \sum_{x_j < x_i} f_j$$

thì $F_n(x_i)$ là một hàm của x_i và được gọi là hàm phân phối thực nghiệm của mẫu hay hàm phân phối mẫu. Chú ý rằng theo luật số lớn (Định lý Béc-nu-li) $F_n(x)$ hội tụ theo xác suất về $F_X(x) = P(X < x)$, trong đó X là biến ngẫu nhiên gốc cảm sinh ra tổng thể (và cả tập mẫu). Như vậy hàm phân phối mẫu có thể dùng để xấp xỉ luật phân phối của tổng thể.

Biểu diễn dữ liệu

Một câu ngạn ngữ Trung Hoa "Một hình ảnh có tác dụng bằng một nghìn lời nói". Để có được một hình ảnh rõ ràng và dễ nhớ về mẫu các giá trị của biến ngẫu nhiên X, ta dùng các đồ thị và các biểu đồ để thể hiện chúng.

(a) Biểu đồ hình cột (bar chart): là biểu đồ nhằm biểu diễn cho dữ liệu được phân nhóm (thường dùng cho dữ liệu định tính) như các tháng trong năm, các nhóm tuổi... Các nhóm được biểu diễn thường xuất hiện theo trục hoành, trục tung là chiều cao của các hình chữ nhật tỷ lệ với giá trị được biểu diễn. Mục tiêu của việc dùng biểu đồ hình cột là đưa ra so sánh giữa các nhóm.

(b) Biểu đồ hình quạt (pie chart): cũng được dùng để biểu diễn dữ liệu được phân nhóm, nhưng các nhóm được biểu diễn bằng các hình quạt trong hình tròn. Số lượng hoặc tỷ lệ của mỗi hạng mục (mỗi nhóm) tỷ lệ với diện tích hình quạt biểu diễn nó. Biểu đồ này thường dùng để phân tích hoặc so sánh ở mức độ tổng thể.

(c) Tổ chức đồ (histogram): thường được dùng để biểu thị tần số hay tần suất các giá trị trong mỗi khoảng giá trị.

1. Nếu độ rộng các khoảng bằng nhau, thì chiều cao của hình chữ nhật dựng trên mỗi khoảng chính là tần số hay tần suất tương ứng của khoảng.
2. Nếu độ rộng các khoảng không bằng nhau, chiều cao của hình chữ nhật dựng trên mỗi khoảng được tính toán sao cho diện tích mỗi hình chữ nhật tỷ lệ với tần số hoặc tần suất của khoảng đó.

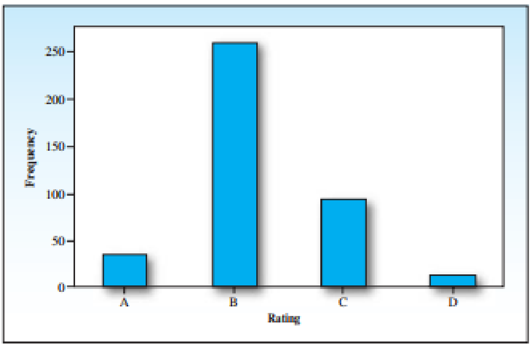
(d) Đa giác tần số, tần suất: dùng khi dữ liệu là liên tục và khoảng dữ liệu rất rộng. Tại mỗi giá trị của dữ liệu x_i và tần số n_i ta chấm một điểm có tọa độ (x_i, n_i). Nối các điểm này với nhau ta được đa giác tần số. Nếu muốn có đa giác tần suất ta thay n_i bằng $f_i = n_i/n$.

Ví dụ 4.7. Khảo sát 400 nhà quản lý giáo dục về đánh giá chất lượng giáo dục công ở Hoa Kỳ, ta nhận được bảng dữ liệu

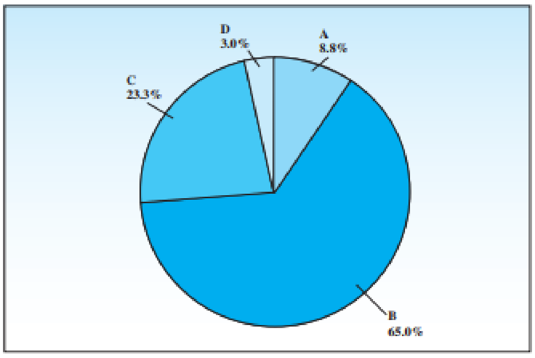
Xếp hạng	A	B	C	D
Tần số	35	260	93	12

- Tổng số nhà quản lý giáo dục được khảo sát $n = 400$.

- 35 người xếp hạng A chiếm 9%; 260 người xếp hạng B chiếm 65%; 93 người xếp hạng C chiếm 23%; 12 người xếp loại C chiếm 3%.
- Biểu đồ hình cột cho tập dữ liệu này biểu diễn ở Hình 4.1
- Biểu đồ hình quạt cho tập dữ liệu này biểu diễn ở Hình 4.2



Hình 4.1: Biểu đồ hình cột cho dữ liệu trong Ví dụ 4.7



Hình 4.2: Biểu đồ hình quạt cho dữ liệu trong Ví dụ 4.7

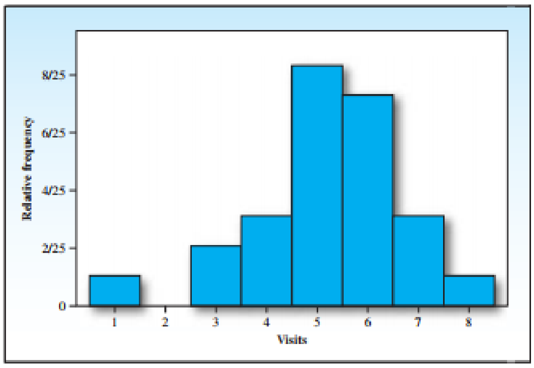
Ví dụ 4.8 (Ví dụ về tổ chức đồ)**.** 25 khách hàng của Starbucks được thăm dò ý kiến trong một cuộc khảo sát tiếp thị "Trong một tuần bạn đến Starbucks bao nhiêu lần?". Số liệu được cho trong bảng sau:

$$\begin{array}{ccccccccccccc} 6 & 7 & 1 & 5 & 6 & 4 & 6 & 4 & 6 & 8 & 6 & 5 & \\ 6 & 3 & 4 & 5 & 5 & 5 & 7 & 6 & 3 & 5 & 7 & 5 & 5 \end{array}$$

Biến được đo lường là "số lần đến Starbucks", một biến rời rạc chỉ nhận các giá trị nguyên. Trong trường hợp này, cách đơn giản nhất là chọn các lớp hoặc khoảng con dưới dạng giá trị nguyên trên phạm vi giá trị quan sát: 1, 2, 3, 4, 5, 6, 7 và 8. Bảng dưới đây cho thấy các lớp và tần số tương ứng của chúng cùng tần số.

Số lượt đến Starbucks	Tần số	Tần suất
1	1	0,04
2	0	0,00
3	2	0,08
4	3	0,12
5	8	0,32
6	7	0,28
7	3	0,12
8	1	0,04

Biểu đồ tần suất tương đối được thể hiện trong Hình 4.3.



Hình 4.3: Biểu đồ tổ chức đồ cho dữ liệu trong Ví dụ 4.8

4.1.4 Đại lượng thống kê và một số thống kê thông dụng

Để nghiên cứu mẫu ngẫu nhiên gốc X, nếu dừng lại ở mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ thì rõ ràng chưa giải quyết được vấn đề gì, bởi các biến ngẫu nhiên X_i có cùng quy luật phân phối xác suất với X mà ta chưa biết hoàn toàn. Vì vậy ta phải liên kết hay tổng hợp các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ lại sao cho biến ngẫu nhiên mới thu được có những tính chất mới, có thể đáp ứng được yêu cầu giải những bài toán khác nhau về biến ngẫu nhiên gốc X.

Đại lượng thống kê

Định nghĩa 4.2 (Thống kê)**.** Trong thống kê toán việc tổng hợp mẫu $W_X = (X_1, X_2, \ldots, X_n)$ được thực hiện dưới dạng hàm của các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$. Ký hiệu

$$\boxed{G = f(X_1, X_2, \ldots, X_n)} \tag{4.1}$$

ở đây f là một hàm nào đó và G được gọi là một thống kê.

Ví dụ 4.9. Cho một mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n. Một ví dụ về thống kê dạng (4.1) là

$$G = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Nhận xét 4.1. **(a)** Thống kê G là một hàm của các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ nên cũng là một biến ngẫu nhiên. Do đó ta có thể tìm ra các "tính chất mới" thông qua việc khảo sát quy luật phân phối xác suất của G và các tham số $E(G)$, $V(G)$...

(b) Nếu mẫu ngẫu nhiên có giá trị $W_x = (x_1, x_2, \ldots, x_2)$ (mẫu cụ thể), ta tính được giá trị cụ thể của G, ký hiệu là $g = f(x_1, x_2, \ldots, x_n)$ hay g_{qs} còn gọi là giá trị quan sát của thống kê G.

Sau đây ta xét một số thống kê thông dụng.

Một số thống kê thông dụng

(a) Trung bình mẫu ngẫu nhiên: Cho mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n được xây dựng từ biến ngẫu nhiên gốc X. Trung bình của nó là một thống kê, ký hiệu là $\overline{X}$ và được định nghĩa bởi hàm sau đây:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i} \tag{4.2}$$

Do $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên nên $\overline{X}$ cũng là biến ngẫu nhiên. Nếu biến ngẫu nhiên gốc X có kỳ vọng $E(X) = \mu$, phương sai $V(X) = \sigma^2$ thì thống kê $\overline{X}$ có kỳ vọng $E(\overline{X}) = \mu$ và phương sai $V(\overline{X}) = \dfrac{\sigma^2}{n}$ nhỏ hơn phương sai của biến ngẫu nhiên gốc n lần, nghĩa là các giá trị có thể có của $\overline{X}$ ổn định quanh kỳ vọng μ hơn các giá trị có thể có của X. Điều này thể hiện "chất lượng mới" của thống kê $\overline{X}$ so với biến ngẫu nhiên gốc X.

(b) Phương sai mẫu ngẫu nhiên và phương sai hiệu chỉnh mẫu ngẫu nhiên: Cho mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n được xây dựng từ biến ngẫu nhiên gốc X. Phương sai của nó là một thống kê, ký hiệu là $\hat{S}^2$ và được xác định bởi hàm

$$\boxed{\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{4.3}$$

trong đó $\overline{X}$ là trung bình của mẫu ngẫu nhiên W_X.

Do $\hat{S}^2$ là biến ngẫu nhiên, nên có thể tính được $E(\hat{S}^2)$ bởi công thức

$$E(\hat{S}^2) = \frac{n-1}{n}\sigma^2, \tag{4.4}$$

trong đó $\sigma^2 = V(X)$.

Để kỳ vọng của phương sai mẫu ngẫu nhiên $\hat{S}^2$ trùng với phương sai của biến ngẫu nhiên gốc X ta cần một sự hiệu chỉnh. Từ (4.4) suy ra

$$E\left(\frac{n}{n-1}\hat{S}^2\right) = \sigma^2.$$

Đặt

$$\boxed{S^2 = \frac{n}{n-1}\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \tag{4.5}$$

và gọi S^2 là phương sai hiệu chỉnh của mẫu ngẫu nhiên (vì nó bằng phương sai mẫu ngẫu nhiên nhân thêm hệ số $\dfrac{n}{n-1}$). Đồng thời ta có $E(S^2) = \sigma^2$.

(c) Độ lệch tiêu chuẩn và độ lệch tiêu chuẩn hiệu chỉnh mẫu ngẫu nhiên: 1. Độ lệch tiêu chuẩn của mẫu ngẫu nhiên được ký hiệu và xác định bởi

$$\boxed{\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}} \tag{4.6}$$

2. Độ lệch tiêu chuẩn hiệu chỉnh của mẫu ngẫu nhiên được ký hiệu và xác định bởi

$$\boxed{S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}} \tag{4.7}$$

(d) Tần suất mẫu ngẫu nhiên: Trường hợp cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không, giả sử p là tần suất có dấu hiệu A của tổng thể. Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân phối Béc-nu-li tham số p có kỳ vọng $E(X) = p$ và phương sai $V(X) = p(1-p)$.

Lấy mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ trong đó X_1, X_2, $\ldots X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối Béc-nu-li với tham số p. Tần số xuất hiện A trong mẫu là $m = \sum_{i=1}^{n} X_i$. Khi đó tần xuất mẫu là một thống kê ký hiệu và xác định bởi

$$\boxed{f = \frac{m}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}} \tag{4.8}$$

Như vậy tần suất mẫu là trung bình mẫu của biến ngẫu nhiên X có phân phối Béc-nu-li $\mathcal{B}(p)$ tham số p. Ngoài ra,

$$\boxed{E(f) = p, \quad V(f) = \frac{p(1-p)}{n}} \tag{4.9}$$

Dưới đây là một số quy luật phân phối xác suất của một số thống kê thông dụng.

Quy luật phân phối xác suất của một số thống kê thông dụng

Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như một biến ngẫu nhiên X có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$. Các tham số này có thể đã biết hoặc chưa biết. Từ tổng thể rút ra một mẫu ngẫu nhiên cỡ n: $W_X = (X_1, X_2, \ldots, X_n)$. Các biến ngẫu nhiên thành phần X_i, $i = 1, \ldots, n$, độc lập có cùng quy luật phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ như X.

Chú ý rằng mọi tổ hợp tuyến tính của các biến ngẫu nhiên có phân phối chuẩn là biến ngẫu nhiên có phân phối chuẩn. Vì vậy ta có các kết quả sau.

(a) Quy luật phân phối xác suất của trung bình mẫu $\overline{X}$: Thống kê trung bình mẫu $\overline{X}$ có phân phối chuẩn và thống kê $U = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$ có phân phối chuẩn tắc

$$\boxed{\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad U = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)} \tag{4.10}$$

Phân vị mức α của U được ký hiệu là $u_{1-\alpha}$. Chẳng hạn phân vị mức $\alpha = 5\%$ của U là $u_{0,95} = 1,645$ vì $\Phi(1,645) = 0,95$; phân vị mức $\alpha = 2,5\%$ của U là $u_{0,975} = 1,96$ vì $\Phi(1,96) = 0,975$.

(b) Quy luật phân phối xác suất của phương sai hiệu chỉnh mẫu ngẫu nhiên: 1. Thống kê $S = \dfrac{n\hat{S}^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2}$ có phân phối khi bình phương với $n-1$ bậc tự do

$$\boxed{S = \frac{n\hat{S}^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{(n-1)}} \tag{4.11}$$

2. Nếu trong (4.11) ta thay $\overline{X}$ bằng μ thì thống kê

$$\boxed{S = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)} \tag{4.12}$$

Phân vị mức α của S ký hiệu là $\chi^2_{1-\alpha}(n)$.

3. Thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n} = \dfrac{\overline{X} - \mu}{\hat{S}}\sqrt{n-1}$ có phân phối Student với $n-1$ bậc tự do:

$$\boxed{T = \frac{\overline{X} - \mu}{S}\sqrt{n} = \frac{\overline{X} - \mu}{\hat{S}}\sqrt{n-1} \sim t(n-1)} \tag{4.13}$$

Phân vị mức α của T ký hiệu là $t_{1-\alpha}(n)$.

Chú ý 4.1. Trong thực hành khi $n \geq 30$ ta có thể không cần đến giả thiết chuẩn của biến ngẫu nhiên gốc, thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$ xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0,1)$.

(c) Phân phối của thống kê tần suất mẫu: Khi n đủ lớn ($np \geq 5$ và $n(1-p) \geq 5$) thì thống kê $U = \dfrac{f-p}{\sqrt{p(1-p)}}\sqrt{n}$ có phân phối xấp xỉ phân phối chuẩn tắc

$$\boxed{U = \frac{f-p}{\sqrt{p(1-p)}}\sqrt{n} \sim \mathcal{N}(0,1)} \tag{4.14}$$

(d) Phân phối Fisher: Nếu hai mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_{n_1}), W_Y = (Y_1, Y_2, \ldots, Y_{n_2})$ kích thước n_1, n_2 được xây dựng từ biến ngẫu nhiên X và Y có phân phối chuẩn $\mathcal{N}(\mu_1; \sigma_1^2)$, $\mathcal{N}(\mu_2; \sigma_2^2)$ tương ứng thì người ta đã chứng minh được rằng:

1. Nếu $\sigma_1^2 = \sigma_2^2$ thì thống kê $F = \dfrac{S_1^2}{S_2^2}$ có phân phối Fisher với $(n_1 - 1, n_2 - 1)$ bậc tự do:

$$\boxed{F = \frac{S_1^2}{S_2^2} \sim \mathcal{F}(n_1 - 1; n_2 - 1)} \tag{4.15}$$

2. Nếu $\sigma_1^2 \neq \sigma_2^2$,

$$\boxed{F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim \mathcal{F}(n_1 - 1; n_2 - 1)} \tag{4.16}$$

Phân vị mức α của F được ký hiệu là $\mathcal{F}_{1-\alpha}(n_1 - 1; n_2 - 1)$.

4.1.5 Cách tính giá trị cụ thể của một số thống kê thông dụng

Bên cạnh việc nghiên cứu quy luật phân phối xác suất của các thống kê, còn cần phải tính toán các giá trị của chúng. Giả sử mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ có một giá trị là $W_x = (x_1, x_2, \ldots, x_n)$. Mẫu cụ thể này có thể cho ở các dạng khác nhau.

(a) Mẫu cho dưới dạng liệt kê. (Tần số của các x_i bằng 1)

(a1) Trung bình mẫu:

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i} \tag{4.17}$$

(a2) Phương sai mẫu:

$$\boxed{\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \tag{4.18}$$

(a3) Phương sai hiệu chỉnh mẫu:

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{n}{n-1}\hat{s}^2} \tag{4.19}$$

(a4) Các độ lệch chuẩn:

$$\boxed{\hat{s} = \sqrt{\hat{s}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}; \quad s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}} \tag{4.20}$$

Để tính các công thức (4.17)–(4.20), ta lập bảng tính toán

x_i	x_i^2
x_1	x_1^2
x_2	x_2^2
$\dots$	$\dots$
x_n	x_n^2
$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} x_i^2$

Ta thấy trung bình mẫu $\overline{x}$ và phương sai mẫu $\hat{s}^2$ là hai giá trị cơ bản nhất đối với mẫu cụ thể này, còn các giá trị s^2, $\hat{s}$, s có thể tính trực tiếp từ $\hat{s}^2$; giá trị của nhiều thống kê khác cũng được tính trên cơ sở đã có $\overline{x}$ và $\hat{s}^2$. Do đó cần cải tiến các công thức tính $\overline{x}$ và $\hat{s}^2$ phù hợp với từng trường hợp số liệu.

(b) Mẫu cho ở dạng rút gọn. (Tần số của các x_i là $n_i > 1$, $\sum_{i=1}^{k} n_i = n$)

(b1) Trung bình mẫu:

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i} \tag{4.21}$$

(b2) Phương sai mẫu:

$$\boxed{\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i\right)^2} \tag{4.22}$$

(b3) Phương sai hiệu chỉnh mẫu:

$$\boxed{s^2 = \frac{n}{n-1}\hat{s}^2} \tag{4.23}$$

(b4) Các độ lệch chuẩn:

$$\boxed{\hat{s} = \sqrt{\hat{s}^2}; \quad s = \sqrt{s^2}} \tag{4.24}$$

Để tính các công thức (4.21)–(4.24), ta lập bảng tính toán

x_i	n_i	$n_i x_i$	$n_i x_i^2$
x_1	n_1	$n_1 x_1$	$n_1 x_1^2$
x_2	n_2	$n_2 x_2$	$n_2 x_2^2$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_k	n_k	$n_k x_k$	$n_k x_k^2$
	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k n_i x_i$	$\sum_{i=1}^k n_i x_i^2$

(c) Phương pháp đổi biến. (Trong trường hợp độ dài các khoảng bằng nhau)

(c1) Trung bình mẫu:

$$\boxed{\overline{x} = x_0 + h\overline{u} = x_0 + \frac{h}{n}\sum_{i=1}^{k} n_i u_i} \tag{4.25}$$

(c2) Phương sai mẫu:

$$\boxed{\hat{s}^2 = h^2\left[\frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{k} n_i u_i\right)^2\right] = h^2\hat{s}_u^2} \tag{4.26}$$

trong đó

x_i là điểm giữa của khoảng thứ i, $i = 1, 2, \ldots, k$;
$u_i = \dfrac{x_i - x_0}{h}$, h là độ dài các khoảng;
$x_0 = x_i$ ứng với n_i lớn nhất.

Để tính các công thức (4.25)–(4.26), ta lập bảng tính toán

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$
x_1	n_1	u_1	$n_1 u_1$	$n_1 u_1^2$
x_2	n_2	u_2	$n_2 u_2$	$n_2 u_2^2$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_k	n_k	u_k	$n_k u_k$	$n_k u_k^2$
	$\sum_{i=1}^k n_i = n$		$\sum_{i=1}^k n_i u_i$	$\sum_{i=1}^k n_i u_i^2$

Tính tham số đặc trưng mẫu trên máy tính CASIO FX570VN PLUS

Bước 1 Chuyển đổi máy tính về chương trình thống kê **MODE** $\to$ **3** $\to$ **AC**

Bước 2 Bật chức năng cột tần số/tần suất **SHIFT** $\to$ **MODE** $\to$ **Mũi tên đi xuống** $\to$ **4(STAT)** $\to$ **1(ON)**

Bước 3 Bật chế độ màn hình để nhập dữ liệu, Nhập số liệu **SHIFT** $\to$ **1** $\to$ **1(TYPE)** $\to$ **1(1-VAR)**

Chú ý nhập xong số liệu thì bấm **AC** để thoát.

Bước 4 Xem kết quả:

- Trung bình mẫu ($\overline{x}$): **SHIFT** $\to$ **1** $\to$ **4(VAR)** $\to$ **2**
- Độ lệch tiêu chuẩn mẫu hiệu chỉnh (s): **SHIFT** $\to$ **1** $\to$ **4** $\to$ **4**

Ví dụ 4.10. Ở một địa điểm thu mua vải, kiểm tra một số vải thấy kết quả sau

Số khuyết tật ở mỗi đơn vị	0	1	2	3	4	5	6
Số đơn vị kiểm tra (10m)	8	20	12	40	30	25	15

Hãy tính kỳ vọng mẫu và độ lệch chuẩn hiệu chỉnh mẫu của mẫu trên.

Lời giải Ví dụ 4.10
Cách 1: Gọi X là số khuyết tật ở mỗi đơn vị. Lập bảng tính toán

x_i	n_i	$n_i x_i$	$n_i x_i^2$
0	8	0	0
1	20	20	20
2	12	24	48
3	40	120	360
4	30	120	480
5	25	125	625
6	15	90	540
$\sum$	$n = 150$	$\sum_i n_i x_i = 499$	$\sum_i n_i x_i^2 = 2073$

Suy ra $\overline{x} = \dfrac{499}{150} = 3,3267$; $\overline{x^2} = \dfrac{2073}{150} = 13,82$; $\hat{s}^2 = \overline{x^2} - (\overline{x})^2 = 13,82 - (3,3267)^2 = 2,7531$; $s^2 = \dfrac{150}{149} \times 2,7531 = 2,7715$; $s = \sqrt{2,7715} = 1,6648$.

Cách 2: Sử dụng máy tính CASIO FX570VN PLUS tính được $\overline{x} = 3,3267$; $s = 1,6648$.

4.2 Ước lượng điểm

Như đã biết, các tham số của dấu hiệu nghiên cứu $\mathcal{X}$ như trung bình, phương sai, cơ cấu của tổng thể theo dấu hiệu $\mathcal{X}$ được sử dụng rất rộng rãi trong phân tích kinh tế - xã hội và nhiều lĩnh vực khác. Song các tham số này thông thường lại chưa biết. Vì vậy đặt ra vấn đề ước lượng chúng nhờ phương pháp mẫu.

Sau khi đã mô hình hóa dấu hiệu $\mathcal{X}$ và cơ cấu tổng thể bằng biến ngẫu nhiên X và quy luật phân phối xác suất của nó, ta có thể phát biểu vấn đề thực tế nêu trên dưới dạng toán học như sau: "Cho biến ngẫu nhiên X, có thể đã biết hoặc chưa biết quy luật phân phối xác suất dạng tổng quát, nhưng chưa biết tham số θ của nó. Hãy ước lượng θ bằng phương pháp mẫu". Đây là một trong những bài toán cơ bản của thống kê toán học.

Dưới đây ta sẽ nghiên cứu các phương pháp tìm ra một số hay một khoảng số để ước lượng θ. Các phương pháp này xuất phát từ cơ sở hợp lý nào đó để tìm ước lượng của θ, chứ không phải là sự chứng minh chặt chẽ.

Phương pháp ước lượng điểm chủ trương dùng giá trị quan sát của một thống kê để ước lượng một tham số (véc tơ tham số) nào đó theo các tiêu chuẩn: vững, không chệch, hiệu quả.

4.2.1 Phương pháp hàm ước lượng

Mô tả phương pháp

Giả sử cần ước lượng tham số θ của biến ngẫu nhiên X. Thông thường X là một biến ngẫu nhiên mà ta muốn biết phân phối xác suất của X. Trong xác suất, biết phân phối của X nghĩa là ta đã có thông tin "đầy đủ" về nó, nói khác đi ta có thể tính được xác suất để biến ngẫu nhiên nhận giá trị trong một miền bất kỳ. Tuy nhiên trên thực tế phân phối xác suất của X thường rất khó nắm bắt. Chính bì vậy ta mong muốn biết được những thông tin chính về X như giá trị trung bình, độ lệch chuẩn, trung vị, mốt, mômen... của X.

Từ X ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n. Chọn lập thống kê $G = f(X_1, X_2, \ldots, X_n)$. Một trong những cách chọn dạng hàm f là tương ứng thống kê đặc trưng của mẫu ngẫu nhiên với tham số cần ước lượng của biến ngẫu nhiên. Phương pháp này gọi là phương pháp mô-men (moment estimation). Chẳng hạn $G = f(X_1, X_2, \ldots, X_n) = \dfrac{1}{n}\sum_{i=1}^n X_i$ nếu là ước lượng cho kỳ vọng $E(X)$, còn $G = \dfrac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$ nếu ước lượng phương sai...

Tiến hành lập mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$. Tính giá trị cụ thể của G ứng với mẫu này, tức là $g = f(x_1, x_2, \ldots, x_n)$. Đây là ước lượng điểm của θ.

Thống kê $G = f(X_1, X_2, \ldots, X_n)$, viết gọn là G, là hàm ước lượng của θ.

Chú ý rằng θ là một số chưa biết, còn $G = f(X_1, X_2, \ldots, X_n)$ là một biến ngẫu nhiên. Như vậy, ở đây ta đã lấy một biến ngẫu nhiên để xấp xỉ cho một số. Câu hỏi đặt ra là

1. Ước lượng đưa ra có "tốt" không?

2. "Ước lượng tốt" được hiểu theo nghĩa nào?

Dưới đây là một số tiêu chuẩn cho ước lượng điểm.

Một số tiêu chuẩn lựa chọn hàm ước lượng

Cùng một mẫu ngẫu nhiên có thể xây dựng nhiều thống kê G khác nhau để ước lượng cho tham số θ. Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau.

(a) Ước lượng không chệch (unbiased estimator): Ước lượng G của θ được gọi là ước lượng không chệch của θ nếu

$$\boxed{E(G) = \theta \quad \text{hay} \quad E(G-\theta) = 0} \tag{4.27}$$

nghĩa là về trung bình "độ chệch" bằng 0. Tính chất này có nghĩa là ước lượng G không có sai số hệ thống mà chỉ có sai số ngẫu nhiên.

Điều kiện (4.27) của ước lượng không chệch có nghĩa là trung bình các giá trị của G bằng θ. Tuy nhiên, không có nghĩa là mọi giá trị của G đều trùng khít với θ mà từng giá trị của G có thể sai lệch rất lớn so với θ. Vì vậy ta tìm ước lượng không chệch sao cho độ sai lệch trung bình là bé nhất.

(b) Ước lượng vững (consistent estimator): Ước lượng G của θ được gọi là ước lượng vững nếu với mọi $\varepsilon > 0$ ta có

$$\lim_{n\to+\infty} P\{|G-\theta| < \varepsilon\} = 1$$

hay

$$\lim_{n\to+\infty} P\{\theta-\varepsilon < G < \theta+\varepsilon\} = 1.$$

Tính chất này đảm bảo cho ước lượng G gần θ tùy ý với xác suất cao khi cỡ mẫu đủ lớn.

(c) Ước lượng hiệu quả (efficient estimator): Đại lượng $G-\theta$ được gọi là sai số (nó là một biến ngẫu nhiên) còn giá trị trung bình của sai số $E(G-\theta) = E(G)-\theta$ được gọi là độ chệch. Ta mong muốn tìm được ước lượng sao cho sai số bình phương trung bình

$$MSE(G) = E(G-\theta)^2$$

nhỏ nhất có thể. Tuy nhiên trong nhiều trường hợp tìm ước lượng thỏa mãn điều kiện này khá khó. Lưu ý rằng

$$E(G-\theta)^2 = [E(G)-\theta]^2 + V(G) = (\text{Độ chệch})^2 + V(G).$$

Vì vậy ta thường chỉ tìm ước lượng tốt nhất trong các ước lượng không chệch tức là ước lượng không chệch có phương sai $V(G)$ nhỏ nhất trong các ước lượng không chệch. Ước lượng đó gọi là ước lượng hiệu quả.

Khi ta không biết ước lượng hiệu quả có tồn tại hay không thì để so sánh các ước lượng không chệch ta sẽ so sánh độ lệch tiêu chuẩn hay phương sai của chúng. Ước lượng không chệch có độ lệch tiêu chuẩn hay phương sai nhỏ hơn sẽ "tốt hơn". Độ lệch tiêu chuẩn của ước lượng điểm G, ký hiệu là σ_G được gọi là sai số tiêu chuẩn (standard error). Ước lượng điểm của sai số tiêu chuẩn được ký hiệu là $\hat{\sigma}_G$. Tổng quát hơn để so sánh hai ước lượng điểm G_1 và G_2 cho tham số θ bất kỳ, ta so sánh sai số bình phương trung bình, ước lượng nào có sai số bình phương trung bình bé hơn là ước lượng tốt hơn. Tức là nếu độ hiệu quả tương đối

$$\frac{MSE(G_1)}{MSE(G_2)}$$

nhỏ hơn 1 thì ta kết luận rằng ước lượng G_1 hiệu quả hơn ước lượng G_2.

Để xét xem ước lượng không chệch G có phải là ước lượng hiệu quả của θ hay không ta cần phải tìm một cận dưới của phương sai của các ước lượng không chệch và so sánh phương sai của G với cận dưới này. Điều này được giải quyết bằng bất đẳng thức Cramer–Rao phát biểu như sau.

Định lý 4.1. Cho mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ được lấy từ tổng thể có dấu hiệu nghiên cứu được mô hình hóa bởi biến ngẫu nhiên X mà hàm mật độ xác suất $f(X, \theta)$ hay hàm phân phối xác suất $F(X, \theta)$ thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế) và $\hat{\theta}$ là ước lượng không chệch bất kỳ của θ thì

$$\boxed{V(G) \geq \frac{1}{nE\left(\frac{\partial(\ln f(X,\theta))}{\partial\theta}\right)^2}} \tag{4.28}$$

4.2.2 Ước lượng điểm cho một số tham số thông dụng

(a) Ước lượng điểm cho kỳ vọng hay giá trị trung bình: Giả sử X là biến ngẫu nhiên với kỳ vọng $E(X) = \mu$ chưa biết, μ được xem là trung bình của tổng thể. Từ X ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ cỡ n. Chọn

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

làm ước lượng điểm cho kỳ vọng $E(X) = \mu$. Ước lượng điểm $\overline{X}$ thỏa mãn cả ba tính chất tốt đã nêu ở trên: không chệch, vững và hiệu quả.

Khi ta có một mẫu cụ thể $W_X = (x_1, x_2, \ldots, x_n)$ thì

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

là một ước lượng điểm của μ.

(b) Ước lượng điểm cho phương sai: Giả sử X là biến ngẫu nhiên với phương sai $V(X) = \sigma^2$ chưa biết, σ^2 được xem là phương sai của tổng thể. Nếu ta có một mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ cỡ n của X thì xuất phát từ công thức tính phương sai, đại lượng

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

được xem xét để dùng làm ước lượng cho σ^2. Tuy nhiên không khó để chỉ ra rằng

$$E(\hat{S}^2) = \frac{n-1}{n}\sigma^2$$

nghĩa là $\hat{S}^2$ là một ước lượng chệch của σ^2. Để thu được ước lượng không chệch cho σ^2 ta "hiệu chỉnh" đại lượng này một chút bằng cách đặt

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}\hat{S}^2.$$

Đại lượng S^2 (chính là phương sai hiệu chỉnh mẫu ngẫu nhiên) là ước lượng không chệch cho σ^2. Người ta đã chứng minh được rằng cả $\hat{S}^2$ và S^2 đều là ước lượng vững cho σ^2. Như vậy ước lượng tốt cho σ^2 là S^2.

Khi có mẫu cụ thể $W_X = (x_1, x_2, \ldots, x_n)$ ta tính được các giá trị cụ thể của $\hat{S}^2$ và S^2, ký hiệu là $\hat{s}^2$ và s^2:

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

đây là các ước lượng điểm của σ^2.

Ví dụ 4.11. Chỉ số IQ của 10 sinh viên được cho như sau:

87, 81, 88, 85, 100, 90, 114, 93, 86, 98

1. Ước lượng điểm cho chỉ số IQ trung bình là trung bình mẫu $\overline{x} = 92,2$.
2. Ước lượng điểm cho độ lệch tiêu chuẩn σ của chỉ số IQ là độ lệch tiêu chuẩn hiệu chỉnh mẫu $s = 9,64$.
3. Sai số tiêu chuẩn của trung bình mẫu là $\sigma_{\overline{x}} = \sigma/\sqrt{n}$, trong đó $n = 10$ là cỡ mẫu. Ước lượng điểm cho sai số tiêu chuẩn của trung bình mẫu là

$$\hat{\sigma}_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{9,64}{\sqrt{10}} = 3,05.$$

Nếu ta dùng $\hat{\mu} = \dfrac{x_1 + x_2}{2}$ như một ước lượng điểm cho chỉ số IQ trung bình thì sai số tiêu chuẩn là $\sigma_{\hat{\mu}} = \sigma/\sqrt{2}$ và ước lượng điểm cho sai số tiêu chuẩn này là

$$\hat{\sigma}_{\overline{\mu}} = \frac{s}{\sqrt{2}} = \frac{9,64}{\sqrt{2}} = 6,81.$$

Rõ ràng ước lượng $\hat{\mu}$ không hiệu quả bằng ước lượng $\overline{x}$.

(c) Ước lượng điểm cho tỷ lệ: Cho p là một tỷ lệ hay xác suất của một sự kiện A trong tổng thể chưa biết. Ta thực hiện n quan sát độc lập và gọi m là số lần xuất hiện A. Khi đó tần suất mẫu

$$f = \frac{m}{n}$$

là ước lượng điểm cho p. Người ta chứng minh được rằng ước lượng này có cả ba tính chất tốt đã nêu ở trên: không chệch, vững và hiệu quả.

Ví dụ 4.12. Trong đợt vận động bầu cử tổng thống người ta phỏng vấn ngẫu nhiên 1600 cử tri thì được biết 960 người sẽ bỏ phiếu cho ứng cử viên A. Hãy chỉ ra ước lượng điểm cho tỷ lệ phiếu thực mà ứng cử viên A sẽ thu được.

Lời giải Ví dụ 4.12 Ước lượng điểm cần tìm là $f = \dfrac{960}{1600} = 0,6 = 60\%$.

4.2.3 Phương pháp ước lượng hợp lý cực đại (maximum-likelihood estimation)

Giả sử ta đã biết quy luật phân phối xác suất dạng tổng quát của biến ngẫu nhiên X, chẳng hạn hàm mật độ xác suất $f_X(x,\theta)$ (có thể hiểu $f_X(x,\theta)$ là công thức xác suất nếu X rời rạc). Cần ước lượng tham số θ của X ta lập mẫu cụ thể $W_x = (x_1, x_2, \dots, x_n)$.

Hàm của đối số θ

$$L\Big(x_1, x_2, \dots, x_n, \theta\Big) = f_X(x_1,\theta) f_X(x_2,\theta) \dots f_X(x_n,\theta) \tag{4.29}$$

và gọi là hàm số hợp lý của tham số θ. Giá trị của hàm hợp lý chính là xác suất hay mật độ xác suất tại điểm $W_x = (x_1, x_2, \dots, x_n)$. Giá trị $g = g(x_1, x_2, \dots, x_n)$ được gọi là ước lượng hợp lý cực đại của θ nếu ứng với giá trị này của θ hàm hợp lý đạt cực đại.

Do hàm L và hàm $\ln L$ đạt cực đại tại cùng một giá trị của θ nên ta có thể tìm giá trị của θ để $\ln L$ đạt cực đại theo các bước sau.

1. Tìm đạo hàm bậc nhất của hàm $\ln L$ theo θ.
2. Lập phương trình $\dfrac{d \ln L}{d\theta} = 0$. Phương trình này gọi là phương trình hợp lý. Giả sử nó có nghiệm $\theta = g = g(x_1, x_2, \dots, x_n)$.
3. Tìm đạo hàm bậc hai $\dfrac{d^2 \ln L}{d\theta^2}$. Nếu tại điểm $\theta = g$ đạo hàm bậc hai âm thì tại điểm này hàm $\ln L$ đạt cực đại. Do đó $g = g(x_1, x_2, \dots, x_n)$ là ước lượng điểm hợp lý tối đa cần tìm.

Ví dụ 4.13. Bằng phương pháp hợp lý cực đại ta tìm được ước lượng của tham số p trong quy luật phân phối nhị thức $\mathcal{B}(n;p)$ là $\frac{\overline{x}}{n}$ và ước lượng của tham số λ trong quy luật phân phối Poisson là $\frac{1}{\overline{x}}$. (Phần chứng minh xem như bài tập).

Ví dụ 4.14. Tìm ước lượng hợp lý cực đại của các tham số μ và σ của biến ngẫu nhiên tuân theo quy luật phân phối chuẩn $\mathcal{N}(\mu;\sigma^2)$.

Lời giải Ví dụ 4.14 Dễ thấy hàm hợp lý có dạng

$$L = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

Lập hệ phương trình hợp lý

$$\begin{cases} \dfrac{d\ln L}{d\mu} = 0, \\ \dfrac{d\ln L}{d\sigma} = 0. \end{cases}$$

Giải hệ này ta nhận được $\mu = \overline{x}$ và $\sigma = s$.

Chú ý 4.2. Đối số của hàm hợp lý là θ chứ không phải $x_1, x_2, \ldots, x_n$. Do vậy nếu thay giá trị mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ bằng mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ thì kết quả và chứng minh trên vẫn đúng. Do đó ta thu được kết quả tổng quát hơn. Chẳng hạn, nếu X tuân theo quy luật phân phối chuẩn thì ước lượng hợp lý của μ là $\overline{X}$ và ước lượng hợp lý của σ là S.

Ngoài ra còn các phương pháp Bayes, phương pháp minimax, phương pháp bootstrap …Các phương pháp này tìm ước lượng điểm cũng như các tiêu chuẩn để kiểm tra các tính chất tốt của một ước lượng điểm nằm ngoài phạm vi của bài học.

BÀI 12 (2 tiết)

4.3 Ước lượng khoảng

Phương pháp ước lượng điểm nói trên có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy cho trường hợp một tham số.

Khái niệm ước lượng khoảng

Giả sử chưa biết đặc trưng θ nào đó của biến ngẫu nhiên X. Ước lượng khoảng của θ là chỉ ra một khoảng số (g_1, g_2) nào đó chứa θ, tức là có thể ước lượng $g_1 < \theta < g_2$.

Phương pháp khoảng ước lượng tin cậy

Để ước lượng tham số θ của biến ngẫu nhiên X, từ biến ngẫu nhiên này ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ cỡ n. Chọn thống kê $G(X, \theta)$ sao cho mặc dù chưa biết giá trị của θ, quy luật phân phối xác suất của G vẫn hoàn toàn xác định. Do đó, với xác suất α khá bé ta tìm được $P(G_1 < \theta < G_2) = 1 - \alpha$. Vì α khá bé, nên $\gamma = 1 - \alpha$ khá lớn (thông thường yêu cầu $1 - \alpha = \gamma \geq 0,95$ để có thể áp dụng nguyên lý xác suất lớn cho sự kiện $(G_1 < \theta < G_2)$). Khi đó, sự kiện $(G_1 < \theta < G_2)$ hầu như chắc chắn xảy ra trong một phép thử. Thực hiện một phép thử đối với mẫu ngẫu nhiên W_X ta thu được mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, từ đó tính được các giá trị của G_1, G_2, ký hiệu là g_1, g_2. Như vậy có thể kết luận: với độ tin cậy $1 - \alpha = \gamma$ tham số θ nằm trong khoảng (g_1, g_2).

(a) (G_1, G_2) được gọi là khoảng tin cậy của θ với độ tin cậy $\gamma = 1 - \alpha$.

(b) $1 - \alpha = \gamma$ được gọi là độ tin cậy của ước lượng.

(c) $I = G_2 - G_1$ được gọi là độ dài khoảng tin cậy.

Các cận G_1 và G_2 phụ thuộc vào mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ nên chúng là các biến ngẫu nhiên. Ta thường mong muốn tìm được các khoảng tin cậy có hai tính chất sau:

1. Có độ tin cậy cao.
2. Độ dài khoảng tin cậy nhỏ, tức là hiệu $G_2 - G_1$ bé.

Đôi khi ta chỉ quan tâm đến giá trị lớn nhất hay giá trị nhỏ nhất của tham số. Ví dụ ta xét hai kết luận sau:

1. Với xác suất 95% thì chiều cao trung bình của người Việt Nam nhỏ hơn 168cm.
2. Với xác suất 95%, chiều cao trung bình của sinh viên Đại học Bách khoa Hà Nội lớn hơn 165cm.

Ở kết luận thứ nhất, với độ tin cậy 95% thì chiều cao trung bình của người Việt Nam nằm trong khoảng $(0; 168)$ và ta chỉ quan tâm đến giá trị lớn nhất. Ở kết luận thứ hai, với độ tin cậy 95% thì chiều cao trung bình của sinh viên Đại học Bách khoa Hà Nội thuộc khoảng $(165; +\infty)$ và ta chỉ quan tâm đến giá trị nhỏ nhất. Các khoảng tin cậy dạng này được gọi là khoảng tin cậy một phía. Khoảng tin cậy mà chỉ quan tâm đến cận trên được gọi là khoảng tin cậy lớn nhất hay khoảng tin cậy trái còn khoảng tin cậy mà chỉ quan tâm tới cận dưới được gọi là khoảng tin cậy nhỏ nhất hay khoảng tin cậy phải. Khoảng tin cậy mà ta quan tâm tới cả hai cận (trên và dưới) được gọi là khoảng tin cậy hai phía (đôi khi gọi là khoảng tin cậy đối xứng vì các cận trên, cận dưới thường đối xứng qua ước lượng điểm của tham số).

Dưới đây là các khoảng tin cậy cho kỳ vọng hay giá trị trung bình, phương sai và tỷ lệ hay xác suất.

4.3.1 Khoảng tin cậy cho kỳ vọng

Bài toán 4.1. Giả sử biến ngẫu nhiên X có kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$, trong đó $E(X) = \mu$ chưa biết. Bài toán đặt ra là tìm ước lượng khoảng cho μ dựa trên các quan sát $W_x = (x_1, x_2, \ldots, x_n)$.

Từ tổng thể, ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ và xét các trường hợp sau.

Trường hợp phân phối chuẩn, phương sai đã biết

Giả sử $X \sim \mathcal{N}(\mu; \sigma^2)$ với σ^2 đã biết. Khi đó $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là biến ngẫu nhiên có phân phối chuẩn với kỳ vọng $E(\overline{X}) = \mu$ và phương sai $V(\overline{X}) = \frac{\sigma^2}{n}$.Do đó

$$\boxed{U = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}} \tag{4.30}$$

có phân phối chuẩn tắc $\mathcal{N}(0; 1)$.

Chọn cặp số không âm α_1, α_2 thỏa mãn $\alpha_1 + \alpha_2 = \alpha$, tìm các phân vị chuẩn tắc $u_{\alpha_1}, u_{1-\alpha_2}$ sao cho $P(U < u_{\alpha_1}) = \alpha_1$; $P(U < u_{1-\alpha_2}) = 1 - \alpha_2$. Do tính chất của phân phối chuẩn tắc $u_{\alpha_1} = -u_{1-\alpha_1}$, suy ra

$$\begin{aligned} P(-u_{1-\alpha_1} < U < u_{1-\alpha_2}) &= P(u_{\alpha_1} < U < u_{1-\alpha_2}) \\ &= P(U < u_{1-\alpha_2}) - P(U < u_{\alpha_1}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha. \end{aligned}$$

Như vậy,

$$1-\alpha = P(-u_{1-\alpha_1} < U < u_{1-\alpha_2}) = P\left(-u_{1-\alpha_1} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < u_{1-\alpha_2}\right).$$

Hay

$$1-\alpha = P\left(\overline{X} - u_{1-\alpha_2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{1-\alpha_1}\frac{\sigma}{\sqrt{n}}\right).$$

Khi có mẫu cụ thể $W_X = (x_1, x_2, \ldots, x_n)$, tính được giá trị cụ thể $\overline{x}$ của $\overline{X}$, khi đó khoảng tin cậy cho μ với độ tin cậy $\gamma = 1-\alpha$ là:

$$\boxed{\left(\overline{x} - u_{1-\alpha_2}\frac{\sigma}{\sqrt{n}} \quad ; \quad \overline{x} + u_{1-\alpha_1}\frac{\sigma}{\sqrt{n}}\right)} \tag{4.31}$$

Như vậy, với độ tin cậy $\gamma = 1-\alpha$ cho trước, có vô số khoảng tin cậy cho μ vì có vô số cặp α_1, α_2 thỏa mãn $\alpha_1 + \alpha_2 = \alpha$. Ở đây ta chỉ xét một số trường hợp đặc biệt.

(a) Khoảng tin cậy hai phía (đối xứng) $(\alpha_1 = \alpha_2 = \alpha/2)$

$$\boxed{\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad ; \quad \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)} \tag{4.32}$$

trong đó $u_{1-\frac{\alpha}{2}}$ được tra từ bảng giá trị hàm phân phối chuẩn tắc (Phụ lục 3) từ hệ thức

$$\boxed{\Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}} \tag{4.33}$$

hoặc tra từ bảng giá trị hàm Laplace (Phụ lục 2) từ hệ thức

$$\boxed{\phi(u_{1-\frac{\alpha}{2}}) = \frac{1-\alpha}{2}} \tag{4.34}$$

(b) Khoảng tin cậy trái $(\alpha_1 = \alpha, \alpha_2 = 0)$:

$$\boxed{\left(-\infty \quad ; \quad \overline{x} + u_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)} \tag{4.35}$$

trong đó $u_{1-\alpha}$ được tra từ bảng giá trị hàm phân phối chuẩn tắc (Phụ lục 3) từ hệ thức

$$\boxed{\Phi(u_{1-\alpha}) = 1-\alpha}. \tag{4.36}$$

hoặc tra từ bảng giá trị hàm Laplace (Phụ lục 2) từ hệ thức

$$\boxed{\phi(u_{1-\alpha}) = \frac{1-2\alpha}{2}}. \tag{4.37}$$

(c) Khoảng tin cậy phải $(\alpha_1 = 0, \alpha_2 = \alpha)$:

$$\boxed{\left(\overline{x} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}} \quad ; \quad +\infty\right)} \tag{4.38}$$

Ví dụ 4.15. Trọng lượng của một loại sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn với độ lệch tiêu chuẩn là 6 gam. Cân thử 36 sản phẩm loại này ta thu được kết quả sau:

177	165	152	174	159	166	160	152	162	175	158	169
166	162	181	168	170	150	173	164	167	177	167	175
165	182	166	158	166	170	168	165	160	160	169	166

1. Với độ tin cậy $1-\alpha = 95\%$, hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình của loại sản phẩm nói trên.
2. Có thể khẳng định trọng lượng trung bình của sản phẩm ít nhất là bao nhiêu với độ tin cậy 99%.
3. Với độ tin cậy 99% hãy tìm khoảng tin cậy cho trọng lượng trung bình của sản phẩm.

Lời giải Ví dụ 4.15 Gọi X là trọng lượng sản phẩm, $X \sim \mathcal{N}(\mu, \sigma^2)$ với $\sigma = 6$. Trọng lượng trung bình của sản phẩm là $E(X) = \mu$ chưa biết cần ước lượng.

1. Đây là bài toán ước lượng bằng khoảng tin cậy đối xứng cho kỳ vọng $E(X) = \mu$ của tổng thể có phân phối chuẩn khi đã biết phương sai với $\sigma = 6$.
   - Chọn thống kê $U = \dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}$. Thống kê $U \sim \mathcal{N}(0;1)$.
   - Với $\alpha = 0,05$, $\Phi(u_{1-\frac{\alpha}{2}}) = \Phi(u_{0,975}) = 0,975$, tra bảng giá trị hàm phân phối chuẩn tắc nhận được $u_{0,975} = 1,96$.
   - Từ số liệu đã cho ta có cỡ mẫu $n = 36$, trung bình mẫu $\overline{x} = 166,22$ (gam), suy ra khoảng tin cậy 95% cho trọng lượng trung bình là

     $$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\,;\, \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \left(166,22 - 1,96 \times \frac{6}{\sqrt{36}}\ \ 166,22 + 1,96 \times \frac{6}{\sqrt{36}}\right)$$

     hay $(164,26 < \mu < 168,18)$.
   - Vậy với độ tin cậy 95%, trọng lượng trung bình của loại sản phẩm nói trên từ 164,26 gam đến 168,18 gam.

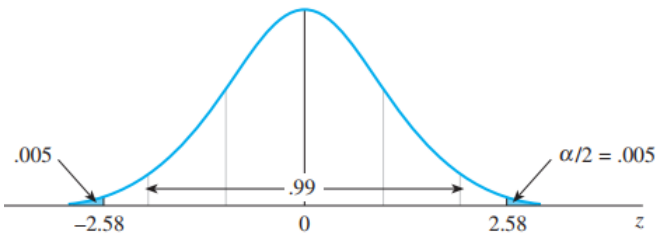
2. Với độ tin cậy $\gamma = 99\%$ suy ra $\alpha = 0,01$, $\Phi(u_{1-\alpha}) = \Phi(u_{0,99}) = 0,99$, tra bảng giá trị hàm phân phối chuẩn tắc nhận được $u_{0,99} = 2,33$ (xem Hình 4.4). Giá trị bé nhất cho trọng lượng trung bình là

   $$\overline{x} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}} = 166,22 - 2,33\frac{6}{\sqrt{36}} = 163,89.$$

   Tức là với xác suất 99% ta có thể khẳng định trọng lượng trung bình của sản phẩm lớn hơn 163,89 gam.

Hình 4.4: Giá trị của $u_{1-\alpha}$ với độ tin cậy 99%

3. Ta có $u_{1-\frac{\alpha}{2}} = 2,58$ (xem Hình 4.5) nên khoảng tin cậy cho trọng lượng trung bình của sản phẩm với độ tin cậy 99% là
$$= \left(166,22 - 2,58 \times \frac{6}{\sqrt{36}} \;\; 166,22 + 2,58 \times \frac{6}{\sqrt{36}}\right) = (163,64\,;\; 168,80).$$
Như vậy, trên cùng một mẫu nếu độ tin cậy càng lớn thì độ dài của khoảng tin cậy càng lớn.



Hình 4.5: Giá trị của $u_{1-\frac{\alpha}{2}}$ với độ tin cậy 99%

Chú ý 4.3. 1. Chú ý rằng không thể viết $P(164,26 < \mu < 168,18) = 0,95$ vì độ tin cậy gắn với khoảng tin cậy ngẫu nhiên chứ không gắn với mẫu cụ thể. Hơn nữa vì μ là một hằng số nên nó chỉ có thể thuộc hoặc không thuộc khoảng (164,26; 168,18) nên $(164,26 < \mu < 168,18)$ không phải là sự kiện ngẫu nhiên.

2. Ta có thể xác định $u_{1-\frac{\alpha}{2}} = 1,96$ ở ý Ví dụ 4.15(1) từ bảng giá trị hàm Laplace (Phụ lục 2) từ hệ thức $\phi(u_{1-\alpha}) = \dfrac{1-\alpha}{2}$.

Trong phần tiếp theo ta chỉ đưa ra công thức áp dụng.

Trường hợp phân phối chuẩn, phương sai chưa biết

Trong nhiều bài toán thực tế ta không biết phương sai của tổng thể. Với giả thiết dữ liệu tuân theo phân phối chuẩn thì người ta đã chứng minh được rằng thống kê

$$\boxed{T = \frac{\overline{X} - \mu}{S}\sqrt{n}} \tag{4.39}$$

có phân phối Student với $n-1$ bậc tự do, $T \sim t(n-1)$. Bằng phương pháp tương tự như đã trình bày ở trên, ta có thể tìm được các khoảng tin cậy cho kỳ vọng $E(X) = \mu$ với độ tin cậy $\gamma = 1-\alpha$.

(a) Khoảng tin cậy hai phía (đối xứng)

$$\boxed{\left(\overline{x} - t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \quad ; \quad \overline{x} + t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)} \tag{4.40}$$

(b) Khoảng tin cậy trái

$$\boxed{\left(-\infty \quad ; \quad \overline{x} + t^{(n-1)}_{1-\alpha}\frac{s}{\sqrt{n}}\right)} \tag{4.41}$$

(c) Khoảng tin cậy phải

$$\boxed{\left(\overline{x} - t^{(n-1)}_{1-\alpha}\frac{s}{\sqrt{n}} \quad ; \quad +\infty\right)} \tag{4.42}$$

trong đó $t^{(n-1)}_{1-\frac{\alpha}{2}}$, $t^{(n-1)}_{1-\alpha}$ được tra từ bảng phân phối Student với $n-1$ bậc tự do (Phụ lục 4).

Ví dụ 4.16. Để kiểm tra sự chính xác của hệ thống đóng gói tự động các bao gạo khi xuất khẩu tại một nhà máy, người ta đã chọn ngẫu nhiên 16 bao và tính được trọng lượng trung bình là 49,75 kg và độ lệch tiêu chuẩn hiệu chỉnh mẫu là 0,5 kg. Giả thiết rằng trọng lượng của những bao gạo có phân phối chuẩn.

1. Tìm khoảng tin cậy cho trọng lượng trung bình của bao gạo với độ tin cậy 95%.
2. Với độ tin cậy 95% có thể khẳng định trọng lượng trung bình của bao gạo cao nhất là bao nhiêu? Dựa trên kết quả thu được có thể khẳng định về trung bình các bao gạo đã bị đóng thiếu hay không nếu biết rằng trọng lượng chuẩn mỗi bao là 50 kg.

Lời giải Ví dụ 4.16 Gọi X là trọng lượng các bao gạo được đóng gói tự động, $X \sim \mathcal{N}(\mu, \sigma^2)$. Trọng lượng trung bình là $E(X) = \mu$ chưa biết, cần ước lượng.

1. Đây là bài toán ước lượng bằng khoảng tin cậy đối xứng cho kỳ vọng $E(X) = \mu$ của tổng thể có phân phối chuẩn khi chưa biết phương sai.
   - Chọn thống kê $T = \dfrac{\overline{X}-\mu}{S}\sqrt{n}$. Thống kê T có phân phối Student, $T \sim t^{(n-1)}$.
   - Với độ tin cậy 99% suy ra $\alpha = 0,05$, $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(15)}_{0,975} = 2,13$, tra bảng phân phối Student (Phụ lục 4).

- Ta có $n = 16$, $\overline{x} = 49,75$, $s = 0,5$, suy ra khoảng tin cậy 95% cho trọng lượng trung bình là

$$\left(\overline{x} - t_{1-\alpha}^{(n-1)}\frac{s}{\sqrt{n}} \quad ; \quad +\infty\right) = \left(49,75 - 2,13 \times \frac{0,5}{\sqrt{16}}\ 49,75 + 2,13 \times \frac{0,5}{\sqrt{16}}\right)$$

hay $(49,48 < \mu < 50,02)$.

- Vậy với độ tin cậy 95%, ta có thể khẳng định trọng lượng trung bình của các bao gạo nói trên từ 49,48 kg đến 50,02 kg.

2. Đây là bài toán ước lượng bằng khoảng tin cậy một phía cho kỳ vọng $E(X) = \mu$ của tổng thể có phân phối chuẩn khi chưa biết phương sai.

   - Với độ tin cậy 95% suy ra $t_{1-\alpha}^{(n-1)} = t_{0,95}^{(15)} = 1,75$, tra bảng phân phối Student (Phụ lục 4).

   - Khoảng tin cậy lớn nhất cho lượng trung bình là

$$\left(0; \overline{x} + t_{1-\alpha}^{(n-1)}\frac{s}{\sqrt{n}}\right) = \left(0; 49,75 + 1,75\frac{0,5}{\sqrt{16}}\right) = (0; 49,97).$$

   - Với độ tin cậy 95% giá trị lớn nhất cho trọng lượng trung bình của các bao gạo là 49,97 kg.

     Ta có thể khẳng định về trung bình các bao gạo đã bị đóng thiếu.

Trường hợp mẫu cỡ lớn ($n \geq 30$)

Khi mẫu cỡ n lớn, trung bình mẫu ngẫu nhiên $\overline{X}$ có phân phối xấp xỉ phân phối chuẩn với trung bình μ và phương sai σ^2/n do đó thống kê

$$\boxed{U = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)} \tag{4.43}$$

Cũng do n lớn nên ta có thể xấp xỉ σ (chưa biết) bởi S do đó thống kê

$$\boxed{U = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim \mathcal{N}(0,1)} \tag{4.44}$$

Trong thực hành cho phép vận dụng với $n \geq 30$.

Các lập luận như đã trình bày ở trên, các khoảng tin cậy cho μ là:

(a) Khoảng tin cậy hai phía (đối xứng)

$$\boxed{\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \quad ; \quad \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)} \tag{4.45}$$

(b) Khoảng tin cậy trái

$$\boxed{\left(-\infty \quad ; \quad \overline{x} + u_{1-\alpha}\frac{s}{\sqrt{n}}\right)} \tag{4.46}$$

(c) Khoảng tin cậy phải

$$\boxed{\left(\overline{x} - u_{1-\alpha}\frac{s}{\sqrt{n}} \quad ; \quad +\infty\right)} \tag{4.47}$$

Ví dụ 4.17. Để ước lượng trọng lượng trung bình của loại trái cây A tại một vùng, người ta thu hoạch ngẫu nhiên 100 trái cây A của vùng đó và thu được kết quả sau

Trọng lượng (gam)	40-42	42-44	44-46	46-48	48-50	50-52
Số trái	7	13	25	35	15	5

Hãy ước lượng trọng lượng trung bình của loại trái cây A trong vùng bằng khoảng tin cậy đối xứng với độ tin cậy 95%. Cho biết trọng lượng loại trái cây A là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Lời giải Ví dụ 4.17 Gọi X là trọng lượng loại trái cây A, $X \sim \mathcal{N}(\mu, \sigma^2)$. Trọng lượng trung bình của loại trái cây A là $E(X) = \mu$ chưa biết, cần ước lượng. Đây là bài toán ước lượng khoảng của kỳ vọng của biến ngẫu nhiên phân phối chuẩn trường hợp chưa biết phương sai, mẫu cỡ $n = 100 > 30$.

- Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 100 > 30$ nên thống kê $U \sim \mathcal{N}(0,1)$.
- Khoảng tin cậy đối xứng cho $E(X) = \mu$ là $\left(\overline{x} - u_{1-\frac{\alpha}{2}}\dfrac{s}{\sqrt{n}} \,;\, \overline{x} + u_{1-\frac{\alpha}{2}}\dfrac{s}{\sqrt{n}}\right)$ trong đó, với $\alpha = 0,05$, $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$ được tra từ bảng giá trị hàm phân phối chuẩn tắc (Phụ lục 3).
- Từ số liệu đã cho tính được $n = 100$, $\overline{x} = 46,06$, $s = 2,48$. Suy ra khoảng tin cậy đối xứng của μ là $\left(46,06 - 1,96 \times \dfrac{2,48}{\sqrt{100}} \,;\, 46,06 + 1,96 \times \dfrac{2,48}{\sqrt{100}}\right)$ hay $(45,573 \,;\, 46,546)$.
- Vậy với độ tin cậy 95%, trọng lượng trung bình của loại trái cây A ở vùng trên từ 45,573 gam đến 46,546 gam.

4.3.2 Khoảng tin cậy cho phương sai (đọc thêm)

Bài toán 4.2. Giả sử X là biến ngẫu nhiên có phân phối chuẩn, $X \sim \mathcal{N}(\mu;\sigma^2)$, với phương sai $V(X) = \sigma^2$ chưa biết. Hãy ước lượng khoảng cho phương sai σ^2 dựa trên số liệu quan sát $W_x = (x_1, x_2, \ldots, x_n)$.

Lập một mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n và xét các trường hợp sau.

Trường hợp kỳ vọng chưa biết

Người ta đã chứng minh được rằng thống kê

$$\boxed{\chi^2 = \frac{(n-1)S^2}{\sigma^2}} \tag{4.48}$$

là biến ngẫu nhiên có phân phối khi bình phương với $(n-1)$ bậc tự do, $\chi^2(n-1)$.

Chọn cặp số không âm α_1, α_2 thỏa mãn $\alpha_1 + \alpha_2 = \alpha$, tìm các phân vị $P(\chi^2 < \chi^2_{(n-1,\alpha_1)}) = \alpha_1$; $P(\chi^2 < \chi^2_{(n-1,1-\alpha_2)}) = 1-\alpha_2$. Từ đó suy ra

$$P\left(\chi^2_{(n-1,\alpha_1)} < \chi^2 < \chi^2_{(n-1,1-\alpha_2)}\right) = P\left(\chi^2_{(n-1,\alpha_1)} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{(n-1,1-\alpha_2)}\right) = 1-\alpha,$$

hay tương đương với

$$P\left(\frac{(n-1)S^2}{\chi^2_{(n-1,1-\alpha_2)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(n-1,\alpha_1)}}\right) = 1-\alpha.$$

Khi có mẫu cụ thể $W_X = (x_1, x_2, \ldots, x_n)$, tính được giá trị cụ thể s^2 của S^2, khi đó khoảng tin cậy cho σ^2 với độ tin cậy $\gamma = 1-\alpha$ là:

$$\boxed{\left(\frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha_2)}} \quad ; \quad \frac{(n-1)s^2}{\chi^2_{(n-1,\alpha_1)}}\right)} \tag{4.49}$$

Ta xét một số trường hợp cụ thể của (4.49).

(a) Khoảng tin cậy hai phía $(\alpha_1 = \alpha_2 = \alpha/2)$

$$\boxed{\left(\frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha/2)}} \quad ; \quad \frac{(n-1)s^2}{\chi^2_{(n-1,\alpha/2)}}\right)} \tag{4.50}$$

(b) Khoảng tin cậy trái $(\alpha_1 = \alpha, \alpha_2 = 0)$:

$$\boxed{\left(0 \quad ; \quad \frac{(n-1)s^2}{\chi^2_{(n-1,\alpha)}}\right)} \tag{4.51}$$

(c) Khoảng tin cậy phải $(\alpha_1 = 0, \alpha_2 = \alpha)$:

$$\boxed{\left(\frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha)}} \quad ; \quad +\infty\right)} \tag{4.52}$$

Chú ý 4.4.

1. Giá trị $\chi^2_{(n-1,1-\alpha/2)}$, $\chi^2_{(n-1,\alpha/2)}$, $\chi^2_{(n-1,\alpha)}$ và $\chi^2_{(n-1,1-\alpha)}$ được tra từ bảng phân phối khi bình phương với $n-1$ bậc tự do (Phụ lục 6).
2. Lấy căn bậc hai các cận trong các khoảng tin cậy cho phương sai ta sẽ thu được các khoảng tin cậy cho các độ lệch chuẩn.

Trường hợp kỳ vọng đã biết

Nếu kỳ vọng $E(X) = \mu_0$ đã biết, người ta đã chứng minh được rằng thống kê

$$\boxed{\chi^2 = \frac{\sum_{i=1}^n \left(X_i - \mu_0\right)^2}{\sigma^2}} \tag{4.53}$$

là biến ngẫu nhiên có phân phối khi bình phương với n bậc tự do, $\chi^2(n)$. Làm giống như trường hợp trên ta nhận được:

(a) Khoảng tin cậy hai phía

$$\boxed{\left(\frac{\sum_{i=1}^n \left(x_i - \mu_0\right)^2}{\chi^2_{(n,1-\alpha/2)}} \quad ; \quad \frac{\sum_{i=1}^n \left(x_i - \mu_0\right)^2}{\chi^2_{(n,\alpha/2)}}\right)} \tag{4.54}$$

(b) Khoảng tin cậy trái

$$\boxed{\left(0 \quad ; \quad \frac{\sum_{i=1}^n \left(x_i - \mu_0\right)^2}{\chi^2_{(n,\alpha)}}\right)} \tag{4.55}$$

(c) Khoảng tin cậy phải

$$\boxed{\left(\frac{\sum_{i=1}^n \left(x_i - \mu_0\right)^2}{\chi^2_{(n,1-\alpha)}} \quad ; \quad +\infty\right)} \tag{4.56}$$

Ví dụ 4.18. Trọng lượng của một loại sản phẩm tuân theo luật phân phối chuẩn. Cân thử từng sản phẩm của một mẫu ngẫu nhiên gồm 25 đơn vị, ta nhận được kết quả sau:

Trọng lượng sản phẩm (gam)	29,3	29,7	30	30,5	30,75
Số sản phẩm	4	5	8	5	3

Với độ tin cậy 95% hãy tìm khoảng tin cậy cho phương sai của trọng lượng sản phẩm trong hai trường hợp

1. Đã biết kỳ vọng $E(X) = 30$.

2. Không biết kỳ vọng.

Lời giải Ví dụ 4.18 Gọi X là trọng lượng sản phẩm, $X \sim \mathcal{N}(\mu, \sigma^2)$. Phương sai của trọng lượng sản phẩm là $V(X) = \sigma^2$ chưa biết, cần ước lượng. Đây là bài toán ước lượng khoảng của phương sai của biến ngẫu nhiên phân phối chuẩn.

1. Trường hợp $E(X) = 30$ đã biết.

- Chọn thống kê $\chi^2 = \dfrac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}$. Thống kê χ^2 có phân phối khi bình phương với n bậc tự do.
- Với độ tin cậy 95% suy ra $\alpha = 0,05$, tra bảng phân phối khi bình phương (Phụ lục 5), $\chi^2_{(n,1-\alpha/2)} = \chi^2_{(25;0,975)} = 40,65$, $\chi^2_{(n,\alpha/2)} = \chi^2_{(25;0,025)} = 13,12$.
- Từ số liệu đã cho tính được $n = 25$, $\sum_{i=1}^5 n_i(x_i-30)^2 = 5,13$. Suy ra khoảng tin cậy cần tìm là

$$\left(\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\chi^2_{(n,1-\alpha/2)}}; \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\chi^2_{(n,\alpha/2)}}\right) = \left(\frac{5,13}{40,65}; \frac{5,13}{13,12}\right)$$

  hay $(0,1262 < \sigma^2 < 0,3910)$.
- Vậy với độ tin cậy 95%, phương sai của trọng lượng sản phẩm từ 0,1262 đến 0,3910.

2. Trường hợp kỳ vọng chưa biết ta sử dụng thống kê

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Tra bảng phân phối khi bình phương với độ tin cậy 95% và bậc tư do $n-1 = 24$ ta được $\chi^2_{(n-1,1-\alpha/2)} = \chi^2_{(24;0,975)} = 39,36$, $\chi^2_{(n-1,\alpha/2)} = \chi^2_{(24;0,025)} = 12,40$.

Tính $\overline{x} = \dfrac{1}{25}\sum_{i=1}^5 n_i x_i = 30,012$, $(n-1)s^2 = \sum_{i=1}^5 n_i(x_i-30,012)^2 = 5,1264$.

Vậy khoảng tin cậy cần tìm là

$$\left(\frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\chi^2_{(n-1,1-\alpha/2)}}; \frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\chi^2_{(n-1,\alpha/2)}}\right) = \left(\frac{5,1264}{39,36}; \frac{5,1264}{12,40}\right)$$

hay $(0,1302 < \sigma^2 < 0,4134)$.

Vậy với độ tin cậy 95% phương sai σ^2 lớn hơn 0,1302 và nhỏ hơn 0,4134 hay độ lệch chuẩn σ nằm từ 0,3608 gam đến 0,6430 gam.

4.3.3 Khoảng tin cậy cho xác suất

Bài toán 4.3. Ta xét một tổng thể mà mỗi cá thể hoặc có tính chất A hoặc không có tính chất A nào đó. Gọi p là tỷ lệ cá thể có tính chất A trong tổng thể. Thông thường p chưa biết. Dựa trên một mẫu được chọn ngẫu nhiên, hãy tìm khoảng tin cậy cho p với độ tin cậy $\gamma = 1-\alpha$ cho trước.

Ta thực hiện n phép thử độc lập, cùng điều kiện. Gọi m là số cá thể có tính chất A trong n cá thể được chọn. Khi đó m là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n;p)$. Khi n lớn thì m có phân phối xấp xỉ phân phối chuẩn và do đó tần suất $f = \dfrac{m}{n}$ cũng có phân phối xấp xỉ phân phối chuẩn với trung bình $E(f) = p$ và phương sai $V(f) = \dfrac{p(1-p)}{n}$. Xấp xỉ này "tốt" nếu $np > 5$ và $n(1-p) > 5$. Tuy nhiên vì p chưa biết nên phương sai $V(f)$ chưa biết. Do f là ước lượng điểm tốt cho p nên phương sai $V(f)$ có thể được xấp xỉ bằng $\dfrac{f(1-f)}{n}$. Khi đó thống kê

$$\boxed{U = \frac{f-p}{\sqrt{f(1-f)}}\sqrt{n}} \tag{4.57}$$

có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0;1)$. Từ đó ta tìm được các khoảng tin cậy cho p khi có mẫu cụ thể là:

(a) Khoảng tin cậy hai phía (đối xứng)

$$\boxed{\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \quad ; \quad f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)} \tag{4.58}$$

(b) Khoảng tin cậy trái

$$\boxed{\left(0 \quad ; \quad f + u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}}\right)} \tag{4.59}$$

(c) Khoảng tin cậy phải

$$\boxed{\left(f - u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}} \quad ; \quad 1\right)} \tag{4.60}$$

Chú ý 4.5. Vì p chưa biết nên ta không kiểm tra được điều kiện $np > 5$ và $n(1-p) > 5$. Trong thực hành ta dùng các điều kiện $nf > 5$ và $n(1-f) > 5$.

Ví dụ 4.19. Điều tra nhu cầu tiêu dùng loại hàng A trong 100 hộ gia đình ở khu dân cư B thấy 60 hộ gia đình có nhu cầu loại hàng trên. Với độ tin cậy $1-\alpha = 95\%$ hãy tìm khoảng tin cậy đối xứng của tỷ lệ hộ gia đình có nhu cầu loại hàng đó. Giả thích kết quả thu được.

Lời giải Ví dụ 4.19 Gọi p là tỷ lệ hộ gia đình ở khu dân cư B có nhu cầu mặt hàng A. Kiểm tra điều kiện $nf = 100 \times 0,6 = 60 > 5$ và $n(1-f) = 100 \times 0,4 = 40 > 5$. Đây là bài toán ước lượng khoảng của tỷ lệ trường hợp mẫu cỡ n đủ lớn.

- Thống kê $U = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$ có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0;1)$.
- Khoảng tin cậy đối xứng của xác suất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \; ; \; f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

  trong đó $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

- Với $n = 100$, $m = 60$, $f = \dfrac{m}{n} = 0,6$, suy ra khoảng tin cậy cần tìm là

$$\left(0,6 - 1,96\sqrt{\frac{0,6\times 0,4}{100}}\ ;\ 0,6 - 1,96\sqrt{\frac{0,6\times 0,4}{100}}\right) = (0,504\ ;\ 0,696).$$

- Vậy tỷ lệ hộ gia đình ở khu dân cư B có nhu cầu loại hàng A là từ 50,4% đến 69,6% với độ tin cậy 95%.

Vì mọi giá trị trong khoảng tin cậy này đều nhỏ hơn 70% nên ta cũng có thể khẳng định rằng có ít hơn 70% hộ gia đình có nhu cầu mặt hàng A.

4.3.4 Xác định kích thước mẫu

Bài toán 4.4. Giả sử θ là một tham số của tổng thể và $G = G_n(X_1, X_2, \ldots, X_n)$ là một ước lượng cho θ dựa trên mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ có kích thước n. Cho trước số $\varepsilon > 0$ và $\gamma \in (0;1)$. Ta nói rằng G_n có độ chính xác (hay sai số) ε với độ tin cậy γ nếu

$$\boxed{P\Big(|\theta - G_n| \le \varepsilon\Big) \ge \gamma} \tag{4.61}$$

Nếu kích thước mẫu n càng lớn thì độ chính xác của ước lượng càng cao, sai số càng nhỏ. Tuy nhiên kích thước mẫu càng lớn thì nhà nghiên cứu càng tốn nhiều thời gian, tiền bạc và công sức cho việc thu thập dữ liệu. Bài toán đặt ra là cần chọn kích thước mẫu tối thiểu là bao nhiêu để đủ đạt được độ chính xác mong muốn.

Ta xét các trường hợp sau đây.

(a) Trường hợp ước lượng cho giá trị trung bình

Giả sử với độ tin cậy γ cho trước, ta muốn có ước lượng cho giá trị trung bình μ với sai số không quá ε. Trong trường hợp X có phân phối chuẩn và phương sai σ^2 đã biết thì

$$\boxed{P\Big(|\overline{x} - \mu| \le u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = \gamma} \tag{4.62}$$

ở đây $\alpha = 1 - \gamma$. Do đó nếu n thỏa mãn

$$u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \le \varepsilon$$

hay tương đương với

$$\boxed{n \ge \frac{\sigma^2(u_{1-\alpha/2})^2}{\varepsilon^2}} \tag{4.63}$$

thì

$$P\Big(|\overline{x} - \mu| \le \varepsilon\Big) \ge P\Big(|\overline{x} - \mu| \le u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = \gamma.$$

Ta sẽ chọn kích thước mẫu n là số nguyên dương bé nhất thỏa mãn điều kiện (4.63) thì sẽ đạt được độ chính xác (hay sai số) với độ tin cậy γ mong muốn.

Tuy nhiên công thức trên chỉ áp dụng được khi σ đã biết. Trong trường hợp X có phân phối chuẩn và phương sai chưa biết thì

$$P\left(|\overline{x}-\mu| \le t_{1-\alpha/2}^{(n-1)}\frac{s}{\sqrt{n}}\right) = \gamma.$$

Do đó nếu n thỏa mãn

$$t_{1-\alpha/2}^{(n-1)}\frac{s}{\sqrt{n}} \le \varepsilon$$

hay tương đương với

$$\boxed{n \ge \frac{s^2(t_{1-\alpha/2}^{(n-1)})^2}{\varepsilon^2}} \tag{4.64}$$

thì ta sẽ có

$$P\left(|\overline{x}-\mu| \le \varepsilon\right) \ge \gamma.$$

Ta không thể tìm n thỏa mãn (4.64) vì cả s và $t_{1-\alpha/2}^{(n-1)}$ đều phụ thuộc vào n. Tuy nhiên, nhìn vào bảng phân phối Student ta thấy khi số bậc tự do lớn thì các phân vị của phân phối Student $t_{1-\alpha/2}^{(n-1)}$ và các phân vị của phân phối chuẩn tắc $u_{1-\alpha/2}$ gần như nhau, vì vậy ta có thể chọn n thỏa mãn

$$\boxed{n \ge \frac{s^2(u_{1-\alpha/2})^2}{\varepsilon^2}} \tag{4.65}$$

Như vậy, nếu tìm được ước lượng cho s thì ta sẽ tìm được n từ bất đẳng thức này. Người ta thường lấy một mẫu sơ bộ cỡ mẫu n đủ lớn ($n \ge 30$) để tính s rồi sau đó tìm n từ bất đẳng thức (4.65).

Ví dụ 4.20. Ta muốn xây dựng một ước lượng với độ tin cậy 95% và độ chính xác 2 dặm cho vận tốc trung bình của ô tô trên đường cao tốc. Một mẫu điều tra sơ bộ với cỡ mẫu 50 cho ta $s = 9$ dặm. Hỏi cần lấy mẫu cỡ tối thiểu là bao nhiêu để đạt được độ chính xác và độ tin cậy đã đặt ra.

Lời giải Ví dụ 4.20 Ta có $\varepsilon = 2$, $s = 9$, $\gamma = 95\%$ nên $u_{1-\alpha/2} = 1,96$. Từ bất đẳng thức (4.65) ta có

$$n \ge \frac{s^2(u_{1-\alpha/2})^2}{\varepsilon^2} = \frac{9^2(1,96)^2}{2^2} = 77,79.$$

Nghĩa là ta phải lấy cỡ mẫu ít nhất là 78. Tuy nhiên vì ta đã có mẫu sơ bộ với cỡ 50, nên thực ra ta chỉ cần bổ sung thêm 28 quan sát nữa.

Chú ý 4.6. 1. Việc lấy mẫu sơ bộ để tìm s là hợp lý vì ta biết rằng s là ước lượng vững cho σ, nghĩa là s hội tụ đến σ khi $n \to \infty$, do đó khi n lớn thì các giá trị của s khá "ổn định" và "gần" σ.

2. Trong ví dụ trên, phương sai của tổng thể dữ liệu được ước lượng thông qua độ phân tán của mẫu sơ bộ. Ta cũng có thể ước lượng phương sai của tổng thể dữ liệu theo những cách khác.

Ví dụ 4.21. Ta muốn ước lượng số năm đi học của những người trưởng thành trong một vùng với độ tin cậy 95% và độ chính xác không quá 1 năm thì cần lấy mẫu tối thiểu là bao nhiêu?

Lời giải Ví dụ 4.21 Ta có $\varepsilon = 1$, $\gamma = 95\%$ nên $u_{1-\alpha/2} = 1,96$. Để ước lượng phương sai của tổng thể dữ liệu, ta nhận thấy rằng số năm học dao động từ) đến 18. Nếu phân phối của số năm đi học là phân phối chuẩn thì hầu hết các giá trị quan sát sẽ thuộc khoảng $(\mu - 3\sigma; \mu + 3\sigma)$. Khoảng này có độ dài 6σ do đó ta dùng ước lượng $6\sigma = 18$ hay $\sigma = 3$. Từ đó ta tìm được

$$n \geq \frac{3^2(1,96)^2}{1^2} = 35.$$

Ta cũng có thể đoán rằng số năm học không thể quá 24 và do đó độ lệch tiêu chuẩn σ không thể quá 4 để có ước lượng thô về σ từ đó tìm n.

(b) Trường hợp ước lượng cho tỷ lệ

Ta xét một tổng thể mà mỗi cá thể hoặc có tính chất A hoặc không có tính chất A nào đó. Gọi p là tỷ lệ cá thể có tính chất A trong tổng thể. Thông thường p chưa biết. Giả sử trên một mẫu ngẫu nhiên cỡ n có m cá thể có tính chất A. Khi đó tần suất $f_n = \dfrac{m}{n}$ có phân phối xấp xỉ phân phối chuẩn với trung bình là p và phương sai $\dfrac{p(1-p)}{n}$. Do đó

$$\boxed{P\left(|f_n - p| \leq u_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = \gamma} \tag{4.66}$$

ở đây $\alpha = 1 - \gamma$. Do đó nếu n thỏa mãn

$$u_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \varepsilon$$

hay tương đương với

$$\boxed{n \geq \frac{(u_{1-\alpha/2})^2 p(1-p)}{\varepsilon^2}} \tag{4.67}$$

thì

$$P\left(|f_n - p| \leq \varepsilon\right) \geq \gamma.$$

Như vậy ta cần lấy n là số nguyên dương nhỏ nhất thỏa mãn (4.67).

Tuy nhiên, vì giá trị của p chưa biết, nên vế phải của (4.67) chưa xác định. Có hai cách để khắc phục tình trạng này.

1. Cách thứ nhất là lấy một mẫu sơ bộ kích thước k để thu được tần suất f_k và lấy f_k làm ước lượng ban đầu cho p. Khi đó bất đẳng thức (4.67) trở thành

$$\boxed{n \geq \frac{(u_{1-\alpha/2})^2 f_k(1-f_k)}{\varepsilon^2}} \tag{4.68}$$

với điều kiện

$$\boxed{kf_k > 5 \quad \text{và} \quad k(1-f_k) > 5} \tag{4.69}$$

Ta sẽ lấy n là số nguyên dương nhỏ nhất thỏa mãn (4.68) và (4.69).

2. Cách thứ hai, sử dụng bất đẳng thức Cauchy $p(1-p) \leq \dfrac{1}{4}$ ta nhận được

$$u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq \frac{u_{1-\alpha/2}}{2\sqrt{n}}.$$

Nếu ta chọn n thỏa mãn điều kiện

$$\boxed{\frac{u_{1-\alpha/2}}{2\sqrt{n}} \quad \text{hay} \quad n \geq \frac{(u_{1-\alpha/2})^2}{4\varepsilon^2}} \tag{4.70}$$

thì n thỏa mãn điều kiện (4.67). Vậy ta sẽ lấy n là số nguyên dương nhỏ nhất thỏa mãn (4.70).

Ví dụ 4.22. Một kỹ sư muốn ước lượng tỷ lệ sản phẩm loại A của một nhà máy với độ tin cậy 90% và sai số không quá 0,02. Hỏi kỹ sư đó phải lấy một mẫu cỡ n bằng bao nhiêu?

Lời giải Ví dụ 4.22

1. Nếu làm theo cách thứ nhất, trước hết người kỹ sư lấy một mẫu cỡ $n = 1000$ sản phẩm kiểm tra thấy có 640 sản phẩm loại A. Khoảng tin cậy của tỷ lệ sản phẩm loại A của nhà máy dựa trên mẫu điều tra này là

$$\begin{aligned}
&f - u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}}\ ;\ f + u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}} \\
&= (0,64 - 1,645\sqrt{\frac{0,64(1-0,64)}{1000}}\ \ 0,64 + 1,645\sqrt{\frac{0,64(1-0,64)}{1000}}) \\
&= (0,64 - ,25\ ;\ 0,64 + 0,25).
\end{aligned}$$

Sai số của ước lượng là 0,25 lớn hơn 0,02. Vậy cần lấy mẫu lớn hơn nữa. Cỡ mẫu n phải thỏa mãn (4.68), tức là

$$n \geq \frac{(1,645)^2(0,64)(0,36)}{(0,02)^2} = 1558,67.$$

Vậy $n = 1559$.

2. Nếu sử dụng cách thứ hai thì phải chọn n thỏa mãn (4.70), tức là

$$n \geq \frac{(1,645)^2}{(4)(0,02)^2} = 1691,266.$$

Do đó $n = 1692$.

Ví dụ 4.23. Một cuộc thăm dò dư luận tại thành phố A được tiến hành để hỏi xem có nên thực hiện một dự án quan trọng với kinh phí lớn trong thành phố hay không. Tỷ lệ người ủng hộ việc thực hiện dự án trên mẫu thăm dò cho ta ước lượng tỷ lệ người ủng hộ việc thực hiện dự án trên toàn bộ số dân của thành phố. Nếu ước lượng tỷ lệ này với độ tin cậy 95% và sai số không quá 0,03 thì phải lấy cỡ mẫu bao nhiêu?

Lời giải Ví dụ 4.23 Nếu sử dụng cách thứ hai, ta phải chọn n thỏa mãn (4.70) hay

$$n \geq \frac{(1,96)^2}{(4)(0,03)^2} = 1067,111.$$

Do đó $n = 1068$. Điều thú vị nhất trong tính toán trên là cỡ mẫu tối thiểu n chỉ phụ thuộc vào độ chính xác và độ tin cậy mà ta mong muốn, chứ không phụ thuộc vào tổng số dân trên thành phố. Nếu hỏi ý kiến 1068 người, thì ta đạt được độ tin cậy là 95% và độ chính xác 0,03 mà không phụ thuộc vào tổng số dân trên thành phố là 5 triệu, là 50 triệu hay 500 triệu!

Chú ý 4.7.

1. Cái khó khăn ở đây là không phải thu thập ý kiến của càng nhiều người càng tốt. Vấn đề lớn là phải đảm bảo đây là mẫu ngẫu nhiên. Chẳng hạn nếu thực hiện ý kiến qua mạng:

    - Gửi email đến n địa chỉ ngẫu nhiên. Giả sử trường hợp tốt nhất, cả n người đều trả lời. Vấn đề là không phải ai cũng dùng email.
    - Ngay cả trường hợp tất cả mọi người đều dùng email, thì không phải ai cũng trả lời. Quyết định trả lời và câu trả lời cũng liên quan đến nhau. Nếu xe hơi của bạn chạy tốt, ít khi bạn trả lời những câu hỏi liên quan đến chất lượng của hãng; nhưng nếu có trục trặc thì khả năng này rất cao. Nếu ta thấy 30% khách trên mạng than thở về chất lượng của xe cũng không nói lên là 30% số người mua xe gặp vấn đề.

2. Nếu p khá gần 0,5 thì sự khác nhau giữa số n tìm được theo hai cách sẽ không nhiều lắm. Nhưng nếu n khá gần 0 hoặc 1 thì sự khác biệt sẽ rất lớn. Do đó, nếu cảm thấy tỷ lệ p rất bé hoặc rất lớn thì nên dùng cách thứ nhất.

Hướng dẫn sử dụng phần mềm thống kê R

R là phần mềm thống kê miễn phí, được phát triển tại phòng thí nghiệm AT&T bởi Rick Becker, John Chambers và các cộng sự. Phiên bản đầu tiên của R được viết vào năm 1976. Tham khảo tại [2].

[2]Nguyễn Văn Tuấn (2015). Phân tích dữ liệu với R. NXB tổng hợp thành phố Hồ Chí Minh.

1. Để tìm khoảng tin cậy cho giá trị trung bình trong trường hợp dữ liệu có phân phối chuẩn và phương sai chưa biết đơn giản nhất là ta dùng hàm *t.test()*.

2. Để tìm khoảng tin cậy cho giá trị trung bình trong trường hợp dữ liệu có phân phối chuẩn và phương sai đã biết ta dùng hàm *z.test()*.

3. Để tìm khoảng tin cậy cho tỷ lệ ta dùng hàm *prop.test()*.

Đối với các khoảng tin cậy một phía ta đặt giá trị "less", "greater" cho tham số *alternative* (giá trị mặc định của tham số này là "two.sided").

Bài tập Chương 4

Ước lượng khoảng cho kỳ vọng

Bài tập 4.1. Xác suất để một sinh viên Đại học Bách khoa Hà Nội thi trượt môn Giải tích 2 là p. Một mẫu lớn n sinh viên được lựa chọn ngẫu nhiên và ký hiệu X là số sinh viên đã trượt môn Giải tích 2 trong mẫu.

(a) Giải thích tại sao có thể sử dụng $\dfrac{X}{n}$ để ước lượng cho p?

(b) Trình bày cách tính xấp xỉ xác suất sự sai khác giữa $\dfrac{X}{n}$ và p nhỏ hơn $0,01$? Áp dụng cho $n = 500$ và $p = 0,2$.

Bài tập 4.2. Tuổi thọ của một loại bóng đèn do một dây chuyền công nghệ sản xuất ra có độ lệch chuẩn là 305 giờ. Người ta lấy ngẫu nhiên ra 45 bóng đèn loại này thấy tuổi thọ trung bình là 2150 giờ. Với độ tin cậy 95% hãy ước lượng tuổi thọ trung bình của loại bóng đèn nói trên.

Bài tập 4.3. Một kỹ sư cho biết trọng lượng tạp chất trong một sản phẩm tuân theo luật phân phối chuẩn với độ lệch chuẩn bằng 3,8gam. Một mẫu ngẫu nhiên gồm 9 sản phẩm được tiến hành kiểm tra và thấy lượng tạp chất như sau (đơn vị tính là gam):

18,2 13,7 15,9 17,4 21,8 16,6 12,3 18,8 16,2

(a) Tìm khoảng tin cậy cho trọng lượng trung bình tạp chất của sản phẩm với độ tin cậy 99%.

(b) Không cần tính toán, nếu độ tin cậy 95% thì khoảng ước lượng trung bình sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

Bài tập 4.4. Giả sử chiều dài của một chi tiết sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn với độ lệch chuẩn là 0,2m. Người ta sản xuất thử nghiệm 35 sản phẩm loại này và tính được chiều dài trung bình là 25m. Với độ tin cậy 95% hãy ước lượng khoảng cho chiều dài trung bình của chi tiết sản phẩm đang được thử nghiệm.

Bài tập 4.5. Để xác định trọng lượng trung bình của các bao gạo được đóng gói bằng máy tự động, người ta chọn ngẫu nhiên ra 20 bao gạo và thấy trung bình mẫu là 49,2kg và độ lệch chuẩn mẫu hiệu chỉnh là 1,8kg. Biết rằng trọng lượng các bao gạo xấp xỉ phân phối chuẩn. Hãy tìm khoảng tin cậy cho trọng lượng trung bình của một bao gạo với độ tin cậy 99%.

Bài tập 4.6. Thời gian đợi phục vụ tại một cửa hàng ăn nhanh là biến ngẫu nhiên xấp xỉ phân phối chuẩn. Người ta khảo sát 16 người thì thấy thời gian đợi trung bình là 4 phút và độ lệch chuẩn mẫu hiệu chỉnh là 1,8 phút. Với độ tin cậy 99% hãy tìm khoảng tin cậy cho thời gian chờ đợi trung bình của một khách hàng tại cửa hàng ăn nhanh này.

Bài tập 4.7. Một mẫu ngẫu nhiên gồm 16 thùng hàng được chọn ra từ tất cả các thùng hàng được sản xuất bởi nhà máy trong một tháng. Trọng lượng của 16 thùng hàng lần lượt như sau (đơn vị tính là kg):

18,6	18,4	19,2	19,8	19,4	19,5	18,9	19,4
19,7	20,1	20,2	20,1	18,6	18,4	19,2	19,8

Tìm khoảng tin cậy cho trọng lượng trung bình tổng thể của tất cả các thùng hàng của nhà máy với độ tin cậy 95%, biết rằng trọng lượng thùng hàng được chọn ngẫu nhiên là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Bài tập 4.8. Để định mức thời gian gia công một chi tiết máy, người ta theo dõi ngẫu nhiên quá trình gia công 35 chi tiết máy và thu được số liệu:

Thời gian (phút)	16-17	17-18	18-19	19-20	20-21	21-22
Số chi tiết máy	3	4	10	9	5	4

Giả sử thời gian gia công chi tiết máy là biến ngẫu nhiên tuân theo luật phân phối chuẩn. Với độ tin cậy 95% hãy ước lượng khoảng tin cậy cho thời gian gia công trung bình một chi tiết máy nói trên.

Bài tập 4.9. Đo áp lực X (tính bằng kg/cm^2) của 18 thùng chứa ta được bảng kết quả sau:

Áp lực (kg/cm^2)	19,6	19,5	19,9	20,0	19,8	20,5	21,0	18,5	19,7
Số thùng	1	2	2	4	2	3	2	1	1

Với độ tin cậy 99% hãy tìm khoảng ước lượng đối xứng của áp lực trung bình của các thùng trên. Biết rằng áp lực là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Bài tập 4.10. Một bài báo trong Nuclear Engineering International (tháng 2 năm 1988, trang 33) mô tả một số đặc điểm của các thanh nhiên liệu được sử dụng trong một lò phản ứng hạt nhân của một công ty điện lực ở Na Uy. Người ta đo tỷ lệ làm giàu của 12 thanh và có được dữ liệu sau:

2,94 3,00 2,90 2,90 2,75 2,95 2,75 3,00 2,95 2,82 2,81 3,05

Giả sử tỷ lệ làm giàu của các thanh nhiên liệu tuân theo luật phân phối chuẩn. Hãy ước lượng khoảng cho tỷ lệ làm giàu trung bình của các thanh nhiên liệu với độ tin cậy 95%.

Bài tập 4.11. Trọng lượng những viên gạch trong một quá trình sản xuất gạch được giả sử là tuân theo luật phân phối chuẩn. Một mẫu ngẫu nhiên gồm 25 viên gạch vừa sản xuất ra trong ngày có trọng lượng trung bình 2,45 kg và độ lệch chuẩn mẫu hiệu chỉnh là 0,15 kg.

(a) Tìm khoảng tin cậy của trọng lượng trung bình của tất cả các viên gạch trong ngày với độ tin cậy 99%.

(b) Không cần tính toán, với độ tin cậy 95% thì khoảng tin cậy trung bình sẽ rộng hơn, hẹp hơn hay bằng với kết quả ý (a)?

(c) Một mẫu ngẫu nhiên gồm 20 viên gạch sẽ được chọn ra trong ngày mai. Không cần tính toán, với độ tin cậy 99% thì khoảng tin cậy cho trọng lượng trung bình của tất cả các viên gạch sản xuất ra trong ngày mai sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

(d) Sự thật rằng, độ lệch chuẩn mẫu của các viên gạch sản xuất trong ngày mai là 0,10kg. Không cần tính toán, với độ tin cậy 99% thì khoảng tin cậy cho trọng lượng trung bình của tất cả các viên gạch sản xuất ra trong ngày mai sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

Bài tập 4.12. Một trường đại học lớn đang quan tâm về lượng thời gian sinh viên tự nghiên cứu mỗi tuần. Người ta tiến hành khảo sát một mẫu ngẫu nhiên gồm 16 sinh viên, dữ liệu cho thấy thời gian nghiên cứu trung bình của một sinh viên là 15,26 giờ/tuần và độ lệch chuẩn hiệu chỉnh là 6,43 giờ. Giả sử thời gian nghiên cứu của sinh viên của trường đại học trên là tuân theo luật phân phối chuẩn.

(a) Tìm khoảng tin cậy cho lượng thời gian tự nghiên cứu trung bình mỗi tuần cho tất cả sinh viên trường đại học này với độ tin cậy 95%.

(b) Không cần tính toán, khoảng tin cậy của trung bình tổng thể khi ước lượng sẽ rộng hơn hay hẹp hơn với ba điều kiện sau:

b1. Mẫu gồm 30 sinh viên được chọn ra, với tất cả các điều kiện khác giống như ý (a)?

b2. Độ lệch chuẩn mẫu là 4,15 giờ, tất cả các điều kiện khác giống như ý (a)?

b3. Độ tin cậy 99%, tất cả các điều kiện khác giống như ý (a)?

Bài tập 4.13. Một kỹ sư nghiên cứu về cường độ nén của bê tông đang được thử nghiệm. Anh ta tiến hành kiểm tra 12 mẫu vật và có được các dữ liệu sau đây:

2216	2234	2225	2301	2278	2255	2249	2204	2286	2263	2275	2295

Giả sử cường độ nén của bê tông đang thử nghiệm tuân theo luật phân phối chuẩn.

(a) Hãy ước lượng khoảng với độ tin cậy 95% cho cường độ nén trung bình của bê tông đang được thử nghiệm.

(b) Hãy ước lượng khoảng tin cậy phải cho cường độ nén trung bình của bê tông đang được thử nghiệm với độ tin cậy 99%.

Bài tập 4.14. Người ta chọn ngẫu nhiên ra 49 sinh viên của một trường đại học và thấy chiều cao trung bình mẫu là 163cm và độ lệch chuẩn mẫu hiệu chỉnh là 12cm. Hãy tìm khoảng ước lượng với độ tin cậy 99% cho chiều cao trung bình của sinh viên của trường đó.

Bài tập 4.15. Một trường đại học tiến hành một nghiên cứu xem trung bình một sinh viên tiêu hết bao nhiêu tiền điện thoại trong một tháng. Họ điều tra 60 sinh viên và cho thấy số tiền trung bình mẫu là 95 nghìn và độ lệch chuẩn mẫu hiệu chỉnh là 36 nghìn. Hãy ước lượng khoảng với độ tin cậy 95% cho số tiền điện thoại trung bình trong một tháng của mỗi sinh viên.

Bài tập 4.16. Người ta điều tra 35 người nghiện thuốc lá được chọn ngẫu nhiên từ số lượng người nghiện hút thuốc lá của một thành phố thấy số điếu thuốc hút trong 5 ngày của họ là:

31	37	48	40	59	97	98	87	80	68	64	45
48	62	74	76	79	85	83	81	93	82	85	79
34	57	95	49	59	63	48	79	50	55	63	

Hãy tìm khoảng ước lượng cho số điếu thuốc hút trung bình trong 5 ngày của những người nghiện thuốc lá của thành phố đó với độ tin cậy 99%.

Bài tập 4.17. Để nghiên cứu về thời gian xem ti vi của một thanh niên từ 18 đến 35 tuổi trong vòng một tuần, người ta tiến hành khảo sát trên 40 người và cho ta bảng số liệu sau:

39	02	43	35	15	54	23	21	25	07	24	33	17	
23	24	43	11	15	17	15	19	06	43	35	25	37	
15	14	08	11	29	12	13	25	15	28	24	06	16	7

Hãy tìm khoảng ước lượng cho thời gian xem ti vi trung bình của thanh niên trong độ tuổi trên trong vòng một tuần với độ tin cậy 99%.

Bài tập 4.18. Để điều tra tiền điện phải trả trong một tháng của một hộ dân cư ở phường A, người ta kiểm tra ngẫu nhiên 200 hộ gia đình ở phường này và được kết quả sau:

Số tiền (nghìn đồng)	[80,180)	[180,280)	[280,380)	[380,480)	[480,580)	[580,680)	[680,780]
Số hộ gia đình	14	25	43	46	39	23	10

Ước lượng khoảng cho số tiền trung bình một hộ dân phải trả ở phường đó với độ tin cậy 95%.

Bài tập 4.19. Để ước lượng số lượng xăng hao phí trên một tuyến đường của một hãng xe khách, người ta tiến hành chạy thử nghiệm 55 lần liên tiếp trên tuyến đường này và có được số liệu:

Lượng xăng hao phí	10,5-11	11-11,5	11,5-12	12-12,5	12,5-13	13-13,5
Tần số	5	12	15	13	6	4

Hãy ước lượng lượng xăng hao phí trung bình cho một xe với độ tin cậy 95%.

Bài tập 4.20. Để xác định giá trung bình đối với một loại hàng hóa trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng thu được số liệu sau:

Giá (nghìn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng	5	8	13	14	30	11	8	6	4	1

Với độ tin cậy 95% hãy ước lượng giá trung bình của loại hàng đó tại thời điểm đang xét. Biết rằng giá hàng hóa là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Ước lượng khoảng cho tỷ lệ hay xác suất

Bài tập 4.21. Để ước lượng cho tỷ lệ những cây bạch đàn có chiều cao đạt chuẩn phục vụ cho việc khai thác ở một nông trường lâm nghiệp, người ta tiến hành đo ngẫu nhiên chiều cao của 135 cây và thấy có 36 cây cao từ 7,5m trở lên. Hãy ước lượng khoảng cho tỷ lệ các cây bạch đàn có chiều cao trên 7,5m với độ tin cậy 95%.

Bài tập 4.22. Để ước lượng số cá có trong hồ người ta bắt từ hồ lên 100 con đánh dấu rồi thả lại vào hồ. Sau đó người ta bắt lên 300 con thì thấy có 32 con bị đánh dấu. Hãy ước lượng khoảng cho số cá có trong hồ với độ tin cậy 99%.

Bài tập 4.23. Để điều tra thị phần xe máy, người ta chọn ngẫu nhiên ra 450 người mua xe máy trong một tháng ở các địa bàn ở một thành phố thì có 275 người mua xe Honda. Tìm khoảng tin cậy cho tỷ lệ người mua xe Honda với độ tin cậy 95%.

Bài tập 4.24. Kiểm tra ngẫu nhiên 400 sản phẩm do một hệ thống máy mới sản xuất thì thấy có 387 chính phẩm. Hãy ước lượng tỷ lệ chính phẩm tối thiểu của hệ thống máy mới với độ tin cậy 95%.

Bài tập 4.25. Thử nghiệm 560 bóng đèn điện tử do một nhà máy sản xuất thì thấy 10 bóng có lỗi kỹ thuật. Hãy tìm ước lượng cho tỷ lệ bóng có lỗi kỹ thuật tối đa với độ tin cậy 95%.

Bài tập 4.26. Mở thử 200 hộp của kho đồ hộp thấy có 10 hộp bị biến chất. Với độ tin cậy 95% hãy ước lượng tỷ lệ hộp bị biến chất tối đa của kho.

Bài tập 4.27. Chọn ngẫu nhiên ra 1000 trường hợp điều trị bệnh ung thư phổi, các bác sĩ thống kê thấy có 823 bệnh nhân bị chết trong vòng 10 năm.

(a) Ước lượng khoảng cho tỷ lệ tử vong của bệnh nhân điều trị bệnh ung thư phổi với độ tin cậy 99%.

(b) Cần phải lấy số lượng mẫu là bao nhiêu để với độ tin cậy 95% các sai số khi dự đoán tỷ lệ bệnh nhân điều trị ung thư phổi tử vong 10 năm là ít hơn 0,03?

Bài tập 4.28. Cần phải lập một mẫu ngẫu nhiên với kích thước là bao nhiêu để tỷ lệ phế phẩm của mẫu là 0,2 và độ dài khoảng tin cậy đối xứng là 0,05 và độ tin cậy của ước lượng là 95%.

Bài tập 4.29. Làm cách nào để ước lượng số thú hiếm trong một khu rừng với độ tin cậy 95%.

Bài tập 4.30. Nghiên cứu về năng suất của loại hoa màu A, người ta kiểm tra năng suất của 64 điểm trồng loại hoa màu này thu được bảng số liệu

Năng suất (tạ/ha)	40–45	45–50	50–55	55-60	60–65	65–70
Số điểm	2	5	15	30	8	4

(a) Hãy ước lượng năng suất trung bình của loại hoa màu A với độ tin cậy 95%; Nếu muốn sai số của ước lượng giảm đi 2 lần thì cần kiểm tra bao nhiêu điểm để đảm bảo yêu cầu nêu trên?

(b) Biết rằng trên toàn miền Bắc có 10.000 điểm trồng loại hoa màu A. Hãy cho biết có khoảng bao nhiêu điểm đạt năng suất trên 60 tạ/ha? Hãy kết luận với độ tin cậy 99%.

(c) Hãy cho biết tỷ lệ những điểm có năng suất trên 60 tạ/ha của loại hoa màu A tối thiểu là bao nhiêu? Hãy kết luận với độ tin cậy 95%?