BỘ TÀI CHÍNH
TRƯỜNG ĐẠI HỌC TÀI CHÍNH - MARKETING

ThS. LÊ TRƯỜNG GIANG

Lý thuyết và Bài tập

XÁC SUẤT - THỐNG KÊ

(Chương trình đại học các ngành kinh tế)

Tp. Hồ Chí Minh, ngày 20, tháng 09, năm 2014

Mục lục

Trang

PHẦN I. LÍ THUYẾT XÁC SUẤT

Tổng quan về lý thuyết xác suất

Xác suất là một bộ phận của toán học nghiên cứu về các hiện tượng ngẫu nhiên. Nói một cách đại khái thì hiện tượng ngẫu nhiên là hiện tượng ta không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát. Tuy nhiên, nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong những hoàn cảnh như nhau, thì trong nhiều trường hợp ta có thể rút ra được những kết luận khoa học về hiện tượng này. Chẳng hạn, ta không thể nói trước một hạt giống có nảy mầm hay không khi gieo xuống đất canh tác, nhưng nếu gieo nhiều hạt thì ta có thể rút ra chất lượng tốt hay xấu của hạt giống. Lý thuyết xác suất là môn khoa học nghiên cứu những quy luật trong những hiện tượng "tưởng chừng" như không có quy luật.

Lý thuyết xác suất ra đời vào nửa cuối thế kỷ thứ 17 ở nước Pháp. Hai nhà toán học vĩ đại của nước Pháp là Blaise Pascal (1623-1662) và Pierre de Fermat (1601-1665) đã trao đổi thư từ với nhau để bàn về một số bài toán liên quan đến trò chơi may rủi. Những bài toán này và các phương pháp giải chúng có thể xem là những nghiên cứu đầu tiên của lý thuyết xác suất. Tuy nhiên, trước đó, ở Italia các nhà toán học Cardano (1501-1576), Pacioli (1445-1509), Tartaglia đã giải một số bài toán riêng lẻ trong trò chơi may rủi.

Những bài toán kiểu Pascal và Fermat ảnh hưởng và khích lệ các nhà toán học trẻ (thời bấy giờ) như Huygens, Bernuolli và De Moivre tiếp tục nghiên cứu xác suất. Đấy là những người có công đầu tiên sáng tạo ra cơ sở toán học của lý thuyết xác suất. Có lẽ Cardano và Huygens (1629-1695) là những người đầu tiên viết sách về xác suất (sách của Cardano xuất bản năm 1663, tức là 100 năm sau khi ông viết xong tác phẩm này; sách của Huygens công bố năm 1657).

Lịch sử thực sự của lý thuyết xác suất bắt nguồn từ các công trình James Bernoulli (1654-1705). Ông là người phát minh ra Luật Số Lớn. Chính vì lý do đó, ngày nay Hội Xác Suất Thông Kê thế giới mang tên Bernoulli.

De Moivre (1667-1754) là tác giả của Định lý Giới Hạn Trung Tâm (trường hợp đối xứng), một trong những thành tựu quan trọng nhất của xác suất.

Năm 1812, Laplace P. S. (1749-1827) công bố cuốn sách "Theorie Analytique des Probabilities" (Lý Thuyết Giải Tích của Xác Suất). Cuốn sách này được xem là một đóng góp to lớn của Laplace trong xác suất. Ông là tác giả của định lý giới hạn trung tâm (trường hợp không đối xứng), và là người đầu tiên áp dụng xác suất vào các vấn đề liên quan đến sai số quan sát.

Poisson S. D (1781-1840) và Gauss C. F. (1777-1855) là những học giả tiếp thêm sức mạnh cho xác suất ứng dụng. Poisson là tác giả của Luật Biến Cố Hiếm nổi tiếng; Gauss là tác giả của lý thuyết sai số và đặc biệt đã sáng tạo ra Phương Pháp Bình Phương Tối Thiểu.

Chebyshev P. L. (1821-1894), Markov A. A. (1856-1922), Liapunov A. M. (1857-1918), Khinchin A. Y. là những nhà toán học người Nga có rất nhiều đóng góp cho sự phát triển của lý thuyết xác suất. Họ đã sử dụng phương pháp moment, hàm đặc trưng để nhận được những định lý giới hạn quan trọng. Đặc biệt Markov là người đã đưa ra mô hình Markov; Khinchin là tác giả của Luật Loga Lặp.

Lý thuyết xác suất hiện đại đi theo hướng tiên đề hoá. Các nhà toán học có công lớn trong hướng này là Berstein (1880-1968), Von Mises (1883-1953), Borel (1871-1956), P. Levy. Tuy nhiên, phải chờ đến sự ra đời của cuốn sách "Foundations of the Theory of Probability, 1933" của Kolmogorov, giới toán học mới công nhận xác suất là một lĩnh vực toán học chặt chẽ. Hệ tiên đề về xác suất của Kolmogorov được hầu hết các nhà toán học thừa nhận. Ngày nay nhiều ngành, nhiều hướng phát triển của lý thuyết xác suất như thống kê toán học, quá trình ngẫu nhiên, lý thuyết các định lí giới hạn, v.v... đều dựa trên hệ tiên đề Kolmogorov và các hệ quả của nó.

Song song với sự phát triển nội tại của lý thuyết xác suất, thống kê toán học ra đời từ các bài toán thực tế, gắn liền với tên tuổi của F. Ganton (1822 - 1911), K. Pierson (1857 - 1936), H. Cramer, R. A. Fisher, v.v... đã giải quyết nhiều bài toán lên quan tới lý thuyết chọn mẫu, lý thuyết ước lượng, lý thuyết kiểm định giả thiết thống kê, lý thuyết tương quan và hồi quy,...

Hiện nay lý thuyết xác suất và thống kê toán học đã và đang có nhiều ứng dụng to lớn trong các lĩnh vực hoạt động khác nhau của con người từ âm nhạc tới vật lý, từ văn học tới thống kê xã hội, từ cơ học tới thị trường chứng khoán, từ dự báo thời tiết tới kinh tế, từ nông học tới y học.

Trong chương trình quy định của Bộ Giáo dục và Đào tạo. Xác suất và Thống kê đã trở thành một môn học cơ sở và được giảng dạy cho hầu hết các chuyên ngành trong các trường Đại học và Cao đẳng.

Để góp phần vào việc giảng dạy môn Xác suất- Thống kê và tạo cho nguồn tài liệu tham khảo môn học thêm phong phú, tác giả đã mạnh dạn biên soạn quyển tài liệu này với hi vọng sẽ giúp ích cho các bạn sinh viên trong hành trình chinh phục tri thức. Cũng xin bày tỏ rằng những bài tập trong tài liệu nhỏ này là sự sưu tầm tổng hợp và trình bày chi tiết lại của các thầy cô đã từng giảng dạy cho tác giả cũng như những bài tập thú vị mà tác giả may mắn đọc được, nghiền ngẫm được.

Chương 1

Những khái niệm cơ bản của lý thuyết xác suất

1.1 Biến cố ngẫu nhiên

1.1.1 Phép thử và biến cố

a. Phép thử là việc thực hiện 1 thí nghiệm hay quan sát một hiện tượng nào đó để xem có xảy ra hay không.

b. Hiện tượng có xảy ra hay không trong phép thử được gọi là biến cố ngẫu nhiên. Biến cố ngẫu nhiên thường được kí hiệu là A, B, C...

1.1.2 Các loại biến cố

a. Không gian mẫu và biến cố sơ cấp

- Trong một phép thử, tập hợp tất cả các kết quả có thể xảy ra được gọi là không gian mẫu kí hiệu là Ω.
- Mỗi phần tử $\omega \in \Omega$ không thể phân nhỏ thành hai biến cố được gọi là biến cố sơ cấp.

b. Biến cố chắc chắn và biến cố không thể

- Trong một phép thử, biến cố nhất định xảy ra là biến cố chắc chắn, kí hiệu Ω.
- Biến cố không thể là biến cố không thể xảy ra khi thực hiện phép thử, ký hiệu $\emptyset$.

c. Số trường hợp đồng khả năng

- Hai hay nhiều biến cố trong một phép thử có khả năng xảy ra như nhau được gọi là đồng khả năng.
- Trong một phép thử mà mọi biến cố sơ cấp đều đồng khả năng thì số phần tử của không gian mẫu được gọi là số trường hợp đồng khả năng của phép thử.

d. Các phép toán

- Tổng của A và B là C, kí hiệu $C = A \cup B$ hay $C = A + B$, xảy ra khi và chỉ khi ít nhất 1 trong hai biến cố A hoặc B xảy ra.
- Tích của A và B là D, kí hiệu $D = A \cap B$ hay $D = AB$, xảy ra khi và chỉ khi cả hai biến cố A và B cùng xảy ra.
- Phần bù của A, kí hiệu: $\overline{A} = \Omega \backslash A = \{\omega \in \Omega \backslash \omega \notin A\}$.
- Kí hiệu $A \subset B$ nghĩa là biến cố A xảy ra thì biến cố B cũng xảy ra (còn được gọi là biến cố A kéo theo biến cố B).
- Biến cố hiệu $A \backslash B$ là biến cố xảy ra khi và chỉ khi xảy ra A nhưng không xảy ra B.

1.1.3 Quan hệ giữa các biến cố

a. Biến cố xung khắc

- Xung khắc: hai biến cố A và B được gọi là xung khắc nếu biến cố này xảy ra thì biến cố kia không xảy ra. Tức là $A \cap B = \emptyset$.
- Họ xung khắc (còn gọi là xung khắc từng đôi): Họ các biến cố $A_1, A_2, ..., A_n$ được gọi là họ xung khắc nếu một biến cố bất kỳ trong họ xảy ra thì các biến cố còn lại không xảy ra. Tức là

$$A_i \cap A_j = \emptyset \quad \forall i, j \ (i \neq j)$$

- Đối lập: hai biến cố A và B được gọi là đối lập nhau nếu chúng thỏa
  + Xung khắc nhau: $A \cap B = \emptyset$,
  + Phải có một trong hai biến cố xảy ra: $A \cup B = \Omega$.

- Họ đầy đủ: Họ các biến cố $A_1, A_2, ..., A_n$ được gọi là một họ đầy đủ nếu chúng thỏa

  + Họ xung khắc: $A_i \cap A_j = \emptyset \quad \forall i, j \ (i \neq j)$.
  + Phải có một biến cố trong họ xảy ra: $A_1 \cup A_2 \cup ... \cup A_n = \Omega$.

b. Biến cố độc lập

- Cho $A, B \subset \Omega$. Ta nói A, B độc lập nếu $P(AB) = P(A)P(B)$.
- Cho $A, B \subset \Omega$ và $P(B) > 0$. Khi đó A, B độc lập $\Leftrightarrow P(A/B) = P(A)$.
- Họ độc lập từng đôi:
  Họ các biến cố $A_1, A_2, ...A_n$ được gọi là họ độc lập từng đôi nếu hai biến cố bất kỳ trong họ là độc lập với nhau, tức là

$$P\left(A_i A_j\right) = P\left(A_i\right) P\left(A_j\right) \quad \forall i, j \ (i \neq j)$$

- Họ độc lập toàn thể:
  Họ các biến cố $A_1, A_2, ..., A_n$ được gọi là họ độc lập toàn thể nếu một biến cố bất kỳ trong họ và tích của một số bất kỳ các biến cố còn lại là độc lập với nhau, tức là

$$P\left(A_i . \prod_{j \in J} A_j\right) = P\left(A_i\right) P\left(\prod_{j \in J} A_j\right) \ , \quad \forall J \subset \{1, 2, ..., n\} \setminus \{i\}$$

1.2 Xác suất của biến cố

1.2.1 Định nghĩa xác suất

a. Định nghĩa xác suất theo quan điểm đồng khả năng(định nghĩa xác suất cổ điển)
Xét một không gian các biến cố sơ cấp có n biến cố sơ cấp đồng khả năng và giả sử có m biến cố sơ cấp thuận lợi cho một biến cố ngẫu nhiên A. Khi đó,

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

Lưu ý hạn chế của cách tính xác suất theo định nghĩa cổ điển là

+ Phép thử phải được phân tích thành những kết cục đồng khả năng.
+ Số trường hợp có thể của phép thử và số trường hợp thuận lợi của biến cố là hữu hạn.

b. Định nghĩa xác suất theo quan điểm hình học

Giả sử một phép thử có không gian các biến cố sơ cấp được biểu diễn bởi miền hình học Ω có số đo hữu hạn và khác 0. Biến cố A được biểu diễn bởi miền A trong Ω. Khi đó, xác suất của biến cố A được xác định bởi công thức

$$P(A) = \frac{mes(A)}{mes(\Omega)}$$

c. Định nghĩa xác suất theo thống kê

Thực hiện n phép thử giống nhau. Giả sử biến cố A xảy ra m lần. Khi đó, m được gọi là tần số xuất hiện biến cố A và tỉ số $\dfrac{m}{n}$ được gọi là tần suất xuất hiện biến cố A. Khi số phép thử tăng lên đủ lớn thì tần suất xuất hiện biến cố A dần về một số cố định được gọi là xác suất của biến cố A.

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

d. Định nghĩa xác suất theo tiên đề

+ Xét một tập Ω gồm tất cả các biến cố sơ cấp và một họ F các tập con của Ω có các tính chất sau:
  - i) $\emptyset \in F \in F$.
  - ii) Nếu $A \in F$ thì $\overline{A} \in F$.
  - iii) Nếu $A_n \in F\,(n = 1, 2, ...)$ thì $\bigcup\limits_{i=1}^{\infty} A_i \in F$ và $\bigcap\limits_{i=1}^{\infty} A_i \in F$. Khi đó F được gọi là một $\sigma-$đại số.
+ Giả sử trên F xác định một hàm tập P có các tính chất sau:
  - i) $P(A) \geq 0, \forall A \in F$.
  - ii) $P(\Omega) = 1$.
  - iii) Nếu $A_n \in F\,(n = 1, 2, ...)$ và $A_i \cap A_j = \emptyset, \forall i \neq j$ thì

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P\left(A_i\right)$$

  Khi đó P được gọi là độ đo xác suất σ- cộng tính hay nói gọn là xác suất trên F.

+ Bộ ba (Ω, F, P) được gọi là một không gian xác suất. Tập $A \in F$ được gọi là biến cố, $P(A)$ được gọi là xác suất của biến cố A. Để đơn giản ta quy ước khi viết xác suất của biến cố nào đó thì biến cố đó thuộc σ - đại số F.

1.2.2 Một số tính chất của xác suất

a. $0 \le P(A) \le 1$.

b. $P(\Omega) = 1$.

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

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

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

f. Nếu $A \subset B$ thì $P(B \setminus A) = P(B) - P(A)$.

1.3 Công thức tính xác suất

1.3.1 Công thức cộng xác suất

$$P(A \cup B) = P(A) + P(B) - P(AB);$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC);$$
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} A_i - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) + ... + (-1)^{n-1} P(A_1A_2...A_n).$$

Chú ý nếu A, B, C xung khắc thì ta có:

$$P(A \cup B) = P(A) + P(B)$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

1.3.2 Công thức nhân xác suất

$$P(A/B) = \frac{P(AB)}{P(B)};$$
$$P(B/B) = 1;$$
$$P(\overline{A}/B) = 1 - P(A/B);$$
$$P(AB) = P(B)P(A/B) = P(A)P(B/A);$$
$$P(ABC) = P(A)P(B/A)P(C/AB);$$
$$P\left(\prod_{i=1}^{n} A_i\right) = P(A_1).P(A_2/A_1).P(A_3/A_1A_2)...P\left(A_n/\prod_{i=1}^{n-1} A_i\right).$$

Chú ý nếu A, B, C độc lập thì ta có:

$$P(A/B) = P(A)\,; \;\; P(B/A) = P(B)$$
$$P(AB) = P(A)\,P(B)$$
$$P(ABC) = P(A)\,P(B)\,P(C)$$

Chú ý nếu A_1 và A_2 xung khắc thì ta có:

$$P(A_1 \cup A_2/B) = P(A_1/B) + P(A_2/B)$$

1.3.3 Công thức siêu bội

Từ một tập hợp gồm N phần tử trong đó có M phần tử có tính chất A, chọn ngẫu nhiên một tập hợp gồm n phần tử. Gọi A_k là biến cố có k phần tử có tính chất A trong n phần tử lấy ra. Khi đó, ta có

$$P(A_k) = \frac{C_M^k . C_{N-M}^{n-k}}{C_N^n}$$

1.3.4 Công thức xác suất toàn phần

Giả sử $A_1, A_2, ..., A_n$ là nhóm biến cố đầy đủ, A là biến cố tùy ý, ta có

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

1.3.5 Công thức Bayes

Giả sử $A_1, A_2, ..., A_n$ là nhóm biến cố đầy đủ, A là biến cố xảy ra cùng với một trong các biến cố A_i . Khi đó, ta có

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

1.3.6 Công thức Bernoulli

Giả sử ta thực hiện một phép thử nào đó n lần độc lập và giống nhau. Trong mỗi phép thử chỉ có một trong hai khả năng xảy ra. Biến cố A xảy ra với xác suất là p, hoặc biến cố A không xảy ra với xác suất là $q = 1 - p$. Khi đó, xác suất để trong n lần thực hiện phép thử biến cố A xảy ra k lần là

$$P_n(x) = C_n^x p^x (1-p)^{n-x}$$

- Xác suất để không có lần thành công nào là $(1-p)^n$,
- Xác suất để có ít nhất một lần thành công là $1-(1-p)^n$,
- Số thành công có khả năng nhất là $[(n+1)p]$.

Các ví dụ

Ví dụ 1.1 .

- Từ một mẫu gồm có người bệnh và cả người không bệnh, chọn ngẫu nhiên một người để kiểm tra là một phép thử. Chọn được người bệnh hay là người không bệnh là một biến cố.
- Bắn một viên đạn vào một mục tiêu là một phép thử. Viên đạn bắn trúng hay bắn trật là một biến cố.

Ví dụ 1.2 Tung một con xúc xắc cân đối và đồng chất. Ta có

- Biến cố con xúc xắc xuất hiện có số chấm nhỏ hơn hoặc bằng 6 là biến cố chắc chắn.
- Biến cố con xúc xắc xuất hiện có số chấm lớn hơn 6 là biến cố không thể.
- Biến cố con xúc xắc xuất hiện có số chấm chẵn là biến cố ngẫu nhiên.

Ví dụ 1.3 Tung một con xúc xắc cân đối và đồng chất. Gọi:

- A là biến cố con xúc xắc xuất hiện số chấm lẻ.
- A_i là biến cố con xúc xắc xuất hiện có số chấm là $i\,(i=\overline{1,6})$.

Ta có $A_1 \subseteq A; A_3 \subseteq A; A_5 \subseteq A$.

Ví dụ 1.4 Có 3 xạ thủ, mỗi người bắn một viên đạn vào một mục tiêu. Gọi:

- A là biến cố có đúng một người bắn trúng,
- B là biến cố có ít nhất một người bắn trúng,
- C là biến cố có số người bắn trúng lớn hơn hoặc bằng 1.

Khi đó, ta có $A \subseteq B$ và $B = C$.

Ví dụ 1.5 Chúng ta biết rằng một người có huyết áp bị hạ sẽ có triệu chứng tim đập yếu hoặc giãn mạch hoặc cả hai triệu chứng đó. Vì vậy, nếu ta gọi:

- A là biến cố huyết áp bị hạ,

- B là biến cố tim đập yếu,
- C là biến cố giãn mạch.

Thì ta sẽ có sự biểu diễn của biến cố A thông qua biến cố B và C là $A = B+C$.

Ví dụ 1.6 Một kỹ sư nông nghiệp trồng thí điểm một loại giống lúa mới ở 5 địa điểm khác nhau. Gọi:

- A_i là biến cố địa điểm thứ i thành công $(i = \overline{1,5})$.
- $\overline{A_i}$ là biến cố địa điểm thứ i không thành công.
- A là biến cố cả 5 địa điểm thành công.
- B là biến có đúng 1 địa điểm không thành công. Khi đó ta có

$$\begin{aligned} A &= A_1.A_2.A_3.A_4.A_5. \\ B &= \overline{A}_1.A_2.A_3.A_4.A_5 + A_1.\overline{A}_2.A_3.A_4.A_5 \\ &+ A_1.A_2.\overline{A}_3.A_4.A_5 + A_1.A_2.A_3.\overline{A}_4.A_5 + A_1.A_2.A_3.A_4.\overline{A}_5. \end{aligned}$$

.

Ví dụ 1.7 Một kho hàng chứa những sản phẩm do 3 nhà máy sản xuất. chọn ngẫu nhiên một sản phẩm từ kho hàng này. Gọi A_i là biến cố chọn được sản phẩm do nhà máy thứ i, $i = 1, 2, 3$ sản xuất. Khi đó, ta có $\{A_1, A_2, A_3\}$ là nhóm biến cố đầy đủ và xung khắc từng đôi vì $A_1 + A_2 + A_3 = \Omega$ và với $i \neq j$ thì $A_i.A_j = \phi;\ i, j \in \{1, 2, 3\}$.

Ví dụ 1.8 Kiểm tra 5 sản phẩm từ một cơ sở sản xuất Gọi:

- A là biến cố có ít nhất 1 phế phẩm,
- A là biến cố không có phế phẩm nào.

Ta có A và B hai biến cố đối lập nhau: $A = \overline{B}$, $B = \overline{A}$.

Ví dụ 1.9 Có 4 lọ thuốc, mỗi lọ đều chứa những viên thuốc tốt và thuốc xấu. Chọn ngẫu nhiên 1 viên thuốc từ mỗi lọ. Gọi A_i là biến cố chọn được viên thuốc tốt từ lọ thứ $i(i = 1, 2, 3, 4)$. Ta có $\{A_1, A_2, A_3, A_4\}$ là nhóm biến cố độc lập toàn phần.

Ví dụ 1.10 Có 3 hộp đựng sản phẩm. Chọn từ mỗi hộp 1 sản phẩm, gọi A_i là biến cố chọn được phế phẩm ở hộp thứ $i, (i = 1, 2, 3)$. Hãy biểu diễn các biến cố sau qua các biến cố A_i:

a) A: chọn được 3 phế phẩm.

b) B: chọn được đúng 1 phế phẩm.

c) C: chọn được ít nhất 2 phế phẩm.

d) D: chọn được đúng 2 phế phẩm, biết hộp thứ nhất đó chọn được phế phẩm.

Ví dụ 1.11 Một danh sách có 10 sinh viên, trong đó có 4 sinh viên khoa X và 6 sinh viên khoa Y. Chọn ngẫu nhiên từ danh sách này 4 sinh viên. Tính xác suất trong các trường hợp sau:

a) Chọn được số sinh viên khoa X bằng số sinh viên khoa Y.

b) Chọn được ít nhất một sinh viên khoa X.

ĐS: a) $\frac{3}{7}$; b) $\frac{13}{14}$.

Ví dụ 1.12 Một hộp có 7 chính phẩm và 3 phế phẩm.

a) Lấy ngẫu nhiên một hộp để kiểm tra, tính xác suất lấy được phế phẩm

b) Lấy ngẫu nhiên có hoàn lại lần lượt từng sản phẩm ra 2 sản phẩm từ hộp (nghĩa là lấy lần đầu 1 sản phẩm ghi kết quả sau đó trả lại hộp rồi lại lấy lần thứ hai), tính xác suất lấy được 2 phế phẩm.

c) Lấy ngẫu nhiên không hoàn lại lần lượt từng sản phẩm ra 2 sản phẩm từ hộp, tính xác suất lấy được 2 phế phẩm.

d) Lấy ngẫu nhiên 2 sản phẩm từ hộp, tính xác suất lấy được 2 phế phẩm.

ĐS: a) 0,3; b) 0,09; c) 0,0667; d) 0,0667.

Ví dụ 1.13 Để nghiên cứu khả năng xuất hiện mặt sấp khi tung một đồng xu cân đối đồng chất, các nhà bác học đã tiến hành tung đồng xu nhiều lần và được kết quả cho ở bảng sau:

Người làm thí nghiệm	Số lần tung	Số lần xuất hiện mặt sấp	Tần suất $f(A)$
Buyffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

Từ thí nghiệm trên ta thấy khi số phép thử tăng lên thì tần suất xuất hiện mặt sấp của đồng xu tiến dần đến 0,5. Trong trường hợp này ta có thể kết luận xác suất xuất hiện mặt sấp của đồng xu khi tung là 0,5.

Ví dụ 1.14 Chúng ta thường nói khi một bà mẹ sinh một đứa con thì khả năng sinh được con trai và con gái là như nhau và bằng 0,5. Chúng ta xem điều này có đúng không qua các sự kiện thống kê sau:

- Người Trung Hoa từ năm 2228 trước công nguyên đã thống kê qua kinh nghiệm đưa ra tỉ số sinh con gái là 0,5.
- Laplace nghiên cứu sinh đẻ ở Luân Đôn, Petecbua và Berlin trong 10 năm và đưa ra tỉ số con gái là $\frac{21}{43}$.
- Dacnon nghiên cứu sinh đẻ ở Pháp cho các số liệu:

Năm	1086	1816	1836	1856	1913	1920
Tần suất sinh con gái	0,485	0,484	0,485	0,487	0,488	0,489

Từ số liệu thống kê này ta có thể nói khi sinh trong hoàn cảnh tự nhiên thì xác suất sinh con gái của bà mẹ là 0,49.

Ví dụ 1.15 Để tìm xác suất tăng giá của một mã chứng khoán, người ta theo dõi qua 2000 phiên thấy có 600 phiên mã chứng khoáng này tăng giá, khi đó người ta coi xác suất tăng giá của mã chứng khoán này là 600/2000.

Ví dụ 1.16 Hãy tìm tỉ lệ phế phẩm của một lô hàng gồm nhiều sản phẩm. Biết rằng người ta kiểm tra ngẫu nhiên lần lượt 1000 sản phẩm của lô hàng này thấy có 5 phế phẩm.
ĐS: 0,5%.

Ví dụ 1.17 (Bài toán hai người gặp nhau)
Hai người hẹn gặp nhau trong khoảng thời gian từ 7h đến 8h tại một địa điểm nào đó. Họ quy ước rằng nếu người này đến điểm hẹn không gặp người kia sẽ đợi 20 phút. Tính xác suất để 2 người hẹn gặp nhau trong cuộc hẹn này.
ĐS: 5/9.

Ví dụ 1.18 Một hộp có 10 bi trong đó có 4 bi đỏ. Lấy ngẫu nhiên 3 bi từ hộp. Tính xác suất sao cho trong 3 bi lấy ra có ít nhất một bi đỏ.
ĐS: 0,8333.

Ví dụ 1.19 Tỷ lệ phế phẩm của một nhà máy là 10%.

a) Chọn ngẫu nhiên 100 sản phẩm của nhà máy, tính xác suất để chọn được ít nhất một phế phẩm.

b) Phải chọn ít nhất bao nhiêu sản phẩm để có xác suất chọn được ít nhất 1 phế phẩm không bé hơn 0.95.

ĐS: a) $1-(0,9)^{100}$; b) 29.

Ví dụ 1.20 Một lớp có 100 sinh viên, trong đó có 50 sinh viên giỏi toán, 50 sinh viên giỏi văn và trong số này có 15 sinh viên vừa giỏi toán vừa giỏi văn. Chọn ngẫu nhiên 1 sinh viên của lớp. tính xác suất để chọn được sinh viên giỏi ít nhất 1 trong 2 môn trên.
ĐS: 85%.

Ví dụ 1.21 Một lớp học có 60 học sinh trong đó có 28 em giỏi Toán, 30 em giỏi Lý, 32 em giỏi Ngoại ngữ, 15 em vừa giỏi Toán vừa giỏi Lý, 10 em vừa giỏi Lý vừa giỏi Ngoại ngữ, 12 em vừa giỏi Toán vừa giỏi Ngoại ngữ, có 2 em giỏi cả 3 môn. Gọi ngẫu nhiên một học sinh của lớp. Tính xác suất gọi được em giỏi ít nhất 1 môn.
ĐS: 0,9167.

Ví dụ 1.22 Một cửa hàng bán một loại TV trong đó tỉ lệ có chất lượng tiếng bị kém là 5%, tỉ lệ có chất lượng hình bị kém là 7%, tỉ lệ kém chất lượng của cả tiếng và hình là 4%. Có người mua một TV của cửa hàng. Tính xác suất mua được TV không bị mắc cả hai lỗi trên.

Ví dụ 1.23 Một mẫu có 10 người, trong đó có 6 người bị bệnh. Chọn ngẫu nhiên 6 người, tính xác suất để chọn được số người bị bệnh nhiều hơn số người không bị bệnh.
ĐS: 23/42.

Ví dụ 1.24 Bỏ ngẫu nhiên 4 lá thư vào 4 phong bì tương ứng đã ghi sẵn địa chỉ. Tính xác suất sao cho có ít nhất một lá thư bỏ đúng phong bì của nó.
ĐS: 0,625.

Ví dụ 1.25 Một hộp có 10 vé trong đó có 3 vé có thưởng. Tính xác suất người thứ hai bốc được vé trúng thưởng, biết rằng người đầu đã bốc được một vé trúng thưởng (mỗi người chỉ được bốc một vé).
ĐS: 0,2222.

Ví dụ 1.26 Một cuộc điều tra về sở thích mua sắm quần áo của dân cư trong một vùng. Trong số 500 người được điều tra có: 240 nam, trong đó có 136 người thích mua sắm, 260 nữ, trong đó có 224 người thích mua sắm.

a) Tính xác suất người được chọn điều tra ngẫu nhiên là nữ.

b) Giả sử chọn được một người nữ của vùng. Tính xác suất người đó thích mua sắm.

Ví dụ 1.27 Để dập tắt nạn dịch sâu bệnh hại lúa, đội bảo vệ thực vật đã tiến hành phun thuốc 3 lần liên tiếp trong 1 tuần. Xác suất sâu bị chết sau lần

phun thứ nhất là 0,5. Nếu sống sót ở lần phun thứ nhất thì khả năng sâu bị chết ở lần phun thứ hai là 0,7. Nếu sống sót ở lần phun thứ hai thì khả năng sâu bị chết ở lần phun thứ ba là 0,9. Tính xác suất sâu bị chết sau đợt phun thuốc.
ĐS: 98,5%.

Ví dụ 1.28 Một xưởng có 2 máy hoạt động độc lập. trog một ngày làm việc, xác suất để 2 máy này bị hỏng tương ứng là 0,1 ; 0,05. Tính xác suất trong một ngày làm việc, xưởng có:

a) 2 máy hỏng.

b) Một máy hỏng.

c) Có máy hỏng.

Ví dụ 1.29 Một hộp có 9 bi trong đó có 3 bi đỏ. Chia ngẫu nhiên số bi trong hộp lần lượt thành 3 phần, mỗi phần 3 bi. Tính xác suất để mỗi phần đều có bi đỏ.

Ví dụ 1.30 Một sản phẩm xuất xưởng phải qua 3 lần kiểm tra. Xác suất để một phế phẩm bị loại ở lần kiểm tra đầu là 0,8 ; nếu lần kiểm tra đầu không không bị loại thì xác suất nó bị loại ở lần kiểm tra thứ hai là 0,9, tương tự nếu lần thứ hai nó cũng không bị loại thì ở lần kiểm tra thứ ba xác suất nó bị loại là 0,95.

a) Tính xác suất để một phế phẩm bị loại qua 3 lần kiểm tra.

b) Giả sử một phế phẩm bị loại sau 3 lần kiểm tra. Tính xác suất để phế phẩm này bị loại ở lần kiểm tra thứ hai.

Ví dụ 1.31 Một lô hàng có 10 sản phẩm, trong đó có 6 sản phẩm tốt. hai người khách hàng lần lượt đến mua mỗi người một sản phẩm. Hỏi khả năng mua được sản phẩm tốt của mỗi người có giống nhau không, tại sao?
ĐS: như nhau.

Ví dụ 1.32 Có 2 chuồng nuôi chuột. chuồng I có 4 con chuột trắng và 3 con chuột đen, chuồng II có chuột trắng và 5 chuột đen. Chọn ngẫu nhiên 2 con từ chuồng I bỏ vào chuồng II, từ chuồng II chọn ngẫu nhiên 1 con. Tính xác suất để con chuột chọn từ chuồng II có màu trắng.
ĐS: 66/189.

Ví dụ 1.33 Một kho hàng chứa cùng một loại sản phẩm do ba nhà máy sản xuất, biết số sản phẩm của nhà máy I chiếm 2/3 sản phẩm của kho hàng, số sản phẩm của nhà máy II chiếm 1/4 số sản phẩm của kho hàng, số sản phẩm còn lại của nhà máy III. Tỷ lệ sản phẩm tốt của mỗi nhà máy lần lượt là 80%,

60% và 40%. Hỏi tỷ lệ sản phẩm tốt của nhà máy là bao nhiêu?
ĐS: 72%.

Ví dụ 1.34 Một mạch điện gồm 2 bộ phận độc lập được mắc nối tiếp, với xác suất bị hỏng trong thời gian nào đó của mỗi bộ phần lần lượt là 0,01 và 0,015. Tại một thời điểm người ta thấy mạch điện bị ngừng làm việc (do ít nhất một bộ phận nào đó bị hỏng). Tìm xác suất để để bộ phận thứ nhất hỏng.
ĐS: 0,396378.

Ví dụ 1.35 Qua thống kê thực tế người ta thấy rằng: tỷ lệ người viêm họng trong số người nghiện thuốc lá là 60% và trong số người không hút thuốc lá là 40%. Giả sử một vùng dân cư hiện có 30% người nghiện thuốc.

a) Tính tỷ lệ người bị viêm họng của vùng dân cư này?

b) Nếu chọn được người không bị viêm họng từ vùng này, tính xác suất để người này là người nghiện thuốc lá.

ĐS: a) 46%; b) 22,2%.

Ví dụ 1.36 Sau đây là dãy các phép thử Bernoulli.

a) Một máy sản xuất lần lượt từng sản phẩm. Xác suất mỗi lần máy sản xuất ra phế phẩm là 0,08. Cho máy sản xuất 15 sản phẩm (15 phép thử Bernoulli).

b) Một hộp có 10 sản phẩm trong đó có 3 phế phẩm. Lấy ngẫu nhiên có hoàn lại lần lượt từng sản phẩm ra 100 sản phẩm (100 phép thử Bernoulli).

c) Xác suất để một mã chứng khoáng tăng giá trong một phiên giao dịch là 10%. Xét 10 phiên giao dịch liên tiếp của mã chứng khoán này (10 phép thử Bernoulli).

Ví dụ 1.37 Một máy sản xuất lần lượt từng sản phẩm. Xác suất mỗi lần máy sản xuất ra sản phẩm loại A là 0,7.

a) Cho máy sản xuất 10 sản phẩm. tính xác suất để trong các sản phẩm này có 2 sản phẩm loại A.

b) Máy cần sản xuất ít nhất bao nhiêu sản phẩm để xác suất của biến cố "trong các sản phẩm máy sản xuất ra có ít nhất một sản phẩm loại A" lớn hơn 0,85.

Ví dụ 1.38 Người ta thống kê tỷ lệ sâu răng ở hai trường tiểu học A và B trong một huyện lần lượt là 20% và 30%. Chọn ngẫu nhiên từ mỗi trường 2 học sinh. Tính xác suất để chọn được đúng 2 học sinh bị sâu răng.
ĐS: 0,362.

Ví dụ 1.39 Một công ty bảo hiểm chia dân cư (đối tượng bảo hiểm) làm 3 loại ít rủi ro; rủi ro trung bình; rủi ro cao. Kinh nghiệm cho thấy tỉ lệ dân gặp rủi ro trong một năm tương ứng với các loại trên là : 0,05 ; 0,08 ; 0,12 và có 60% dân cư thuộc loại rủi ro ; 30% dân cư thuộc loại rủi ro trung bình và 10% dân cư thuộc loại rủi ro cao.

a) Tìm tỉ lệ dân cư gặp rủi ro sau một năm cố định nào đó của vùng.

b) Nếu một người gặp rủi ro trong năm thì khả năng người đó thuộc loại ít rủi ro là bao nhiêu?

Ví dụ 1.40 Có 2 hộp sản phẩm, hộp thứ nhất có 10 sản phẩm, trong đó có 3 phế phẩm ; hộp thứ hai có 12 sản phẩm, trong đó có 4 phế phẩm. Một khách hàng lấy ngẫu nhiên một hộp, từ đó lấy ngẫu nhiên ra 2 sản phẩm để kiểm tra, nếu toàn là chính phẩm thì mua hộp đó.

a) Tính xác suất khách hàng mua hộp sản phẩm.

b) Giả sử khách hàng đã mua hộp sản phẩm. Khả năng khách hàng mua hộp nào nhiều hơn ?

Ví dụ 1.41 Một thiết bị gồm 3 loại linh kiên loại 1, 2, 3. Chúng chiếm tương ứng 35%, 25%, 40% tổng số linh kiện của thiết bị. Tỉ lệ thiết bị bị hỏng tương ứng là 15%, 25%, 5%. Thiết bị đang hoạt động bỗng nhiên bị hỏng. theo bạn, nhiều khả năng nhất linh kiện nào bị hỏng (giả sử các linh kiện không hỏng đồng thời)?

1.4 Bài tập có hướng dẫn

Bài 1.1 Huy động vốn dự án tạo nên sự gia tăng lớn về nguồn tài chính cho các công ty trong những năm gần đây. Theo Venture Economics (*Investor's Business Daily,* 28, 4, 2000) có đến 2374 giải ngân vốn dự án được thực hiện vào năm 1999. Trong đó, 1434 là các công ty ở California, 390 là các công ty ở Massachusetts và 217 là các công ty ở Colorado. Trong số đó, 20% các công ty nhận được nguồn tài chính vào đầu giai đoạn phát triển và 55% công ty vào giai đoạn mở rộng. Giả sử rằng bạn muốn chọn ngẫu nhiên một trong số các công ty đó để biết họ sử dụng những nguồn vốn đó như thế nào.

a. Tính xác suất công ty được chọn sẽ đến từ California?

b. Tính xác suất công ty được chọn sẽ không đến từ một trong bốn đề cập trên?

c. Tính xác suất công ty sẽ không ở trong giai đoạn đầu phát triển?

d. Giả sử rằng các công ty ở giai đoạn đầu phát triển phân phối đều ở các đất nước, hãy cho biết có bao nhiêu công ty Massachusetts nhận được nguồn tài chính từ vốn dự án ở giai đoạn đầu phát triển?

e. Tổng lượng nguồn tài chính đầu tư là 32,4 tỉ USD. Hãy ước lượng con số này mà Colorado nhận được?

Giải.

a. Gọi A biến cố "chọn công ty đến từ California", xác suất công ty được chọn sẽ đến từ California là

$$P(A) = \frac{1434}{2374} \approx 0,6040.$$

b. Gọi B là biến cố chọn công ty không đến từ bốn bốn đề cặp trên, xác suất tương tứng được tính như sau

$$P(B) = \frac{2374 - 1434 - 390 - 217 - 112}{2374} = \frac{221}{2374} \approx 0,0931.$$

c. Gọi C là biến cố công ty được chọn không ở trong giai đoạn đầu phát triển, xác suất tương ứng là

$$P(C) = 1 - 0,22 = 0,78.$$

d. Số công ty Massachusetts là 390, tỉ lệ công ty nhận được nguồn tài chính ở giai đoạn đầu phát triển bằng tỉ lệ chung là 0,22. Do đó, số công ty Massachusetts nhận được nguồn tài chính từ vốn dự án ở giai đoạn đầu phát triển là 0,22.390=85,8.

e. Tổng lượng nguồn tài chính đầu tư là 32,4 tỉ USD. Ước lượng con số này mà Colorado nhận được là $\dfrac{32.400.000.000}{2374} \times 112 = 1.528.559.393$USD.

Bài 1.2 Cho bảng phân phối phí hàng tháng của một mẫu gồm 100 khách hàng của hãng ga và điện Montana như sau

Phí phải trả (USD)	0-49	50-99	100-149	150-199	200-249
Số khách hàng	13	22	34	26	5

a. Đặt A là biến cố phí phải trả của khách hàng bằng hoặc lớn hơn 150 USD. Tính P(A)?

b. Đặt B là biến cố phí phải trả của khách hàng nhỏ hơn 150 USD. Tính P(B)?

Giải.

a. Số khách hàng có phí phải đóng lớn hơn hoặc bằng 150 USD là 26+5=31. Từ đó tính P(A)
$$P(A) = \frac{31}{100} = 0,31.$$

b. Xác suất phí phải trả của khách hàng thấp hơn 150 USD là
$$P(B) = 1 - P(A) = 1 - 0,31 = 0,69.$$

Bài 1.3 Dữ liệu gồm 30 quỹ đầu tư trái phiếu quy định thời gian đáo hạn 1 năm và 5 năm kì hạn cuối là 31 tháng 3 năm 2000 (*Theo tờ The Wall Street Journal*, ngày 10 tháng 4 năm 2000). Giả sử chúng ta chỉ xét trái phiếu có đáo hạn 1 năm với lãi suất cao nhất là 2% và đáo 5 năm với lãi suất cao nhất là 44%. Biết rằng một nửa quỹ đầu tư trái phiếu có đáo hạn 1 năm với lãi suất cao nhất 2%, 12 quỹ đầu tư trái phiếu có thời gian đáo hạn 5 năm với lãi suất 44% và 6 quỹ đầu tư trái phiếu có thời gian đáo hạn cả 1 năm và 5 năm.

a. Tính xác suất của một quỹ đầu tư trái phiếu có thời gian đáo hạn là 1 năm, xác suất của một quỹ đầu tư trái phiếu có thời gian đáo hạn là 5 năm và xác suất của một quỹ đầu tư trái phiếu có thời gian đáo hạn cả 1 năm và 5 năm?

b. Tính xác suất của một quỹ đầu tư trái phiếu hoặc có thời gian đáo hạn 1 năm hoặc có thời gian đáo hạn 5 năm hoặc có thời gian đáo hạn cả 1 năm và 5 năm

c. Tính xác suất của một quỹ đầu tư trái phiếu có thời gian đáo hạn không phải 1 năm cũng không phải là 5 năm?

Giải. Theo dữ liệu đã cho số quỹ đầu trái phiếu có thời hạn 1 năm là $30 : 2 = 15$ trong số này có $15 - 6 = 9$ quỹ đầu tư cổ phiếu chỉ có thời gian đáo hạn là 1 năm. Số quỹ đầu tư trái phiếu chỉ có thời hạn 5 năm là $12 - 6 = 6$. Từ đó ta có thể tính các xác suất sau.

a. Gọi A, B và C tương ứng là biến cố cho thông tin một quỹ đầu tư trái phiếu có thời gian đáo hạn là 1 năm, 5 năm và cả 1 năm và 5 năm. Tính các xác suất tương ứng:
$$P(A) = \frac{15}{30} = 0,5;\ P(B) = \frac{12}{30} = 0,4 \text{ và } P(C) = \frac{6}{30} = 0,2.$$

b. Gọi D là biến ngẫu nhiên chỉ một quỹ đầu tư trái phiếu hoặc có thời gian đáo hạn 1 năm hoặc có thời gian đáo hạn 5 năm hoặc có thời gian đáo hạn cả 1 năm và 5 năm
$$P(D) = \frac{9+6+6}{30} = 0,7.$$

c. Xác suất của một quỹ đầu tư trái phiếu có thời gian đáo hạn không phải 1 năm cũng không phải là 5 năm

$$P(\overline{D}) = 1 - P(D) = 1 - 0,7 = 0,3.$$

Bài 1.4 Gọi A là biến cố chỉ phương tiện giao thông chính để con người đi làm việc là ô tô và B là sự kiện mà phương tiện giao thông chính là xe buýt. Ta giả thiết rằng $P(A) = 0,45$ và $P(B) = 0,35$.

a. Hai sự kiện A và B có rời nhau không? Tính xác suất để một người sử dụng hoặc ô tô hoặc xe buýt để đi làm việc?

b. Tính xác suất phương tiện đi làm việc chính của con người là một phương tiện khác ngoài xe buýt?

Giải.

a. Hai sự kiện A và B là rời nhau vì trong một thời gian một người không thể nào vừa đi ô tô lại vừa đi xe buýt. Xác suất để một người sử dụng hoặc ô tô hoặc xe buýt để đi làm việc là

$$P(A + B) = P(A) + P(B) = 0,8.$$

b. Xác suất phương tiện đi làm việc chính của con người là một phương tiện khác ngoài xe buýt

$$P(\overline{B}) = 1 - P(B) = 0,65.$$

Bài 1.5 Cho hai biến cố A và B, biết $P(A) = 0,5$, $P(B) = 0,6$ và $P(AB) = 0,4$.

a. Tính $P(A|B)$?

b. Tính $P(B|A)$?

c. Hai sự kiện A và B có độc lập với nhau không? Giải thích?

Giải.

a. Theo công thức xác suất có điều kiện ta có

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{0,4}{0,6} = \frac{2}{3}.$$

b. Tương tự câu a, ta có

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{0,4}{0,5} = 0,8.$$

c. Hai sự kiện A và B không độc lập với nhau không vì

$$P(AB) = 0,4 > P(A).P(B) = 0,3.$$

Bài 1.6 Kiểm tra một số tài khoản của Sun Bank, phân loại theo thời gian hoạt động của tài khoản và số dư tài khoản. Kiểm toán viên sẽ chọn ngẫu nhiên một số tài khoản từ 1000 tài khoản được cho trong bảng sau

Độ tuổi	Số tài khoản (USD)		
	0-99	500-999	≥1000
Nhỏ hơn 2 năm	120	240	90
Từ 2 năm trở lên	75	275	200

a. Tính xác suất một tài khoản có thời gian hoạt động thấp hơn 2 năm?

b. Tính xác suất một tài khoản có số dư từ 1000 USD trở lên?

c. Tính xác suất hai tài khoản mà cả hai đều có số dư từ 1000 USD trở lên?

d. Tính xác suất một tài khoản có số dư từ 500 - 999 (USD) biết rằng tài khoản đó hoạt động từ 2 năm trở lên?

e. Tính xác suất một tài khoản hoạt động chưa đến 2 năm và có số dư từ 1000 USD trở lên?

f. Tính xác suất một tài khoản hoạt động ít nhất 2 năm với điều kiện có số dư là 500-999 USD?

Giải.

a. Gọi A là biến cố một tài khoản có thời gian hoạt động thấp hơn 2 năm

$$P(A) = \frac{120 + 240 + 90}{1000} = 0,45.$$

b. Gọi B là biến cố một tài khoản có số dư từ 1000 USD trở lên

$$P(B) = \frac{200 + 90}{1000} = 0,29.$$

c. Gọi C là biến cố hai tài khoản mà cả hai đều có số dư từ 1000 USD trở lên

$$P(C) = \frac{C_{290}^2}{C_{1000}^2} \approx 0,0839.$$

d. Số tài khoản hoạt động từ 2 năm trở lên là $75 + 275 + 200 = 550$. Gọi D là biến cố một tài khoản có số dư 500 - 999 (USD) biết tài khoản có thời gian hoạt động từ 2 năm trở lên

$$P(D) = \frac{275}{550} = 0,5.$$

e. Gọi E là biến cố một tài khoản hoạt động chưa đến 2 năm và có số dư từ 1000 USD trở lên

$$P(E) = \frac{90}{1000} = 0,09.$$

f. Gọi F là biến cố một tài khoản hoạt động ít nhất 2 năm với điều kiện có số dư là 500-999 USD

$$P(F) = \frac{275}{240 + 275} = \frac{55}{103} \approx 0,533981.$$

Bài 1.7 Để đánh giá về chương trình chương trình huấn luyện bán hàng, một công ty nhận thấy rằng 50 người bán hàng nhận tiền thưởng năm rồi, 20 người trong số đó đã tham gia chương trình huấn luyện đặc biệt. Tổng số nhân viên bán hàng của công ty là 200 người. Gọi B là biến cố người bán hàng nhận được tiền thưởng và S là biến cố người bán hàng có tham gia chương trình huấn luyện bán hàng.

a. Tìm $P(B), P(S|B)$ và $P(SB)$.

b. Giả sử có 40% người tham gia chương trình huấn luyện bán hàng. Tính xác suất để 1 người bán hàng nhận được tiền thưởng nếu người đó có tham gia chương trình huấn luyện, $P(B|S)$?

c. Nếu công ty đánh giá chương trình huấn luyện dựa trên tác động của nó đến khả năng một người bán hàng nhận được tiền thưởng. Bạn đánh giá chương trình huấn luyện này như thế nào? Nhận xét về sự độc lập, phụ thuộc của hai biến ngẫu nhiên S và B?

Giải.

a. Xác suất để người bán hàng nhận được tiền thưởng là

+ $P(B) = \dfrac{50}{200} = 0,25.$
+ $P(S|B) = \frac{20}{50} = 0,4.$
+ $P(SB) = P(B).P(S|B) = 0,25.0,4 = 0,1.$

b. Xác suất để 1 người bán hàng nhận được tiền thưởng nếu người đó có tham gia chương trình huấn luyện

$$P(B|S) = \frac{P(BS)}{P(S)} = \frac{0,1}{0,4} = 0,25.$$

c. Từ câu b ta thấy xác suất để 1 người bán hàng nhận được tiền thưởng nếu người đó có tham gia chương trình huấn luyện là $P(B|S) = 25\%$ bằng với xác suất 1 người bán hàng bất kì nhận được tiền thưởng là $P(B) = 25\%$. Từ đó kết luận chương trình huấn luyện bán hàng không tác động đến khả năng nhận tiền thưởng của người bán hàng.
Nhận xét về sự độc lập, phụ thuộc của hai biến ngẫu nhiên S và B, ta có $P(BS) = 0,1$, $P(B) = 0,25$, $P(S) = 0,4$ suy ra $P(BS) = P(B).P(S)$. Từ đó suy ra S và B là độc lập.

Bài 1.8 Một công ty nghiên cứu về số tai nạn làm mất thời gian xảy ra tại nhà máy Brownsville, Texas. Thống kê cho thấy có 6% người lao động gặp tai nạn làm mất thời gian vào năm vừa rồi. Nhà quản lý tin rằng, với một chương trình huấn luyện đặc biệt về an toàn sẽ làm giảm tai nạn xuống còn 5% trong năm nay. Hơn nữa, ước lượng rằng 15% người lao động gặp tai nạn mất thời gian năm vừa rồi sẽ có tai nạn mất thời gian vào năm nay.

a. Xác định tỉ lệ người lao động sẽ gặp tai nạn mất thời gian trong cả 2 năm?

b. Xác định tỉ lệ người lao động sẽ gặp ít nhất 1 lần tai nạn mất thời gian trong 2 năm?

Giải.
Gọi A, B lần lượt là biến cố một người lao động gặp tai nạn năm ngoái, gặp tai nạn năm nay. Theo giả thiết ta có $P(A) = 0,06$ và $P(B) = 0,05$. Xác suất người lao động gặp tai nạn năm nay với điều kiện người đó gặp tai nạn năm ngoái là $P(B|A) = 0,15$.

a. Xác suất một người lao động gặp tai nạn hai năm là

$$P(AB) = P(A).P(B|A) = 0,06.0,15 = 0,009.$$

Vậy tỉ lệ người lao động gặp tai nạn cả hai năm là 0,9%.

b. Xác suất một người lao động gặp tai nạn ít nhất một lần trong hai năm là

$$P(A+B) = P(A) + P(B) - P(AB) = 0,06 + 0,05 - 0,009 = 0,101$$

Vậy tỉ lệ người lao động gặp ít nhất 1 lần tai nạn mất thời gian trong 2 năm là 10,1%.

Bài 1.9 Xác suất tiên nghiệm của biến cố A_1, A_2 và A_3 lần lượt là $P(A_1) = 0,2; P(A_2) = 0,5; P(A_3) = 0,3$. Xác suất điều kiện của biến cố B đối với biến cố A_1, A_2, A_3 lần lượt là $P(B|A_1) = 0,5; P(B|A_2) = 0,4; P(B|A_3) = 0,3$.

a. Tính $P(BA_1); P(BA_2)$ và $P(BA_3)$.

b. Vận dụng định lý Bayes, hãy tính xác suất hậu nghiệm $P(A_2|B)$?

c. Sử dụng định lý Bayes bằng cách lập bảng để tính $P(A_1|B); P(A_2|B); P(A_3|B)$?

Giải.

a. Các xác suất được tính lần lượt

$$P(B \cap A_1) = P(A_1).P(B|A_1) = 0,2.0,5 = 0,1.$$

$$P(B \cap A_2) = P(A_2).P(B|A_2) = 0,5.0,4 = 0,2.$$

$$P(B \cap A_3) = P(A_3).P(B|A_3) = 0,3.0,3 = 0,09.$$

b. Vận dụng định lý Bayes, ta có

$$\begin{aligned} P(A_2|B) &= \frac{P(A_2).P(B|A_2}{P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + P(A_3).P(B|A_3)} \\ &= \frac{0,2}{0,1+0,2+0,09} = \frac{20}{39} \approx 0,5128. \end{aligned}$$

c. Vận dụng định lý Bayes bằng cách lập bảng

A_i	$P(A_i)$	$P(B\|A_i)$	$P(A_iB)$	$P(A_i\|B)$
A_1	$0,2$	$0,5$	$0,1$	$0,1 : 0,39 \approx 0,2564$
A_2	$0,5$	$0,4$	$0,2$	$0,2 : 0,39 \approx 0,5128$
A_3	$0,3$	$0,3$	$0,09$	$0,09 : 0,39 \approx 0,2308$
			$P(B) = 0,39$	

Bài 1.10 Một công ty dầu chọn mua một vùng đất ở Alaska. Gọi A là biến cố tìm thấy dầu chất lượng cao, B là sự kiện tìm thấy dầu chất lượng trung bình, C là sự kiện không tìm thấy dầu. Nghiên cứu sơ bộ ban đầu về địa chất tại đó cho kết quả với xác suất tiên nghiệm như sau: $P(A) = 0,5, P(B) = 0,2, P(C) = 0,3$.

a. Xác suất tìm thấy dầu là bao nhiêu?

b. Sau khi khoan sâu khoảng 61m đầu tiên, thực hiện kiểm tra đất. Xác suất tìm thấy loại đất đặc biệt tương ứng với loại dầu qua kiểm tra là

$$P(D|A) = 0,2; P(D|B) = 0,2; P(D|C) = 0,3$$

Với D là biến cố chỉ một loại đất.
Công ty nhận xét như thế nào về đất kiểm tra? Tính các xác suất $P(A|D)$, $P(B|D)$, $P(C|D)$ và xác suất tìm thấy dầu mới.

Giải.

a. Xác suất tìm thấy dầu là $P(A) + P(B) = 0,5 + 0,2 = 0,7$.

b. Ta có tính xác suất $P(D)$, áp dụng công thức Bayes từ hệ đầy đủ các biến cố A, B, C

$$\begin{aligned} P(D) &= P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) \\ &= 0,5.0,2 + 0,2.0,2 + 0,3.0,3 = 0,23. \end{aligned}$$

Tính các xác suất $P(A|D)$, $P(B|D)$, $P(C|D)$

$$P(A|D) = \frac{P(A).P(D|A)}{P(D)} = \frac{0,5.0,2}{0,23} = \frac{10}{23} \approx 0,4348.$$

$$P(B|D) = \frac{P(B).P(D|B)}{P(D)} = \frac{0,2.0,2}{0,23} = \frac{4}{23} \approx 0,1739.$$

$$P(C|D) = \frac{P(C).P(D|C)}{P(D)} = \frac{0,3.0,3}{0,23} = \frac{9}{23} \approx 0,3913.$$

Nhận xét: Xác suất tìm thấy dầu mới (Sau khi khoan sâu khoảng 61m đầu tiên) là $\frac{10}{23} + \frac{4}{23} = \frac{14}{23} \approx 0,6087$. Như vậy, khả năng tìm thấy dầu sau khi khoan sâu 61m là 60,87% thấp hơn xác suất tìm thấy dầu lúc ban đầu là 70%. Xác suất dầu cao ban đầu là 50% giảm xuống còn 42,48% và xác suất dầu trung bình ban đầu là 20% cũng giảm xuống còn 17,39%.

Bài 1.11. Gieo hai con xúc xắc cân đối và đồng chất. Tính $P(A)$, nếu A là biến cố ngẫu nhiên chỉ tổng số điểm trên hai con xúc xắc chia hết cho 5.

Giải

Không gian các biến cố sơ cấp

$$\Omega = \left\{(i, j) / i = \overline{1,6}; j = \overline{1,6}\right\}$$

gồm 36 biến cố sơ cấp. Giả sử các biến cố sơ cấp là đồng khả năng. Khi đó, nếu A là biến cố ngẫu nhiên chỉ tổng số điểm chia hết cho 5, tức là

$$A = \{(1,4); (4,1); (2,3); (3,2); (5;5); (4,6); (6,4)\}$$

thì ta có

$$P(A) = \frac{7}{36}.$$

Bài 1.12. Một khối lập phương có các mặt quét sơn được chia thành 1000 khối lập phương con đều nhau. Trộn đều và rút ngẫu nhiên một khối. Tính xác suất:

a) Rút được khối có hai mặt đã quét sơn.

b) Rút được khối có ba mặt đã quét sơn.

Giải

Dễ thấy, không gian các biến cố sơ cấp có 1000 phần tử. Để chia khối lập phương thành 1000 khối con đều nhau thì mỗi cạnh phải chia 10 đoạn bằng nhau. Khi đó có $8.12 = 96$ khối lập phương con có hai mặt quét sơn và có 8 khối lập phương con có ba mặt quét sơn.

a) Gọi A là biến cố rút được khối có 2 mặt quét sơn, khi đó

$$P(A) = \frac{96}{1000} = 0,096.$$

b)Gọi B là biến cố rút được khối có 3 mặt quét sơn, khi đó

$$P(B) = \frac{8}{1000} = 0,008.$$

Bài 1.13. Trong 100 vé xổ số có 10 vé trúng thưởng. Một người mua 5 vé. Tính xác suất:

a) Có 3 vé trúng thưởng.

b) Có 5 vé trúng thưởng.

c) Có ít nhất 1 vé trúng thưởng.

Giải

a) Xác suất để có 3 vé trúng thưởng là

$$p_a = \frac{C_{10}^3.C_{90}^2}{C_{100}^5}.$$

b) a) Xác suất để có 5 vé trúng thưởng là

$$p_b = \frac{C_{10}^5}{C_{100}^5}.$$

c) Xác suất để có ít nhất một vé trúng thưởng là

$$p_c = 1 - \frac{C_{90}^5}{C_{100}^5}.$$

Bài 1.14. Giả sử A và B là hai biến cố và $A \subset B$. Chứng minh rằng hoặc $P(A) = 0$ hoặc $P(B) = 1$.

Giải

Do $A \subset B$ nên $A \cap B = A$. Từ giả thiết A và B độc lập, ta có

$$P(A \cap B) = P(A) = P(A).P(B),$$

suy ra $P(A)[1 - P(B)] = 0$ hay $P(A) = 0$ hoặc $P(B) = 1$.

Bài 1.15. Một hộp chứa 3 bi trắng, 7 bi đỏ và 15 bi xanh. Một hộp khác chứa 10 bi trắng, 6 bi đỏ và 9 bi xanh. Lấy ngẫu nhiên từ mỗi hộp một bi. Tìm xác suất để hai bi rút ra cùng màu.

Giải

Gọi A là biến cố hai bi rút ra cùng màu, khi đó

$$P(A) = \frac{3}{25}\frac{10}{25} + \frac{7}{25}\frac{6}{25} + \frac{15}{25}\frac{9}{25} = \frac{207}{625}.$$

Bài 1.16. Trong một vùng dân cư tỉ lệ người mắc bệnh tim là 0,09; mắc bệnh khớp là 0,12 và mắc cả hai bệnh là 0,07. Khám ngẫu nhiên một người trong vùng đó, tính xác suất để người đó không mắc cả bệnh tim và bệnh khớp.

Giải

Gọi A và B lần lượt là các biến cố người dân mắc bệnh tim và bệnh khớp. Ta có $P(A) = 0,09$ và $P(B) = 0,12$. Khi đó, xác suất để người dân không mắc cả 2 bệnh tim và khớp là

$$\begin{aligned} P\left(\overline{A \cup B}\right) &= 1 - P\left(A \cup B\right) \\ &= 1 - P\left(A\right) - P\left(B\right) + P\left(A \cap B\right) \\ &= 1 - 0,09 - 0,12 + 0,07 \\ &= 0,86. \end{aligned}$$

Bài 1.17. Một sinh viên khi vào thi môn xác suất thống kê chỉ thuộc 12 trong số 20 câu hỏi thi. Tính xác suất để sinh viên đó trả lời được cả 4 câu hỏi trong phiếu thi mà anh ta phải trả lời.

Giải

Gọi A là biến cố sinh viên trả lời được cả 4 câu hỏi trong phiếu thi mà anh ta phải trả lời, khi đó

$$P\left(A\right) = \frac{C_{12}^4}{C_{20}^4} \approx 0,102.$$

Bài 1.18. Ba người chơi trò gieo đồng xu liên tiếp, người nào gieo được mặt sấp thì thắng. Tính xác suất thắng của từng người.

Giải

Gọi xác suất thắng của từng người là: p_1, p_2, p_3. Khi đó ta có $p_1 + p_2 + p_3 = 1$. Dễ thấy do người thứ nhất được gieo trước nên $p_1 = 2p_2 = 4p_3$. Khi đó, $p_1 = \frac{4}{7}; p_2 = \frac{2}{7}; p_3 = \frac{1}{7}$.

Bài 1.19. Một hộp có 100 tấm thẻ như nhau được đánh số từ 1 tới 100. Rút ngẫu nhiên hai thẻ rồi đặt theo thứ tự từ trái qua phải. Tính xác suất để:

a) Rút được hai thẻ lập nên một số có hai chữ số.

b) Rút được hai thẻ lập nên một số chia hết cho 5

Giải

a) Gọi A là biến cố hai thẻ rút được có hai chữ số. Số biến cố sơ cấp thuận lợi cho biến cố A là $k = A_9^2$ và biến cố sơ cấp đồng khả năng của không gian Ω là $n = A_{100}^2$. Khi đó,

$$P(A) = \frac{k}{n} = \frac{A_9^2}{A_{100}^2} \approx 0,0073.$$

b) Gọi B là biến cố hai thẻ rút được tạo thành một số chia hết cho 5, khi đó để có biến cố sơ cấp thuận lợi cho biến cố B, ta rút thẻ thứ hai một cách tùy ý trong 20 thẻ mang các số 5, 10, 15, 20,..., 95, 100 và rút 1 trong 99 thẻ còn lại đặt vào vị trí đầu. Do đó, $k = 99.20$. Ngoài ra $n = A_{100}^2$. Khi đó

$$P(B) = \frac{k}{n} = \frac{99.20}{A_{100}^2} = 0,2.$$

Bài 1.20. Xác suất tiêu thụ điện năng trong mỗi ngày không vượt quá mức quy định ở một nhà máy là $p = 0,75$. Tính xác suất trong 6 ngày liên tiếp có 4 ngày lượng điện không vượt mức quy định.

Giải

Từ bài toán ta nhận được $n = 6$, $k = 4$, $p = 0,75$, $q = 0,25$. Khi đó,

$$P_6(4) = C_6^4(0,75)^4(0,25)^2 = 0,3.$$

Bài 1.21. Một người bắn liên tiếp 5 viên đạn vào mục tiêu với xác suất trúng đích của mỗi viên đạn là 0,2. Để phá hủy mục tiêu thì cần từ 3 viên trúng mục tiêu trở lên. Tính xác suất để mục tiêu bị tiêu diệt.

Giải

Theo giả thiết ta có $n = 5$, $p = 0,2$. Khi đó

$$P_5(3;5) = P_5(3) + P_5(4) + P_5(5) \approx 0,0579.$$

Bài 1.22. Trong một vùng dân cư có 20% người dân mắc một loại bệnh truyền nhiễm. Khám ngẫu nhiên 500 người. Tính số người mắc bệnh có khả năng lớn nhất và xác suất tương ứng.

Giải

Ta có

$$(n+1)p = (500+1).0,2 = 100,2.$$

nên số người dân có khả năng mắc bệnh lớn nhất là 100, với xác suất tương ứng là

$$P_{500}(100) = C_{500}^{100}(0,2)^{100}(0,8)^{400}.$$

Bài 1.23. Tung hai lần một con xúc xắc cân đối. Giả sử A là biến cố tổng số điểm hai lần tung bé hơn 4, còn B là biến cố số điểm lần tung đầu tiên là 1. Hãy tính $P(A/B)$.

Giải

+ Không gian các biến cố sơ cấp

$$\Omega = \left\{(i,j)\,/i = \overline{1,6}; j = \overline{1,6}\right\}$$

+ Các biến cố sơ cấp thuận lợi cho biến cố A là (1,1); (1,2); (2,1).

+ Các biến cố sơ cấp thuận lợi cho biến cố B là (1,1); (1,2); (1,3); (1,4); (1,5); (1,6).

+ Các biến cố sơ cấp thuận lợi cho biến cố $A \cap B$ là (1,1); (1,2).

Vậy

$$P\left(A/B\right) = \frac{P\left(A \cap B\right)}{P\left(B\right)} = \frac{2/36}{3/36} = \frac{2}{3}.$$

Bài 1.24. Có 10 khách bước ngẫu nhiên vào một cửa hàng có 3 quầy. Tìm xác suất để có 3 người đến quầy số 1.

Giải

Theo giả thiết bài toán ta có thể xem việc mỗi hành khách đến các quầy là một giai đoạn. Khi đó ta có 10 giai đoạn, mỗi giai đoạn có 3 cách. Theo quy tắc nhân ta có số trường hợp đồng khả năng là $n = 3^{10}$.

Gọi A là biến cố có 3 người đến quầy số 1. Khi đó số trường hợp thuận lợi cho A là số cách chọn ngẫu nhiên 3 người từ 10 người để xếp vào quầy số 1 và 7 người còn lại sẽ đến ngẫu nhiên vào quầy 2 và quầy 3. Tức là $m = C_{10}^3.2^7$.

Vậy

$$P\left(A\right) = \frac{C_{10}^3.2^7}{3^{10}} \approx 0,2601.$$

Bài 1.25. Đoạn AB có độ dài l bị bẻ gãy làm 3 đoạn tại 2 điểm ngẫu nhiên. Tính xác suất để 3 đoạn này ghép lại thành một tam giác.

Giải

Giả sử x và y là độ dài các đoạn thứ nhất và thứ hai. Khi đó, độ dài đoạn còn lại là $l - x - y$. Không gian biến cố sơ cấp

$$\Omega = \left\{(x,y) \in R^2 : 0 < x < l; 0 < y < l; 0 < x + y < l\right\}.$$

Dựa vào tính chất của tam giác: tổng 2 cạnh không nhỏ hơn cạnh thứ ba, ta có biến cố để ba đoạn tạo thành tam giác sẽ là

$$C = \left\{(x,y) \in R^2 : 0 < x < \frac{l}{2}; 0 < y < \frac{l}{2}; x + y > \frac{l}{2}\right\}.$$

Bằng hình học dễ thấy tỉ lệ giữa diện tích hình C và Ω là $\frac{1}{4}$. Điều này dẫn tới $P(C) = \frac{1}{4}$.

Bài 1.26. Tìm xác suất của điểm M rơi vào hình tròn nội tiếp tam giác đều cạnh $a = 2cm$.

Giải

Gọi A là biến cố điểm M rơi vào hình tròn nội tiếp tam giác đều cạnh $a = 2cm$. Gọi h là đường cao trong tam giác đều, khi đó $h = a.\frac{\sqrt{3}}{2} = \sqrt{3}$

+ Ta có diện tích của tam giác đều đã cho là

$$mes\,(\Omega) = \frac{1}{2}ah = \sqrt{3}$$

+ Ta có diện tích hình tròn đã cho là

$$mes\,(A) = \pi.R^2 = \frac{1}{9}\pi h^2 = \frac{\pi}{3}.$$

Vậy theo công thức xác suất hình học ta có

$$P\,(A) = \frac{mes\,(A)}{mes\,(\Omega)} = \frac{\pi/3}{\sqrt{3}} \approx 0,605.$$

Bài 1.27. Hai số thực a và b được chọn ngẫu nhiên sao cho $-2 \leq b \leq 0$ và $0 \leq a \leq 3$. Tính xác suất để khoảng cách giữa a và b lớn hơn 3.

Giải

Không gian mẫu của phép thử: $\{(a, b) \in \mathbb{R}^2 : -2 \leq b \leq 0 \,\text{và}\, 0 \leq a \leq 3\}$ được biểu diễn trong mặt phẳng tọa độ bởi hình chữ nhật (G) có 2 đỉnh đối diện là (0,0) và (3,-2). (G) có diện tích $mes(G) = 2.3 = 6$.

Biến cố "khoảng cách giữa a và b lớn hơn 3" được biểu diễn bởi miền tam giác (H) gồm những điểm trong mặt phẳng tọa độ (a,b) sao cho $a - b > 3$. Đó là những điểm thuộc (G) và ở phía dưới đường thẳng $x - y - 3 = 0$. (H) có diện tích $mes(H) = (2.2)/2$.

Vậy theo công thức xác suất hình học ta có

$$p = \frac{mes\,(H)}{mes\,(G)} = \frac{1}{3}.$$

Bài 1.28. Giả sử A và B hẹn gặp nhau trong khoảng thời gian [0;60] với điều kiện người thứ nhất tới sẽ đợi người kia trong 20 đơn vị thời gian, sau đó đi khỏi. Tính xác suất để A và B gặp nhau.

Giải

Giả sử x là thời gian đến của A, y là thời gian đến của B. Khi đó, không gian các biến cố sơ cấp tương ứng với phép thử sẽ là tập hợp Ω có dạng

$$\{(a,b) \in \mathbb{R}^2 : 0 \leq x \leq 60 \, \text{và} \, 0 \leq y \leq 60\}.$$

Kí hiệu C là biến cố 2 người gặp nhau, khi đó

$$C = \{(x,y) \in \Omega; |x-y| \leq 20\}.$$

Vậy theo công thức xác suất hình học ta có

$$P(C) = \frac{60^2 - (60-20)^2}{60^2} = \frac{5}{9}.$$

Bài 1.29. Cho hai biến cố A, B và $P(A) = 0.2; P(B) = 0.5; P(A+B) = 0.6$. Tính $P(AB)$.

Giải

áp dụng công thức cộng xác suất ta có $P(AB) = 0,1$.

Bài 1.30. Cho ba biến cố A, B, C và $P(A) = 0.4; P(B) = 0.2; P(c) = 0.3$. Tính xác suất $P(A+B+C)$, biết rằng A, B, C độc lập nhau.

Giải

áp dụng công thức cộng xác suất ta có $P(A+B+C) = 0,664$.

Bài 1.31. Ba xạ thủ bắn vào mục tiêu. Xác suất trúng đích của xạ thủ thứ nhất là 0,7; của xạ thủ thứ hai là 0,8; của xạ thủ thứ ba là 0,9. Xác định xác suất để cả ba xạ thủ cùng bắn trúng đích.

Giải

Xác suất để cả ba xạ thủ cùng bắn trúng mục tiêu là $0,7.0,8.0,9 = 0,504$.

Bài 1.32. Trong hộp có 7 bi trắng và 8 bi đen. Lấy từ hộp ra lần lượt 2 bi (bi lấy xong không hoàn lại hộp). Tìm xác suất để 2 bi đều trắng.

Giải

Xác suất để 2 bi đều trắng là $\frac{7}{15}.\frac{6}{14} = 0,2$.

Bài 1.33. Từ một hộp chứa 8 viên bi đỏ và 5 viên bi trắng người ta lấy ngẫu nhiên 2 lần, mỗi lần 1 viên bi, không hoàn lại. Tính xác suất để lấy được:

a) 2 viên bi đỏ

b) 2 viên bi khác màu

c) Viên bi thứ hai là bi trắng.

Giải

Với $i \in \{1, 2\}$, đặt:

T_i: "viên bi lấy ra lần thứ i là bi trắng",

D_i: "viên bi lấy ra lần thứ i là bi đỏ".

a) Đặt A: "lấy được 2 viên bi đỏ", chúng ta có:

$$P(A) = P(D_1D_2) = P(D_1).P(D_2/D_1) = \frac{8}{13}.\frac{7}{12} = \frac{14}{39}$$

b) Đặt B: "lấy được hai viên bi khác màu", chúng ta có:

$$\begin{aligned} P(B) &= P(T_1D_2 + D_1T_2) = P(T_1D_2) + P(D_1T_2) \\ &= P(T_1)P(D_2/T_1) + P(D_1)P(T_2/D_1) \\ &= \frac{5}{13}\frac{8}{12} + \frac{8}{13}\frac{5}{12} \\ &= \frac{20}{39} \end{aligned}$$

c) tương tự ta có $P(T_2) = \dfrac{5}{13}$.

Bài 1.34. Rút ngẫu nhiên đồng thời 3 lá bài từ bộ bài 52 lá. Tính xác xuất để rút được ít nhất 1 lá bài át.

Giải

Gọi A là biến cố rút được ít nhất 1 lá át. Khi đó, $\overline{A}$ là biến cố không rút được lá át nào. ta có

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{C_{48}^3}{C_{52}^3} \approx 0,2174$$

Bài 1.35. Một lô hàng có 50 sản phẩm, trong đó có 4 phế phẩm. Một khách hàng kiểm tra bằng cách lấy ngẫu nhiên 10 sản phẩm từ lô hàng này. Nếu trong 10 sản phẩm lấy ra có không quá 1 phế phẩm thì người này sẽ mua lô hàng. Tính xác xuất để lô hàng được mua.

Giải

Gọi A: biến cố lô hàng được mua. A_i biến cố trong 10 sản phẩm lấy ra có i phế phẩm, $i = \overline{0, 10}$.

Khi đó, ta có: $A = A_0 + A_1$ và $\{A_0, A_1\}$ xung khắc

$$\Rightarrow P(A) = P(A_0) + P(A_1) \approx 0,82584.$$

Bài 1.36. Hai xạ thủ cùng bắn vào một tấm bia một cách độc lập, mỗi người bắn một phát. Xác suất để xạ thủ thứ nhất và thứ hai bắn trúng bia lần lượt là 0,7 và 0,8. Tính xác suất để:

a) Cả hai xạ thủ bắn trúng bia.

b) Chỉ có xạ thủ thứ nhất bắn trúng bia.

Giải

Gọi A_i là biến cố thứ i bắn trúng bia, $i = 1, 2$. Ta có $\{A_1, A_2\}$ độc lập.

a) Gọi A là biến cố cả hai xạ thủ bắn trúng bia. Khi đó, ta có

$$P(A) = P(A_1A_2) = P(A_1).P(A_2) = 0,7.0,8 = 0,56.$$

b) Gọi B là biến cố chỉ có xạ thủ thứ nhất bắn trúng bia. Khi đó, ta có

$$P(A) = P(A_1\overline{A_2}) = P(A_1).P(\overline{A_2}) = 0,7.0,2 = 0,14$$

Bài 1.37. Hai người công nhân cùng sản xuất một sản phẩm. Xác suất để người thứ nhất làm ra phế phẩm là 0,02 và người thứ hai làm ra phế phẩm là 0,03. Rút một sản phẩm từ số sản phẩm của hai người. Tính xác suất để sản phẩm lấy ra không là phế phẩm.

Giải

Gọi B_i là biến cố sản phẩm lấy ra là do người công nhân thứ i làm, $i = \overline{1, 2}$. Dễ thấy B_1 và B_2 tạo ra nhóm đầy đủ các biến cố.

Gọi A là biến cố sản phẩm lấy ra không là phế phẩm. Khi đó

$$P(A) = P(B_1).P(A/B_1) + P(B_2).P(A/B_2) = 0,975.$$

Bài 1.38. Có 3 hộp: mỗi hộp đựng 5 viên bi, trong đó hộp thứ i có i viên bi trắng ($i = \overline{1, 3}$). Lấy ngẫu nhiên từ mỗi hộp ra một viên bi.

a) Tìm xác suất để lấy được 3 viên bi trắng.

b) Nếu trong 3 bi lấy ra có một bi trắng, tìm xác suất để viên bi trắng đó là của hộp thứ nhất.

Giải

Gọi A_i là biến cố lấy được bi trắng từ hộp thứ $i, i = 1, 2, 3$.

Ta có $\{A_1, A_2, A_3\}$ là hệ độc lập toàn phần.

a) Gọi A là biến cố lấy được 3 viên bi trắng

$$\Rightarrow A = A_1.A_2.A_3$$
$$\Rightarrow P(A) = P(A_1).P(A_2).P(A_3) = \tfrac{1}{5}.\tfrac{2}{5}.\tfrac{3}{5} = 0,048$$

b) Gọi B là biến cố trong 3 bi lấy ra có 1 bi trắng, ta cần tính

$$P(A_1/B) = \frac{P(A_1.B)}{P(B)} = \frac{P\left(A_1.\overline{A_2}.\overline{A_3}\right)}{P(B)} = \frac{P(A_1)P\left(\overline{A_2}\right)P\left(\overline{A_3}\right)}{P(B)} = \frac{3}{29}$$

Trong đó

$$P(B) = P\left(A_1.\overline{A_2}.\overline{A_3}\right) + P\left(\overline{A_1}.A_2.\overline{A_3}\right) + P\left(\overline{A_1}.\overline{A_2}.A_3\right) = \frac{58}{125}$$

Bài 1.39. Ba người cùng vào một cửa hàng. Mỗi người muốn mua cùng một cái Tivi, nhưng cửa hàng chỉ còn hai cái Tivi. Người bán hàng làm 3 lá thăm, trong đó có hai lá được đánh dấu. Mỗi người lần lượt rút một lá thăm. Nếu ai rút được lá có đánh dấu thì được mua Tivi. Chứng minh rằng cách làm trên là công bằng cho cả ba người mua hàng.

Giải

Với $i \in \{1, 2, 3\}$, đặt A_i: "người thứ i rút được lá thăm có đánh dấu", chúng ta có:

- $P(A_1) = \frac{2}{3}$,
- $A_2 = A_1A_2 + \overline{A_1}A_2$, nên

$$P(A_2) = P(A_1).P(A_2/A_1) + P\left(\overline{A_1}\right).P\left(A_2/\overline{A_1}\right) = \frac{2}{3}.\frac{1}{2} + \frac{1}{3}.1 = \frac{2}{3}$$

- $A_3 = \overline{A_1}A_2A_3 + A_1\overline{A_2}A_3 \Rightarrow P(A_3) = P\left(\overline{A_1}A_2A_3\right) + P\left(A_1\overline{A_2}A\right) = \frac{2}{3}$

Vậy cách làm trên là công bằng cho cả 3 người mua hàng.

Bài 1.40. Có hai hộp thuốc. Hộp thứ nhất đựng 8 lọ thuốc, trong đó có 3 lọ kém chất lượng; hộp thứ hai đựng 6 lọ thuốc, trong đó có hai lọ kém chất lượng.

a) Lấy ngẫu nhiên từ mỗi hộp ra một lọ. Tính xác suất để được 1 lọ tốt và 1 lọ kém chất lượng.

b) Lấy ngẫu nhiên một hộp rồi từ hộp đó lấy ra một lọ thì được lọ kém chất lượng. Tính xác suất để lọ kém chất lượng đó thuộc hộp 2.

Giải

a) Lấy ngẫu nhiên mỗi hộp ra một lọ:

Với $i \in \{1, 2\}$, đặt tên các biến cố:

T_i: "Lọ thuốc lấy ra từ hộp thứ i là lọ tốt",

K_i: "Lọ thuốc lấy ra từ lọ thứ i là lọ kém chất lượng".

Khi đó ta có

$$\begin{aligned}
&A = K_1T_2 + T_1K_2\\
&\Rightarrow P(A) = P(K_1T_2 + T_1K_2) = P(K_1T_2) + P(T_1K_2)\\
&= P(K_1)P(T_2) + P(T_1)P(K_2) = \frac{11}{24}
\end{aligned}$$

b) Lấy ngẫu nhiên một hộp rồi từ hộp đó lấy ra một lọ:

Đặt H_i: "lấy được hộp thứ i" ($i = \overline{1, 2}$) và B: "lấy được lọ kém chất lượng".

Chúng ta có:

$$\begin{aligned}
&B = H_1B + H_2B\\
&P(H_1) = P(H_2) = \tfrac{1}{2};\\
&P(B/H_1) = \tfrac{3}{8};\\
&P(B/H_2) = \tfrac{2}{6}
\end{aligned}$$

Vậy xác suất cần tìm là

$$\begin{aligned}
P(H_2/B) &= \frac{P(H_2).P(B/H_2)}{P(B)}\\
&= \frac{P(H_2).P(B/H_2)}{P(H_1).P(B/H_1) + P(H_2).P(B/H_2)}\\
&= \frac{8}{17}
\end{aligned}$$

Bài 1.41. Một lô hạt giống gồm 3 loại để lẫn lộn. Loại 1 chiếm 2/3 số hạt, loại 2 chiếm 1/4 , còn lại là loại 3. Tỉ lệ nẩy mầm của hạt loại 1, hạt loại 2 và hạt loại 3 theo thứ tự là 80%, 60% và 40%. Lấy ngẫu nhiên một hạt từ lô hạt giống.

a) Tính xác suất để hạt giống lấy ra là nảy mầm được. ý nghĩa của xác suất này đối với lô hạt giống là gì?

b) Giả sử hạt giống lấy ra là nẩy mầm được. Tính xác suất để hạt giống đó thuộc loại 2?

c) Giả sử hạt giống lấy ra là không nẩy mầm được. Nhiều khả năng nhất là hạt giống đó thuộc loại nào? Tại sao?

Giải

a) Gọi A là biến cố "hạt giống lấy ra là hạt nẩy mầm được" và H_i là biến cố "hạt giống lấy ra thuộc loại i" $(i = 1, 2, 3)$,(khi đó H_1, H_2, H_3 lập thành một hệ đầy đủ) chúng ta có:

$$P(H_1) = \frac{2}{3};\ P(H_2) = \frac{1}{4};\ P(H_3) = \frac{1}{12}$$
$$P(A/H_1) = 0,8;\ \ P(A/H_2) = 0,6;\ P(A/H_3) = 0,4$$

áp dụng công thức xác suất toàn phần ta có

$$P(A) = \sum_{i=1}^{3} P(H_i).P(A/H_i) \approx 0,7167$$

Xác suất $P(A)$ chính là tỉ lệ nẩy mầm chung của lô hạt giống.

b)Xác suất phải tính là $P(H_2/A)$.

Theo công thức Bayes,

$$P(H_2/A) = \frac{P(H_2).P(A/H_2)}{P(A)} = \frac{3}{14}$$

c) Giả sử hạt giống lấy ra là không nẩy mầm được, chúng ta phải so sánh các xác suất $P(H_i/\overline{A})$, với $i \in \{1, 2, 3\}$ (bạn đọc tự giải).

Bài 1.42. Một đám đông có số đàn ông bằng nửa số đàn bà. Xác suất để đàn ông bị bệnh tim là 0,06 và xác suất để đàn bà bị bệnh tim là 0,036. Chọn ngẫu nhiên 1 người từ đám đông, tính xác suất để người này bị bệnh tim.

Giải

Gọi

+ A là biến cố người được chọn bị bệnh tim.

+ H_1 là biến cố người được chọn là đàn bà.

+ H_2 là biến cố người được chọn là đàn ông.

Ta có $H_1; H_2$ là hệ đầy đủ, khi đó áp dụng công thức xác suất toàn phần ta có:

$$P(A) = P(H_1).P(A/H_1) + P(H_2).P(A/H_2) = 0,044.$$

Bài 1.43. Có hai hộp đựng bi: hộp I có 5 bi trắng và 7 bi đen; hộp II có 6 bi trắng và 4 bi đen. Lấy ngẫu nhiên 1 bi từ hộp I bỏ sang hộp II, rồi từ hộp II lấy ngẫu nhiên ra 1 bi. Tính xác suất:

a) Bi lấy từ hộp II là bi trắng.

b) Bi lấy từ hộp I sang hộp II là bi trắng, biết rằng bi lấy từ hộp II là bi trắng.

Giải

Gọi T_i: "biến cố bi lấy từ hộp thứ i là bi trắng", $i = 1, 2$.

a) Ta cần tính $P(T_2)$. Ta có $\{T_1, \overline{T_1}\}$ là hệ đầy đủ và xung khắc.

áp dụng CTXSTP:

$$\begin{aligned} P(T_2) &= P(T_1).P(T_2/T_1) + P\left(\overline{T_1}\right).P\left(T_2/\overline{T_1}\right) \\ &= \frac{5}{12}.\frac{7}{11} + \frac{7}{12}.\frac{6}{11} = \frac{7}{12} \approx 0,5833. \end{aligned}$$

b) Ta cần tính:

$$P(T_1/T_2) = \frac{P(T_1.T_2)}{P(T_2)} = \frac{P(T_1).P(T_2/T_1)}{P(T_2)} = \frac{5}{11} \approx 0,4545.$$

Bài 1.44. Điều tra về giới tính của sinh viên ở một trường học, người ta thấy rằng có 65% nam và 35% nữ. Trong đó, tỷ lệ học sinh nữ và nam thích chơi game tương ứng là 20% và 25%. Chọn ngẫu nhiên một sinh viên của trường này, tính xác suất:

a) Sinh viên được chọn thích chơi game.

b) Sinh viên được chọn là nam, biết rằng sinh viên này thích chơi game.

Giải

a) Gọi A là biến cố chọn được sinh viên thích chơi game; A_1 là biến cố chọn được sinh viên nữ; A_2 là biến cố chọn được sinh viên nam.

Ta có $P(A_1) = 35\%; P(A_2) = 65\%$ và $\{A_1, A_2\}$ là hệ đầy đủ

áp dụng CTXSTP:

$$P(A) = P(A_1).P(A/A_1) + P(A_2).P(A/A_2) = 35\%.20\% + 65\%.25\% = 0.2325.$$

b) Ta cần tính:

$$P(A_2/A) = \frac{P(A_2.A)}{P(A)} = \frac{P(A_2).P(A/A_2)}{P(A)} = 0,6989.$$

Bài 1.45. Một nhà máy có hai phân xưởng. Sản lượng của phân xưởng I gấp 3 lần sản lượng của phân xưởng II. Tỷ lệ phế phẩm của phân xưởng I và II lần lượt là 3% và 5%. Chọn ngẫu nhiên 1 sản phẩm của nhà máy, tính xác suất:

a) Chọn được sản phẩm tốt do phân xưởng I sản xuất.

b) Chọn được 1 phế phẩm.

c) Giả sử chọn được sản phẩm tốt, tính xác suất để sản phẩm này do phân xưởng I sản xuất.

Giải

Gọi A là biến cố chọn được sản phẩm tốt; A_i là biến cố chọn được sản phẩm do phân xưởng thứ i sản xuất, $i = 1, 2$.

$$\Rightarrow P(A_1) = \frac{3}{4}; P(A_2) = \frac{1}{4}$$

a) Ta cần tính:

$$P(A.A_1) = P(A_1).P(A/A_1) = \frac{3}{4}.(1 - 3\%) = 0,7275$$

b) Ta cần tính: $P(\overline{A})$. Ta có: $\{A_1, A_2\}$ là hệ đầy đủ

áp dụng CTXSTP:

$$P(\overline{A}) = P(A_1).P(\overline{A}/A_1) + P(A_2).P(\overline{A}/A_2) = \frac{3}{4}.3\% + \frac{1}{4}.5\% = 0.035.$$

c) Ta cần tính:

$$P\left(A_1/A\right) = \frac{P\left(A_1.A\right)}{P\left(A\right)} = \frac{P\left(A_1\right).P\left(A/A_1\right)}{P\left(A\right)}.$$

áp dụng CTXSTP ta có:

$$P\left(A\right) = P\left(A_1\right).P\left(A/A_1\right) + P\left(A_2\right).P\left(A/A_2\right) = \frac{3}{4}.97\% + \frac{1}{4}.95\% = 0,965.$$

Khi đó, $P\left(A_1/A\right) = \dfrac{\frac{3}{4}.97\%}{0,965} \approx 0,7539.$

Bài 1.46. Gieo 100 hạt đậu tương. Xác suất nảy mầm của mỗi hạt là 0,9. Tính xác suất để trong 100 hạt:

a) Có 85 hạt nảy mầm.

b) Có ít nhất 99 hạt nảy mầm.

Giải

a) Gọi A là biến cố có 85 hạt nảy mầm, khi đó ta có

$$P\left(A\right) = P_{100}\left(85\right) = C_{100}^{85}(0,9)^{85}(0,1)^{15} = 0,0327$$

b) Gọi B là biến cố có ít nhất 99 hạt nảy mầm. Ta cần tính $P(B)$. Gọi C là biến cố có 99 hạt nảy mầm và D là biến cố có 100 hạt nảy mầm.

Ta có $B = C + D$ (C và D xung khắc)

$$\begin{aligned} \Rightarrow P\left(B\right) &= P\left(C\right) + P\left(D\right) \\ &= P_{100}\left(99\right) + P_{100}\left(100\right) \\ &= C_{100}^{99}(0,9)^{99}(0,1)^{1} + C_{100}^{100}(0,9)^{100}(0,1)^{0} \\ &= 0,0003. \end{aligned}$$

Bài 1.47. Một lô hàng chứa rất nhiều sản phẩm với tỷ lệ phế phẩm là 0,02. Lấy ngẫu nhiên 150 sản phẩm để kiểm tra. Tính xác suất lấy được ít nhất 1 phế phẩm.

Giải

Gọi A là biến cố lấy ít nhất một phế phẩm, khi đó ta có

$$P(A) = 1 - P(B) = 1 - C_{150}^{0}(0,02)^{0}(0,98)^{150} = 0,9517$$

Bài 1.48. Ngân hàng đề thi môn Xác Suất Thống Kê có 500 câu hỏi. Thầy Hùng chọn ngẫu nhiên 20 câu hỏi để làm đề thi cuối kỳ. Mỗi câu hỏi có 4 phương án trả lời trong đó chỉ có một phương án đúng. Mỗi câu trả lời đúng đạt 0,5 điểm. Bạn Nam làm bài thi bằng cách chọn hú họa một phương án cho mỗi câu hỏi. Tính xác suất để bạn Hậu đạt 8 điểm.

Giải

Gọi A là biến cố bạn Hậu đạt 8 điểm.

Theo đề bài ta có lược đồ Bernoulli với:

+ Số phép thử : $n = 20$.

+ Xác suất để bạn Hậu trả lời đúng 1 câu hỏi: $p = 0,25$.

Ta có

$$P(A) = P_{20}(16) = C_{20}^{16}(0,25)^{16}(0,75)^{4} = 0,357.10^{-6}$$

Bài 1.49. Tỉ lệ sản xuất ra phế phẩm của một máy là 8%. Xem một lô hàng gồm 75 sản phẩm do máy đó sản xuất ra.

a) Tính xác suất để trong lô hàng có 10 phế phẩm.

b) Trong lô hàng, nhiều khả năng nhất có bao nhiêu phế phẩm? Tính xác suất tương ứng.

Giải

Nếu xem việc máy sản xuất ra một sản phẩm là một phép thử Bernoulli, với xác suất cho "thành công" là $p = 0,08$, thì khi máy đó sản xuất 75 sản phẩm, nó đã thực hiện quá trình $B(75; 0,08)$.

a) Xác suất phải tính:

$$P_{75}(10) = C_{75}^{10}(0,08)^{10}.(0,92)^{65} \approx 0,03941$$

b) Số phế phẩm nhiều khả năng nhất trong lô hàng là: $[(75+1).0,08] = 6$,

với xác suất tương ứng:

$$P_{75}(6) = C_{75}^{6}(0,08)^{6}.(0,92)^{69} \approx 0,16745$$

Bài 1.50. Có n chiếc bình và r viên bi. Người ta lần lượt bỏ một cách ngẫu nhiên từng viên bi vào một bình bất kỳ trong n bình đó, cho đến khi hết bi... Tính xác suất để bình thứ k có đúng m viên bi ($m < r$).

Giải

Bỏ ngẫu nhiên một viên bi vào một bình là một phép thử. Khi viên bi được bỏ vào bình thứ k thì xem như đó là một phép thử thành công. Xác suất để một phép thử thành công là $p = \frac{1}{n}$. Vậy, chúng ta có quá trình $B(r, \frac{1}{n})$ và xác suất phải tính là:

$$P_r(m) = C_r^m \left(\frac{1}{n}\right)^m \left(\frac{n-1}{n}\right)^{r-m}$$

Bài 1.51. Người ta muốn lấy ngẫu nhiên một số hạt từ một lô hạt giống có tỉ lệ hạt lép là 3% để nghiên cứu. Hỏi phải lấy ít nhất bao nhiêu hạt sao cho xác suất để có ít nhất một hạt lép không bé hơn 95%.

Giải

Gọi n là số hạt phải lấy, chúng ta có $B(n; 0.03)$. Xác suất để có ít nhất một hạt lép là $1 - (1 - 0,03)^n = 1 - (0,97)^n$.

Theo giả thiết, chúng ta có:

$$1 - (0,97)^n \geq 0,95 \Leftrightarrow (0,97)^n \leq 0,05 \Leftrightarrow n \geq 98,3523$$

Vậy, phải lấy ít nhất 99 hạt giống.

Bài 1.52. Cho 3 hộp linh kiện máy tính mà khả năng lựa chọn của mỗi hộp là như nhau. Hộp thứ nhất có 30 linh kiện, trong đó có 20 linh kiện tốt và 10 linh kiện xấu. Hộp thứ hai có 30 linh kiện đều tốt. Hộp thứ ba có 30 linh kiện, trong có 15 linh kiện tốt và 15 linh kiện xấu. Chọn ngẫu nhiên một hộp và từ đó lấy ngẫu nhiên 1 linh kiện.

a) Tính xác suất linh kiện lấy ra là linh kiện tốt.

b) Giả sử linh kiện lấy ra là tốt. Tìm xác suất để linh kiện đó là của hộp thứ 3.

Giải

a) Gọi A là biến cố lấy ra là tốt và B_i là biến cố linh kiện lấy ra từ hộp thứ $i; i = 1, 2, 3$.

Dễ thấy B_1, B_2, B_3 lập thành một hệ đầy đủ các biến cố. Theo công thức xác suất đầy đủ, ta có:

$$\begin{aligned}
P(A) &= P(B_1).P(A/B_1) + P(B_2).P(A/B_2) + P(B_3).P(A/B_3) \\
&= \frac{1}{3}\left(\frac{20}{30} + \frac{30}{30} + \frac{15}{30}\right) = \frac{13}{18}.
\end{aligned}$$

b) Xác suất để linh kiện tốt lấy ra là của hộp thứ 3:

$$P(B_3/A) = \frac{P(B_3).P(A/B_3)}{P(A)} = 0,23.$$

Bài 1.53. Hộp thứ nhất có 5 quả cầu trắng và 10 quả cầu đỏ , hộp thứ hai có 3 quả cầu trắng và 7 quả cầu đỏ. Từ mỗi hộp lấy ngẫu nhiên một quả cầu, rồi từ 2 quả cầu đó lấy ngẫu nhiên một quả. Tìm xác suất để quả cầu lấy ra sau là cầu trắng.

Giải

Gọi A là biến cố quả cầu lấy ra sau cùng là cầu trắng, B_{TT} là biến cố hai quả cầu lấy ra từ hai hộp là trắng, B_{DD} là biến cố hai quả cầu lấy ra từ hai hộp là đỏ, B_{TD} là biến cố quả cầu lấy ra từ hộp thứ nhất là trắng và quả cầu lấy ra từ hộp thứ hai là đỏ, B_{DT} là biến cố quả cầu lấy ra từ hộp thứ nhất là đỏ và quả cầu lấy ra từ hộp thứ hai là trắng.

Dễ thấy $B_{TT}; B_{DT}; B_{TD}$ và B_{DD} là các biến cố lập nên nhóm đầy đủ.

Theo công thức xác suất đầy đủ ta có

$$\begin{aligned}
P(A) &= P(B_{TT}).P(A/B_{TT}) + P(B_{DD}).P(A/B_{DD}) \\
&+ P(B_{DT}).P(A/B_{DT}) + P(B_{TD}).P(A/B_{TD}) \\
&= \left(\frac{5}{15}.\frac{3}{10}.1\right) + \left(\frac{10}{15}.\frac{7}{10}.0\right) + \left(\frac{5}{15}.\frac{7}{10}.\frac{1}{2}\right) + \left(\frac{10}{15}.\frac{3}{10}.\frac{1}{2}\right) \\
&= 0,31667.
\end{aligned}$$

Bài 1.54. Có 2 chuồng thỏ thí nghiệm, trong đó chuồng I có 12 con thỏ trắng và 3 con thỏ nâu, chuồng II có 16 con thỏ trắng và 4 con thỏ nâu. Tình cờ, 1 con thỏ từ chuồng II nhảy qua chuồng I. Từ chuồng I bắt ngẫu nhiên 1 con thỏ. Tính xác suất để con thỏ bắt ra là thỏ trắng.

Giải

Gọi A là biến cố con thỏ nhảy từ chuồng II sang chuồng I là thỏ trắng

$$P\left(A\right) = \frac{16}{20} = \frac{4}{5}$$

Khi đó, biến cố phủ định $\overline{A}$ là biến cố con thỏ nhảy từ chuồng II sang chuồng I là thỏ nâu

$$P\left(\overline{A}\right) = 1 - P\left(A\right) = \frac{1}{5}$$

Dể thấy các biến cố A và $\overline{A}$ tạo thành nhóm đầy đủ các biến cố.

Gọi B là biến cố bắt ra được thỏ trắng từ chuồng I.

Khi đó, B/A (tương tự $B/\overline{A}$) là biến cố bắt được thỏ trắng khi con thỏ nhảy sang là thỏ trắng (tương tự, thỏ nâu), với:

$$P(B/A) = \frac{13}{16}; P(B/\overline{A}) = \frac{12}{16}.$$

áp dụng công thức xác suất đầy đủ, với

$$P\left(B\right) = P\left(A\right).P\left(B/A\right) + P\left(\overline{A}\right).P\left(B/\overline{A}\right) = \frac{4}{5}.\frac{13}{16} + \frac{1}{5}.\frac{12}{16} = 0,8.$$

Ta suy ra xác suất để bắt được thỏ nâu là

$$P(\overline{B}) = 1 - P(B) = 0,2$$

Bài 1.55. Tỉ lệ người dân nghiện thuốc lá ở một vùng là 30%. Biết rằng tỉ lệ người bị viêm họng trong số người nghiện thuốc lá là 60%, còn tỉ lệ người bị viêm họng trong số người không hút thuốc là là 40%.

a) Chọn ngẫu nhiên một người, biết rằng người đó viêm họng, tính xác suất để người đó nghiện thuốc lá.

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

Giải

a) Gọi A là biến cố người được chọn là nghiện thuốc lá: $P(A) = 0,3$.

Khi đó, biến cố phủ định $\overline{A}$ là biến cố người được chọn không nghiện thuốc: $P(\overline{A}) = 1 - P(A) = 0,7$.

Dễ thấy A và $\overline{A}$ tạo thành một nhóm đầy đủ các biến cố.

Gọi B là biến cố người được chọn bị viêm họng. Khi đó, B/A (tương tự, $B/\overline{A}$) là biến cố người viêm họng được chọn có nghiện thuốc là (tương tự, không nghiện thuốc lá) với xác suất:

$$P(B/A) = 0,6; P(B/\overline{A}) = 0,4$$

Theo công thức xác suất đầy đủ, ta có

$$P(B) = P(A).P(B/A) + P(\overline{A}).P(B/\overline{A}) = (0,3.0,6) + (0,7.0,4) = 0,46.$$

Do A/B là biến cố người được chọn có nghiện thuốc lá, biết rằng người đó viêm họng. áp dụng công thức Bayes:

$$P(A/B) = \frac{P(A).P(B/A)}{P(B)} = \frac{0,3.0,6}{0,46} = 0,39.$$

Vậy xác suất để người, biết rằng viêm họng, được chọn có nghiện thuốc lá bằng 0,39.

Ta cũng có thể suy ra xác suất để người, biết rằng viêm họng, được chọn không nghiện thuốc lá:

$$P(\overline{A}/B) = 1 - P(A/B) = 0,61.$$

b) tương tự câu a) ta tính được xác suất cần tìm là $P(\overline{A}/\overline{B}) = 0,78$ (bạn đọc tự giải).

Bài 1.56. Trong một bệnh viện, tỉ lệ bệnh nhân các tỉnh như sau: tỉnh A có 25%, tỉnh B có 35%, tỉnh C có 40%. Biết rằng tỉ lệ bệnh nhân là giáo viên các tỉnh là: tỉnh A có 2%, tỉnh B có 3% và tỉnh C có 3,5%. Chọn ngẫu nhiên một bệnh nhân, tính xác suất để bệnh nhân đó là giáo viên.

Giải

Gọi X, Y, Z lần lượt là biến cố bệnh nhân được chọn thuộc tỉnh A, B, C. Các biến cố này tạo thành nhóm biến cố đầy đủ với xác suất:

$$P(X) = \frac{1}{4}; P(Y) = \frac{7}{20}; P(Z) = \frac{2}{5}$$

Gọi U là biến cố bệnh nhân được chọn là giáo viên, khi đó ta có

$$P(U/X) = \frac{2}{100}; P(U/Y) = \frac{3}{100}; P(U/Z) = \frac{7}{200}$$

Theo công thức xác suất đầy đủ, ta có

$$P(U) = P(X).P(U/X) + P(Y).P(U/Y) + P(Z).P(U/Z) = 0,0295.$$

Bài 1.57. Trong một vùng dân cư mà tỉ lệ nữ là 55%, có một đợt dịch truyền nhiễm với tỉ lệ mắc bệnh của nam là 6% và của nữ là 2%.

a) Tính tỉ lệ mắc bệnh chung của vùng dân cư đó.

b) Chọn ngẫu nhiên một người từ vùng dân cư đó thì gặp phải người bị mắc bệnh. Tính khả năng người bị mắc bệnh đó là nam giới.

Giải

a) Gọi A là biến cố người dân được chọn là nam. Biến cố phủ định $\overline{A}$ là biến cố người dân được chọn là nữ. Các biến cố A và $\overline{A}$ tạo thành nhóm biến cố đầy đủ, với xác suất:

$$P(A) = 0,45; P(\overline{A}) = 0,55.$$

Gọi B là biến cố người được chọn mắc bệnh dịch.

Khi đó, B/A (tương tự, $B/\overline{A}$) là biến cố người nam (tương tự, nữ) được chọn là mắc bệnh:

$$P(B/A) = 0,06; P(B/\overline{A}) = 0,02.$$

Theo công thức xác suất đầy đủ, ta có

$$P(B) = P(A).P(B/A) + P\left(\overline{A}\right).P\left(B/\overline{A}\right) = 0,038$$

Vậy tỉ lệ mắc bệnh chung của vùng là 0,038.

b) (A/B) là biến cố bệnh nhân được chọn là nam.

áp dụng công thức Bayes, ta có:

$$P(A/B) = \frac{P(A).P(B/A)}{P(B)} = \frac{0,45.0,06}{0,038} = 0,71$$

Ta suy ra xác suất để bệnh nhân được chọn là nữ:

$$P(\overline{A}/B) = 1 - P(A/B) = 0,29.$$

Bài 1.58. Trong số 10 xạ thủ có 5 người bắn trúng bia với xác suất 0,9; có 3 người bắn trúng bia với xác suất 0,8 và có 2 người bắn trúng bia với xác suất 0,7. Chọn ngẫu nhiên một xạ thủ và cho anh ta bắn một viên đạn nhưng kết quả không trúng bia. Hỏi xạ thủ ấy có khả năng thuộc nhóm nào?

Giải

Gọi A_1, A_2 và A_3 lần lượt là các biến cố chỉ xạ thủ thuộc nhóm thứ nhất, thứ hai và thứ ba.

Dễ thấy A_1, A_2 và A_3 tạo thành nhóm đầy đủ các biến cố, với

$$P(A_1) = \frac{5}{10}; P(A_2) = \frac{3}{10}; P(A_3) = \frac{2}{10}.$$

Gọi B là biến cố xạ thủ chọn ra bắn không trúng bia. Khi đó

$$\begin{aligned} P(B) &= P(A_1).P(B/A_1) + P(A_2).P(B/A_2) + P(A_3).P(B/A_3) \\ &= \frac{5}{10}.\frac{1}{10} + \frac{3}{10}.\frac{2}{10} + \frac{2}{10}.\frac{3}{10} \\ &= \frac{17}{100}. \end{aligned}$$

Theo công thức Bayes ta có

$$P(A_1/B) = \frac{5}{17};$$
$$P(A_2/B) = \frac{6}{17};$$
$$P(A_3/B) = \frac{6}{17}.$$

Xạ thủ được chọn ra có khả năng thuộc nhóm thứ hai và thứ ba.

1.5 Bài tập đề nghị

Bài 1.59. Bài toán chiếc kim Buffon: Trên mặt phẳng kẻ các đường thẳng song song cách nhau khoảng 2a. Tung ngẫu nhiên một cái kim có chiều dài $2l\,(l < a)$. Tính xác suất để cái kim cắt một đường thẳng bất kỳ trên mặt phẳng.

ĐS: $\frac{2l}{a\pi}$.

Bài 1.60. Cho A, B là hai biến cố độc lập và $P(A) = 0.3$ và $P(B) = 0.7$. Tính xác suất:

a) $P(AB)$.

b) $P(A/B)$.

c) $P(B/A)$.

ĐS: $0,21; 0,3; 0,7$.

Bài 1.61. Trong một hộp có 10 bi trắng và 20 bi đen. Hỏi xác suất để trong hai bi lấy ra có một bi trắng, còn bi kia đen? (bi lấy ra không hoàn lại hộp).

a) Lấy đồng thời 2 bi.

b) Lấy lần lượt từng viên.

ĐS: $0,4598$.

Bài 1.62. Nam nộp hồ sơ đi dự thi vào trường đại học A và trường cao đẳng B. Khả năng Nam thi đậu vào trường đại học là 0,6 và trường cao đẳng là 0,8. Khả năng Nam không thi đậu vào ít nhất một trong hai trường là 0,3. Tính xác suất để Nam thi đậu vào ít nhất một trong hai trường A và B.

ĐS: $0,7$.

Bài 1.63. Gieo một con xúc xắc cân đối và đồng chất. Tính xác suất để xuất hiện mặt lớn hơn 4 chấm.

ĐS: $\dfrac{1}{3}$.

Bài 1.64. Một lô hàng gồm 50 máy vi tính, trong đó có 20 máy do công ty A sản xuất và 30 máy do công ty B sản xuất. Một khách hàng chọn ngẫu nhiên 5 máy từ lô hàng này. Tính xác suất để:

a) Cả 5 máy do công ty A sản xuất.

b) Có 3 máy do công ty B sản xuất.

Bài 1.65. Chọn ngẫu nhiên 2 viên bi trong một bình có 3 bi xanh, 4 bi đỏ và 5 bi vàng. Tính xác suất để chọn được đúng 2 viên bi đỏ.

Bài 1.66. Một hộp có chứa 7 viên bi trắng và 3 viên bi đen cùng kích thước. Lấy ngẫu nhiên cùng một lúc 4 viên bi. Tìm xác suất để trong 4 viên bi lấy ra được có:

a) 2 viên bi đen.

b) ít nhất 2 viên bi đen.

c) Toàn là bi trắng.

Bài 1.67. Một khách sạn có 6 phòng đơn. Có 10 khách đến thuê phòng, trong đó có 6 nam và 4 nữ. Người quản lý chọn ngẫu nhiên 6 người. Tính xác suất để:

a) Cả 6 người đều là nam.

b) Có 4 nam và 2 nữ.

c) Có ít nhất hai nữ.

Bài 1.68. Từ một nhóm bạn gồm 5 người: Nam, Ngọc, Tú, Quyên và Hải. Chọn ngẫu nhiên 3 bạn từ nhóm này, tính xác suất để trong đó có bạn Quyên.

ĐS: $0,6$.

Bài 1.69. Trong một tuần lễ có 7 tai nạn giao thông. Tính xác suất để mỗi ngày xảy ra đúng 1 tai nạn?

ĐS: $\dfrac{7!}{7^7}$.

Bài 1.70. Một kiện hàng có 10 sản phẩm, trong đó 8 sản phẩm loại A. Kiểm tra lần lượt không hoàn lại 2 sản phẩm của kiện hàng. Xác suất để 2 sản phẩm kiểm tra có không quá 1 sản phẩm loại A là bao nhiêu?

ĐS: $\dfrac{17}{45}$.

Bài 1.71. Trong một xưởng có 3 máy làm việc. Trong một ca, máy thứ nhất có thể cần sửa chữa với xác suất 0,12; máy thứ hai với xác suất 0,15 và máy thứ ba với xác suất 0,18. Tìm xác suất sao cho trong một ca làm việc có ít nhất một máy không cần sửa chữa.

ĐS: 0,99676.

Bài 1.72. Có 4 linh kiện điện trong một mạch điện, chúng có thể bị hỏng một cách độc lập trong khoảng thời gian T với xác suất tương ứng: 0,2; 0,3; 0,4; 0,5. Tìm xác suất để mạch bị hỏng trong thời gian T nếu mạch mắc song song. Câu hỏi tương tự cho trường hợp mắc nối tiếp.

Bài 1.73. Có hai kiện hàng, kiện thứ nhất có 8 sản phẩm tốt và 2 sản phẩm xấu; Kiện thứ hai có 9 sản phẩm tốt và 1 sản phẩm xấu. Lấy ngẫu nhiên 2 sản phẩm từ kiện thứ nhất bỏ vào kiện thứ hai, rồi sau đó lấy ngẫu nhiên 3 sản phẩm từ kiện thứ hai. Tìm xác suất để trong 3 sản phẩm lấy ra có 2 sản phẩm tốt.

Bài 1.74. Có 3 hộp thuốc, hộp thứ nhất có 5 ống thuốc tốt và 2 ống thuốc xấu; hộp thứ hai có 4 ống thuốc tốt và 1 ống thuốc xấu; hộp thứ ba có 3 ống thuốc tốt. Lấy ngẫu nhiên 1 hộp và từ hộp đó rút ngẫu nhiên 2 ống thuốc.

a) Tính xác suất của biến cố rút được 1 ống tốt và 1 ống xấu.

b) Giả sử rút được 2 ống tốt, tính xác suất của biến cố 2 ống đó thuộc hộp thứ hai.

Bài 1.75. Một nhà máy sản xuất bóng đèn có 2 phân xưởng, biết rằng xưởng thứ nhất sản xuất gấp 4 lần xưởng thứ hai, tỉ lệ bóng đèn hư phân xưởng thứ nhất là 10%, tỉ lệ bóng đèn hư phân xưởng thứ hai là 20%, chọn ngẫu nhiên 1 bóng đèn của nhà máy.

a) Tính xác suất của bóng đèn bị hư.

b) Giả sử chọn được bóng đèn hư, nhiều khả năng đèn này do phân xưởng nào sản xuất? Tại sao?

Bài 1.76. Tỉ lệ sản xuất ra phế phẩm của một máy là 6%. Một lô hàng gồm 60 sản phẩm do máy đó sản xuất ra.

a) Tính xác suất để trong lô hàng có 5 phế phẩm.

b) Tính xác suất để trong lô hàng không có phế phẩm nào.

c) Tính xác suất để trong lô hàng có ít nhất 1 phế phẩm.

d) Trong lô hàng, nhiều khả năng nhất có bao nhiêu phế phẩm? Tính xác suất tương ứng.

Bài 1.77. Trong một lô hàng có 100 sản phẩm, trong đó có 20 sản phẩm loại A. Lấy ngẫu nhiên (liên tiếp từng sản phẩm một) ba sản phẩm. Tính xác suất để cả 3 sản phẩm đều là loại A.

Bài 1.78. Một hộp thuốc có 5 ống thuốc tốt và 3 ống thuốc kém chất lượng. Chọn ngẫu nhiên lần lượt không trả lại 2 ống. Tính xác suất để:

a) Cả hai ống chọn được đều tốt.

b) Chỉ ống thuốc chọn ra đầu tiên là tốt.

c) Trong hai ống có ít nhất một ống thuốc tốt.

Bài 1.79. Một lô sản phẩm gồm hai loại, trong đó số sản phẩm do máy thứ nhất sản xuất gấp hai lần số sản phẩm do máy thứ hai sản suất. Tỉ lệ phế phẩm của máy thứ nhất là 0,04 và của máy thứ hai là 0,06. Lấy ngẫu nhiên một sản phẩm. Tìm xác suất để sản phẩm chọn được là tốt.

Bài 1.80. Một hộp đựng 15 quả bóng bàn trong đó có 9 quả còn mới. Lần đầu người ta lấy ngẫu nhiên 3 quả để thi đấu, sau đó lại trả vào hộp. Lần thứ hai lấy ngẫu nhiên 3 quả. Tìm xác suất để cả 3 quả lấy ra lần sau đều mới.

Bài 1.81. Có hai chuồng thỏ. Chuồng thứ nhất có 5 con thỏ đen và 10 con thỏ trắng. Chuồng thứ hai có 7 thỏ đen và 3 con thỏ trắng. Từ chuồng thứ nhất ta bắt ngẫu nhiên 2 con thỏ bỏ vào chuồng thứ hai, rồi sau đó từ chuồng thứ hai bắt ra một con thỏ, tính xác suất để con thỏ này là thỏ trắng.

Bài 1.82. Một chuồng gà có 9 con mái và 1 con trống. Chuồng kia có 1 con mái và 5 con trống. Từ mỗi chuồng ta bắt ra ngẫu nhiên một con làm thịt. Các con gà còn lại được dồn vào một chuồng thứ ba. Từ chuồng thứ ba này lại bắt ngẫu nhiên một con gà. Tính xác suất để ta bắt được gà trống.

ĐS: $\dfrac{304}{840}$.

Bài 1.83. Có 3 hộp: mỗi hộp đựng 5 viên bi, trong đó hộp thứ i có i bi trắng . Chọn ngẫu nhiên một hộp, rồi từ hộp đã chọn lấy ngẫu nhiên đồng thời 3 viên bi.

a) Tìm xác suất để lấy được 3 viên bi trắng.

b) Nếu trong 3 bi lấy ra có một bi trắng, tìm xác suất để viên bi trắng đó là của hộp thứ nhất.

ĐS: $0,048; \dfrac{3}{29}$.

Bài 1.84. Một xạ thủ bắn 10 viên đạn vào 1 tấm bia một cách độc lập. Xác suất bắn trúng bia của mỗi viên đạn bằng nhau và bằng 0,6. Tìm xác suất có từ 5 đến 7 viên đạn trúng đích.

ĐS: $0,6665$.

Bài 1.85. Trong kho có 20 bao hàng, trong đó có 12 bao loại I (chứa 60% hàng tốt) và 8 bao loại II (chứa 40% hàng tốt). Lấy ngẫu nhiên 1 bao và từ bao đó lấy ngẫu nhiên một sản phẩm.

a) Tính xác suất để sản phẩm đó là hàng tốt.

b) Giả sử sản phẩm đó là hàng tốt. Theo ý bạn thì bao lấy ra thuộc loại gì? Vì sao?

Bài 1.86. Trong 18 xạ thủ có 5 người có khả năng bắn trúng bia với xác suất 0,8; 7 người có khả năng bắn trúng bia với xác suất 0,7; 4 người có khả năng bắn trúng bia với xác suất 0,6 và 2 người bắn trúng bia với xác suất 0,5. Chọn ngẫu nhiên một xạ thủ và anh ta đã bắn không trúng đích. Hỏi anh ta có khả năng thuộc nhóm nào nhiều hơn?

Bài 1.87. Trong một thành phố nào đó, tỉ lệ người thích xem bóng đá là 60%. Chọn ngẫu nhiên một nhóm gồm 12 người.

a) Tính xác suất để trong nhóm đó có đúng 5 người thích xem bóng đá.

b) Tính xác suất để trong nhóm đó không có người nào thích xem bóng đá.

c) Tính xác suất để trong nhóm đó có ít nhất một người thích xem bóng đá.

d) Trong nhóm được chọn nói trên, theo bạn nhiều khả năng nhất là có bao nhiêu người thích xem bóng đá? Tính xác suất tương ứng.

Bài 1.88. Trong một nhà máy sản xuất thức ăn đóng hộp có tỉ lệ hộp không đạt tiêu chuẩn chất lượng là 4%. Ban thẩm định chất lượng chọn ngẫu nhiên một số hộp để kiểm tra. Hỏi phải lấy ít nhất bao nhiêu hộp để kiểm tra sao cho xác suất để có ít nhất một hộp không đạt chất lượng không bé hơn 90%.

Bài 1.89. Một nhân viên bán hàng mỗi ngày đi chào hàng ở 20 nơi với xác suất bán được hàng ở mỗi nơi là 0,3.

a) Tìm xác suất để người đó bán được hàng ở 4 nơi.

b) Tìm xác suất để người đó bán được hàng ở ít nhất một nơi.

Bài 1.90. Trong một hộp có 9 viên bi, trong đó có 3 bi đỏ còn lại là bi xanh. Lần lượt lấy ngẫu nhiên có hoàn lại 3 bi. Tìm xác suất để:

a) Lấy được 2 bi xanh.

b) Lấy được ít nhất 1 bi đỏ.

ĐS: 0,4444; 0,7037.

Bài 1.91. Trong 100 vé xổ số có 8 vé trúng thưởng. Một người mua 3 vé.

a) Tính xác suất để người đó trúng được 1 vé.

b) Tính xác suất để người đó trúng được 2 vé.

c) Tính xác suất để người đó trúng được ít nhất một vé.

Bài 1.92. Một khối lập phương có 6 mặt quét sơn được chia thành 1000 khối lập phương con đều nhau. Lấy ngẫu nhiên đồng thời ra 3 khối

a) Tính xác suất lấy được 1 khối có hai mặt quét sơn và 2 khối có 3 mặt quét sơn.

b) Tính xác suất lấy được cả 3 khối có 3 mặt quét sơn.

Bài 1.93. Gieo ngẫu nhiên hai con xúc xắc cân đối đồng chất. Hãy tính xác suất

a) Mặt 6 xuất hiện đúng một lần.

b) Mặt chẵn xuất hiện cả hai lần.

c) Tổng các mặt xuất hiện là 4.

d) Tổng các mặt xuất hiện chia hết cho 3.

Bài 1.94. Một lô hàng gồm 14 chiếc máy tính, trong đó có 3 chiếc hiệu IBM, 4 chiếc hiệu HP, 4 chiếc hiệu Samsung và 3 chiếc hiệu Sony. Chọn ngẫu nhiên đồng thời ra 4 chiếc máy tính. Hãy tính xác suất

a) Có đủ 4 nhãn hiệu của 4 hãng IBM, HP, Samsung và Sony.

b) Có 2 chiếc nhãn hiệu HP và 2 chiếc nhãn hiệu Samsung.

c) Không có chiếc nào có nhãn hiệu IBM và Sony.

Bài 1.95. Một hộp có 18 sản phẩm, trong đó các loại A, B và C lần lượt là 7, 6 và 5. Lấy ngẫu nhiên ra 8 sản phẩm từ hộp đó. Tính xác suất để trong 8 sản phẩm lấy ra mỗi loại A, B và C có ít nhất một sản phẩm.

Bài 1.96. Hộp thứ nhất có 2 viên bi xanh và 3 viên bi đỏ. Hộp thứ hai có 1 viên bi xanh và 2 viên bi đỏ. Từ mỗi hộp lấy ngẫu nhiên ra 1 viên bi. Sau đó chọn ngẫu nhiên 1 viên bi từ 2 viên bi lấy ra từ hộp. Tính xác suất để viên bi được chọn là bi đỏ.

Bài 1.97. Hộp thứ nhất có 4 viên bi xanh và 3 viên bi đỏ. Hộp thứ hai có 1 viên bi xanh và 2 viên bi đỏ. Từ hộp thứ nhất lấy ngẫu nhiên ra 1 viên bi bỏ qua hộp thứ hai. Sau đó chọn ngẫu nhiên 2 viên bi từ hộp thứ hai. Tính xác suất để lấy được 2 viên bi cùng màu.

Bài 1.98. Một hộp có 5 viên bi xanh và 10 viên bi đỏ. Thực hiện 9 lần lấy ra (có hoàn lại)một viên bi để xác định màu

a) Tính xác suất có 5 lần lấy được bi đỏ.

b) Tìm số bi đỏ có khả năng nhất.

Bài 1.99. Hộp thứ nhất có 2 sản phẩm loại A và 3 sản phẩm loại B. Hộp thứ hai có 1 sản phẩm loại A và 4 sản phẩm loại B. Lấy ngẫu nhiên 2 sản phẩm từ hộp thứ nhất bỏ sang hộp thứ hai. Sau đó lấy từ hộp thứ hai ra 2 sản phẩm

a) Tính xác suất để 2 sản phẩm lấy ra có 1 sản phẩm loại A và 1 sản phẩm loại B.

b) Tính xác suất để 2 sản phẩm lấy ra là 2 sản phẩm loại A.

c) Tính xác suất để có ít nhất một sản phẩm loại B.

Bài 1.100. Gieo 25 hạt giống có xác suất nảy mầm là 0.75

a) Tính xác suất để có 15 hạt nảy mầm.

b) Số hạt giống có khả năng nảy mầm nhất là bao nhiêu? Tính xác suất tương ứng.

c) Phải gieo bao nhiêu hạt giống loại đó để xác suất có một hạt nảy mầm không bé hơn 0.999.

Bài 1.101. Có hai chuồng thỏ thí nghiệm, chuồng thứ nhất có 2 thỏ trắng và 3 thỏ nâu, chuồng thứ hai có 1 thỏ trắng và 2 thỏ nâu. Tình cờ, 2 con thỏ từ chuồng thứ nhất chạy sang chuồng thứ hai. Bắt từ chuồng thứ hai ra 2 con thỏ

a) Tính xác suất để 2 con thỏ bắt ra là thỏ trắng.

b) Tính xác suất bắt ra 2 con thỏ khác màu.

c) Tính xác suất bắt được ít nhất một thỏ trắng.

Bài 1.102. Một bộ bài 52 quân bỏ đi 2 quân bài một cách ngẫu nhiên và chia đều cho 5 người chơi. Tính xác suất một người chơi được chia 2 quân át.

Bài 1.103. Một người có 3 chổ câu cá với xác suất để một lần thả câu sẽ câu được một con cá ở mỗi chỗ lần lược là 0.8, 0.7 và 0.9. Người đó chọn ngẫu nhiên một chổ để câu bằng cách gieo ngẫu nhiên 2 đồng xu cân đối và đồng chất. Nếu 2 đồng xu sấp thì chọn chỗ câu thứ nhất, nếu 2 đồng xu ngữa thì chọn chỗ câu thứ hai,trường hợp còn lại sẽ chọn chỗ câu thứ ba.

a) Nếu chọn được một chỗ để câu, tính xác suất để sau một lần thả câu người đó câu được một con cá.

b) Tại một chỗ câu đã chọn, người đó thả câu 3 lần thì được một con cá. Hỏi khả năng người đó đã câu ở địa điểm nào?

Bài 1.104. Có 3 người A, B và C lần lượt bắt 3 lá thăm trong đó có 2 lá thăm tốt và 1 lá thăm không tốt. Hỏi khả năng bắt được lá thăm tốt của A, B và C có như nhau không? tại sao?

Bài 1.105. Phải gieo một con xúc xắc cân đối đồng chất bao nhiêu lần để xác suất có ít nhất một lần mặt "lục" xuất hiện lớn hơn $\frac{1}{2}$.

Bài 1.106. Giả sử tỷ lệ làm ra chính phẩm của một máy là 99%. Hỏi phải làm ra bao nhiêu sản phẩm để xác suất máy đó làm ra ít nhất một chính phẩm là 95%.

Bài 1.107. Xác suất để thời tiết thuận lợi cho giống lúa A là 0.9. Nếu thời tiết thuận lợi thì xác suất để giống lúa A đạt năng suất cao là 0.85. Nếu thời tiết không thuận lợi thì với xác suất 0.2 giống lúa A sẽ đạt năng suất cao.

a) Tính xác suất để giống lúa A đạt năng suất cao.

b) Giả sử giống lúa A không đạt năng suất cao, hãy tính xác suất để thời tiết không thuận lợi cho giống lúa A.

Bài 1.108. Ba bức điện được truyền theo một kênh thông tin với mức độ chính xác khác nhau. Cụ thể mỗi bức điện đều có một trong ba khả năng sau

a) A_1: bức điện được truyền đúng.

b) A_2: bức điện được truyền sai lệch một phần.

c) A_3: bức điện được truyền sai lệch hoàn toàn.

Với xác suất p_1, p_2 và p_3 mà $p_j \in (0, 1), p_1 + p_2 + p_3 = 1$. Giả sử các bức điện được truyền đúng hay sai lệch là độc lập với nhau. Hãy tính các xác suất

a) Cả 3 bức điện đều được truyền đúng.

b) Có ít nhất một bức điện bị truyền sai lệch hoàn toàn.

Bài 1.109. Có hai chuồng gà nằm cạnh nhau. Chuồng thứ nhất có 18 gà mái và 2 gà trống, chuồng thứ 2 có 15 gà mái và 5 gà trống. Bất ngờ 2 con gà từ chuồng thứ hai nhảy sang chuồng thứ nhất. Người ta bắt ngẫu nhiên 2 con gà từ chuồng thứ nhất bỏ vào lại chuồng thứ hai.

a) Tính xác suất bắt được 2 con gà trống.

b) Tính xác suất bắt được 1 gà trống và một gà mái.

c) Tính xác suất bắt được ít nhất 1 con gà trống.

Bài 1.110. Tại một nhà máy sản suất bóng đèn điện tử xác suất làm ra một bóng đèn đạt tiêu chuẩn là 0.8. Trước khi xuất xưởng các bóng đèn cần phải đóng dấu kiểm tra chất lượng. Vì kiểm tra không chặt nên mỗi bóng đạt tiêu chuẩn có xác suất được đóng dấu là 0.9 còn mỗi bóng không đạt tiêu chuẩn có xác suất được đóng dấu là 0.05

a) Chọn ngẫu nhiên một bóng đèn. Tính xác suất bóng đèn đó được đóng dấu kiểm tra chất lượng.

b) Giả sử chọn được một bóng đã có đóng dấu kiểm tra chất lượng. Tính xác suất để bóng đèn đó là loại bóng đèn đạt tiêu chuẩn.

Bài 1.111. Một ca thợ gồm 3 công nhân sản xuất cùng một loại sản phẩm với số sản phẩm làm ra tỷ lệ với 3:4:5 và với tỷ lệ phế phẩm tương ứng là 2%:2.5%:3%. Chọn ngẫu nhiên ra một sản phẩm do ca thợ đó sản suất

a) Tìm xác suất để sản phẩm chọn ra là phế phẩm.

b) Nếu kiểm tra thấy đó là phế phẩm, hãy tính xác suất để phế phẩm do công nhân thứ i sản xuất ra ($i = 1, 2, 3$).

Bài 1.112. Xác suất để một xạ thủ bắn trúng mục tiêu ở mỗi lần bắn là 0.4. Tính xác suất mục tiêu bị tiêu diệt sau 3 lần bắn độc lập, biết rằng xác suất mục tiêu bị tiêu diệt khi trúng 1, 2, 3 phát lần lượt là 0.2, 0.5 và 0.8.

Bài 1.113. Từ một lô hàng có 5 sản phẩm loại I và 2 sản phẩm loại II. Lấy ngẫu nhiên đồng thời từ lô hàng đó ra 3 sản phẩm.

a) Tìm xác suất để trong 3 sản phẩm lấy ra có 2 sản phẩm loại I và 1 sản phẩm loại II.

b) Tìm xác suất để trong 3 sản phẩm lấy ra có ít nhất 1 sản phẩm loại I.

Bài 1.114. Một em bé trong túi trái có 5 bi đỏ và 4 bi xanh, trong túi phải có 6 bi đỏ và 3 bi xanh. Lấy ngẫu nhiên hai viên bi từ túi trái bỏ qua túi phải, rồi lấy ngẫu nhiên từ túi phải ra hai viên bi.

a) Tính xác suất để 2 viên bi lấy ra lần sau là 2 bi đỏ.

b) Tính xác suất để 2 viên bi lấy ra lần sau là 2 bi cùng màu.

a) Tính xác suất để 2 viên bi lấy ra lần sau là 2 bi khác màu.

Bài 1.115. Có 2 lô sản phẩm. Lô thứ nhất có 8 sản phẩm loại I và 3 sản phẩm loại II; lô thứ hai có 15 sản phẩm loại I và 5 sản phẩm loại II. Từ mỗi lô lấy ngẫu nhiên ra một sản phẩm. Sau đó lấy ngẫu nhiên 1 sản phẩm từ 2 sản phẩm vừa lấy ra trước. Tìm khả năng để sản phẩm lấy ra sau cùng là sản phẩm loại I.

Bài 1.116. Xác suất để một xạ thủ bắn trúng bia là 0.95. Hỏi xạ thủ này phải bắn ít nhất bao nhiêu viên để với xác suất không bé hơn 0.99 xạ thủ này bắn trúng bia ít nhất một viên.

Bài 1.117. Một máy sản xuất xác chi tiết điện tử với xác suất phế phẩm là 0.015.

a) Hãy tính xác suất để sản xuất 300 chi tiết điện tử, phát hiện ra 5 phế phẩm.

b) Tính số phế phẩm có khả năng nhất.

c) Phải sản xuất bao nhiêu chi tiết điện tử loại đó để xác suất có ít nhất một chi tiết không bị hỏng không bé hơn 0.9999.

Chương 2

Biến ngẫu nhiên và các đặc trưng

2.1 Biến ngẫu nhiên

✠ Một trong những khái niệm cơ sở của lí thuyết xác suất là khái niệm về biến ngẫu nhiên (BNN) (còn được gọi là đại lượng ngẫu nhiên). Ta có thể hiểu về BNN như sau:

"Biến" là cái có thể thay đổi "Ngẫu nhiên" là khi người ta chưa xác định được cái gì đó, thì người ta gọi nó là ngẫu nhiên. Còn khi xác định được thì trở thành "định tính" hết ngẫu nhiên. Một biến có thể là ngẫu nhiên với người này, nhưng không ngẫu nhiên với người khác tùy vào lượng thông tin nhận được. "Biến ngẫu nhiên" là những biến mà giá trị của nó được xác định một cách ngẫu nhiên.

✠ Tuy nhiên đó là cách hiểu thông thường, đại khái và chưa chặt chẽ về mặt toán học. Ta sẽ định nghĩa BNN theo ngôn ngữ toán học như sau:

BNN là một số được gán cho từng kết quả của phép thử. Do các kết quả của phép thử được ghi bởi các biến cố sơ cấp, nên BNN có thể coi như hàm $X = X(\omega)$ trên không gian các biến cố sơ cấp Ω và nhận mỗi giá trị tương ứng với một xác suất nào đó. Cụ thể BNN X của một phép thử với không gian mẫu Ω là một ánh xạ

$$
\begin{aligned}
X: \ \Omega &\to \mathbb{R} \\
\omega &\mapsto X(\omega).
\end{aligned}
$$

✠ BNN coi như được xác định nếu như ta biết được 2 yếu tố sau:

- Tập các giá trị của BNN.
- Các xác suất mà BNN nhận giá trị thuộc tập đó.

✠ Căn cứ vào tập các giá trị của BNN mà ta chia BNN thành 2 loại

- BNN rời rạc nếu tập các giá trị của nó là hữu hạn hoặc vô hạn đếm được (tập giá trị có thể đánh số thứ tự).
- BNN liên tục nếu tập các giá trị của nó là vô hạn không đếm được (tập giá trị lấp đầy một khoảng số thực nào đó).

✠ BNN thường được kí hiệu là $X, Y, Z, ...$ hoặc $X1, X2, ...$

✠ Chú ý: Không phải mọi hàm xác định trên không gian các biến cố sơ cấp Ω là BNN mà thực chất vấn đề là cần trả lời câu hỏi: xác suất để giá trị của BNN X thuộc tập hợp này hay tập hợp khác trên $\mathbb{R}$, tức là xét $P(\omega : X(\omega) \in B)$ với B là lớp các tập thuộc . Từ đó dẫn đến các định nghĩa chính xác hơn của BNN mà nội dung của nó được trình bày trong các tài liệu chuyên ngành.

2.2 Luật phân phối xác suất của biến ngẫu nhiên

\+ Luật phân phối xác suất của biến ngẫu nhiên (BNN) là một biểu đồ, trong đó chỉ ra

- Các giá trị có thể nhận được của BNN.
- Xác suất tương ứng để BNN nhận các giá trị đó.

\+ Luật phân phối xác suất thường được thể hiện dưới hai hình thức: hàm mật độ xác suất và hàm phân phối xác suất.

\+ Hàm mật độ xác suất của BNN rời rạc thường được thể hiện dưới dạng bảng phân phối xác xuất, như sau

X	x_1	x_2	x_3	.....	x_n
p	p_1	p_2	p_3	.....	p_n

với x_i là các giá trị của X; $p_i = P\left(X = x_i\right) : \sum\limits_{i=1}^{n} p_i = 1$ và $p_i > 0,\ \ i = \overline{1, n}$.
Nếu $c \notin \{x_1, x_2, ..., x_n\}$ thì $P(X = c) = 0$.

\+ Hàm mật độ xác suất của BNN liên tục được biểu thị bởi đồ thị của hàm số $y = f(x)$ xác định trên $(-\infty; +\infty)$, thỏa mãn

- $f\left(x\right) \geq 0, \forall x \in \mathbb{R}$
- $\int\limits_{-\infty}^{+\infty} f\left(x\right) dx = 1$

\+ Tính chất của hàm mật độ

- Nếu f liên tục tại điểm x_0 thì $F'(x_0) = f(x_0)$.
- Với mọi số thực a và b thõa $a < b$ ta có

$$P\left(a < X < b\right) = P\left(a \leq X < b\right) = P\left(a < X \leq b\right) = P\left(a \leq X \leq b\right) = \int\limits_a^b f\left(t\right) dt$$

- Với mọi $a \in \mathbb{R}$, $P(X = a) = 0$.

2.3 Hàm phân phối

1. Định nghĩa

Hàm F được xác định bởi: $\forall x \in \mathbb{R},\ F\left(x\right) = P\left(X < x\right)$ được gọi là hàm phân phối xác suất của BNN X, hay gọi ngắn gọn là hàm phân phối.

a. Với X là BNN rời rạc có bảng phân phối xác suất là

X	x_1	x_2	x_3	.....	x_n
p	p_1	p_2	p_3	.....	p_n

thì hàm phân phối $F(x) = \sum\limits_{x_i < x} P(X = x_i)$. Cụ thể

$$F(x) = \begin{cases} 0 & \text{nếu} & x \le x_1 \\ p_1 & \text{nếu} & x_1 < x \le x_2 \\ p_1 + p_2 & \text{nếu} & x_2 < x \le x_3 \\ \dots & \dots & \dots \\ p_1 + p_2 + \dots + p_k & \text{nếu} & x_k < x \le x_{k+1} \\ \dots & \dots & \dots \\ p_1 + p_2 + \dots + p_{n-1} & \text{nếu} & x_{n-1} < x \le x_n \\ \sum_{i=1}^{n} p_i = 1 & \text{nếu} & x_n \le x \end{cases}$$

b. Với X là BNN liên tục có hàm mật độ xác suất là $f(x)$ thì hàm phân phối xác suất $F(x) = \int\limits_{-\infty}^{x} f(t)\, dt$. Cụ thể

- Nếu X có $f(x) = \begin{cases} \varphi(x) & \text{khi} & x \in [a,\, b] \\ 0 & \text{khi} & x \notin [a,\, b] \end{cases}$

$$\text{thì} \quad F(x) = \begin{cases} 0 & \text{khi} & x < a \\ \int\limits_a^x \varphi(t)\, dt & \text{khi} & a \le x \le b \\ 1 & \text{khi} & x > b \end{cases}$$

- Nếu X có $f(x) = \begin{cases} 0 & \text{khi} & x < a \\ \varphi(x) & \text{khi} & x \ge a \end{cases}$

$$\text{thì} \quad F(x) = \begin{cases} 0 & \text{khi} & x < a \\ \int\limits_a^x \varphi(t)\, dt & \text{khi} & x \ge a \end{cases}$$

2. Tính Chất

a. $\forall (x_1, x_2) \in \mathbb{R}^2, (x_1 \le x_2 \Rightarrow F(x_1) \le F(x_2))$

b. $\begin{cases} \lim\limits_{x \to -\infty} F(x) = 0 \\ \lim\limits_{x \to +\infty} F(x) = 1 \end{cases}$

c. $F(x)$ liên tục trái trên $\mathbb{R}$.

d. $F(x)$ là hàm đơn điệu không giảm.

e. Cho X là BNN rời rạc, với mọi a, b thõa $a < b$ ta có

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

ii. $P(X = x_i) = F(x_{i+1}) - F(x_i)$

f. Cho X là BNN liên tục, với mọi a, b thõa $a < b$ ta có

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

g. $P(X = a) = F(a^+) - F(a)$ với mọi $a \in \mathbb{R}$, trong đó $F(a^+) = \lim\limits_{x \to a^+} f(x)$.

h. Nếu X là BNN liên tục có hàm mật độ xác suất $f(x)$ thì $F'(x) = f(x)$.

2.4 Các số đặc trưng của một biến ngẫu nhiên

1. Kỳ vọng (kí hiệu $E(X)$)

a. Định nghĩa
Ta định nghĩa kỳ vọng cho từng loại biến ngẫu nhiên như sau

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

X	x_1	x_2	x_3	.....	x_n
p	p_1	p_2	p_3	.....	p_n

thì kỳ vọng của X được xác định bằng công thức

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

- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là $f(x)$ thì kỳ vọng của X được xác định bằng công thức

$$E\left(X\right) = \int\limits_{-\infty}^{+\infty} x f\left(x\right) dx$$

b. Tính chất

Cho C là một hằng số, X và Y là hai biến ngẫu nhiên. Từ định nghĩa kỳ vọng ta rút ra được các tính chất sau

- $E(C) = C$.
- $E(C.X) = C.E(X)$.
- $E(X \pm Y) = E(X) \pm E(Y)$.
- Nếu X và Y là hai biến ngẫu nhiên độc lập thì $E(X.Y) = E(X).E(Y)$.
- Nếu $X \leq Y$ thì $E(X) \leq E(Y)$.

c. ý nghĩa

Kỳ vọng của một biến ngẫu nhiên là con số đặc trưng cho giá trị trung bình của biến ngẫu nhiên đó.

2. Phương sai (kí hiệu $V(X)$ hoặc $V(X)$)

a. Định nghĩa

Phương sai của biến ngẫu nhiên X bằng trung bình của bình phương sự chênh lệch của những giá trị của X so với trung bình của nó.

- Phương sai của X được xác định bằng biểu thức sau

$$D(X) = E[X - E(X)]^2 = E\left(X^2\right) - [E(X)]^2$$

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

X	x_1	x_2	x_3	.....	x_n
p	p_1	p_2	p_3	.....	p_n

thì phương sai của X được xác định bằng công thức

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

- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là $f(x)$ thì phương sai của X được xác định bằng công thức

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

b. Tính chất

Cho C là một hằng số, X và Y là hai biến ngẫu nhiên. Từ định nghĩa kỳ vọng ta rút ra được các tính chất sau

- $D(C) = 0$.
- $D(C.X) = C^2.D(X)$.
- Nếu X và Y là hai biến ngẫu nhiên độc lập thì $D(X \pm Y) = D(X) \pm D(Y)$.

c. ý nghĩa
Phương sai của một biến ngẫu nhiên là con số đặc trưng cho sự phân tán của biến ngẫu nhiên quanh kỳ vọng của nó.

3. Độ lệch chuẩn (kí hiệu $\sigma(X)$)
Dù rằng phương sai biểu thị sự phân tán của các biến ngẫu nhiên, tuy nhiên lại không cùng đơn vị với các biến ngẫu nhiên đó. Chính vì thế người ta đưa ra một tham số mới cũng có ý nghĩa giống như phương sai, nhưng cùng đơn vị với biến ngẫu nhiên. Đại lượng này gọi là độ lệch chuẩn. Biểu thức xác định độ lệch chuẩn
$$\sigma(X) = \sqrt{D(X)}$$

4. Mode (kí hiệu $Mod(X)$)

a. Nếu X là biến ngẫu nhiên rời rạc thì Mode của X là giá trị mà tại đó $P[X = Mot(X)]$ là lớn nhất.

b. Nếu X là biến ngẫu nhiên liên tục thì Mode của X là giá trị mà tại đó hàm mật độ $f(x)$ đạt cực đại.

c. Mode của biến ngẫu nhiên X được hiểu theo nghĩa là giá trị tin chắc nhất. Để tìm Mode của một biến ngẫu nhiên liên tục ta tìm cực đại của hàm mật độ xác suất của biến ngẫu hiên đó. Do đó biến ngẫu nhiên có thể có 1 hay nhiều giá trị Mode và có thể không có Mode nào.

5. Median (kí hiệu $Med(X)$)
Median hay còn gọi là trung vị là giá trị nằm chính giữa phân phối xác suất của biến ngẫu nhiên. Nói cách khác đó là giá trị chia phân phối của biến ngẫu nhiên thành 2 phần bằng nhau.

a. Với X là biến ngẫu nhiên rời rạc thì $Med(X) = x_i$ nếu
$$F(x_i) \leq \frac{1}{2} \leq F(x_{i+1}),\ x_i \in X(\Omega).$$

b. Với X là biến ngẫu nhiên liên tục thì $Med(X) = x_0$ nếu
$$F(x_0) = \int\limits_{-\infty}^{x_0} f(x)\,dx = \frac{1}{2}$$

c. Chú ý:

- Median là không duy nhất, có thể có nhiều Median.
- Nếu phân phối của biến ngẫu nhiên X đối xứng và có một Mode thì cả 3 đặc trưng: kỳ vọng, Median và Mode là trùng nhau.
- Nếu phân phối của X đối xứng, hoặc gần đối xứng thì dùng kỳ vọng định vị là tốt nhất.
- Nếu phân phối của X quá lệch thì dùng Median và Mode định vị sẽ tốt hơn.

2.5 Một số phân phối xác suất thường gặp

2.5.1 Phân phối nhị thức

a) Biến ngẫu nhiên X được gọi là có phân phối nhị thức với các tham số n, p $(n \in Z_+^*; 0 < p < 1)$, kí hiệu $X \sim B(n, p)$, nếu X nhận các giá trị $k = 0, 1, 2, \ldots, n$ với xác suất tương ứng là:

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

b) Hàm đặc trưng: $\varphi(t) = \left[1 + p\left(e^{it} - 1\right)\right]^n$.

c) Kỳ vọng và phương sai:

$$\mathbb{E}(X) = np, \quad D(X) = np(1-p).$$

2.5.2 Phân phối Poisson-binomial

a) Cho X_j là biến ngẫu nhiên Bernoulli với xác suất thành công tương ứng là $P(X_j = 1) = 1 - P(X_j = 0) = P_j; (0 < p_j < 1)$. Đặt $S_n = \sum\limits_{j=1}^{n} X_j$, nếu các X_j độc lập thì S_n được gọi là biến ngẫu nhiên tuân theo phân phối Poisson-binomial với tham số $\lambda_n = \sum\limits_{j=1}^{n} p_j$. Kí hiệu $X \sim Poisson - binomial\,(\lambda_n)$.

b) Hàm đặc trưng: $\varphi(t) = \prod\limits_{j=1}^{n} \left[1 + p_j\left(e^{it} - 1\right)\right]$.

c) Kỳ vọng và phương sai:

$$\mathbb{E}(S_n) = \sum_{j=1}^{n} p_j, \quad D(S_n) = \sum_{j=1}^{n} p_j(1-p_j).$$

Chú ý 2.1. *Khi $p_1 = p_2 = \ldots = p_n = p$ thì biến ngẫu nhiên S_n có phân phối nhị thức với tham số n, p. Hay nói cách khác phân phối nhị thức là trường hợp đặc biệt của phân phối Poisson-binomial.*

2.5.3 Phân phối Poisson

a) Biến ngẫu nhiên X được gọi là có phân phối Poisson với tham số dương λ, kí hiệu $X \sim Poisson\,(\lambda)$, nếu X nhận các giá trị $k = 0, 1, 2, \ldots$ với xác suất tương ứng là:

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

b) Hàm đặc trưng: $\varphi\,(t) = e^{\lambda\left(e^{it}-1\right)}$.

c) Kỳ vọng và phương sai:

$$\mathbb{E}\,(X) = D\,(X) = \lambda.$$

Chú ý 2.2. *Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên rời rạc, độc lập nhau có phân phối Poisson với tham số $\lambda_1, \lambda_2, \ldots, \lambda_n$ thì biến ngẫu nhiên $X = X_1 + X_2 + \ldots + X_n$ cũng có phân phối Poisson với tham số $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.*

2.5.4 Phân phối Poisson phức hợp

a) Cho dãy $(X_j)_{j\geq 1}$ là các biến ngẫu nhiên độc lập cùng phân phối có kỳ vọng μ và phương sai σ^2. Cho W là biến ngẫu nhiên có phân phối Poisson với tham số λ, khi đó đại lượng $S = \sum\limits_{j=1}^{W} X_j$ sẽ có phân phối được gọi là phân phối Poisson phức hợp.

b) Hàm đặc trưng: $\varphi\,(t) = e^{\lambda(\psi(t)-1)}$, trong đó $\psi(t)$ là hàm đặc trưng của X_j.

c) Kỳ vọng và phương sai:

$$E\,(S) = \lambda\mu, \quad D\,(S) = \lambda\left(\sigma^2 + \mu^2\right).$$

2.5.5 Phân phối hình học dạng thứ nhất

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối hình học với tham số dương $p (0 < p < 1)$, kí hiệu $X \sim Geometry(p)$, nếu X nhận các giá trị $k = 1, 2, ...$ với xác suất tương ứng là:

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

b) Hàm đặc trưng và hàm sinh

$$\varphi_X(t) = \frac{pe^{it}}{1-(1-p)\,e^{it}}; \quad \psi_X(t) = \frac{pt}{1-(1-p)\,t}$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{1}{p}; \;\; D(X) = \frac{1-p}{p^2}$$

d) Ý nghĩa: giá trị $P(X=k)$ là xác suất để xuất hiện lần thành công đầu tiên ở phép thử thứ k. Ở đây ta hiểu k là số phép thử để nhận được lần thành công đầu tiên.

2.5.6 Phân phối hình học dạng thứ hai

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối hình học với tham số dương $p(0<p<1)$,kí hiệu $X \sim Geometry(p)$, nếu X nhận các giá trị $k = 1, 2, ...$ với xác suất tương ứng là:

$$P(X=k) = p(1-p)^k.$$

b) Hàm đặc trưng và hàm sinh

$$\varphi_X(t) = \frac{p}{1-(1-p)\,e^{it}}; \quad \psi_X(t) = \frac{p}{1-(1-p)\,t}$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{1-p}{p}; \;\; D(X) = \frac{1-p}{p^2}$$

d) Ý nghĩa: giá trị $P(X=k)$ là xác suất để xuất hiện lần thành công đầu tiên ở phép thử thứ $k+1$. Ở đây ta hiểu k là số lần thất bại trước lần thành công đầu tiên.

2.5.7 Phân phối nhị thức âm dạng thứ nhất

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối nhị thức âm với tham số dương $r, p(r = 1, 2, ...; 0 < p < 1)$, kí hiệu $X \sim NB(r,p)$, nếu X nhận các giá trị $k = r, r+1, ...$ với xác suất tương ứng là:

$$P(X=k) = C_{k-1}^{r-1} p^r (1-p)^{k-r}, \;\; k \geq r$$

b) Hàm đặc trưng và hàm sinh

$$\varphi_X(t) = \left[\frac{pe^{it}}{1-(1-p)\,e^{it}}\right]^r; \quad \psi_X(t) = \left[\frac{pt}{1-(1-p)\,t}\right]^r$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{r}{p}; \;\; D(X) = \frac{r(1-p)}{p^2}$$

d) Ý nghĩa: giá trị $P(X = k)$ là xác suất để nhận được đủ rlần thành công ở phép thử thứ k. Ở đây ta hiểu k là số phép thử để nhận được đủ r lần thành công.

e) Chú ý: khi $r = 1$ thì phân phối nhị thức âm dạng thứ nhất chính là phân phối hình học dạng thứ nhất. Các vấn đề liên quan đến tổng hình học được các nhà toán học sử dụng dạng phân phối này.

2.5.8 Phân phối nhị thức âm dạng thứ hai

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối nhị thức âm với tham số dương $r, p (r = 1, 2, ...; 0 < p < 1)$, kí hiệu $X \sim NB(r, p)$, nếu X nhận các giá trị $k = 1, 2, ...$ với xác suất tương ứng là:

$$P(X = k) = C_{r+k-1}^k p^r (1-p)^k$$

b) Hàm đặc trưng và hàm sinh

$$\varphi_X(t) = \left[\frac{p}{1-(1-p)\,e^{it}}\right]^r; \quad \psi_X(t) = \left[\frac{p}{1-(1-p)\,t}\right]^r$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{r(1-p)}{p}; \;\; D(X) = \frac{r(1-p)}{p^2}$$

d) Ý nghĩa: giá trị $P(X = k)$ là xác suất để nhận được đủ rlần thành công ở phép thử thứ $k + r$. Ở đây ta hiểu k là số lần thất bại trước lần thành công thứ r.

e) Chú ý: khi $r = 1$ thì phân phối nhị thức âm dạng thứ hai chính là phân phối hình học dạng thứ hai. Các vấn đề liên quan đến bài toán xấp xỉ Poisson được các nhà toán học sử dụng dạng phân phối này.

2.5.9 Phân phối Gamma

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối Gamma với hai tham số dương r và λ, kí hiệu $X \sim G(r, \lambda)$, nếu X có hàm mật độ là:

$$f(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & \text{khi} \quad x > 0 \\ 0 & \text{khi} \quad x \le 0 \end{cases}$$

Trong đó hàm Γ được xác định như sau

+ $\Gamma(1) = 1$.

+ Nếu $r > 1$ thì $\Gamma(r) = \int\limits_0^{+\infty} x^{r-1} e^{-x} dx = (r-1)\Gamma(r-1) = (r-1)!$.

+ Ngoài ra $\Gamma(\dfrac{1}{2}) = \sqrt{\pi}$.

b) Hàm đặc trưng

$$\varphi_X(t) = \left[\frac{\lambda}{\lambda - it}\right]^r$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{r}{\lambda}; \quad D(X) = \frac{r}{\lambda^2}.$$

d) Chú ý: Chúng ta có một số kết quả sau

+ Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối Gamma với tham số r_i và $\lambda, i = 1, 2, ..., n$ thì biến ngẫu nhiên $S_n = X_1 + X_2 + ... + X_n$ có phân phối Gamma với tham số $r_1 + r_2 + ... + r_n$ và λ.

+ Khi $r = 1$ thì phân phối Gamma được gọi là phân phối mũ.

+ Khi $r = \frac{n}{2}$ và $\lambda = \frac{1}{2}$ thì phân phối Gamma được gọi là phân phối chi bình phương.

2.5.10 Phân phối mũ

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối mũ với tham số dương λ, kí hiệu $X \sim Exp(\lambda)$, nếu X có hàm mật độ là:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{khi} \quad x \ge 0 \\ 0 & \text{khi} \quad x < 0 \end{cases}$$

b) Hàm đặc trưng

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{1}{\lambda}; \quad D(X) = \frac{1}{\lambda^2}.$$

2.5.11 Phân phối chi bình phương

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối chi bình phương với n bậc tự do, kí hiệu $X \sim \chi^2(n)$, nếu X có hàm mật độ là:

$$f(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} & \text{khi } \; x \geq 0 \\ 0 & \text{khi } \; x < 0 \end{cases}$$

b) Hàm đặc trưng

$$\varphi_X(t) = \left[\frac{1}{1 - 2it}\right]^{\frac{n}{2}}$$

c) Kỳ vọng và phương sai

$$E(X) = n; \quad D(X) = 2n.$$

2.5.12 Phân phối Laplace dạng đối xứng

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối Laplace với các tham số dương α, λ, kí hiệu $X \sim L(\alpha, \lambda)$, nếu X có hàm mật độ là:

$$f(x) = \frac{\lambda}{2}e^{-\lambda|x-\alpha|} \quad -\infty < x < +\infty.$$

b) Hàm đặc trưng

$$\varphi_X(t) = \frac{\lambda^2}{t^2 + \lambda^2}e^{i\alpha t}.$$

c) Kỳ vọng và phương sai

$$E(X) = \alpha; \quad D(X) = \frac{2}{\lambda^2}.$$

2.5.13 Phân phối Laplace dạng không đối xứng

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối Laplace với các tham số dương m, a, σ, kí hiệu $X \sim L(m, a, \sigma)$, nếu X có hàm đặc trưng là:

$$\varphi_X(t) = \frac{e^{imt}}{1 - iat + \frac{\sigma^2 t^2}{2}}$$

b) Kỳ vọng và phương sai

$$E(X) = m + a; \quad D(X) = a^2 + \sigma^2.$$

2.5.14 Phân phối chuẩn

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối chuẩn với các tham số dương μ, σ^2, kí hiệu $X \sim N(\mu, \sigma^2)$, nếu X có hàm mật độ là

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{-(x-\mu)^2}{2\sigma^2} \right\} \qquad -\infty < x < \infty$$

b) Hàm đặc trưng

$$\varphi(t) = \exp\left\{ i\mu t - \frac{t^2\sigma^2}{2} \right\}$$

c) Kỳ vọng và phương sai

$$E(X) = \mu; \quad D(X) = \sigma^2.$$

2.5.15 Phân phối đều

a) Định nghĩa: biến ngẫu nhiên X được gọi là có phân phối đều trên đoạn $[a, b]$ $(a < b)$, kí hiệu là $X \sim U[a, b]$ nếu X có hàm mật độ là

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

b) Hàm đặc trưng

$$\varphi(t) = \frac{1}{(b-a)\, it} \left[\exp(itb) - \exp(ita) \right]$$

c) Kỳ vọng và phương sai

$$E(X) = \frac{b^2 - a^2}{2(b-a)}; \quad D(X) = \frac{(b-a)^2}{12}.$$

Một số ví dụ

Ví dụ 2.1 Các biến sau là biến ngẫu nhiên rời rạc:

- Số chấm xuất hiện trên mặt con xúc xắc khi tung một con xúc xắc.
- Số học sinh vắng mặt trong một buổi .
- Số sản phẩm tốt khi mua một lô hàng.

Ví dụ 2.2 Các biến sau là biến ngẫu nhiên liên tục:

- Nhiệt độ không khí tại một thời điểm nào đó.
- Sai số khi đo lường một đại lượng vật lý nào đó.
- Khoảng thời gian giữa 2 ca cấp cứu của một bệnh viện.

Ví dụ 2.3 Trên một cái kệ có 6 cuốn toán và 4 cuốn lý, chọn ngẫu nhiên 3 cuốn sách. Lập bảng phân phối xác suất của số sách toán chọn được.

Ví dụ 2.4 Giả sử tuổi thọ của một loại côn trùng là biến ngẫu nhiên liên tục X (Đơn vị hàng tháng) có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} 0 & \text{khi} \quad x \notin [0;2] \\ m(x-2) & \text{khi} \quad x \in [0;2] \end{cases}$$

a) Tìm tham số m.

b) Tính tỷ lệ côn trùng chết trước khi nó được 1 tháng tuổi.

ĐS: a) -1/2; b) 75%.

Ví dụ 2.5 Một người hằng ngày từ nhà đến cơ quan phải qua 4 ngã tư. Xác suất gặp đèn đỏ ở mỗi ngã tư là 25%. Lập hàm phân phối xác suất số lần gặp đèn đỏ của người đó.

Ví dụ 2.6 Cho biến ngẫu nhiên liên tục có hàm mật độ xác suất:

$$f(x) = \begin{cases} 0 & \text{khi} \quad x \le 0 \\ x & \text{khi} \quad 0 < x \le 1 \\ 2-x & \text{khi} \quad 1 < x \le 2 \\ 1 & \text{khi} \quad x > 2 \end{cases}$$

Tìm hàm phân phối xác suất của biến ngẫu nhiên X .

Ví dụ 2.7 Trong một trò chơi, mỗi người chơi được bắn 2 viên đạn. Nếu bắn trúng mỗi viên sẽ được 20000 đồng, nếu bắn trật mỗi viên sẽ bị mất 10000 đồng. Giả sử xác suất bắn trúng mỗi viên đạn là 40%. Nếu bạn tham gia cuộc thi này thì số tiền mà bạn nhiều khả năng có nhất là bao nhiêu?
ĐS: 10000.

Ví dụ 2.8 Xét lại ví dụ 2.7, hãy tìm số tiền trung bình mà một người có được qua cuộc thi trên.
ĐS: 4000.

Ví dụ 2.9 Giả sử thời gian sống của một loài sinh vật là biến ngẫu nhiên (đơn vị tính bằng giờ) có hàm mật độ xác suất

$$f(x) = \begin{cases} 0 & \text{khi } x \notin [0;2] \\ m(x-2) & \text{khi } x \in [0;2] \end{cases}$$

Tính thời gian sống trung bình của loài sinh vật đó.
ĐS: $\frac{2}{3}$.

Ví dụ 2.10 Năng suất của hai máy tương ứng là các biến ngẫu nhiên X, Y (đơn vị : sản phẩm/phút) có bảng phân phối xác suất như sau:

X	0	1	2	3
P	0,3	0,1	0,15	0,35

Y	1	2	3
P	0,1	0,5	0,4

Nếu phải chọn mua một trong hai máy này ta nên chọn máy nào?
ĐS: máy Y.

Ví dụ 2.11 Trọng lượng của một loại sản phẩm là X (đơn vị kg) có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} \frac{3}{2}\left(1 - x^2\right) & \text{khi} \quad x \in [0,1] \\ 0 & \text{khi} \quad x \notin [0,1] \end{cases}$$

Tìm phương sai và độ chênh lệch chuẩn của X .

Ví dụ 2.12 Ở một trạm xe buýt, chiếc xe đầu tiên khởi hành lúc 7 giờ và cứ 15 phút sẽ có một xe khác khởi hành. Một hành khách đến trạm xe buýt trong khoảng thời gian từ 8 giờ 30 phút. Tính xác suất để hành khách này phải đợi ít hơn 5 phút.
ĐS: $\frac{1}{3}$.

Ví dụ 2.13 Giả sử tuổi thọ của một loại mạch điện tử là biến ngẫu nhiên có phân phối mũ với tuổi thọ trung bình là 5 năm, thời gian bảo hành là 2 năm. Tính tỷ lệ những mạch điện tử bị thay thế trong thời gian bảo hành.
ĐS: 0,33.

Ví dụ 2.14 Giả sử trọng lượng sản phẩm do một nhà máy sản xuất là biến ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình 5 gam, phương sai $0.01gam^2$.

a) Tính tỷ lệ sản phẩm có trọng lượng từ 4,8 gam đến 5,1 gam.

b) Những sản phẩm có trọng lượng sai lệch so với trung bình nhỏ hơn 0,2 gam được xem là loại tốt. tính tỷ lệ sản phẩm loại tốt của nhà máy.

ĐS: a) 81,85%; b) 95,44%.

Ví dụ 2.15 Trọng lượng trẻ sơ sinh là biến ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình là 3kg, độ lệch chuẩn 0,2kg. biết trọng lượng trẻ sinh ra tối thiểu là 1.5kg.

a) Trẻ sơ sinh thiếu cân nếu trọng lượng nhỏ hơn 2,5kg. tính tỷ lệ trẻ thiếu cân.

b) Người ta muốn có chế độ chăm sóc đặc biệt cho 10% tổng số trẻ nhẹ cân nhất. tính trọng lượng tối đa những đứa trẻ được chăm sóc đặc biệt.

ĐS: a) 0,006; b) 2,744.

Ví dụ 2.16 Tỷ lệ những trái bưởi năm roi đạt tiêu chuẩn xuất khẩu ở khu vườn của ông A chỉ là 20%.

a) Nếu chọn ngẫu nhiên 3 trái bưởi trong khu vườn của ông A. tính xác suất để chọn được 1 trái bưởi đạt tiêu chuẩn.

b) Nếu chọn 100 trái. Tính xác suất trong các trường hợp sau:

   i. Có đúng 10 trái bưởi đạt tiêu chuẩn xuất khẩu.

   ii. Có từ 10 đến 40 trái bưởi đạt tiêu chuẩn xuất khẩu.

c) Nếu chọn 1000 trái bưởi, thì trung bình có bao nhiêu trái bưởi đạt tiêu chuẩn xuất khẩu.

ĐS: a) 0,386; b) 0,0044; 0,0228; c) 200.

Ví dụ 2.17 Một công ty sản xuất dược phẩm với tỷ lệ phế phẩm của những viên thuốc là 0,002. Nếu công ty sản xuất 1000 viên thuốc. Tính xác suất để có không quá 2 viên thuốc là phế phẩm.
ĐS: 0,6808.

Ví dụ 2.18 Một trường đại học gồm 10000 sinh viên, trong đó có 1000 sinh viên học kém. Chọn ngẫu nhiên 100 sinh viên để kiểm tra. Tính xác suất trong các trường hợp sau:

a) Chọn được 20 sinh viên học kém.

b) Chọn được từ 10 đến 60 sinh viên học kém.

ĐS: a) 0,00057; b) 0,5.

Ví dụ 2.19 Một hộp có 10 sản phẩm trong đó có 4 phế phẩm. lấy ngẫu nhiên 2 sản phẩm từ hộp để kiểm tra. Gọi X là số phế phẩm lấy được. Tìm phân phối xác suất của X . Tính $P\{-1 \leq X \leq 1\}$; $P\{-3 \leq X < 0\}$; $P\{X \geq 0,1\}$.
ĐS: $\dfrac{39}{45}$; $\dfrac{2}{3}$.

Ví dụ 2.20 Cho biến ngẫu nhiên X có hàm mật độ xác suất:

$$f(x) = \begin{cases} 0 & \text{khi} \quad x < 1 \\ \dfrac{c}{x^4} & \text{khi} \quad x \geq 1. \end{cases}$$

Tìm c. tính $P\{-1 < X \leq 2\}$; $P\{X < 2\}$.
ĐS: 3; 7/8; 7/8.

Ví dụ 2.21 Một phân xưởng có 2 máy hoạt động động lập. Xác suất trong một ngày làm việc các máy đó hỏng tương ứng là 0,1 ; 0,2. Gọi X là số máy hỏng trong một ngày làm việc. lập hàm phân phối xác suất của X, vẽ đồ thị của nó.

Ví dụ 2.22 Tuổi thọ X của một thiết bị (đơn vị : giờ) có hàm mật độ xác suất:

$$f(x) = \begin{cases} \dfrac{100}{x^2} & \text{khi} \quad x \geq 100 \\ 0 & \text{khi} \quad x < 100 \end{cases}$$

a) Tìm hàm phân phối của X.

b) Thiết bị được gọi là loại A nếu tuổi thọ của nó kéo dài ít nhất 400 giờ. Tính tỉ lệ thiết bị loại A.

ĐS: b) 25%.

Ví dụ 2.23 Tuổi thọ của một loại máy là biến ngẫu nhiên X (đơn vị: năm) liên tục có hàm phân phối xác suất:

$$F(x) = \begin{cases} 0 & \text{khi} \quad x \leq 0 \\ ax^4 & \text{khi} \quad x \in (0;3) \\ 1 & \text{khi} \quad x \geq 3 \end{cases}$$

a) Tìm a và hàm mật độ xác suất của X.

b) Tìm xác suất để trong 3 máy loại này được sử dụng có 2 máy mà tuổi thọ của chúng không quá 2 năm.

ĐS: a) 1/81; b) 0,0939.

Ví dụ 2.24 Cho biến ngẫu nhiên X với hàm mật độ xác suất

$$f(x) = \begin{cases} \dfrac{2}{x^2} & \text{khi} \quad x \in [1,2] \\ 0 & \text{khi} \quad x \notin [1,2] \end{cases}$$

a) Tìm $E(X)$.

b) Tìm kỳ vọng của $Y = X^5 - \dfrac{2}{X}$.

ĐS: a) 2ln2; b) 6.

Ví dụ 2.25 Năng suất của hai máy tương ứng là các biến ngẫu nhiên X, Y (đơn vị: sản phẩm/phút) có bảng phân phối xác suất như sau:

X	1	2	3	4
P	0,3	0,1	0,5	0,1

Y	2	3	4	5
P	0,1	0,4	0,4	0,1

Nếu phải chọn mua một trong hai máy này ta nên chọn máy nào?
ĐS: máy Y.

Ví dụ 2.26 Trọng lượng của một loại sản phẩm là X (đơn vị kg) có hàm mật độ xác suất:

$$f(x) = \begin{cases} \frac{3}{16}\left(x^2 - 1\right) & \text{khi } 2 \le x \le 3 \\ 0 & \text{khi } x \notin [2,3]. \end{cases}$$

Tính trọng lượng trung bình và độ lệch tiêu chuẩn của X.
ĐS: a) 2,5781; b) 0,28.

Ví dụ 2.27 Cho biến ngẫu nhiên X có bảng phân phối xác suất:

X	-1	0	1	2
P	0,25	0,15	0,3	0,3

Tìm $Med(X)$; $P\{|X - EX| < 1\}$.
ĐS: 1; 0,7.

Ví dụ 2.28 Thời gian X (tính bằng phút) của một khách hàng chờ để được phục vụ tại một quầy hàng là biến ngẫu nhiên X có phân phối chuẩn với trung bình 4,5 và phương sai 1,21.

a) Tính tỉ lệ khách hàng phải chờ để được phục vụ từ 3,5 phút đến 6 phút; không quá 3,5 phút; quá 6 phút.

b) Thời gian phải chờ tối thiểu là bao nhiêu, nếu tỉ lệ khách hàng phải chờ phục vụ vượt quá thời gian đó không quá 5%?

Ví dụ 2.29 Một doanh nghiệp cần mua một loại trục máy có đường kính từ 18cm đến 1,22cm. Có hai nhà máy bán loại trục máy này và đường kính của các loại trục máy được sản xuất ra là biến ngẫu nhiên có phân phối chuẩn với các số đặc trưng cho trong bảng:

	Đường kính trung bình (cm)	Độ lệch chuẩn	Giá bán
Nhà máy 1	1,2	0,01	3triệu /1hộp/100chiếc
Nhà máy 2	1,2	0,015	2,7triệu/1hộp/100chiếc

Vậy doanh nghiệp nên mua trục máy của nhà máy nào?
ĐS: nhà máy 1.

Ví dụ 2.30 Cho Z có phân phối chuẩn chuẩn tắc (Z $N(0,1)$).

a) Tìm giá trị tới hạn mức 5% của Z.

b) Tìm mức xác suất α và giá trị tới hạn z_x của Z sao cho:

$$P\{-z_x < Z < z_x\} = 0,98.$$

ĐS: a) 1,65; b) 2,33.

Ví dụ 2.31 Một công ty bán 3 loại hàng : A, B, C với giá bán một đơn vị tương ứng là 21,2 ; 21,35 ; 21,5 (USD). Gọi $X - 1, X_2, X_3$ tương ứng là số đơn vị hàng bán của các loại hàng A, B, C trong 1 tuần. $X - 1, X_2, X_3$ là các biến ngẫu nhiên có phân phối chuẩn với trung bình $\mu_1 = 1000$, $\mu_2 = 500\, \mu_3 = 300$ và độ lệch chuẩn $\sigma_1 = 100$, $\sigma_2 = 80$, $\sigma_3 = 50$. Tính xác suất doanh thu Y của công ti trong 1 tuần vượt 45000 USD.
ĐS: 0,0513.

Ví dụ 2.33 Trọng lượng của một loại sản phẩm là biến ngẫu nhiên có trung bình 50g, độ lệch tiêu chuẩn 10g. Các sản phẩm được đóng thành hộp, mỗi hộp 100 sản phẩm. Hộp có trọng lượng trên 4,85kg là đạt tiêu chuẩn. Tính tỉ lệ hộp đạt tiêu chuẩn.
ĐS: 93,32%.

Ví dụ 2.34 Tuổi thọ X (tính bằng năm) của một mạch điện tử trong máy tính là biến ngẫu nhiên có phân phối mũ, trung bình 6,25. Thời gian bảo hành của mạch điện tử là 5 năm. Tính tỉ lệ mạch điện tử bán ra phải thay thế.
ĐS: 55,056%.

Ví dụ 2.35 Số khách đến mua hàng tại một quầy hàng là biến ngẫu nhiên có phân phối Poisson với trung bình 3 phút có 4 người.

a) Tính xác suất trong khoảng thời gian 3 phút có 2 khách đến quầy .

b) Tính xác suất trong khoảng thời gian 3 phút có khách đến quầy mua hàng.

c) Tìm số khách có thể hi vọng và số khách trung bình đến quầy mua hàng trong khoảng thời gian 3 phút.

d) Tính xác suất trong khoảng thời gian 30 giây có 2 khách đến quầy.

Ví dụ 2.36 Một người mỗi ngày đi bán hàng ở 5 nơi khác nhau. Xác suất bán được hàng ở mỗi nơi của người đó là 0,3. Tính :

a) Xác suất người đó bán được hàng tại 2 nơi trong một ngày.

b) Xác suất người đó bán được hàng trong một ngày.

c) Tìm số nơi đáng tin nhất và số nơi trung bình người đó bán được hàng trong một ngày.

d) Mỗi năm người đó đi bán hàng 300 ngày, tìm trung bình số ngày người đó bán được hàng trong một năm.

Ví dụ 2.37 Theo thống kê thì cứ 1000 sản phẩm thủy tinh được sản xuất ra bởi một dây chuyền công nghệ có 1 sản phẩm bị hỏng. Tính xác suất khi dây chuyền công nghệ này sản xuất ra 8000 sản phẩm có 30 sản phẩm bị hỏng.

Ví dụ 2.38 Một khách sạn nhận đặt phòng của 325 khách hàng cho 300 phòng vào ngày 8 tháng 3 vì theo kinh nghiệm của những năm trước cho thấy có 10% khách đặt phòng nhưng không đến. tính xác suất:

a) Có 280 khách dặt phòng và đến vào ngày 8 tháng 3 để nhận phòng.

b) Tất cả các khách đặt phòng đến vào ngày 8 tháng 3 đều nhận được phòng.

Ví dụ 2.39 Một sọt có 30 trái cam trong đó có 5 trái bị hỏng.

a) Tính xác suất trong 4 trái cam mua ngẫu nhiên từ sọt có 3 trái không hỏng.

b) Tính xác suất trong 10 trái cam mua ngẫu nhiên từ sọt có 6 trái không bị hỏng.

c) Tìm số cam bị hỏng trung bình khi mua 10 trái ngẫu nhiên từ sọt và phương sai của số cam hỏng này.

Ví dụ 2.40 Một nhà sản xuất thiết bị gạt nước ôtô cho rằng trong 5000 thiết bị loại này cung cấp cho một đại lí có 1000 thiết bị có khuyết tật. Tính xác suất trong 10 thiết bị được mua ngẫu nhiên tại đại lý có 3 thiết bị có khuyết tật. ĐS: 0,2013.

2.6 Bài tập có hướng dẫn

Bài 2.1 Cho bảng phân phối xác suất về lợi nhuận (nghìn USD) trong mười năm hoạt động của công ty MRA

x	-100	0	50	100	150	200
$f(x)$	0,10	0,20	0,30	0,25	0,10	

Lưu ý: Giá trị âm biểu thị giá trị thua lỗ.

a. Tìm giá trị $f(200)$ còn thiếu trong bảng và giải thích?

b. Tính xác suất công ty MRA có lãi?

c. Tính xác suất công ty MRA có lãi ít nhất là 100.000USD?

Giải.

a. Ta có $\sum_i f(x_i) = 1$. Từ đó suy ra

$$f(200) = 1 - f(-100) - f(0) - f(50) - f(100) - f(150) = 0,05.$$

b. Xác suất công ty MRA có lãi là

$$\sum_{\{i|\ x_i>0\}} f(x_i) = f(50) + f(100) + f(150) + f(200) = 0,70.$$

c. Xác suất công ty MRA có lãi ít nhất là 100.000USD cho cho bởi

$$\sum_{\{i|\ x_i\geq 100\}} f(x_i) = f(100) + f(150) + f(200) = 0,40.$$

Bài 2.2. Xét một phép thử Nhị thức với $n = 10$ và $p = 0,10$. Sử dụng bảng xác suất cho sẵn hãy tính

a. Tính $f(0)$.

b. Tính $f(2)$.

c. Tính $P(X \leq 2)$

d. Tính $P(X \geq 1)$.

e. Tính $E(X)$.

f. Tính $Var(X)$ và σ.

Giải. Dựa vào bảng tra $n = 10$ và $p = 0,10$ ta có

a. $f(0) = 0,3487$.

b. $f(2) = 0,1937$.

c. $P(X \leq 2) = f(0) + f(1) + f(2) = 0,3487 + 0,3874 + 0,1937 = 0,9298$

d. $P(X \geq 1) = 1 - P(X < 1) = 1 - f(0) = 1 - 0,3487 = 0,6513$.

e. $E(X) = np = 10.0,10 = 1$.

f. $Var(X) = np(1-p) = 10.0,1.0,9 = 0,9$ và $\sigma = \sqrt{Var(X)} = \sqrt{0,9} \approx 0,9487$.

Bài 2.3. Một cuộc khảo sát của Harris Interactive dành cho Khu nghỉ dưỡng và khách sạn InterContinental hỏi những người tham gia như sau "Khi đi du lịch quốc tế, bạn thường tự mình thám hiểm dựa trên kiến thức văn hóa của bản thân hay đi theo nhóm du lịch của bạn với một lịch trình nhất định". Cuộc khảo sát cho thấy có 23% người được hỏi trả lời sẽ đi theo nhóm du lịch của họ với một lịch trình. (Theo *USA Today*, ngày 21 tháng 1 năm 2004).

a. Trong một mẫu có 6 khách du lịch quốc tế, tính xác suất có đúng 2 người trả lời là đi theo nhóm du lịch của họ?

b. Trong một mẫu có 6 khách du lịch quốc tế, tính xác suất có ít nhất 2 người trả lời đi theo nhóm du lịch của họ?

c. Trong một mẫu có 10 khách du lịch quốc tế, tính xác suất không ai trả lời đi theo nhóm du lịch của họ?

Giải. Một khách du lịch được hỏi sẽ trả lời một trong hai đáp án, hoặc sẽ đi theo nhóm (kết quả thành công, xác suất p=0,23) hoặc tự thám hiểm (kết quả thất bại, xác suất 1-p=0,77). Như vậy, kết quả trả lời của một khách du lịch được xem như kết quả của một phép thử Bernolli.

a. Số người trả lời là đi theo nhóm du lịch của họ trong một mẫu có 6 khách du lịch quốc tế chính là số thành công trong phép thử Nhị thức với $n = 6$. Xác suất có đúng x người trả lời là đi theo nhóm du lịch của họ được cho bởi

$$f(x) = C_6^x.0,23^x.0,77^{6-x} \quad x = 0,1,2,\ldots,6.$$

Với $x = 2$ ta có $f(2) = 0,2789$

b. Lập luận tương tự câu a, xác suất có ít nhất 2 người trả lời đi theo nhóm du lịch của họ được cho bởi

$$\sum_{\{x_i | x_i \geq 2\}} f(x_i) = 1 - f(0) - f(1) = 0,4181$$

c. Tương tự câu a trong trường hợp $n = 10$, trong một mẫu có 10 khách du lịch quốc tế, xác suất không ai trả lời đi theo nhóm du lịch của họ là

$$f(0) = C_{10}^0.0,23^0.0,77^{10} = 0,0733$$

Bài 2.4. Một máy được đánh giá là hoạt động tốt nếu chỉ 3% số sản phẩm sản xuất ra bị lỗi. Giả sử chúng thực hiện một phép thử bằng cách lấy ngẫu nhiên 2 lần mỗi lần một sản phẩm do máy sản xuất để kiểm tra và quan tâm đến số sản phẩm bị lỗi trong hai sản phẩm đó.

a. Với những điều kiện nào thì phép thử trên là phép thử Nhị Thức?

b. Có bao nhiêu kết quả của phép thử có 1 sản phẩm bị lỗi?

c. Tính xác suất có 0 sản phẩm bị lỗi, 1 sản phẩm bị lỗi, 2 sản phẩm bị lỗi?

Giải.

a. Để phép thử trên là phép thử Nhị Thức thì cần thỏa điều kiện sau: (1) Xác suất để lấy ngẫu nhiên 1 sản phẩm lỗi luôn không đổi là 0,03; (2) kết quả của các lần lấy sản phẩm là độc lập nhau.

b. Có 2 kết quả của phép thử xuất hiện 1 sản phẩm bị lỗi đó là (sản phẩm tốt lần lấy thứ nhất, sản phẩm bị lỗi lần lấy thứ hai) và (sản phẩm bị lỗi lần lấy thứ nhất, sản phẩm tốt lần lấy thứ hai).

c. Gọi X là biến ngẫu nhiên chỉ số sản phẩm bị lỗi trong 2 sản phẩm lấy ra, khi đó X có phân phối Nhị thức với $p = 0,03$. Xác suất cần tính
$P(X = 0) = C_2^0 0,03^0.0,97^2 = 0,9409$; $P(X = 1) = C_2^1 0,03^1.0,97^1 = 0,0582$;
$P(X = 2) = C_2^2 0,03^2.0,97^0 = 0,0009$.

Bài 2.5. Có hơn 50 triệu lượt khách truy cập vào trang mạng B&Bs (Bed and Breakfast) năm vừa rồi. Trang mạng B&Bs *http://www.bestinns.com* thuộc miền Bắc Mỹ, trung bình mỗi phút có khoảng 7 lượt khách truy cập. (Theo Time, tháng 9 năm 2001).

a. Tính xác suất để không có khách truy cập trang trong khoảng thời gian 1 phút?

b. Tính xác suất để có từ 2 khách trở lên truy cập trang trong khoảng thời gian 1 phút?

c. Tính xác suất để có từ 1 khách trở lên truy cập trang trong khoảng thời gian 30 giây?

d. Tính xác suất để có từ 5 khách trở lên truy cập trang trong khoảng thời gian 1 phút?

Giải. Gọi X là biến ngẫu nhiên chỉ số lượt truy cập vào trang mạng B&Bs trong một khoảng thời gian, khi đó X có phân phối Poisson với tham số λ

$$f(x) = \frac{\lambda^x.e^{-\lambda}}{x!},$$

với $x = 0, 1, 2, \ldots$, và λ là trung bình số lần lượt truy cập trang mạng trong khoảng thời gian đó.

a. Theo lập luận trên ta có X có phân phối poisson với $\lambda = 7$. Xác suất để không có khách truy cập trang trong khoảng thời gian 1 phút được cho bởi

$$f(0) = \frac{7^0.e^{-7}}{0!} = 0,0009.$$

b. Tương tự câu a, xác suất để có từ 2 khách trở lên truy cập trang trong khoảng thời gian 1 phút là

$$P(X \geq 2) = 1 - P(X < 2) = 1 - f(0) - f(1) = 1 - \frac{7^0.e^{-7}}{0!} - \frac{7^1.e^{-7}}{1!} = 0,9927.$$

c. Gọi X là biến ngẫu nhiên chỉ số lượt khách truy cập trang trong khoảng thời gian 30 giây, khi đó X có phân phối Poisson với tham số $\lambda = \frac{7}{60}.30 = 3,5$. Xác suất để có từ 1 khách trở lên truy cập trang trong khoảng thời gian 30 giây là

$$P(X \geq 1) = 1 - P(X < 1) = 1 - f(0) = 1 - \frac{3,5^0.e^{-3,5}}{0!} = 0,9698.$$

d. Theo câu a và b, xác suất để có từ 5 khách trở lên truy cập trang trong khoảng thời gian 1 phút được tính

$$P(X \geq 5) = 1 - P(X < 5) = 1 - f(0) - f(1) - f(2) - f(3) - f(4) = 0,8271.$$

Bài 2.6. Tại một sân bay quốc tế lớn, những hành khách đến chỗ "kiểm tra hành khách" một cách ngẫu nhiên và độc lập. Trung bình có khoảng 10 hành khách đến trong mỗi phút.

a. Tính xác suất không có hành khách đến khoảng thời gian 1 phút?

b. Tính xác suất có 3 hành khách hoặc ít hơn đến trong khoảng thời gian 1 phút?

c. Tính xác suất không có hành khách đến trong khoảng thời gian 15 giây?

d. Tính xác suất có ít nhất 1 hành khách đến trong khoảng thời gian 15 giây?

Giải. Gọi X là biến ngẫu nhiên chỉ số hành khách đến nơi kiểm tra trong mỗi khoảng thời gian nhất định, khi đó X có phân phối Poisson với tham số λ là số trung bình của lượng khách đến trong khoảng thời gian đó.

a. Dựa theo lập luận trên, trong trường hợp này X có phân phối Poisson với tham số $\lambda = 10$. Xác suất không có hành khách đến khoảng thời gian 1 phút là

$$f(0) = \frac{10^0.e^{-10}}{0!} = 0,000045.$$

b. Tương tự câu a, xác suất có 3 hành khách hoặc ít hơn đến trong khoảng thời gian 1 phút

$$P(X \leq 3) = \sum_{x=0}^{3} f(x) = \frac{10^0.e^{-10}}{0!} + \frac{10^1.e^{-10}}{1!} + \frac{10^2.e^{-10}}{2!} + \frac{10^3.e^{-10}}{3!} = 0,0103.$$

c. Biến ngẫu nhiên X có phân phối Poisson với tham số $\lambda = \frac{10}{60}.15 = 2,5$. Xác suất không có hành khách đến trong khoảng thời gian 15 giây là

$$f(0) = \frac{2,5^0.e^{-2,5}}{0!} \approx 0,0821.$$

d. Tương tự câu c, xác suất có ít nhất 1 hành khách đến trong khoảng thời gian 15 giây là

$$P(X \geq 1) = 1 - P(X < 1) = 1 - f(0) = 0,9179.$$

Bài 2.7. Đặt X là số chi tiết cần thiết để sản xuất một sản phẩm mới. Giả sử X là biến ngẫu nhiên được ước lượng bởi phân phối chuẩn với $\mu = 200$ và $\sigma = 40$, tính các xác suất sau:

a. $P(180 \leq X \leq 220)$

b. $P(X \geq 250)$

c. $P(X \leq 100)$

d. $P(225 \leq X \leq 250)$

Giải. Chuẩn hóa biến ngẫu nhiên X dạng $Z = \frac{X - \mu}{\sigma}$ có phân phối chuẩn tắc với hàm phân phối $P(X \leq x) = \Phi(x)$.

a.

$$\begin{aligned} P(180 \leq X \leq 220) &= P\left(\frac{180-200}{40} \leq Z \leq \frac{220-200}{40}\right) \\ &= \Phi(0,5) - \Phi(-0,5) \\ &= 2.\Phi(0,5) - 1 = 2.0,6915 - 1 = 0,383. \end{aligned}$$

b. $P(X \geq 250) = P\left(Z \geq \frac{250-200}{40}\right) = 1 - \Phi(1,25) = 1 - 0,8944 = 0,1056.$

c. $P(X \leq 100) = P\left(Z \leq \frac{100-200}{40}\right) = \Phi(-2,5) = 1 - \Phi(2,5) = 0,0062.$

d. $P(225 \leq X \leq 250) = P\left(\frac{225-200}{40} \leq Z \leq \frac{250-200}{40}\right) \approx \Phi(1,25) - \Phi(0,625) \approx 0,160335755.$

Bài 2.8. Mức giá trung bình để các công ty lập cổ phiếu S&P 500 là 30 USD, độ lệch chuẩn là 8,20 USD. (Theo *Business Week,* Special Annual Issue, Spring 2003). Giả sử rằng giá cổ phiếu trên có phân phối chuẩn.

a. Tính xác suất một công ty sẽ có giá một cổ phiếu thấp nhất là 40 USD?

b. Tính xác suất một công ty sẽ có giá một cổ phiếu không cao hơn 20 USD?

c. Giá một cổ phiếu là bao nhiêu để đưa một công ty vào top 10 phần trăm?

Giải. Gọi X là biến ngẫu nhiên chỉ giá cổ phiếu S&P 500 của một công ty, theo giả thiết X có phân phối chuẩn.

a. Xác suất một công ty sẽ có giá một cổ phiếu thấp nhất là 40 USD

$$\begin{aligned} P(X \geq 40) &= P\left(\frac{X-30}{8,20} \geq \frac{40-30}{8,20}\right) \\ &= 1 - P\left(\frac{X-30}{8,20} \leq \frac{50}{41}\right) \approx 1 - \Phi(1,22) = 0,1112. \end{aligned}$$

b. Xác suất một công ty sẽ có giá một cổ phiếu không cao hơn 20 USD

$$\begin{aligned} P(X \leq 20) &= P\left(\frac{X-30}{8,20} \leq \frac{20-30}{8,20}\right) \\ &= P\left(\frac{X-30}{8,20} \leq -\frac{50}{41}\right) \approx \Phi(-1,22) = 0,1112. \end{aligned}$$

c. Gọi x_0 là mức giá cổ phiếu cần tìm, ta có điều kiện

$$\begin{aligned} P(X > x_0) &= P\left(\frac{X-30}{8,20} > \frac{x_0-30}{8,20}\right) \\ &= 1 - P\left(\frac{X-30}{8,20} \leq \frac{x_0-30}{8,20}\right) \\ &= 1 - \Phi\left(\frac{x_0-30}{8,2}\right) = 0,1. \end{aligned}$$

Tra bảng $\Phi\left(\frac{x_0-30}{8,2}\right) = 0,9$ ta được $\frac{x_0-30}{8,2} = 1,2816$ suy ra $x_0 = 40,50912.$
Vậy giá một cổ phiếu cần tìm là 40,51USD.

Bài 2.9. Một máy đổ khuôn thành những sản phẩm riêng biệt. Độ lệch chuẩn của những khối sản phẩm được đổ ra theo tính toán trước đó là 0,6 Ounces. Nếu chỉ có 2% số sản phẩm luôn nhỏ hơn 18 Ounces, hãy tính khối lượng trung một sản phẩm do máy sản xuất. Giả thiết khối lượng của các khối sản phẩm được đóng khuông có phân phối chuẩn.

Giải. Gọi X là biến ngẫu nhiên chỉ khối lượng của một sản sản phẩm do máy đổ ra, kỳ vọng của X là μ, độ lệch chuẩn là σ ta có chuẩn hóa của X là $\dfrac{X-\mu}{\sigma}$ có luật chuẩn tắc. Ta có

$$P(X \leq 18) = 0,02$$

tương đương với

$$P\left(\frac{X-\mu}{0,6} \leq \frac{18-\mu}{0,6}\right) = 0,02.$$

Tra bảng giá trị $\Phi\left(\dfrac{18-\mu}{0,6}\right) = 0,02$ ta được $\dfrac{18-\mu}{0,6} = -2,054$ hay $\mu = 19,2324$.

Bài 2.10. Cho biến ngẫu nhiên X có phân phối mũ với hàm mật độ xác suất

$$f(x) = \frac{1}{3}e^{-\frac{x}{3}} \qquad \text{với } x \geq 0.$$

a. Tìm công thức cho $P(X \leq x_0)$.

b. Tính $P(X \leq 2)$.

c. Tính $P(X \geq 3)$.

d. Tính $P(X \leq 5)$.

e. Tính $P(2 \leq X \leq 2)$.

Giải.

a. Đặt $F(x_0) = P(X \leq x_0)$, ta có $F(x_0) = \displaystyle\int_0^{x_0} f(x)dx = \int_0^{x_0} \frac{1}{3}e^{-\frac{x}{3}}dx = 1 - e^{-\frac{x_0}{3}}$.

b. Theo câu a ta có $P(X \leq 2) = F(2) = 1 - e^{-\frac{2}{3}} \approx 0,4866$.

c. Ta có $P(X \geq 3) = 1 - P(X < 3) = 1 - F(3) = \dfrac{1}{e} \approx 0,3679$.

d. Ta có $P(X \leq 5) = F(5) = 1 - e^{-\frac{5}{3}} \approx 0,8111$.

e. Tính $P(2 \leq X \leq 5) = \displaystyle\int_2^5 f(x)dx = F(5) - F(2) = \frac{1}{e^{\frac{2}{3}}} - \frac{1}{e^{\frac{5}{3}}} \approx 0,3245$.

Bài 2.11. Tuổi thọ (giờ) của một thiết bị điện tử là biến ngẫu nhiên có phân phối mũ với hàm mật độ xác suất

$$f(x) = \frac{1}{50}e^{-\frac{x}{50}} \qquad \text{với } x \geq 0.$$

a. Tính tuổi thọ trung bình của thiết bị điện tử đó?

b. Tính xác suất thiết bị sẽ xuất hiện hư hỏng trong 25 đầu giờ hoạt động?

c. Tính xác suất thiết bị hoạt động từ 100 trở lên trước lúc xuất hiện hư hỏng?

Giải.

a. Tuổi thọ trung bình của thiết bị điện tử đó là $\mu = 50$(giờ).

b. Xác suất thiết bị sẽ xuất hiện hư hỏng trong 25 giờ đầu hoạt động là

$$P(0 \leq X \leq 25) = \int_0^{25} \frac{1}{50}e^{-\frac{x}{50}}dx = 1 - e^{-\frac{1}{2}} \approx 0,3935.$$

c. Xác suất thiết bị hoạt động từ 100giờ trở lên trước lúc xuất hiện hư hỏng

$$P(X \geq 100) = 1 - P(X < 100) = 1 - \int_0^{100} \frac{1}{50}e^{-\frac{x}{50}}dx = e^{-2} \approx 0,1353.$$

Bài 2.12. Thời gian (phút) giữa hai cuộc điện thoại gọi đến một văn phòng bồi thường bảo hiểm có phân phối mũ với phân phối xác suất

$$f(x) = 0,50.e^{-0,50x} \qquad \text{với } x \geq 0.$$

a. Thời gian trung bình giữa các cuộc gọi là bao nhiêu?

b. Tính xác suất thời gian giữa các cuộc gọi là bằng hoặc ngắn hơn 30 giây?

c. Tính xác suất thời gian giữa các cuộc gọi là bằng hoặc ngắn hơn 1 phút?

d. Tính xác suất không có cuộc gọi nào trong khoảng từ 5 phút hoặc hơn?

Giải.

a. Thời gian trung bình giữa các cuộc gọi là $\mu = \dfrac{1}{0,5} = 2$ phút.

b. Xác suất thời gian giữa các cuộc gọi là bằng hoặc ngắn hơn 30 giây là

$$P(X \leq 0,5) = \int_0^{0,5} 0,50.e^{-0,50x}dx = 1 - e^{-0,25} = 0,2212.$$

c. Xác suất thời gian giữa các cuộc gọi là bằng hoặc ngắn hơn 1 phút

$$P(X \leq 1) = \int_0^1 0,50.e^{-0,50x}dx = 1 - e^{-0,5} = 0,3935.$$

d. Xác suất không có cuộc gọi nào trong khoảng từ 5 phút hoặc hơn

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \int_0^5 0,50.e^{-0,50x}dx = e^{-2,5} \approx 0,0821.$$

Bài 2.13. Một kiện hàng có 10 sản phẩm, trong đó có 6 sản phẩm loại I và 4 sản phẩm loại II. Chọn ngẫu nhiên (đồng thời) từ kiện hàng ra 2 sản phẩm. Gọi X là số sản phẩm loại II được lấy ra. Lập bảng phân phối xác suất của X.

Giải

Ta có $X = 0, 1, 2.$

Tính xác suất:

$$P\left(X = k\right) = \frac{C_4^k.C_6^{2-k}}{C_{10}^2}, \;\; k = 0, 1, 2$$

Vậy bảng PPXS của X là:

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

Bài 2.14. Lô hàng I có 10 sản phẩm tốt và 2 phế phẩm, lô hàng II có 14 sản phẩm tốt và 5 phế phẩm. Chọn ngẫu nhiên từ lô hàng I ra 1 sản phẩm và bỏ vào lô hàng II. Sau đó, từ lô hàng II chọn ngẫu nhiên ra 3 sản phẩm. Gọi X là số sản phẩm tốt trong 3 sản phẩm lấy ra từ lô hàng II.

a) Lập bảng phân phối xác suất của X.

b) Tính $P(1 < X \leq 4)$.

Giải

a) Lập bảng PPXS của X:

Ta có: $X = 0, 1, 2, 3.$

Gọi A là biến cố sản phẩm lấy từ lô hàng I bỏ vào lô hàng II là sản phẩm tốt.

$\Rightarrow \{A, \overline{A}\}$ là hệ đầy đủ, do đó áp dụng công thức xác suất toàn phần ta có

$$\begin{aligned} P(X=k) &= P(A).P(X=k/A) + P(\overline{A}).P(X=k/\overline{A}) \\ &= \frac{10}{12}.\frac{C_{15}^{k}.C_{5}^{3-k}}{C_{20}^{3}} + \frac{2}{12}.\frac{C_{14}^{k}.C_{6}^{3-k}}{C_{20}^{3}},\ k=\overline{0,3} \end{aligned}$$

Vậy bảng phân phối xác suất của X là:

X	0	1	2	3
p	$\frac{7}{684}$	$\frac{8}{57}$	$\frac{1057}{2280}$	$\frac{2639}{6840}$

b) Dựa vào bảng phân phối xác suất, ta có:

$$P(1 < X \leq 4) = P(X=2) + P(X=3) = \frac{581}{684} \approx 0,8494.$$

Bài 2.15. Kiện hàng I có 12 sản phẩm trong đó có 3 phế phẩm, kiện hàng II có 15 sản phẩm trong đó có 5 phế phẩm. Chọn ngẫu nhiên từ mỗi kiện hàng ra 1 sản phẩm. Gọi X là số sản phẩm tốt chọn được. Lập bảng phân phối xác suất của X.

Giải

Ta có $X = 0, 1, 2.$

Gọi A_i là biến cố sản phẩm lấy ra từ kiện hàng I là sản phẩm tốt, $i = 1, 2.$

$\Rightarrow \{A, \overline{A}\}$ là hệ độc lập.

áp dụng công thức nhân:

$$P(X=0) = P(\overline{A_1.A_2}) = P(\overline{A_1}).P(\overline{A_2}) = \tfrac{3}{12}.\tfrac{5}{15} = \tfrac{1}{12},$$

$$P(X=2) = P(A_1.A_2) = P(A_1).P(A_2) = \tfrac{9}{12}.\tfrac{10}{15} = \tfrac{1}{2},$$

$$P(X=1) = 1 - P(X=0) - P(X=2) = \tfrac{5}{12}.$$

Vậy bảng PPXS của X là:

X	0	1	2
p	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{1}{2}$

Bài 2.16. Một xạ thủ có 4 viên đạn, bắn lần lượt từng viên vào một mục tiêu một cách độc lập. Xác suất trúng mục tiêu ở mỗi lần bắn là 0,7. Nếu có 1 viên trúng mục tiêu hoặc hết đạn thì dừng. Gọi X là số viên đạn đã bắn.

a) Lập bảng phân phối xác suất của X.

b) Tính $P(2 \leq X < 4)$.

Giải

a) Lập bảng PPXS của X:

Ta có $X = \overline{1,4}$.

Gọi A_i là biến cố viên đạn thứ i trúng mục tiêu, $i = \overline{1,4}$.

Theo giả thiết, $\{A_1, A_2, A_3\}$ là hệ độc lập toàn phần.

Ta tính các xác suất:

$$P\left(X=1\right) = P\left(A_1\right) = 0,7;$$

$$P\left(X=2\right) = P\left(\overline{A_1}.A_2\right) = P\left(\overline{A_1}\right).P\left(A_2\right) = 0,3.0,7 = 0,21;$$

$$P\left(X=3\right) = P\left(\overline{A_1}.A_2.A_3\right) = P\left(\overline{A_1}\right).P\left(A_2\right).P\left(A_3\right) = 0,3.0,3.0,7 = 0,063;$$

$$P\left(X=4\right) = 1 - P\left(X=1\right) - P\left(X=2\right) - P\left(X=3\right) = 0,027.$$

Vậy bảng PPXS của X là:

X	1	2	3	4
p	0,7	0,21	0,063	0,027

b) Dựa vào bảng PPXS: $P(2 \leq X < 4) = P(X=2) + P(X=3) = 0,273$.

Bài 2.17. Cho X là BNN rời rạc có bảng phân phối xác suất là

X	0	1	2
p	0.6	0.3	0.1

Tìm hàm phân phối xác suất của X.

Giải

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x \le 0 \\ 0,6 & \text{nếu} \quad 0 < x \le 1 \\ 0,9 & \text{nếu} \quad 1 < x \le 2 \\ 1 & \text{nếu} \quad x > 2 \end{cases}$$

Bài 2.18. Cho X là BNN rời rạc có bảng phân phối xác suất là

X	-2	0	1	2	3
p	0.1	0.2	0.1	0.5	0.1

a) Tìm hàm phân phối xác suất của X.

b) Tính các xác suất $P(0 \le X < 3)$; $P(-2 < X \le 3)$.

Giải

a)

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x \ge -2 \\ 0,1 & \text{nếu} \quad -2 < x \le 0 \\ 0,3 & \text{nếu} \quad 0 < x \le 1 \\ 0,4 & \text{nếu} \quad 1 < x \le 2 \\ 0,9 & \text{nếu} \quad 2 < x \le 3 \\ 1 & \text{nếu} \quad x > 3 \end{cases}$$

b) Dựa vào bảng PPXS:

$$P(0 \le X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0,8;$$
$$P(-2 < X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0,9.$$

Bài 2.19. Một lô hàng có 100 sản phẩm, trong đó có 90 sản phẩm tốt. Lấy ngẫu nhiên đồng thời 15 sản phẩm từ lô hàng này. Tìm số sản phẩm tốt trung bình và phương sai của số sản phẩm tốt trong 15 sản phẩm được lấy ra.

Giải

Gọi X là số sản phẩm tốt trong 15 sản phẩm lấy ra, ta có: $X \sim H(100, 90, 15)$.

- Số sản phẩm tốt trung bình: $E(X) = np = 15.\dfrac{90}{100} = 13,5.$

- Phương sai: $D(X) = npq\dfrac{N-n}{N-1} = 15.\dfrac{90}{100}.\dfrac{10}{100}.\dfrac{85}{99} \approx 1,1591$

Bài 2.20. Có 20 chi tiết máy, trong đó có 15 chi tiết máy tốt. Lấy ngẫu nhiên đồng thời 4 chi tiết máy. Gọi X là số chi tiết máy tốt trong số 4 chi tiết máy được lấy ra.

a) Xác định quy luật phân phối xác suất của X.

b) Tính xác suất để lấy được 3 chi tiết máy tốt.

c) Tính trung bình số chi tiết máy tốt được lấy ra và phương sai của X.

Giải

a) X tuân theo phân phối siêu bội $X \sim H(20, 15, 4)$.

b) Ta có

$$P(X = k) = \frac{C_{15}^{k}.C_{5}^{4-k}}{C_{20}^{4}} \quad \left(k = \overline{0,4}\right)$$

Do đó,

$$P(X = 3) = \frac{C_{15}^{3}.C_{5}^{1}}{C_{20}^{4}} = \frac{455}{969} \approx 0,4696.$$

c) Trung bình số chi tiết tốt được lấy ra: $E(X) = np = 4.\frac{15}{20} = 3.$

Phương sai: $D(X) = npq\dfrac{N-n}{N-1} = 4.\dfrac{15}{20}.\dfrac{5}{20}.\dfrac{16}{19} \approx 0,6316$

Bài 2.21. Một ông chủ vườn lan đã để nhầm 20 chậu lan có hoa màu đỏ vào cùng với 100 chậu lan có hoa màu tím (lan chưa nở hoa). Một khách hàng chọn mua ngẫu nhiên đồng thời 15 chậu từ 120 chậu lan này.

a) Tính xác suất khách hàng mua được từ 5 đến 6 chậu lan có hoa màu đỏ.

b) Gọi X là số chậu lan có hoa màu đỏ mà khách hàng chọn được. Tính trung bình và phương sai của X.

Giải

Gọi X là số chậu lan có hoa màu đỏ mà khách hàng mua được, ta có

$$X \sim H(120, 20, 15) \Rightarrow P(X = k) = \frac{C_{20}^{k}.C_{100}^{15-k}}{C_{120}^{15}}, \;\; k = \overline{0,15}$$

a)

$$P(5 \leq X \leq 6) = P(X = 5) + P(X = 6) = \frac{C_{20}^{5}.C_{100}^{10}}{C_{120}^{15}} + \frac{C_{20}^{6}.C_{100}^{9}}{C_{120}^{15}} \approx 0,07232$$

b)
Trung bình: $E(X) = np = 15.\frac{20}{120} = 2,5$.

Phương sai: $D(X) = npq\dfrac{N-n}{N-1} = 15.\dfrac{20}{120}.\dfrac{100}{120}.\dfrac{105}{119} \approx 1,8382$

Bài 2.22. Tung ba lần một con xúc xắc. Tính xác suất để có:

a) Hai lần xuất hiện mặt 1 chấm.

b) ít nhất một lần xuất hiện mặt 1 chấm.

Giải

Gọi X là số lần xuất hiện mặt 1 chấm trong ba lần tung con xúc xắc. Ta có $X \sim B(n,p)$ với $n = 3$ và $p = \frac{1}{6}$.

Khi đó,

$$P(X = k) = C_3^k.\left(\frac{1}{6}\right)^k.\left(\frac{5}{6}\right)^{3-k}, \; k = \overline{0,3}$$

a) Ta tính $P(X = 2) = C_3^2.\left(\frac{1}{6}\right)^2.\left(\frac{5}{6}\right)^1 = \frac{5}{72} \approx 0,0694$.

b) Ta có

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - C_3^0.\left(\frac{1}{6}\right)^0.\left(\frac{5}{6}\right)^3 \approx 0,4213$$

Bài 2.23. Một đề thi có 10 câu hỏi, mỗi câu có 4 phương án trả lời trong đó chỉ có 1 phương án đúng. Sinh viên A trả lời một cách ngẫu nhiên tất cả các câu hỏi. Gọi X là số câu trả lời đúng trong 10 câu.

a) Xác định quy luật phân phối của X.

b) Tính xác suất để sinh viên A trả lời đúng từ 2 đến 3 câu hỏi.

c) Tính xác suất để sinh viên A trả lời đúng ít nhất một câu hỏi.

d) Tính trung bình số câu hỏi được trả lời đúng và phương sai của X.

e) Tính số câu hỏi mà sinh viên A có khả năng trả lời đúng lớn nhất.

Giải

a) X tuân theo phân phối nhị thức: $X \sim B(10; 0,25)$.

b) Ta có

$$\begin{aligned}
&P(2 \le X \le 3) = P(X = 2) + P(X = 3) \\
&= C_{10}^{2}(0,25)^{2}(0,75)^{8} + C_{10}^{3}(0,25)^{3}(0,75)^{7} \approx 0,531.
\end{aligned}$$

c) Ta có

$$P(X \ge 1) = 1 - P(X < 1) = 1 - P(X = 0) \approx 0,943.$$

d)Kỳ vọng và phương sai

$$\begin{aligned}
&E(X) = np = 10.0,25 = 2,5; \\
&D(X) = npq = 10.0,25.0,75 = 1,875.
\end{aligned}$$

e) Số câu hỏi mà sinh viên có khả năng trả lời đúng nhất chính là $Mod(X)$, ta có

$$np - q \le Mod(X) \le np + p \Leftrightarrow 1,75 \le Mod(X) \le 2,75$$

Vậy $Mod(X) = 2$.

Bài 2.24. một người nuôi 160 con gà mái cùng loại. Xác suất để 1 con gà đẻ trứng trong ngày là 0,8.Giả sử mỗi trứng bán được 2200 đồng, tiền cho mỗi con gà ăn trong ngày là 1000 đồng, tính số tiền lãi trung bình thu được trong ngày.

Giải

a) Gọi X là số trứng thu được trong ngày, khi đó $X \sim B(160; 0,8)$. Gọi Y là số tiền thu được trong ngày. Ta cần tính $E(Y)$.

Ta có $Y = 2200.X - 1000.160 = 2200X - 160000$. Từ đây suy ra

$$E(Y) = 2200.E(X) - 160000 = 2200.160.0,8 - 160000 = 121600 \text{ đồng}.$$

Bài 2.25. Một nhà vườn trồng 121 cây mai với xác suất nở hoa của mỗi cây mai trong dịp tết là 0,75. Giá bán 1 cây mai nở hoa là 500 ngàn đồng.

a) Tính số cây mai trung bình nở hoa trong dịp tết.

b) Giả sử nhà vườn bán hết những cây mai nở hoa, tính số tiền trung bình mà nhà vườn thu được.

Giải

Gọi X là số cây mai nở hoa trong dịp tết. Ta có: $X \sim B(121; 0,75)$.

a) Ta tính $E(X) = np = 121.0,75 = 90,75$ (cây).

b)Gọi Y là số tiền nhà vườn thu được khi bán những cây mai. Ta có: $Y = 500X$. Suy ra số tiền trung bình nhà vườn thu được là

$$E(Y) = 500E(X) = 500.90,75 = 45,375 \,(\text{triệu đồng})$$

Bài 2.26. Tại một bến cảng, trung bình mỗi ngày có 5 tàu cập bến. Tính xác suất để trong một ngày:

a) Có 3 tàu cập bến.

b) Có ít nhất 4 tàu cập bến.

c) Có đúng 5 tàu cập bến.

d) Có từ 3 đến 7 tàu cập bến.

Giải

Gọi X là số tàu cập bến trong một ngày thì $X \sim P(5)$. Do đó,

$$P(X = k) = \frac{e^{-5}.5^k}{k!} \quad (k = 0, 1, 2, ...)$$

Ta có:

a) $P(X = 3) = \dfrac{e^{-5}.5^3}{3!} = \dfrac{e^{-5}.125}{6} \approx 0,1404$

b) $P(X \geq 4) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \approx 0,735$

c) $P(X = 5) = \dfrac{e^{-5}.5^5}{5!} = \dfrac{e^{-5}.625}{24} \approx 0,175.$

d) $P(3 \leq X \leq 7) = P(X = 3) + .. + P(X = 7) \approx 0,742.$

Bài 2.27. Tại một siêu thị, trung bình cứ 5 phút thì có 10 khách đến quầy tính tiền.

a) Tính xác suất để trong 1 phút có 3 khách đến quầy tính tiền.

b) Tính xác suất để trong 1 phút có từ 1 đến 3 khách đến quầy tính tiền.

c) Tính số khách có khả năng đến quầy tính tiền lớn nhất trong 1 giờ.

Giải

a) Gọi X là số khách đến quầy tính tiền trong 1 phút thì $X \sim P(\lambda_1)$ với λ_1 là trung bình số khách đến quầy tính tiền trong 1 phút: $\lambda_1 = \dfrac{1.10}{5} = 2$. Ta có:

$$P(X = 3) = \frac{e^{-\lambda_1}(\lambda_1)^3}{3!} = \frac{e^{-2}.2^3}{3!} \approx 0,18044$$

b) Ta có

$$P(1 \le X \le 3) = P(X = 1) + P(X = 2) + P(X = 3) \approx 0.722$$

c) Gọi Y là số khách đến quầy tính tiền trong 1 giờ thì $Y \sim P(\lambda_2)$ với λ_2 là trung bình số khách đến quầy tính tiền trong 1 giờ: $\lambda_2 = \dfrac{60.10}{5} = 120$.

Số khách có khả năng đến quầy tính tiền lớn nhất trong 1 giờ chính là $Mod(Y)$.

Ta có:

$$\lambda_2 - 1 \le Mod(Y) \le \lambda_2 \Leftrightarrow 119 \le Mod(Y) \le 120$$

Vậy $Mod(Y) = 119$ hoặc 120.

Bài 2.28. Bộ phận quản lý của một siêu thị cho biết : trong những ngày đông khách, nhân viên tính tiền của siêu thị trung bình trong 5 phút tính tiền xong cho 2 khách. Tính xác suất để trong ngày đông khách, một khách hàng phải chờ quá 5 phút để người khách kế trước được tính tiền xong.

Giải

Gọi X là số khách được nhân viên tính tiền xong trong 5 phút. Ta có $X \sim P(2)$.

Ta cần tính:

$$P(X < 1) = P(X = 0) = \frac{e^{-2}.2^0}{0!} = e^{-2}$$

Bài 2.29. Một người có 4 xe ôtô cho thuê. Hàng ngày, chi phí cho mỗi xe là 10usd (cho dù xe có được thuê hay không). Giá cho thuê mỗi xe là 70usd. Giả sử yêu cầu thuê xe mỗi ngày là BNN có phần phối Poisson với tham số $\lambda = 2,8$. Tính số tiền trung bình người này thu được trong một ngày.

Giải

Gọi X là số yêu cầu thuê xe mỗi ngày.

Theo giả thiết, $X \sim P(2,8)$ nên $E(X) = 2,8$.

Gọi Y là số tiền thu được trong một ngày, ta có

$$Y = 70X - 10.4 = 70X - 40$$

Vậy số tiền trung bình người này thu được trong một ngày là

$$E(Y) = 70E(X) - 40 = 70.2,8 - 40 = 156$$

Bài 2.30. Cho BNN rời rạc X có bảng phân phối xác suất:

X	1	2	3	4
p	0.15	0.25	0.4	0.2

a) Tìm $P(-1 < X \leq 2)$.

b) Tìm kỳ vọng và phương sai của X.

c) Tìm hàm phân phối xác suất của X.

Giải

a) Dựa vào bảng PPXS:

$$P(-1 < X \leq 2) = P(X = 1) + P(X = 2) = 0,15 + 0,25 = 0,4$$

b) Kỳ vọng và phương sai

$$E(X) = \sum_{i=1}^{4} x_i p_i = 2,65;$$
$$D(X) = \sum_{i=1}^{4} x_i^2 p_i - [E(X)]^2 = 7,95 - 2,65^2$$

c) Hàm phân phối xác suất

$$F(x) = \begin{cases} 0 & \text{nếu} & x \leq 1 \\ 0,15 & \text{nếu} & 1 < x \leq 2 \\ 0,4 & \text{nếu} & 2 < x \leq 3 \\ 0,8 & \text{nếu} & 3 < x \leq 4 \\ 1 & \text{nếu} & x > 4 \end{cases}$$

Bài 2.31. Cho ĐLNN X rời rạc có bảng phân phối xác suất là:

X	0	0.1	0.3	0.4	0.7
p	a	0.2	b	0.2	0.1

a) Tìm a, b sao cho $E(X) = 0,2$.

b) Tìm a, b sao cho $DX = 0,0503$.

Giải

a) Ta có

$$E(X) = \sum_{i=1}^{5} x_i p_i = 0,3b + 0,17$$

Theo giả thiết ta có $E(X) = 2$. Suy ra

$$0,3b + 0,17 = 0,2 \Leftrightarrow b = 0,1$$

Mặc khác, ta có

$$a + 0,2 + b + 0,2 + 0,1 = 1 \Leftrightarrow a + b = 0,5$$

Vậy $a = 0,4$ và $b = 0,1$.

b) Ta có

$$\begin{aligned} D(X) &= \sum_{i=1}^{5} x_i^2 . p_i - [E(X)]^2 \\ &= 0,09b + 0,083 - (0,3b + 0,17)^2 \\ &= -0,09b^2 - 0,012b + 0,0541 \end{aligned}$$

Theo giả thiết, ta có

$$D(X) = 0,0503 \Leftrightarrow -0,09b^2 - 0,012b + 0,0541 = 0,0503 \Leftrightarrow b \approx 0,1494$$

Suy ra $a = 0,5 - b \approx 0,3506$.

Bài 2.32. người quản lý của một công ty xem xét tình trạng hoạt động của các máy trong công ty này. Dựa vào các thông tin trong quá khứ và bằng kinh nghiệm của mình, người quản lí đưa ra bảng phân phối xác suất của số máy hỏng X trong tuần như sau:

X	0	1	2	3	4
p	0.1	0.24	0.42	0.16	0.08

a) Tìm số máy bị hỏng trung bình trong 1 tuần và độ lệch chuẩn của số máy bị hỏng.

b) Chi phí thiệt hại do mỗi máy bị hỏng trong 1 tuần là 3 triệu đồng. Tính chi phí thiệt hại do các máy bị hỏng trung bình của công ty này trong một tuần.

Giải

a) Số máy hỏng trung bình :

$$E(X) = \sum_{i=1}^{5} x_i.p_i = 1,88$$

Phương sai và độ lệch chuẩn:

$$D(X) = \sum_{i=1}^{5} x_i^2.p_i - [E(X)]^2 = 4,64 - 1,88^2 = 1,1056$$
$$\Rightarrow \sigma = \sqrt{D(X)} = \sqrt{1,1056} \approx 0,0515.$$

b) Gọi Y (triệu đồng) là số tiền thiệt hại do máy bị hỏng của công ty trong 1 tuần. Ta có: $Y = 3X$.

Như vậy chi phí thiệt hại do các máy bị hỏng trung bình của công ty này trong một tuần là $E(Y) = 3E(X) = 3.1,88 = 5,64$ (triệu đồng).

Bài 2.33. Có 3 kiện hàng: kiện 1 chứa 10 sản phẩm tốt và 2 phế phẩm, kiện 2 chứa 9 sản phẩm tốt và 3 phế phẩm, kiện 3 chứa 7 sản phẩm tốt và 5 phế phẩm. Các sản phẩm giống hệt nhau về kích thước và trọng lượng. Lấy ngẫu nhiên mỗi kiện hàng 1 sản phẩm. Gọi X là số phế phẩm được lấy ra.

a) Lập bảng phân phối xác suất của X.

b) Tìm $Mod(X), Med(X)$, kỳ vọng và phương sai của X.

Giải

a)Bảng PPXS của X

X	0	1	2	3
p	$\frac{35}{96}$	$\frac{131}{288}$	$\frac{47}{288}$	$\frac{5}{288}$

b) Ta có

$$Mod\,(X) = 1;$$
$$Med\,(X) = 1;$$
$$E\,(X) = \sum_{i=1}^{4} x_i.p_i = \tfrac{5}{6} \approx 0,8333;$$
$$D\,(X) = \sum_{i=1}^{5} x_i^2.p_i - [E\,(X)]^2 = \tfrac{41}{72} \approx 0,5684.$$

Bài 2.34. Gọi X là số lần mặt nhất xuất hiện sau 3 lần tung một con xúc xắc.

a) Lập bảng phân phối xác suất của X.

b) Tìm xác suất có ít nhất một lần được mặt nhất.

c) Tính xác suất có tối đa hai lần được mặt nhất.

d) Tính $E(X)$ và $D(X)$.

Giải

a) Bảng PPXS của BNN X:

X	0	1	2	3
p	$\frac{125}{216}$	$\frac{25}{72}$	$\frac{5}{72}$	$\frac{1}{216}$

b) Xác suất có ít nhất 1 lần được mặt nhất là:

$$P\,(X \geq 1) = 1 - P\,(X = 0) = 1 - C_3^0.\left(\frac{1}{6}\right)^0.\left(\frac{5}{6}\right)^3 = \frac{91}{216}$$

c) Xác suất có tối đa hai mặt nhất là

$$P\,(X \leq 2) = 1 - P\,(X = 3) = 1 - C_3^3.\left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^0 = \frac{215}{216}$$

d) Vì $X \sim B(3, \frac{1}{6})$ nên

$$E\,(X) = np = 0,5;$$
$$D\,(X) = npq \approx 0,4167.$$

Bài 2.35. Một gia đình có 10 người con. Giả sử xác suất sinh con trai, con gái trong mỗi lần sinh là bằng nhau. Tính xác suất

a) Không có con trai.

b) Có 5 con trai và 5 con gái.

c) Số con trai ít hơn số con gái.

Giải

Gọi X là số con trai trong số 10 người con thì $X \sim B(10; 0,5)$. Ta có:

$$P(X = k) = C_{10}^{k}.(0,5)^{k}.(0,5)^{10-k}, \ \ k = \overline{0,10}$$

a) Xác suất không có con trai:

$$P(X = 0) = C_{10}^{0}.(0,5)^{0}.(0,5)^{10} = \frac{1}{1024} \approx 0,0098$$

b)Xác suất có 5 con trai và 5 con gái:

$$P(X = 5) = C_{10}^{5}.(0,5)^{5}.(0,5)^{5} = \frac{63}{256} \approx 0,2461$$

c) Xác suất để số con trai ít hơn sô con gái:

$P(X < 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \approx 0,37695.$

Bài 2.36. Một người nuôi 100 con gà mái đẻ. Xác suất để 1 con gà đẻ trứng trong ngày là 60.

a) Tìm số trứng gà trung bình thu được trong ngày.

b) Nếu muốn trung bình mỗi ngày thu được 120 trứng gà thì cần nuôi bao nhiêu con gà?

Giải

Gọi X là số trứng gà thu được trong ngày. Ta có: $X \sim B(100; 0,6)$.

a) Số trứng gà trung bình thu được trong ngày:

$$E(X) = np = 100.0,6 = 60.$$

b) Gọi n là số con gà mái đẻ cần nuôi thì $X \sim B(n; 0,6)$. Suy ra $E(X) = np = 0,6n$.

Như vậy, ta có

$$E(X) = 120 \Leftrightarrow 0,6n = 120 \Leftrightarrow n = 200.$$

Bài 2.37. Một máy sản xuất lần lượt từng sản phẩm với xác suất có một phế phẩm là 2. Cho máy sản xuất ra 15 sản phẩm.

a) Tính xác suất để trong 15 sản phẩm sản xuất ra có đúng 3 phế phẩm.

b) Số phế phẩm có khả năng cao nhất được sản xuất ra là bao nhiêu?

c) Tìm số sản phẩm tốt trung bình trong 15 sản phẩm được sản xuất ra.

Giải

Gọi X là số phế phẩm trong 15 sản phẩm được sản xuất ra, ta có $X \sim B(15; 0,02)$.

a)Ta có:

$$P\left(X=3\right)=C_n^3.p^3q^{n-3}=C_{15}^3.(0,02)^3.(0,98)^{12}\approx 0,0029.$$

b) Số phế phẩm có khả năng nhất được sản xuất ra là $Mod(X)$. Ta có

$$np-q\le Mod\left(X\right)\le np+p\Leftrightarrow -0,68\le Mod\left(X\right)\le 0,32\Rightarrow Mod\left(X\right)=0.$$

c) Gọi Y là số sản phẩm tốt trong 15 sản phẩm thì $Y \sim B(15; 0,98)$. Như vậy

$$E\left(Y\right)=np=15.0,98=14,7.$$

Bài 2.38. Nhà máy dệt muốn tuyển dụng người biết rành về một loại sợi. Một người có thể phân biệt sợi thật hay giả với xác suất đúng là 80, xin dự tuyển. Giám đốc nhân sự thử thách người này 7 lần: mỗi lần ông đem ra 2 loại sợi, trong đó có 1 loại sợi thật. Nếu người dự tuyển nói đúng ít nhất 6 lần thì sẽ được tuyển dụng. Tính xác suất người này được tuyển dụng.

Giải

Gọi X là số lần người dự tuyển trả lời đúng trong 7 lần thử thách thì xác suất để người này được tuyển dụng chính là

$$P(X\ge 6)=P(X=6)+P(X=7)$$

Mặc khác, ta có $X \sim B(7;p)$ với p là xác suất người này trả lời đúng trong mỗi lần thử thách.

Theo giả thiết $p = 0,8$. Vậy $X \sim B(7; 0,8)$

Như vậy

$$P\left(X\ge 6\right)=C_7^6.(0,8)^6.(0,2)^1+C_7^7.(0,8)^7.(0,2)^0\approx 0,5767.$$

Bài 2.39. ở một tổng đài điện thoại, các cú điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và trung bình có 2 cuộc gọi đến trong một phút. Tìm xác suất để:

a) Có đúng 5 cú điện thoại gọi đến trong 2 phút.

b) Không có cú điện thoại nào gọi đến trong 30 giây.

c) Có ít nhất một cú điện thoại gọi đến trong 10 giây.

Giải

Gọi X là số cuộc điện thoại gọi đến trong 2 phút. Ta suy ra $X \sim P(\lambda_1)$ với λ_1 là trung bình số cuộc điện thoại gọi đến trong 2 phút.

Theo giả thiết, 1 phút có trung bình 2 cuộc gọi. Suy ra 2 phút có trung bình $\lambda_1 = \frac{2.2}{1} = 4$. Ta suy ra $X \sim P(4)$. Vậy

$$P(X=5) = \frac{e^{-\lambda_1}.\lambda_1^5}{5!} = \frac{e^{-\lambda_1}.4^5}{5!} = \frac{128e^{-4}}{15} \approx 0,1563$$

b) Gọi Y là số cuộc gọi điện thoại gọi đến trong 30 giây. Khi đó $Y \sim P(\lambda_2)$ với $\lambda_2 = \frac{0,5.2}{1} = 1$. Vậy

$$P(Y=0) = \frac{e^{-\lambda_2}.\lambda_2^0}{0!} = \frac{e^{-1}.1^0}{0!} \approx 0,3679.$$

c) Gọi Z là số cuộc điện thoại gọi đến trong 10 giây. Khi đó $Y \sim P(\lambda_3)$ với $\lambda_3 = \frac{2.\frac{1}{6}}{1} = \frac{1}{3}$. Vậy

$$P(Z \geq 1) = 1 - P(Z=0) = 1 - \frac{e^{-\lambda_3}.\lambda_3^0}{0!} \approx 0,2835$$

Bài 2.40. Quan sát thấy trung bình cứ một phút thì có 2 ôtô qua trạm thu phí. Tính xác suất:

a) Có 5 ôtô đi qua trạm thu phí trong 3 phút.

b) Có nhiều hơn 1 ôtô đi qua trạm thu phí trong 30 giây.

Giải

a)Gọi X là số ôtô đi qua trạm thu phí trong 3 phút, khi đó theo giả thiết ta suy ra $X \sim P(6)$. Vậy

$$P(Z=5) = \frac{e^{-\lambda_1}.\lambda_1^5}{5!} \approx 0,1606.$$

b)Gọi Y là số ôtô đi qua trạm thu phí trong 30 giây, theo giả thiết ta suy ra $X \sim P(1)$. Vậy

$$P(Y>1) = 1 - P(Y=0) - P(Y=1) \approx 0,2642.$$

Bài 2.41. Một cửa hàng trong một khu phố nhập về mỗi ngày 34kg loại thực phẩm này với giá 2500 đồng/kg và bán ra với giá 4000 đồng/kg. Nếu bị ế thì cuối cùng cửa hàng phải bán hạ giá còn 15000 đồng/kg mới hết hàng. Tính tiền lời trung bình của cửa hàng này về loại thực phẩm nói trên trong một ngày. Cho biết nhu cầu hằng ngày của người dân ở một khu phố về một loại thực phẩm tươi sống là BNN X có bảng phân phối xác suất như sau:

$X(kg)$	31	32	33	34
p	0.15	0.25	0.45	0.15

Giải

Gọi Y là số tiền lời cửa hàng thu được trong một ngày. Ta có $Y = 40000X + 15000(34 - X) - 25000.34 = 25000X - 340000$.

Từ đây ta suy ra

$$E(Y) = E(25000X - 340000) = 25000E(X) - 340000.$$

Theo giả thiết bài toán ta có

$$E\left(X\right) = \sum_{i=1}^{4} x_i.p_i = 32,6 \ \ (kg)$$

Vậy $E(Y) = 25000.32,6 - 340000 = 475000$ (đồng).

Bài 2.42. Theo số liệu thống kê ở một cửa hàng đậu tương, người ta thấy lượng đậu bán ra là BNN X có bảng phân phối xác suất như sau:

$X(kg)$	10	13	16	19	22
p	0.15	0.2	0.35	0.2	0.1

Giả sử giá đậu nhập vào là 10000 đồng/kg thì cửa hàng sẽ lãi 5000 đồng/kg; nếu đến cuối ngày không bán được sẽ lỗ 8000 đồng/kg. Vậy mỗi ngày cửa hàng nên nhập bao nhiêu kg đậu để thu được tiền lãi trung bình nhiều nhất?

Giải

Gọi Y là tiền lãi cửa hàng thu được trong một ngày; Z là khối lượng đậu mà cửa hàng nên nhập vào mỗi ngày. Ta đi tìm Z sao cho $E(Y)$ nhiều nhất.

Ta nhận thấy rằng Z nhận một trong các giá trị: 10,13,16,19,22.

Ta có

$$\begin{aligned} Y &= 5000X - 8000\,(Z - X) \\ &= 13000X - 8000Z \end{aligned}$$

$\Rightarrow E\,(Y) = 13000X - 8000Z$

Theo giả thiết ta có

$$E\,(X) = \sum_{i=1}^{4} x_i.p_i = 15,7\ \ (kg)$$

Suy ra $E(Y) = 13000.15,7 - 8000Z = 204100 - 8000Z$.

Ta lập bảng:

$X(kg)$	10	13	16	19	22
$E(Y)$	124100	108100	76100	52100	28100

Vậy cửa hàng nên nhập 10 (kg) đậu tương mỗi ngày.

Bài 2.43. Nghiên cứu về mức lương (triệu đồng/năm) của 400 công nhân ngành may, người ta được bảng số liệu sau:

Lương	12	13.2	15.6	18	20.4	24
Số công nhân	16	60	160	100	40	24

a) Gọi X là thu nhập trong một năm của công nhân may. Lập bảng phân phối xác suất của X.

b) Tính mức lương trung bình của một công nhân trong một năm.

c) Tính phương sai của X.

Giải

a) Bảng PPXS của X:

X	12	13.2	15.6	18	20.4	24
p	0,04	0,15	0,4	0,25	0,1	0,06

b) Mức lương trung bình của một công nhân trong một năm

$$E\left(X\right)=\sum_{i=1}^{6}x_i.p_i=16,68.$$

c) Phương sai:

$$D\left(X\right)=\sum_{i=1}^{6}x_i^2.p_i-\left[E\left(X\right)\right]^2=8,1936.$$

Bài 2.44. Cho hàm số

$$f\left(x\right)=\begin{cases}\dfrac{3x^2}{8} & \text{nếu} \quad x\in[0,2]\\ 0 & \text{nếu} \quad x\notin[0,2]\end{cases}$$

a) Chứng minh rằng $f(x)$ là hàm mật độ của BNN X.

b) Tính các xác suất $P(1\leq X\leq \dfrac{3}{2})$ và $P(1<X<3)$.

Giải

a) Ta nhận thấy $f(x)\geq 0, \forall x$ nên $f(x)$ là hàm mật độ xác suất tương đương

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

Vì

$$\int_{-\infty}^{+\infty}f\left(x\right)dx=\int_{0}^{2}\frac{3x^2}{8}dx=\left(\frac{x^3}{8}\right)^2=1 \quad \text{đpcm}$$

b) áp dụng công thức $P\left(a\leq X\leq b\right)=\int\limits_a^b f\left(x\right)dx$ ta có:

$$P\left(1\leq X\leq \tfrac{3}{2}\right)=\int_1^{\frac{3}{2}}f\left(x\right)dx=\int_1^{\frac{3}{2}}\frac{3x^2}{8}dx=\frac{19}{64};$$

$$P\left(1\leq X\leq 3\right)=\int_1^{3}f\left(x\right)dx=\int_1^{2}\frac{3x^2}{8}dx=\frac{7}{8}.$$

Bài 2.45. Chứng minh rằng hàm số sau là hàm mật độ xác suất của BNN liên tục X nào đó:

$$f\left(x\right)=\begin{cases}\dfrac{1}{x^2} & \text{nếu} \quad x\geq 1\\ 0 & \text{nếu} \quad x<1\end{cases}$$

Khi đó, tính: $P(X \le 3)$; $P(1 \le X \le 2)$; $P(-1 < X < 2)$; $P(X > 5)$.

Giải

a) Ta nhận thấy $f(x) \ge 0, \forall x$ nên $f(x)$ là hàm mật độ xác suất tương đương

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

Vì

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{1}^{+\infty} \frac{1}{x^2}dx = 1$$

b) áp dụng công thức $P(a \le X \le b) = \int_a^b f(x)\,dx$ ta có:

$$P(X \le 3) = \int_{-\infty}^{3} f(x)\,dx = \int_{1}^{3} \frac{1}{x^2}dx = \frac{2}{3};$$
$$P(1 \le X \le 2) = \int_{1}^{2} f(x)\,dx = \int_{1}^{2} \frac{1}{x^2}dx = \frac{1}{2};$$
$$P(-1 \le X \le 2) = \int_{-1}^{2} f(x)\,dx = \int_{1}^{2} \frac{1}{x^2}dx = \frac{1}{2};$$
$$P(X > 5) = \int_{5}^{+\infty} f(x)\,dx = \int_{5}^{+\infty} \frac{1}{x^2}dx = \frac{1}{5}.$$

Bài 2.46. Cho BNN X liên tục có hàm mật độ xác suất xác định bởi:

$$f(x) = \begin{cases} c\left(1 - x^2\right) & \text{nếu } x \in [-1, 1] \\ 0 & \text{nếu } x \notin [-1, 1] \end{cases}$$

a) Xác định hằng số c.

b) Tính xác suất $P(-0,5 \le X \le 0,8)$

Giải

a) Theo giả thiết, $f(x)$ là hàm mật độ xác suất nên ta có $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

Mặc khác ta có

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-1}^{1} c\left(1-x^2\right)dx = 2\int_{0}^{1} c\left(1-x^2\right)dx = \left[2c\left(x-\frac{x^3}{3}\right)\right]_0^1 = \frac{4c}{3}$$

Ta có

$$\frac{4c}{3} = 1 \Leftrightarrow c = \frac{3}{4}.$$

b) Ta có:

$$P(-0,5 \le X \le 0,8) = \int_{-0,5}^{0,8} f(x)\,dx = \int_{-0,5}^{0,8} \frac{3}{4}\left(1-x^2\right)dx = 0,81575.$$

Bài 2.47. Cho BNN X liên tục có hàm mật độ xác suất là

$$f(x) = \begin{cases} ax+b & \text{nếu} \quad x \in (0,1) \\ 0 & \text{nếu} \quad x \notin (0,1) \end{cases}$$

Giả $P\left(X > \frac{1}{2}\right) = \frac{1}{3}$, hãy xác định hằng số a, b.

Giải

\+ Do $f(x)$ là hàm mật độ xác suất nên ta có

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \int_{-1}^{1} (ax+b)\,dx = 1 \Leftrightarrow \frac{a}{2} + b = 1$$

\+ Mặc khác, ta có

$$P\left(X > \frac{1}{2}\right) = \frac{1}{3} \Leftrightarrow \int_{1/2}^{+\infty} f(x)\,dx = \frac{1}{3} \Leftrightarrow \int_{1/2}^{1} (ax+b)\,dx = \frac{1}{3} \Leftrightarrow \frac{3a}{8} + \frac{b}{2} = \frac{1}{3}$$

Vậy ta có hệ phương trình

$$\begin{cases} a + 2b = 2 \\ 9a + 12b = 8 \end{cases} \Leftrightarrow \begin{cases} a = -\frac{4}{3} \\ b = \frac{5}{3} \end{cases}$$

Bài 2.48. Cho BNN Cauchy X có hàm phân phối xác suất

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x.$$

a) Tính xác suất $P(0 < X < 1)$.

b) Tìm hàm mật độ của X.

Giải

a) Ta có

$$P(0 < X < 1) = F(1) - F(0) = \left(\frac{1}{2} + \frac{\arctan 1}{\pi}\right) - \left(\frac{1}{2} + \frac{\arctan 0}{\pi}\right) = \frac{1}{4}.$$

b) Hàm mật độ của BNN X

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{\pi(1 + x^2)}.$$

Bài 2.49. Giả sử X là đại lượng ngẫu nhiên có hàm mật độ xác suất

$$f(x) = \frac{C}{(1 + x^2)} \quad \forall x : -\infty < x < +\infty.$$

a) Xác định hệ số C.

b) Tính xác suất $P(0 < X < 1)$.

c) Tìm hàm phân phối xác suất của X.

Giải

a) Dựa vào tính chuẩn hóa của hàm mật độ, ta có

$$\begin{aligned}
&\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow C \int\limits_{-\infty}^{+\infty} \tfrac{1}{1+x^2}dx = 1 \\
&\Leftrightarrow C\left[\lim_{a\to-\infty} \int\limits_a^0 \tfrac{1}{1+x^2}dx + \lim_{b\to+\infty} \int\limits_0^b \tfrac{1}{1+x^2}dx\right] = 1 \\
&\Leftrightarrow C = \tfrac{1}{\pi}.
\end{aligned}$$

b) Dễ thấy

$$P(0 < X < 1) = \frac{1}{\pi}\int\limits_0^1 \frac{1}{1 + x^2}dx = \frac{1}{4}.$$

c) Hàm phân phối xác suất của BNN X là

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x.$$

Bài 2.50. Cho X là BNN có hàm phân phối xác suất

$$F\left(x \right) = A + B\arctan \frac{x}{C},$$

trong đó A, B và C là các hằng số. Xác định các hằng số A, B, C và tìm hàm mật độ.

Giải

+ Từ tính chất của hàm phân phối xác suất, ta có hệ phương trình sau

$$\left\{ \begin{array}{l} \lim\limits_{x \to +\infty} F\left(x \right) = 1 \\ \lim\limits_{x \to -\infty} F\left(x \right) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} A + B\frac{\pi}{2} = 1 \\ A - B\frac{\pi}{2} = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} A = \frac{1}{2} \\ B = \frac{1}{\pi} \end{array} \right.$$

+ Khi đó, hàm phân phối và hàm mật độ của X có dạng

$$\begin{array}{l} F\left(x \right) = \frac{1}{2} + \frac{1}{\pi}\arctan \frac{x}{C}; \\ f\left(x \right) = \frac{dF(x)}{dx} = \frac{C}{\pi(x^2 + C^2)}. \end{array}$$

Do tính không âm của hàm mật độ nên $C > 0$.

Bài 2.51. Giả sử hàm phân phối của BNN X có dạng

$$F\left(x \right) = \left\{ \begin{array}{ccc} 0 & \text{nếu} & x \le 1 \\ a{(x - 1)}^2 & \text{nếu} & 1 < x \le 3 \\ 1 & \text{nếu} & x > 3 \end{array} \right.$$

Tính $E(X)$ và $D(X)$.

Giải

Dễ dàng tính được hệ số $a = \frac{1}{4}$.

Khi đó, kì vọng và phương sai của X được tính theo công thức

$$E\left(X \right) = \frac{1}{2}\int\limits_1^3 x\left(x - 1 \right)dx = \frac{7}{3};$$

$$D\left(X \right) = \frac{1}{2}\int\limits_1^3 \left(x - \frac{7}{3} \right)^2 \left(x - 1 \right)dx = \frac{2}{9}.$$

Bài 2.52. Biến ngẫu nhiên X có hàm mật độ phân phối xác định bởi

$$f\left(x\right)=\begin{cases} a\cos x & \text{nếu } |x|\le \frac{\pi}{2} \\ 0 & \text{nếu } |x|>\frac{\pi}{2} \end{cases}$$

Tính $E(X)$ và $D(X)$.

Giải

Từ tính chuẩn hóa của hàm mật độ, suy ra $a=\frac{1}{2}$.

Khi đó, kì vọng và phương sai được tính như sau:

$$E\left(X\right)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x\cos x dx=0;$$

$$D\left(X\right)=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x^2\cos x dx=\frac{\pi^2}{4}.$$

Bài 2.53. Chiều cao của một loại cây lấy gỗ là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn với chiều cao trung bình là 20m và độ lệch chuẩn là 2,5m. Cây đạt tiêu chuẩn khai thác là cây có chiều cao tối thiểu là 15m. Tính tỷ lệ cây đạt tiêu chuẩn khai thác.

Giải

Gọi X là chiều cao của cây. Ta có $X\sim N(\mu,\sigma^2)$ với $\mu=20m;\sigma=2,5m$

Tỷ lệ cây đạt tiêu chuẩn khai thác chính bằng $P(X\ge 15)$, ta có

$$P\left(X\ge 15\right)=0,5-\Phi_0\left(\frac{15-20}{2,5}\right)=0,5-\Phi_0\left(-2\right)=0,5+0,4772=0,9772.$$

Bài 2.54. Chiều cao của các sinh viên ở một trường đại học là BNN có phân phối chuẩn với chiều cao trung bình là 158cm và độ lệch chuẩn là 7,5cm. Nếu chọn ra 10% sinh viên có chiều cao cao nhất thì chiều cao tối thiểu của sinh viên trong nhóm này là bao nhiêu?

Giải

Gọi X là chiều cao của sinh viên thì $X \sim N(\mu, \sigma^2)$ với $\mu = 158; \sigma = 7,5$. Gọi a là chiều cao tối thiểu trong nhóm sinh viên có chiều cao cao nhất. Ta cần tìm a sao cho $P(X \geq a) = 10\% = 0,1$.

Ta có:

$$P\left(X \geq a\right) = 0,1 \Leftrightarrow 0,5 - \Phi_0\left(\frac{a - 158}{7,5}\right) = 0,1 \Leftrightarrow \Phi_0\left(\frac{a - 158}{7,5}\right) = 0,4$$

Ta suy ra:

$$\frac{a - 158}{7,5} = 1,29 \Leftrightarrow a = 167,675.$$

Bài 2.55. Điểm thi Toeic của sinh viên năm cuối ở một trường đại học là BNN X có phân phối chuẩn với giá trị trung bình là 560 điểm và độ lệch chuẩn là 78 điểm. Tính:

a) Tỷ lệ sinh viên có điểm từ 600 đến 700 điểm.

b) Tỷ lệ sinh viên có điểm thi trên 500 điểm.

c) Giả sử nhà trường muốn xác định điểm Toeic tối thiểu để sinh viên có thể ra trường với tỷ lệ 80%, tính điểm Toeic tối thiểu này.

Giải

Gọi X là điểm thi Toeic của sinh viên. Ta có $X \sim N(\mu, \sigma^2)$ với $\mu = 560; \sigma = 78$.

a) Tỷ lệ sinh viên có điểm từ 600 đến 700 điểm là:

$$\begin{aligned}
P\left(600 < X < 700\right) &= \Phi_0\left(\frac{700 - \mu}{\sigma}\right) - \Phi_0\left(\frac{600 - \mu}{\sigma}\right) \\
&= \Phi_0\left(\frac{700 - 560}{78}\right) - \Phi_0\left(\frac{600 - 560}{78}\right) \\
&= \Phi_0\left(1,79\right) - \Phi_0\left(1,28\right) \\
&= 0,4633 - 0,3997 = 0,0636.
\end{aligned}$$

b) Tỷ lệ sinh viên có điểm thi trên 500 điểm là:

$$\begin{aligned} P(X > 500) &= 0,5 - \Phi_0\left(\frac{500-\mu}{\sigma}\right) \\ &= 0,5 - \Phi_0\left(\frac{500-560}{78}\right) \\ &= 0,5 - \Phi_0(-0,77) \\ &= 0,5 + 0,2794 = 0,7794. \end{aligned}$$

c) Gọi a là mức điểm tối thiểu mà sinh viên cần đạt được. Ta cần tìm a sao cho $P(X \geq a) = 0,8$, ta có

$$\begin{aligned} P(X \geq a) = 0,8 &\Leftrightarrow 0,5 - \Phi_0\left(\frac{a-\mu}{\sigma}\right) = 0,8 \\ &\Leftrightarrow 0,5 - \Phi_0\left(\frac{a-560}{78}\right) = 0,8 \\ &\Leftrightarrow \Phi_0\left(\frac{560-a}{78}\right) = 0,3 \end{aligned}$$

Ta suy ra:

$$\frac{560-a}{78} = 0,85 \Leftrightarrow a = 493,7$$

Bài 2.56. Giả sử tuổi thọ của một thiết bị điện tử là BNN liên tục X có hàm mật độ xác suất là

$$f(x) = \begin{cases} ce^{-cx} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$

a) Xác định hằng số c.

b) Tính xác suất $P(X \leq 10)$.

c) Nếu $P(X \leq 10) = \frac{1}{2}$ thì giá trị của c là bao nhiêu?

Giải

a) Ta nhận thấy $f(x) \geq 0, \forall x$

Mặc khác ta có

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_0^{+\infty} ce^{-cx}dx = \lim_{a\to\infty}\left[-e^{-cx}\right]_0^a = 1$$

Vậy hàm số đã cho là hàm mật độ xác suất với mọi giá trị của c.

b)

$$P\left(X \le 10\right) = \int\limits_{-\infty}^{10} f\left(x\right)dx = \int\limits_{0}^{10} ce^{-cx}dx = \left[-e^{-cx}\right]_0^{10} = 1 - e^{-10c}.$$

c) Ta có:

$$P\left(X \le 10\right) = \frac{1}{2} \Leftrightarrow 1 - e^{-10c} = \frac{1}{2} \Leftrightarrow c = \frac{\ln 2}{10}.$$

Bài 2.57. Tuổi thọ của một bộ phận trong một dây chuyền sản xuất là BNN X (tháng) có hàm mật độ như sau:

$$f\left(x\right) = \begin{cases} \dfrac{25}{2{(x+10)}^2} & \text{nếu} \quad x \in (0, 40) \\ 0 & \text{nếu} \quad x \notin (0, 40) \end{cases}$$

a) Tìm hàm phân phối xác suất của X.

b) Tìm xác suất để tuổi thọ của thiết bị nhỏ hơn 1 năm.

Giải

a)

$$F\left(x\right) = \begin{cases} 0 & \text{nếu} \quad x \ge 0 \\ \int\limits_{-\infty}^{x} f\left(t\right)dt & \text{nếu} \quad 0 < x \le 40 \\ 1 & \text{nếu} \quad x > 40 \end{cases} = \begin{cases} 0 & \text{nếu} \quad x \ge 0 \\ \frac{5x}{4(x+10)} & \text{nếu} \quad 0 < x \le 40 \\ 1 & \text{nếu} \quad x > 40 \end{cases}$$

Thật vậy, với $0 < x \le 40$ thì

$$\int\limits_{-\infty}^{x} f\left(t\right)dt = \int\limits_{0}^{x} \frac{25}{2{(t+10)}^2}dt = \left[-\frac{25}{2\left(t+10\right)}\right]_0^x = \frac{5x}{4\left(x+10\right)}$$

b) Ta có:

$$P\left(0 < X < 12\right) = \int\limits_{0}^{12} f\left(x\right)dx = \int\limits_{0}^{12} \frac{25}{2{(t+10)}^2}dt = \frac{15}{22}$$

Bài 2.58. Cho BNN X có hàm mật độ:

$$f\left(x\right) = \begin{cases} 0 & \text{nếu} \quad x < 2 \\ \dfrac{2}{x^2} & \text{nếu} \quad x \ge 2 \end{cases}$$

a) Tìm hàm phân phối xác suất của X.

b) Tìm $P(-3 < X < 5)$

Giải

a)

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x < 2 \\ \int\limits_{-\infty}^{x} f(t)\, dt & \text{nếu} \quad x \geq 2 \end{cases} = \begin{cases} 0 & \text{nếu} \quad x < 2 \\ 1 - \dfrac{2}{x} & \text{nếu} \quad x \geq 2 \end{cases}$$

Thật vậy, với $x \geq 2$ thì

$$\int\limits_{-\infty}^{x} f(t)\, dt = \int\limits_{2}^{x} \frac{2}{t^2} dt = \left[-\frac{2}{t}\right]_2^x = 1 - \frac{2}{x}$$

b) Ta có

$$P(-3 < X < 5) = F(5) - F(-3) = 0,6.$$

Bài 2.59. Cho BNN X liên tục có hàm mật độ xác suất là

$$f(x) = \begin{cases} \dfrac{3x^2}{16} & \text{khi} \quad x \in (-2, 2) \\ 0 & \text{khi} \quad x \notin (-2, 2) \end{cases}$$

Tìm kỳ vọng và phương sai của X.

Giải

+ Kỳ vọng:

$$E(X) = \int\limits_{-\infty}^{+\infty} xf(x)\, dx = \int\limits_{-2}^{2} x.\frac{3x^2}{16} dx = \int\limits_{-2}^{2} \frac{3x^3}{16} dx = 0$$

+ Phương sai:

$$D(X) = \int\limits_{-\infty}^{+\infty} x^2 f(x)\, dx - [E(X)]^2 = \int\limits_{-2}^{2} x^2.\frac{3x^2}{16} dx = \frac{3}{8}\int\limits_{0}^{2} x^4 dx = 2,4.$$

Bài 2.60. BNN X có hàm phân phối xác suất

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

a) Tìm hàm mật độ xác suất.

b) Tìm hệ số m.

Giải

a) Hàm mật độ xác suất của X là

$$f(x) = \begin{cases} 3mx^2 - 6x + 2 & \text{khi} \quad x \in [0,1] \\ 0 & \text{khi} \quad x \notin [0,1] \end{cases}$$

b) Xác định m, ta có

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \int_0^1 \left(3mx^2 - 6x + 2\right)dx = 1 \Leftrightarrow m = 2$$

Bài 2.61. Tuổi thọ của một loại linh kiện điện tử (đơn vị tính ngàn giờ) là biến ngẫu nhiên X có hàm mật độ f được xác định bởi:

$$f(x) = \begin{cases} 10^{-5}kx^2(10 - x)^2 & \text{nếu} \quad x \in [0;10] \\ 0 & \text{nếu} \quad x \notin [0;10] \end{cases}$$

Trong đó k là số thực cho trước. Hãy xác định k và tính tuổi thọ trung bình của loại linh kiện đang khảo sát.

Giải

Theo giả thiết chúng ta có $k \geq 0$ và

$$1 = \int_{-\infty}^{+\infty} f(x)\,dx = \int_0^{10} 10^{-5}kx^2(10 - x)^2dx$$

Từ đó chúng ta tính được $k = 30$.

Tuổi thọ trung bình của một linh kiện:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx = \int_0^{10} 10^{-5}.30x^3(10 - x)^2dx = 5$$

Bài 2.62. Giả sử BNN X có hàm mật độ xác suất

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$

Tính $E(X); D(X)$ và $E(X^k)$ với $k \geq 1$.

Giải

Dùng định nghĩa tính trực tiếp, ta có:

$$E(X) = \lambda^{-1}; \quad D(X) = \lambda^{-2}$$

và bằng quy nạp toán học có

$$E\left(X^k\right) = \frac{k!}{\lambda^k} \ \ \forall k \geq 1.$$

Bài 2.63. Cho BNN X nhận giá trị trong khoảng $(0; +\infty)$, có hàm phân phối xác suất

$$F(x) = \begin{cases} 1 - e^{-\frac{x^2}{2}} & \text{nếu} \quad x > 0 \\ 0 & \text{nếu} \quad x \leq 0 \end{cases}$$

Tìm hàm mật độ, kì vọng, phương sai, median và mode của BNN X.

Giải

+ Hàm mật độ

$$f(x) = \begin{cases} xe^{-\frac{x^2}{2}} & \text{nếu} \quad x > 0 \\ 0 & \text{nếu} \quad x \leq 0 \end{cases}$$

+ Kì vọng

$$E(X) = \int\limits_0^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \sqrt{\frac{\pi}{2}}.$$

+ Phương sai

$$D(X) = \frac{4-\pi}{2}.$$

+ Median $Med(X) = \sqrt{2log2}$

+ Mode $Mod(X) = 1$.

Bài 2.64. Giả sử X là một BNN liên tục có hàm mật độ

$$f(x) = \begin{cases} K(1-x) & \text{nếu} \quad x \in [0,1] \\ 0 & \text{nếu} \quad x \notin [0,1] \end{cases}$$

Tìm hằng số K, kỳ vọng, median và phương sai của X.

Giải

Dễ thấy bằng tính toán trực tiếp

$$\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \int\limits_{0}^{1} K(1-x)\,dx = 1 \Leftrightarrow K = 2;$$
$$E(X) = \int\limits_{-\infty}^{+\infty} xf(x)dx = 2\int\limits_{0}^{1} x(1-x)\,dx = \tfrac{1}{3};$$
$$D(X) = 2\int\limits_{0}^{1} x^2(1-x)\,dx - \left(\tfrac{1}{3}\right)^2 = \tfrac{1}{18};$$
$$Med(X) = \tfrac{2-\sqrt{2}}{2}.$$

Bài 2.65. Cho BNN X có hàm mật độ với $a > 0$,

$$f(x) = \begin{cases} \dfrac{x}{a^2}e^{-\frac{x^2}{2a^2}} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$

Tìm median của BNN X.

Giải

Ta có hàm phân phối xác suất của BNN X là

$$F(x) = \begin{cases} 1 - e^{-\frac{x^2}{2a^2}} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$

Khi đó từ phương trình $F[Med(X)] = \frac{1}{2}$, suy ra $Mod(X) = \sqrt{2ln2}$.

Bài 2.66. Xác định hằng số a để hàm

$$f(x) = \begin{cases} A.\sin 2x & \text{nếu} \quad x \in \left(0, \frac{\pi}{2}\right) \\ 0 & \text{nếu} \quad x \notin \left(0, \frac{\pi}{2}\right) \end{cases}$$

Là hàm mật độ của X.

a) Xác định hệ số A?

b) Tính kì vọng và phương sai của X?

c) Tính $P\left(0 \le X \le \frac{\pi}{4}\right)$.

Giải

a) Từ tính chất chuẩn hoá của hàm mật độ, ta có

$$\int\limits_0^{\pi/2} A\sin 2x dx = 1 \Leftrightarrow A = 1$$

b) Tính kì vọng và phương sai

$$\begin{array}{l} E\left(X\right) = \int\limits_0^{\pi/2} x\sin 2x dx = \frac{1}{4}; \\ D\left(X\right) = \int\limits_0^{\pi/2} x^2\sin 2x dx - \left(\frac{1}{4}\right)^2 = \frac{\pi^2}{8} - \frac{9}{16}. \end{array}$$

c)

$$P\left(0 \le X \le \frac{\pi}{4}\right) = \int\limits_0^{\pi/4} f\left(x\right)dx = \int\limits_0^{\pi/4} \sin x dx = 1 - \frac{\sqrt{2}}{2}$$

Bài 2.67. Tuổi thọ của dân cư của một quốc gia là BNN X có hàm mật độ như sau:

$$f\left(x\right) = \begin{cases} kx^2(100 - x)^2 & \text{khi} \quad x \in [0, 100] \\ 0 & \text{khi} \quad x \notin [0, 100] \end{cases}$$

a) Xác định hằng số k.

b) Tuổi thọ trung bình của dân cư quốc gia trên là bao nhiêu?

c) Tìm tỉ lệ người có tuổi thọ từ 60 đến 70 tuổi.

Giải

a) Xác định k, ta có

$$\int\limits_{-\infty}^{+\infty} f\left(x\right)dx = \int\limits_0^{100} kx^2(100 - x)^2 dx = \frac{10^9}{3}k$$

Mặc khác,

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \frac{10^9}{3}k = 1 \Leftrightarrow k = \frac{3}{10^9}$$

b) Tuổi thọ trung bình của người dân

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx = \frac{3}{10^9}\int_0^{100} x^3(100-x)^2dx = 50$$

c) Tỉ lệ người có tuổi thọ từ 60 đến 70 tuổi là

$$P(60 \le x \le 70) = \int_{60}^{70} kx^2(100-x)^2dx = \frac{3}{10^9}\int_{60}^{70} x^2(100-x)^2dx \approx 0,154$$

Vậy tỉ lệ người có tuổi thọ từ 60 đến 70 là 15,4%.

Bài 2.68. BNN X có hàm mật độ xác suất

$$f(x) = \begin{cases} kx^2(4-x) & \text{khi} \quad x \in [0,4] \\ 0 & \text{khi} \quad x \notin [0,4] \end{cases}$$

a) Xác định k?

b) Tính $P(X \le 2)$

Giải

a) Xác định k, ta có

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \int_0^4 kx^2(4-x)\,dx = 1 \Leftrightarrow k = \frac{3}{64}.$$

b)

$$P(X \le 2) = \frac{3}{64}\int_0^2 x^2(4-x)\,dx = \frac{5}{16}.$$

Bài 2.69. Thời gian học rành nghề sửa tivi của một người là BNN X (năm) có hàm mật độ xác suất là

$$f(x) = \begin{cases} \dfrac{9}{40}x^2 + \dfrac{1}{5} & \text{khi} \quad x \in (0,2) \\ 0 & \text{khi} \quad x \notin (0,2) \end{cases}$$

a) Tính xác suất để một người học rành nghề sửa tivi trước một năm rưỡi.

b) Tính $E(13X+5)$ và $E(X^2)$

c) Tìm phương sai của X.

Giải

a) Xác suất để một người học rành nghề sửa tivi trước 1 năm rưỡi:

$$P\left(0 < X \le 1,5\right) = \int_0^{1,5} f\left(x\right)dx = \int_0^{1,5} \left(\frac{9}{40}x^2 + \frac{1}{5}\right)dx \approx 0,553125$$

b) Ta có $E(13X+5) = 13E(X) + 5$.

Mà

$$E\left(X\right) = \int_{-\infty}^{+\infty} xf\left(x\right)dx = \int_0^2 x\left(\frac{9}{40}x^2 + \frac{1}{5}\right)dx = 1,3.$$

Suy ra $E(13X+5) = 21,9$.

+ Ta có

$$E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2 f\left(x\right)dx = \int_0^2 x^2\left(\frac{9}{40}x^2 + \frac{1}{5}\right)dx \approx 1,9733.$$

c) Phương sai của BNN X

$$D\left(X\right) = \int_{-\infty}^{+\infty} x^2 f\left(x\right)dx - \left[E\left(X\right)\right]^2 = \int_0^2 x^2\left(\frac{9}{40}x^2 + \frac{1}{5}\right)dx - \left(1,3\right)^2 \approx 0,2833.$$

Bài 2.70. Tuổi thọ của một loại bóng đèn (đơn vị: giờ) là BNN có phân phối chuẩn với trung bình 900 giờ và độ lệch chuẩn là 100 giờ. Tính tỷ lệ bóng đèn có tuổi thọ trong khoảng 775 giờ đến 975 giờ.

Giải

Gọi X là tuổi thọ của bóng đèn. Theo giả thiết ta có $N \sim N(900; 100^2)$. Ta có

$$P\left(775 \le X \le 975\right) = \Phi_0\left(\frac{975-900}{100}\right) - \Phi_0\left(\frac{775-900}{100}\right) \approx 0,6678.$$

Bài 2.71. Chỉ số IQ của những người trong một câu lạc bộ là BNN tuân theo phân phối chuẩn với trung bình $\mu = 105$ và độ lệch chuẩn $\sigma = 10$. Chọn ngẫu nhiên một người trong nhóm này. Tính xác suất để người được chọn:

a) Có chỉ số IQ ít nhất bằng 50.

b) Có chỉ số IQ nhiều nhất là 80.

c) Có chỉ số IQ trong khoảng từ 95 đến 125.

Giải

Gọi X là chỉ số IQ của những người trong câu lạc bộ. Theo giả thiết, $X \sim N(105, 20^2)$, ta có:

$$
\begin{aligned}
&a)\ P(X \geq 50) = 0,5 - \Phi_0\left(\frac{50-105}{20}\right) = 0,997;\\
&b)\ P(0 \leq X \leq 80) = \Phi_0\left(\frac{80-105}{20}\right) - \Phi_0\left(\frac{0-105}{20}\right) = 1,1056;\\
&c)\ P(95 \leq X \leq 125) = \Phi_0\left(\frac{125-105}{20}\right) - \Phi_0\left(\frac{95-105}{20}\right) = 0,5328.
\end{aligned}
$$

Bài 2.72. Tuổi thọ của một loại sản phẩm là BNN có phân phối chuẩn với trung bình là 11 năm và độ lệch chuẩn là 2 năm.

a) Nếu quy định thời gian bảo hành là 10 năm thì tỷ lệ sản phẩm phải bảo hành là bao nhiêu?

b) Nếu muốn tỷ lệ sản phẩm phải bảo hành là 10% thì phải quy định thời gian bảo hành là bao nhiêu?

Giải

Gọi X (năm) là tuổi thọ của loại sản phẩm này. Ta có $X \sim N(11, 4)$.

a) Tỉ lệ sản phẩm phải bảo hành

$$P(0 \leq X \leq 10) = \Phi_0\left(\frac{10-11}{2}\right) - \Phi_0\left(\frac{0-11}{2}\right) = 0,3085.$$

b) Gọi a là thời gian bảo hành. Ta cần tìm a sao cho $P(0 \le X \le a) = 10\%$. Ta có

$$\begin{aligned}
P(0 \le X \le a) = 10\% &\Leftrightarrow \Phi_0\left(\frac{a-11}{2}\right) - \Phi_0\left(\frac{0-11}{2}\right) = 0,1 \\
&\Leftrightarrow \Phi_0\left(\frac{a-11}{2}\right) + 0,5 = 0,1 \\
&\Leftrightarrow \Phi_0\left(\frac{a-11}{2}\right) = -0,4 \\
&\Leftrightarrow \Phi_0\left(\frac{11-a}{2}\right) = 0,4 = \Phi_0(1,29) \\
&\Leftrightarrow \frac{11-a}{2} = 1,29 \\
&\Leftrightarrow a = 8,42
\end{aligned}$$

Bài 2.73. Một loại chi tiết máy được gọi là đạt tiêu chuẩn nếu đường kính của nó sai lệch không quá 0,33mm về giá trị tuyệt đối so với đường kính thiết kế. Cho biết đường kính của loại chi tiết máy này là BNN phân phối theo quy luật chuẩn, với độ lệch tiêu chuẩn là 0,3mm.

a) Tìm xác suất để sản xuất được một chi tiết máy đạt tiêu chuẩn.

b) Tìm trung bình số chi tiết máy đạt tiêu chuẩn khi sản xuất 100 chi tiết.

Giải

Gọi X là đường kính của chi tiết máy. Theo giả thiết ta có $X \sim N(\mu, \sigma^2)$ với $\sigma = 0,3$.

a) Xác suất để sản xuất được 1 chi tiết máy đạt tiêu chuẩn là

$$P(|X-\mu| \le 0,33) = 2\Phi_0\left(\frac{0,33}{0,3}\right) = 2\Phi_0(1,1) = 0,7286.$$

b) Gọi Y là số chi tiết đạt tiêu chuẩn khi sản xuất 100 chi tiết. Khi đó $Y \sim B(n,p)$ với $n = 100$ và $p = 0,7286$. Trung bình số chi tiết đạt tiêu chuẩn khi sản xuất 100 chi tiết là

$$E(X) = np = 100.0,7286 = 72,86.$$

Bài 2.74. Trọng lượng của một sản phẩm X (đơn vị: gam) do một máy tự động sản xuất ra với $X \sim N(100, 2)$. Sản phẩm được coi là đạt kĩ thuật nếu trọng lượng của nó đạt từ 98 đến 103 gam.

a) Tìm tỉ lệ sản phẩm không đạt kĩ thuật của máy.

b) Cho máy sản xuất 100 sản phẩm. Tính xác suất không quá 15 sản phẩm không đạt kĩ thuật trong 100 sản phẩm này.

Bài 2.75. Cho X là BNN rời rạc tuân theo phân phối Poisson với tham số λ. Hãy tìm kỳ vọng và phương sai của X.

Giải

áp dụng khai triển Maclaurin

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

+ Kỳ vọng

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

+ Phương sai

$$D\left(X\right) = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

Ta có

$$\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2 P\left(X = k\right) \\
&= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \left[\frac{k\left(k-1\right)\lambda^k}{k!} + \frac{k\lambda^k}{k!}\right] \\
&= \lambda^2 e^{-\lambda} \sum_{k-2=0}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} \\
&= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda
\end{aligned}$$

Vậy

$$D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Bài 2.76. Cho X là BNN rời rạc tuân theo phân phối hình học với tham số p. Hãy tìm kỳ vọng và phương sai của X.

Giải

+ Kỳ vọng

$$E\left(X\right)=\sum_{k=0}^{\infty}k.P\left(X=k\right)=\sum_{k=0}^{\infty}k.p.q^{k}=pq\sum_{k=1}^{\infty}kq^{k-1}=pq.\frac{1}{\left(1-q\right)^{2}}=\frac{1-p}{p}.$$

+ Phương sai

$$D\left(X\right)=E\left(X^{2}\right)-\left[E\left(X\right)\right]^{2}$$

Ta có

$$E\left(X^{2}\right)=\sum_{k=0}^{\infty}k^{2}P\left(X=k\right)=\sum_{k=0}^{\infty}k^{2}pq^{k}=pq^{2}\sum_{k=0}^{\infty}k^{2}q^{k-2}.$$

Mà

$$\begin{aligned}\sum_{k=0}^{\infty}k^{2}q^{k-2}&=\sum_{k=0}^{\infty}k\left(k-1\right)q^{k-2}+\sum_{k=0}^{\infty}kq^{k-2}\\&=\sum_{k=2}^{\infty}k\left(k-1\right)q^{k-2}+\frac{1}{q}\sum_{k=1}^{\infty}kq^{k-1}\\&=\frac{2}{p^{3}}+\frac{1}{qp^{2}}.\end{aligned}$$

Suy ra

$$E\left(X^{2}\right)=\frac{2q^{2}+qp}{p^{2}}.$$

Vậy

$$D(X)=\frac{1-p}{p^{2}}.$$

Bài 2.77. Cho X là BNN liên tục có phân phối đều với hàm mật độ

$$f\left(x\right)=\begin{cases}\dfrac{1}{b-a} & \text{nếu} \quad x\in\left(a,b\right)\\ 0 & \text{nếu} \quad x\notin\left(a,b\right)\end{cases}$$

Hãy tìm kỳ vọng và phương sai của X.

Giải

+ Kỳ vọng

$$E\left(X\right)=\int_{-\infty}^{+\infty}xf\left(x\right)dx=\int_{a}^{b}xf\left(x\right)dx=\frac{1}{b-a}\int_{a}^{b}xdx=\frac{a+b}{2}.$$

+ Phương sai

$$D(X) = E\left(X^2\right) - [E(X)]^2 = \int_a^b x^2 f(x)\,dx - \left(\frac{a+b}{2}\right)^2 = \frac{(a-b)^2}{12}.$$

Bài 2.78. Cho X là BNN liên tục có phân phối mũ với hàm mật độ

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$

Hãy tìm kỳ vọng và phương sai của X.

Giải

+ Kỳ vọng

$$E(X) = \int_0^{+\infty} x f(x)\,dx = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = \lambda \lim_{a\to+\infty} \int_0^a x e^{-\lambda x} dx = \lambda \lim_{a\to+\infty} I$$

Theo công thức tích phân từng phần ta tính được

$$I = -\frac{a}{\lambda}e^{-\lambda a} - \frac{1}{\lambda^2}e^{-a\lambda} + \frac{1}{\lambda^2}$$

Vậy $E(X) = \dfrac{1}{\lambda}$.

+ Phương sai

$$D(X) = E\left(X^2\right) - [E(X)]^2$$

Mà $E(X^2) = \dfrac{2}{\lambda^2}$.

Vậy $D(X) = \dfrac{1}{\lambda^2}$.

2.7 Bài tập đề nghị

Bài 2.79. Có 3 hộp: mỗi hộp đựng 15 viên bi, trong đó hộp thứ nhất có 3 viên bi trắng, hộp thứ hai có 5 viên bi trắng, hộp thứ ba có 2 viên bi trắng. Chọn ngẫu nhiên một hộp, rồi từ hộp đó lấy ngẫu nhiên đồng thời 3 viên bi. Gọi X là số viên bi trắng trong 3 viên bi lấy ra.

a) Lập bảng phân phối xác suất của X.

b) Tính các xác suất $P(0 < X < 3)$, $P(-1 \le X \le 2)$, $P(X > 1)$.

ĐS:

a)

X	0	1	2	3
p	$\frac{626}{1365}$	$\frac{193}{455}$	$\frac{149}{1365}$	$\frac{11}{1365}$

b) 0,5333; 0,9919; 0,1172.

Bài 2.80. Cho X là BNN rời rạc có bảng phân phối xác suất là

X	-1	0	2	4	5
p	0.15	0.1	0.45	0.05	0.25

a) Tìm $ModX, MedX, E(X)$ và $D(X)$.

b) Tìm $E(2X + 5), D(3X + 9)$.

ĐS:

a) 2; 2; 2,2; 4,16.

b) 9,4; 37,44.

Bài 2.81. Cho X là BNN rời rạc có bảng phân phối xác suất là

X	1	2	3	4
p	0.15	a	0.35	b

Tìm giá trị của hằng số a, b để $E(X) = 2,8$.

ĐS: $a = 0,2; b = 0,3$.

Bài 2.82. Một lô hàng có 8 sản phẩm, trong đó có 5 sản phẩm loại I và 3 sản phẩm loại II. Lấy ngẫu nhiên đồng thời 3 sản phẩm từ lô hàng này.

a) Tính số sản phẩm loại I tin chắc nhất trong 3 sản phẩm lấy ra.

b) Tính trung bình số sản phẩm loại I trong 3 sản phẩm lấy ra.

c) Gọi Y là số sản phẩm loại II trong 3 sản phẩm lấy ra. Tính $E(Y)$ và $D(Y)$.

ĐS: 2; 1,875; 0,5022.

Bài 2.83. Một xí nghiệp có hai ô tô vận tải hoạt động. Xác suất trong ngày làm việc các ô tô bị hỏng tương ứng bằng 0,1 và 0,2. Gọi X là số ô tô bị hỏng trong thời gian làm việc.

a) Tìm quy luật phân phối xác suất của X.

b) Thiết lập hàm phân bố xác suất của X.

Bài 2.84. Một thiết bị gồm 3 bộ phận hoạt động độc lập với nhau. Xác suất trong thời gian t các bộ phận bị hỏng tương ứng là 0,4; 0,2 và 0,3.

a) Tìm quy luật phân phối xác suất của số bộ phận bị hỏng X.

b) Thiết lập hàm phân số xác suất của X.

c) Tính xác suất trong thời gian t có không quá hai bộ phận bị hỏng.

d) Tìm mốt và trung vị của BNN X.

Bài 2.85. Một hộp có 3 quả cầu trắng và 2 quả cầu đen. Lấy ngẫu nhiên lần lượt từng quả cầu cho đến khi lấy được cầu trắng. Tìm quy luật phân phối xác suất của số cầu được lấy ra.

Bài 2.86. Tại một tuyến giao thông, người ta thống kê được số hành khách trên một chuyến xe buýt (SKTMCXB) thay đổi như sau:

SKTMCXB	20	25	30	35	40
Xác suất	0.25	0.2	0.15	0.3	0.1

a) Tính số hành khách trung bình, phương sai của số hành khách trên một chuyến xe buýt.

b) Chi phí cho một chuyến xe là 114 ngàn đồng (không phụ thuộc số khách trên xe). Để công ty xe buýt thu được lợi nhuận trung bình cho một chuyến xe là 60 ngàn đồng thì phải qui định giá vé cho mỗi hành khách là bao nhiêu?

Bài 2.87. Một lô sản phẩm gồm 12 sản phẩm trong đó có 8 sản phẩm tốt và 4 phế phẩm. Lấy ngẫu nhiên từ lô hàng 3 sản phẩm. Gọi X là sản phẩm tốt trong 3 sản phẩm lấy ra. Tìm phân phối xác suất của X. Viết hàm phân phối. Tính $P(1 \le X < 3)$.

Bài 2.88. Bắn liên tiếp 3 viên đạn độc lập vào một mục tiêu. Xác suất trúng đích của mỗi viên đạn là 0,5. Gọi X là số viên đạn trúng đích trong 3 viên. Tìm hàm phân phối xác suất của X. Viết hàm phân phối của X. Tính xác suất $P(X \geq 1)$. Tính kỳ vọng, phương sai và độ lệch chuẩn của X.

Bài 2.89. Một lô sản phẩm gồm 12 sản phẩm trong đó có 10 sản phẩm tốt và 2 phế phẩm . Lấy ngẫu nhiên 2 sản phẩm từ lô hàng. Gọi X là số phế phẩm trong 2 sản phẩm lấy ra. Tìm phân phối xác suất của X. Tính kỳ vọng, phương sai và độ lệch chuẩn của X.

Bài 2.90. Cho phân phối xác suất của BNN X là

x_i	0	1
$P(X = x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

Tính: $E(2X+1), E\left(X^2\right), E\left(X^3\right), E\left(e^X\right), D(X)$.

Bài 2.91. Cho phân phối xác suất của BNN X là

X	-1	0	1
p	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Tìm mode và median của X.

Bài 2.92. Một hộp có 10 sản phẩm. Các sản phẩm trong hộp gồm hai loại: loại A và loại B. Gọi X là BNN chỉ số sản phẩm loại A có trong hộp. Cho biết phân phối xác suất của X như sau:

x_i	1	2	3
$P(X = x_i)$	0.2	0.5	0.3

Lấy ngẫu nhiên từ hộp ra 3 sản phẩm (không hoàn lại). Đặt Y biểu thị số sản phẩm loại A có trong 3 sản phẩm lấy ra.

a) Tìm luật phân phối xác suất của Y.

b) Tính $E(Y), D(Y)$.

Bài 2.93. Sản phẩm của một nhà máy được đóng thành từng hộp, mỗi hộp chứa 10 sản phẩm. Gọi X là BNN chỉ số sản phẩm loại 1 có trong hộp. Cho biết X có phân phối xác suất như sau:

x_i	7	8	9	10
$P(X = x_i)$	0.2	0.3	0.3	0.2

Tiến hành kiểm tra 300 hộp theo cách sau: từ mỗi hộp chon ngẫu nhiên ra 3 sản phẩm, nếu có ít nhất 2 sản phẩm loại 1 thì nhận hộp đó.

a) Tính xác suất để có ít nhất 275 hộp được nhận.

b) Tìm số hộp được nhận nhiều khả năng nhất.

Bài 2.94. Một kiện hàng có 10 sản phẩm, trong đó có 4 sản phẩm loại I, 3 sản phẩm loại II và 3 sản phẩm loại III. Lấy ngẫu nhiên từ kiện ra 2 sản phẩm. Gọi X và Y, theo thứ tự, là BNN chỉ số sản phẩm loại I và số sản phẩm loại II có trong 2 sản phẩm lấy ra

a) Lập bảng phân phối xác suất đồng thời của X và Y. Lập bảng phân phối xác suất lề của X và của Y.

b) Lập bảng phân phối xác suất điều kiện của Y, với điều kiện X lấy giá trị 1.

c) X và Y có độc lập không? Tại sao?

Bài 2.95. Cho BNN X có phân phối xác suất:

x_i	-2	0	1	2	3
$P(X = x_i)$	0.1	0.2	0.1	0.5	0.1

Hãy tính $E\left(X\right); D\left(X\right); Mod\left(X\right); m\left(X\right)$.

Bài 2.96. Một hộp có 10 sản phẩm. Các sản phẩm trong hộp gồm hai loại: loại A và loại B. Gọi X là BNN chỉ số sản phẩm loại A có trong hộp. Cho biết phân phối xác suất của X như sau:

x_i	1	2	3
$P(X = x_i)$	0.2	0.5	0.3

Lấy ngẫu nhiên từ hộp ra 3 sản phẩm (không hoàn lại). Đặt Y biểu thị số sản phẩm loại A có trong 3 sản phẩm lấy ra.

a) Tìm luật phân phối xác suất của Y.

b) Tính $E(Y), D(Y)$.

Bài 2.97. Một kiện hàng có 10 sản phẩm. Các sản phẩm trong kiện gồm hai loại: loại I và loại II. Gọi Y là BNN chỉ số sản phẩm loại II có trong kiện hàng. Cho biết phân phối xác suất của Y như sau:

y_i	3	4	5
$P(Y = y_i)$	0.4	0.2	0.4

Lấy ngẫu nhiên từ kiện hàng ra 4 sản phẩm (không hoàn lại). Đặt X biểu thị số sản phẩm loại II có trong 4 sản phẩm lấy ra.

a) Tìm luật phân phối xác suất của X.

b) Tính $E(X), D(X)$.

Bài 2.98. Một kiện hàng có 10 sản phẩm, trong đó có 4 sản phẩm loại I, 3 sản phẩm loại II và 3 sản phẩm loại III. Lấy ngẫu nhiên từ kiện ra 2 sản phẩm. Gọi X và Y, theo thứ tự, là BNN chỉ số sản phẩm loại I và số sản phẩm loại II có trong 2 sản phẩm lấy ra

a) Lập bảng phân phối xác suất đồng thời của X và Y. Lập bảng phân phối xác suất lề của X và của Y

b) Lập bảng phân phối xác suất điều kiện của Y, với điều kiện X lấy giá trị 1.

c) X và Y có độc lập không? Tại sao?

Bài 2.99. Lô hàng có 10 sản phẩm trong đó có 2 sản phẩm loại A, 3 sản phẩm loại B và 5 sản phẩm loại C. Lấy ngẫu nhiên ra 2 sản phẩm. Gọi X, Y lần lượt là BNN chỉ số sản phẩm loại A, loại B có trong 2 sản phẩm lấy ra.

a) Lập bảng phân phối xác suất đồng thời của X và Y. Lập bảng phân phối xác suất lề của X và của Y.

b) Lập bảng phân phối xác suất điều kiện của X, với điều kiện Y lấy giá trị 0.

c) X và Y có độc lập không? Tại sao?

Bài 2.100. Có 3 kiện hàng: kiện 1 chứa 10 sản phẩm tốt và 2 phế phẩm, kiện 2 chứa 9 sản phẩm tốt và 3 phế phẩm, kiện 3 chứa 7 sản phẩm tốt và 5 phế phẩm. Các sản phẩm giống hệt nhau về kích thước và trọng lượng. Chọn ngẫu nhiên một kiện hàng, rồi từ đó lấy ra 3 sản phẩm. Gọi X là số sản phẩm tốt trong 3 sản phẩm lấy ra.

a) Lập bảng phân phối xác suất của X.

b) Tìm $ModX, MedX$, kỳ vọng và phương sai của X.

ĐS:

a)

X	0	1	2	3
p	$\frac{1}{60}$	$\frac{107}{660}$	$\frac{101}{220}$	$\frac{239}{660}$

b) 2; 2; 2,1667; 0,5631.

Bài 2.101. Có hai chiếc bình: mỗi bình chứa 10 bi, trong đó có 4 bi đỏ, Lấy từ bình I ra 2 bi bỏ sang bình II. Sau đó lấy ngẫu nhiên đồng thời từ bình II ra 3 bi. Gọi X là số bi đỏ lấy được từ bình II.

a) Tìm bảng phân phối xác suất của X.

b) Tìm kỳ vọng và phương sai của X.

ĐS:

a)

X	0	1	2	3
p	$\frac{7}{55}$	$\frac{2}{5}$	$\frac{46}{165}$	$\frac{7}{165}$

b) 1,0848; 0,7201.

Bài 2.102. Một lô hàng có 14 sản phẩm tốt và 5 phế phẩm. Chọn ngẫu nhiên ra 3 sản phẩm. Gọi X là số sản phẩm tốt trong 3 sản phẩm lấy ra.

a) Lập bảng phân phối xác suất của X.

b) Tính trung bình số sản phẩm tốt lấy ra và phương sai của X.

c) Cho $Y = 9X + 12$, tính $E(Y), D(Y)$.

ĐS:

a)

X	0	1	2	3
p	$\frac{10}{969}$	$\frac{140}{969}$	$\frac{455}{969}$	$\frac{364}{969}$

b) 2,2105; 0,5171.

c) 31,8947; 41,8837..

Bài 2.103. Một lô hàng có 9 sản phẩm, trong đó có 6 sản phẩm loại I và 3 sản phẩm loại II. Lần lượt lấy ngẫu nhiên có hoàn lại 3 sản phẩm từ lô hàng này. Gọi X là số sản phẩm loại I lấy ra.

a) Lập bảng phân phối xác suất của X.

b) Tính trung bình số sản phẩm loại I trong 3 sản phẩm lấy ra và $D(X)$.

c) Giả sử $Y = 10X + D(X)$. Tính $E(Y), D(Y)$.

ĐS:

a)

X	0	1	2	3
p	$\frac{1}{27}$	$\frac{2}{9}$	$\frac{4}{9}$	$\frac{8}{27}$

b) 0,6667.

c) 20,6667; 66,6667.

Bài 2.104. Một kiện hàng có 6 sản phẩm loại I và 4 sản phẩm loại II. Tiền lời khi bán 1 sản phẩm loại I là 50000 đồng, loại II là 30000 đồng. Lấy ngẫu nhiên 3 sản phẩm để bán.

a) Tìm luật phân phối xác suất của số tiền lời thu được do bán 3 sản phẩm trên.

b) Tìm số tiền lời trung bình khi bán 3 sản phẩm trên.

ĐS:

a)

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

b) 126000.

Bài 2.105. Một thúng mận có 100 trái, trong đó có 10 trái bị hư. Một nam sinh viên đến mua 10 trái và nhờ người bán lựa dùm. Người bán bảo mận rất ngon và lấy ngẫu nhiên những trái mận để bán. Gọi X là trái mận hư chọn phải

a) Tính xác suất để không có trái mận nào bị hư.

b) Tìm trung bình số trái mận bị hư mà sinh viên này mua phải.

c) Tìm phương sai $D(X)$.

ĐS: 0,3305; 1; 0,8182.

Bài 2.106. Trong một lô hàng có 800 sản phẩm loại I và 200 sản phẩm loại II. Lần lượt lấy ngẫu nhiên có hoàn lại 3 sản phẩm từ lô hàng này. Gọi X là số sản phẩm loại I lấy được

a) Tìm quy luật phân phối xác suất của X.

b) Tính $E(5X + DX)$ và $D(6X + EX)$.

c) Tính xác suất lấy được nhiều nhất 2 sản phẩm loại I.

ĐS:

a) $X \sim B(3; 0,8)$.

b) 12,48; 17,28.

c) 0,448.

Bài 2.107. Trong một lô hàng có 800 sản phẩm loại I và 200 sản phẩm loại II. Lần lượt lấy ngẫu nhiên không hoàn lại 3 sản phẩm từ lô hàng này. Gọi X là số sản phẩm loại I lấy được

a) Tìm quy luật phân phối xác suất của X.

b) Tính $E(X + 10)$ và $D(5X + 2)$.

ĐS:

a)

X	0	1	2	3
p	0,0079	0,0958	0,3847	0,5116

b) 12,4; 11,095.

Bài 2.108. Có 3 lô hàng: mỗi lô có 1000 sản phẩm. Tỷ lệ sản phẩm loại II trong từng lô hàng lần lượt là 10, 20 và 30. Một người mua lấy ngẫu nhiên từ mỗi lô hàng ra 10 sản phẩm để kiểm tra. Người này thỏa thuận rằng: nếu lô hàng nào mà trong 10 sản phẩm lấy ra có không quá 2 sản phẩm loại II thì mua lô hàng đó. Tìm xác suất để có ít nhất một lô hàng được mua.

ĐS: 0,98604.

Bài 2.109. Cho X là BNN rời rạc tuân theo phân phối siêu hình học với các tham số N, M và n . Hãy tìm kỳ vọng và phương sai của X.

Bài 3.102 Cho X là BNN liên tục tuân theo phân phối chuẩn với hàm mật độ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma}\right)$$

Hãy tìm kỳ vọng và phương sai của X.

Bài 2.110. Cho một BNN X có hàm mật độ

$$f(x) = \frac{C}{e^{-x} + e^{x}}$$

a) Xác định hệ số C.

b) Xác định hàm phân phối xác suất $F(x)$.

c) Phải quan sát BNN X bao nhiêu lần để xác suất có ít nhất một lần biến $X \in (ln\frac{1}{\sqrt{3}}; ln\sqrt{3})$ không bé hơn 0.999.

Bài 2.111. Một cặp trẻ sinh đôi có thể do cùng trứng (sinh đôi thật), hay do hai trứng khác nhau sinh ra (sinh đôi giả). Các cặp sinh đôi thật luôn có cùng giới tính. Đối với cặp sinh đôi giả thì cùng giới tính có xác suất 0.5. Thống kê cho thấy 34% cặp sinh đôi đều là trai, 30% cặp sinh đôi đều là gái và 36% cặp sinh đôi có giới tính khác nhau.

a) Tìm tỉ lệ cặp sinh đôi thât.

b) Chọn ngẫu nhiên một cặp sinh đôi thì được một cặp cùng chứng. Tính xác suất để đó là cặp sinh đôi thật.

Bài 2.112. Một công ty kinh doanh mặt hàng A dự định sẽ áp dụng 1 trong 2 phương án kinh doanh. Gọi:

X_1 (triệu đồng/tháng) là lợi nhuận thu được khi áp dụng phương án 1
X_2 (triệu đồng/tháng) là lợi nhuận thu được khi áp dụng phương án 2

Thì $X_1 \sim N\,(140,\ 2500)$ và $X_2 \sim N\,(200,\ 3600)$. Nếu biết rằng: để công ty tồn tại và phát triển thì lợi nhuận thu được từ mặt hàng A phải đạt ít nhất 80 triệu đồng/tháng. Hãy cho biết công ty nên áp dụng phương án nào để kinh doanh mặt hàng A? Vì sao?

ĐS: $88,49\%; 97,72\%$ suy ra chọn phương án 2.

Bài 2.113. Tuổi thọ của người dân ở một địa phương là BNN X có hàm phân phối xác suất cho như sau

$$F\,(x) = \begin{cases} 0 & \text{khi} \quad x \le 0 \\ 1 - e^{-0,013x} & \text{khi} \quad x > 0 \end{cases}$$

a) Tính tỷ lệ người dân thọ từ 60 đến 70 tuổi.

b) Xác định hàm mật độ xác suất của X.

ĐS:

a) 0,0559.

b)

$$f(x) = \begin{cases} 0 & x \le 0 \\ 0,013e^{-0,013x} & x > 0 \end{cases}$$

Bài 2.114. 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).

$$f(x) = \begin{cases} k(30 - x) & \text{nếu} \quad x \in (0, 30) \\ 0 & \text{nếu} \quad x \notin (0, 30) \end{cases}$$

a) Tìm k.

b) Tìm xác suất để nhu cầu về loại hàng đó không vượt quá 12000 sản phẩm trong một năm.

c) Tìm nhu cầu trung bình hàng năm về loại hàng đó.

Bài 2.115. Cho BNN X liên tục có hàm mật độ xác suất là

$$f(x) = \begin{cases} a\left(3x - x^3\right) & \text{khi} \quad x \in \left[0, \sqrt{3}\right] \\ 0 & \text{khi} \quad x \notin \left[0, \sqrt{3}\right] \end{cases}$$

a) Xác định hằng số a.

b) Tìm $ModX$.

c) Tìm kỳ vọng và phương sai của X.

d) Tìm kỳ vọng và phương sai của $Y = 3X + 5$.

ĐS: $\frac{4}{9}; 1; 0,9238; 0,1467; 7,7713; 1,32.$

Bài 2.116. Thời gian xếp hàng chờ mua hàng của khách là biến ngẫu nhiên liên tục với hàm phân bố xác suất như sau (đơn vị: phút)

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x \le 0 \\ ax^3 - 3x^2 + 2x & \text{nếu} \quad 0 < x \le 1 \\ 1 & \text{nếu} \quad x > 1 \end{cases}$$

a) Tìm hệ số a.

b) Tìm thời gian xếp hàng trung bình.

c) Tìm xác suất để trong 3 người xếp hàng thì có không quá 2 người phải chờ quá 0,5 phút.

ĐS: 2; 0,5; 0,875.

Bài 2.117. Cho biến ngẫu nhiên liên tục có hàm phân bố xác suất sau đây:

$$F\left(x\right)=\frac{1}{2}+\frac{1}{\pi}\arctan x$$

a) Tìm $P(0<X<1)$.

b) Tìm hàm mật độ xác suất của X.

ĐS: $\frac{1}{4};\frac{1}{\pi(1+x^2)}$.

Bài 2.118. Biến ngẫu nhiên X có hàm mật độ

$$f\left(x\right)=\begin{cases}0 & \text{nếu} \quad x\notin\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\\ \frac{2}{\pi}\cos^2 x & \text{nếu} \quad x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$$

Tìm $E(X), D(X)$.

Bài 2.119. Cho BNN X có hàm mật độ xác suất như sau:

$$f\left(x\right)=\begin{cases}k\sin x & \text{nếu} \quad x\in[0;\pi]\\ 0 & \text{nếu} \quad x\notin[0;\pi]\end{cases}$$

a) Xác định k ?

b) Tính $P\left(X>\frac{\pi}{2}\right)$; $P\left(\frac{\pi}{4}<X<\frac{\pi}{2}\right)$.

c) Tính kì vọng và phương sai của X.

Bài 2.120. Cho BNN có hàm phân phối xác suất như sau:

$$F\left(x\right)=\begin{cases}0 & \text{nếu} \quad x\le a\\ A+B\arcsin\frac{x}{a} & \text{nếu} \quad -a<x<a\\ 1 & \text{nếu} \quad x\ge a\end{cases}$$

a) Tìm A, B sao cho $F(x)$ là liên tục.

b) Tìm hàm mật độ xác suất $F(x)$.

c) Tính $P\left(-\dfrac{a}{2} < X < \dfrac{a}{2}\right)$

Bài 2.121. Giả sử hàm phân phối của BNN X là : $F\left(x\right) = \frac{1}{\pi}\arctan x + \frac{1}{2}$. Tìm hàm mật độ của X và tính xác suất $P\left(-1 \le X < 1\right)$.

Bài 2.122. Giả sử hàm mật độ của BNN X là

$$f\left(x\right) = \begin{cases} Ae^{-\lambda x} & \text{nếu} \quad x > 0; \lambda > 0 \\ 0 & \text{nếu} \quad x \le 0 \end{cases}$$

Tìm A. Tìm hàm phân phối của X. Tính tuổi thọ trung bình (kỳ vọng của X).

Bài 2.123. Biến ngẫu nhiên X có hàm mật độ phụ thuộc vào hằng số A:

$$f\left(x\right) = \begin{cases} A\left(x-2\right)\left(4-x\right) & \text{nếu} \quad 2 \le x \le 4 \\ 0 & \text{nếu} \quad x < 2, x > 4 \end{cases}$$

a) Tìm A.

b) Tìm $F(x)$.

Bài 2.124. Cho X có phân phối Reley, đặc trưng cho tâm sai các chi tiết máy với hàm phân phối :

$$F\left(x\right) = \begin{cases} 1 - e^{\frac{-x^2}{2\sigma^2}} & \text{nếu} \quad x \ge 0\left(\sigma > 0\right) \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$

a) Tìm $f(x)$.

b) Tính$P\left(0 \le X < \sigma\right)$.

Bài 2.125. BNN X được gọi là có phân phối Cauchy nếu:

$$F\left(X\right) = A + B\arctan\frac{x}{a}, x \in \left(-\infty, +\infty\right),$$

với A, B, a là các hằng số $(a > 0)$.

a) Hãy xác định A, B.

b) Tìm hàm mật độ $f(x)$.

Bài 2.126. Hàm phân phối của biến ngẫu nhiên liên tục tuyệt đối X có dạng:

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x \le 0 \\ \frac{1}{2} - \frac{1}{2}\cos x & \text{nếu} \quad 0 < x \le \pi \\ 1 & \text{nếu} \quad x > \pi \end{cases}$$

Tìm hàm mật độ $f(x)$.

Bài 2.127. Cho hàm số

$$f(x) = \begin{cases} x - \dfrac{x^3}{4} & \text{khi} \quad x \in (0,2) \\ 0 & \text{khi} \quad x \notin (0,2) \end{cases}$$

a) Chứng minh rằng $f(x)$ là hàm mật độ xác suất của BNN liên tục X nào đó.

b) Tìm $ModX$.

c) Tìm $P(3/2 < X < 5/2)$.

d) Tìm kỳ vọng và phương sai của X.

ĐS: 1; 1,1547; 0,1914; 1,0667; 0,1956.

Bài 2.128. Gọi X là tuổi thọ của con người trong một vùng nào đó. Một công trình nghiên cứu cho biết hàm mật độ xác suất của X là:

$$f(x) = \begin{cases} cx^2(100 - x)^2 & \text{khi} \quad x \in [0,100] \\ 0 & \text{khi} \quad x \notin [0,100] \end{cases}$$

a) Xác định hằng số c.

b) Tìm trung bình và phương sai của X.

c) Tìm xác suất để một người có tuổi thọ từ 60 tuổi trở lên.

d) Tìm xác suất để một người có tuổi thọ từ 60 tuổi, biết rằng hiện nay người này đã 50 tuổi.

ĐS: $3.10^{-9}; 50; \frac{2500}{7}; 0,31744; 0,63488$.

Bài 2.129. Người ta muốn lấy một số hạt lúa từ một kho lúa có tỉ lệ hạt lép là 10% để kiểm tra. Biết rằng kho lúa có rất nhiều hạt.

a) Phải lấy ít nhất bao nhiêu hạt lúa để xác suất có một hạt lép không bé hơn 90%?

b) Lấy ngẫu nhiên 150 hạt lúa, tính xác suất để trong đó có 20 hạt lép; có từ 10 đến 45 hạt lép.

Bài 2.130. Sản phẩm trong nhà máy được đóng thành từng kiện, mỗi kiện gồm 14 sản phẩm trong đó có 8 sản phẩm loại A và 6 sản phẩm loại B. Khách hàng chọn cách kiểm tra như sau: từ mỗi kiện lấy ra 4 sản phẩm; nếu thấy số sản phẩm thuộc loại A nhiều hơn số sản phẩm thuộc loại B thì mới nhận kiện đó; ngược lại thì loại kiện đó. Kiểm tra 100 kiện (trong rất nhiều kiện). Tính xác suất để:

a) Có 38 kiện được nhận.

b) Có từ 40 đến 43 kiện được nhận.

c) Có ít nhất 25 kiện được nhận.

Bài 2.131. Sản phẩm trong nhà máy được đóng thành từng kiện, mỗi kiện gồm 20 sản phẩm trong đó có 14 sản phẩm loại A và 6 sản phẩm loại B. Khách hàng chọn cách kiểm tra như sau: từ mỗi kiện lấy ra 3 sản phẩm; nếu có ít nhất 1 sản phẩm loại B thì không nhận kiện hàng đó. Kiểm tra 100 kiện (trong rất nhiều kiện). Tính xác suất để:

a) Có 40 kiện được nhận

b) Có từ 45 đến 48 kiện được nhận.

c) Có ít nhất 50 kiện được nhận.

Bài 2.132. Trong ngày hội thi, mỗi chiến sĩ sẽ chọn ngẫu nhiên một trong hai loại súng và với khẩu súng chọn được sẽ bắn 100 viên đạn. Nếu có từ 65 viên trở lên trúng bia thì được thưởng. Giả sử đối với chiến sĩ A, xác suất bắn 1 viên trúng bia bằng khẩu súng loại I là 60% và bằng khẩu súng loại II là 50%.

a) Tính xác suất để chiến sĩ A được thưởng.

b) Giả sử chiến sĩ A dự thi 10 lần. Hỏi số lần được thưởng tin chắc nhất là bao nhiêu?

c) Chiến sĩ A phải tham gia hội thi ít nhất bao nhiêu lần để xác suất có ít nhất một lần được thưởng không nhỏ hơn 98%.

Bài 2.133. Một xí nghiệp có 2 máy I và II. Trong ngày hội thi, mỗi công nhân dự thi sẽ chọn ngẫu nhiên 1 máy và với máy đó sản xuất 100 sản phẩm. Nếu số sản phẩm loại tốt sản suất được không ít hơn 70 thì được thưởng. Giả sử với một công nhân A, xác suất sản xuất được sản phẩm loại tốt với hai máy lần lượt là 65% và 70%.

a) Tính xác suất công nhân A được thưởng.

b) Giả sử A dự thi 20 lần. Số lần được thưởng có nhiều khả năng nhất là bao nhiêu?

c) Công nhân A phải dự thi ít nhất bao nhiêu lần để xác suất có ít nhất một lần được thưởng không nhỏ hơn 95%.

Bài 2.134. Sản phẩm của một nhà máy được đóng thành từng hộp, mỗi hộp chứa 10 sản phẩm. Gọi X là BNN chỉ số sản phẩm loại 1 có trong hộp. Cho biết X có phân phối xác suất như sau:

x_i	7	8	9	10
$P(X = x_i)$	0.2	0.3	0.3	0.2

Tiến hành kiểm tra 300 hộp theo cách sau: từ mỗi hộp chon ngẫu nhiên ra 3 sản phẩm, nếu có ít nhất 2 sản phẩm loại 1 thì nhận hộp đó.

a) Tính xác suất để có ít nhất 275 hộp được nhận.

b) Tìm số hộp được nhận nhiều khả năng nhất.

Bài 2.135. Thời gian để sản xuất một sản phẩm loại A là một BNN tuân theo phân phối chuẩn với các tham số $\mu = 10, \sigma = 1$ (đơn vị phút)

a) Tính xác suất để một sản phẩm loại A nào đó được sản suất trong khoảng thời gian từ 8 phút đến 14 phút.

b) Tính thời gian cần thiết để sản xuất một sản phẩm loại A bất kỳ.

Bài 2.136. Cho BNN X tuân theo luật phân phối $N(\mu, \sigma^2)$. Biết rằng X lấy giá trị nhỏ hơn 60 với xác suất 0,1003 và lấy giá trị lớn hơn 90 với xác suất 0,0516. Hãy tính μ, σ .

Bài 2.137. Thời gian để máy M sản suất ra một sản phẩm S là một biến ngẫu nhiên tuân theo luật phân phối chuẩn. Sau nhiều lần quan sát, người ta thấy rằng: Thời gian để sản xuất S mất hơn 2 giờ chiếm tỉ lệ 15,87% và mất hơn 2,5 giờ chiếm tỉ lệ 2,28%. Tìm khoảng thời gian cần thiết, hầu như chắc chắn, để sản xuất một sản phẩm S.

Bài 2.138. Thời gian (tính bằng tháng) từ lúc vay đến lúc trả tiền của một khách hàng tại một ngân hàng là BNN X có phân phối chuẩn trung bình 18 tháng, độ lệch tiêu chuẩn 4 tháng. Tính tỉ lệ khách hàng trả tiền cho ngân hàng trong khoảng từ 10 đến 19 tháng; không ít hơn một năm; ít hơn 9 tháng.

Bài 2.139. Có 3 giống lúa, sản lượng của chúng (đơn vị tính bằng tấn/ha) là 3 đại lượng ngẫu nhiên: $X_1 \sim N(8;0,8)\,; X_2 \sim N(10;0,6)\,; X_3 \sim (10;0,5)$. Cần chọn một trong ba giống đó để trồng. Theo bạn thì nên chọn giống nào? Tại sao?

Bài 2.140. Đường kính của một loại trục máy do một máy tiện sản xuất ra là một BNN có phân phối chuẩn với trung bình 250mm và phương sai 25 mm^2 . Trục máy được gọi là hợp quy cách nếu đường kính từ 245mm đến 255 mm. Cho máy sản xuất 100 trục. Tính xác suất để có:

a) 50 trục hợp quy cách.

b) Không quá 80 trục hợp quy cách.

Bài 2.141. Sản phẩm được đóng thành hộp. Mỗi hộp có 10 sản phẩm trong đó có 7 sản phẩm loại A. Người mua hàng quy định cách kiểm tra như sau: từ hộp lấy ngẫu nhiên 3 sản phẩm, nếu thấy cả 3 sản phẩm đều là loại A thì nhận hộp đó. Nếu ngược lại thì loại hộp đó. Giả sử kiểm tra 200 hộp trong rất nhiều hộp, tính xác suất để:

a) Có 50 hộp được nhận.

b) không quá 60 hộp được nhận.

c) Phải kiểm tra ít nhất bao nhiêu hộp để xác suất có ít nhất 1 hộp được nhận $\geq 95\%$.

Bài 2.142. Xác suất sinh được một em bé gái là 0,52. Tính xác suất sao cho trong 300 em bé sắp sinh: Có 170 bé trai; số bé trai vào khoảng từ 150 đến 170; số bé trai ít nhất là 170.

Bài 2.143. Một cơ sờ sản suất, trung bình trong một tuần, nhận được 4 đơn đặt hàng. Biết rằng số đơn đặt hàng X mà cơ sở nhận được trong một tuần là một BNN có phân phối Poisson. Tính xác suất để cơ sở đó

a) Nhận được hơn 6 đơn đặt hàng trong một tuần.

b) Nhận được 7 đơn đặt hàng trong hai tuần liên tiếp.

Bài 2.144. Trọng lượng X của một loại sản phẩm (đơn vị: gam) có phân phối chuẩn. Biết 65% số sản phẩm có trọng lượng lớn hơn 20 gam và 8% số sản phẩm có trọng lượng lớn hơn 30 gam.

a) Nấu sản phẩm được chấp nhận có trọng lượng nhỏ hơn 25 gam thì tỉ lệ sản phẩm bị loại là bao nhiêu?

b) Cần quy định trọng lượng tối thiểu là bao nhiêu để tỉ lệ sản phẩm bị loại nhỏ hơn 2%.

Bài 2.145. Tuổi thọ X của một loại bóng đèn (đơn vị: năm) là biến ngẫu nhiên có phân phối chuẩn trung bình 4,2 năm, phương sai 2.25 $(\text{năm})^2$. Khi bán một bóng đèn thì lãi 100 ngàn đồng, song nếu đèn phải bảo hành thì lỗ 300 ngàn đồng. Vậy để tiền lãi trung bình khi bán một bóng đèn là 30 ngàn đồng thì phải quy định thời gian bảo hành là bao nhiêu?

ĐS: 2,8 (năm).

Bài 2.146. Một trạm cho thuê xe có 3 xe taxi. Hàng ngày phải nộp thuế 8 USD/1 xe dù xe có được thuê hay không. Mỗi chiếc xe taxi được thuê với giá 20 USD/1 nhày. Giả sử yêu cầu thuê xe của trạm là X có phân phối Poisson với tham số $\lambda = 2.8$.

a) Gọi Y là số tiền lời trong 1 ngày của trạm. Tính số tiền lời trung bình của trạm thu được trong một ngày.

b) Giải bài toán trong trường hợp trạm có 4 xe.

c) Đưa ra kết luận trạm nên có 3 hay 4 xe.

ĐS: 20,772; 18,932; 3.

Bài 2.147. Một lô hàng có 12 thùng sản phẩm, mỗi thùng có 10 sản phẩm, trong đó có 3 phế phẩm. Lấy ngẫu nhiên từ mỗi thùng một sản phẩm để kiểm tra. Tìm số phế phẩm tin chắc nhất và số phế phẩm trung bình có trong các sản phẩm lấy ra.

ĐS: 3; 3,6.

Bài 2.148. Biến cố nào trong các biến cố sau có sác suất lớn hơn.

a) Có ít nhất 2 lần xuất hiện mặt 6 chấm khi gieo con súc sắc 12 lần.

b) Có ít nhất 3 lần xuất hiện mặt 6 chấm khi gieo con súc sắc 18 lần.

ĐS: 0,6187; 0,5973.

Bài 2.149. Trong một nhà máy có 3 phân xưởng dệt. Mỗi phân xưởng có 100 máy dệt hoạt động độc lập với nhau. Xác suất trong một ca sản xuất mỗi máy dệt hỏng là như nhau và = 2,5%.

a) Tìm luật phân phối xác suất của số máy hỏng trong một ca sản xuất của một phân xưởng.

b) Trung bình trong một ca sản xuất toàn nhà máy có bao nhiêu máy dệt bị hỏng?

c) Nếu mỗi kĩ sư có thể chữa tối đa được 2 máy dệt bị hỏng trong 1 ca sản xuất thì nhà máy cần bố trí trực sửa chữa mỗi ca bao nhiêu kĩ sư là hợp lí nhất.

ĐS: $X \sim B(100; 0,025); 2,5; 7,5; 3$.

Chương 3

Vector ngẫu nhiên

3.1 Vector ngẫu nhiên

Trong nhiều mô hình xác suất, ta buộc phải khảo sát đến đồng thời nhiều đại lượng ngẫu nhiên trên cùng một đối tượng. Chẳng hạn, khi quan tâm đến chỉ số BMI (Body mass index) để biết thông tin về độ béo phì ở người ta cần khảo sát đồng thời hai đại lượng ngẫu nhiên X là chiều cao và Y là cân nặng. Như vậy chỉ số BMI được cho bởi một bộ hai biến ngẫu nhiên (X, Y), ta gọi là một véc tơ ngẫu nhiên. Trong trường hợp này, véc tơ ngẫu nhiên có hai thành phần nên được gọi là véc tơ ngẫu nhiên hai chiều. Trường hợp tổng quát, một véc tơ ngẫu nhiên có n thành phần được gọi là véc tơ ngẫu nhiên n chiều.

3.1.1 Định nghĩa

Giả sử $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên trên không gian xác suất $(\Omega, \mathcal{A}, P)$. Hàm X được xác định như sau:

$$\begin{aligned} X : \Omega &\longrightarrow R^n \\ \omega &\longmapsto X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \end{aligned}$$

được gọi là một véc tơ ngẫu nhiên n chiều trên Ω. Kí hiệu $X = (X_1, X_2, \ldots, X_n)$.

3.1.2 Hàm phân phối đồng thời

Cho biến ngẫu nhiên n chiều $X = (X_1, X_2, \ldots, X_n)$ trên không gian xác suất $(\Omega, \mathcal{A}, P)$. Hàm F_X được xác định như sau:

$$\begin{aligned} F_X : R^n &\longrightarrow R \\ (x_1, x_2, \ldots, x_n) &\longmapsto F_X(x_1, x_2, \ldots, x_n) = P\left(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n\right) \end{aligned}$$

được gọi là hàm phân phối (tích lũy) đồng thời của các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ hay gọi là hàm phân phối của véc tơ ngẫu nhiên X.

Tính chất. Đặt véc tơ ngẫu nhiên $X = (X_1, X_2, \ldots, X_n)$.

a. Cho $a = (a_1, a_2, \ldots, a_n), b = (b_1, b_2, \ldots, b_n) \in R^n$.

$$\text{Nếu } a_i \leq b_i, i = \overline{1, n} \text{ thì } F_X(a) \leq F_X(b).$$

b. Hàm F_X là hàm liên tục trái theo mỗi biến, nghĩa là

$$\lim_{x_i \to x_i^{0-}} F_X(x_1, x_2, \ldots, x_i, \ldots, x_n) = F_X(x_1, x_2, \ldots, x_i^0, \ldots, x_n).$$

c. Với mỗi $i \in \{1, 2, \ldots, n\}$ ta có

$$\lim_{x_i \to -\infty} F_X(x_1, x_2, \ldots, x_i, \ldots, x_n) = 0.$$

Với mọi $i \in \{1, 2, \ldots, n\}$

$$\lim_{\substack{x_1 \to +\infty \\ \cdots \\ x_n \to +\infty}} F_X(x_1, x_2, \ldots, x_n) = 1.$$

d. Cho $a = (a_1, a_2, \ldots, a_n), b = (b_1, b_2, \ldots, b_n) \in R^n$. Với $a_i \leq b_i, i = \overline{1, n}$ ta có

$$\begin{aligned} & F_X\left(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \ldots, a_n \leq X_n \leq b_n\right) = \\ = \; & F_X(b_1, b_2, \ldots b_n) - \sum_{i=1}^{n} P_i + \sum_{i<j} P_{ij} + \cdots + (-1)^n F_X(a_1, a_2, \ldots, a_n), \end{aligned}$$

trong đó $P_{ij\ldots k} = F_X(c_1, c_2, \ldots, c_n), c_i = a_i, c_j = a_j, \ldots, c_n = a_n$ và $c_{s \neq i,j,k} = b_s$.

Chẳng hạn

$$P\left(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2\right) = F_X(b_1, b_2) - F_X(a_1, b_2) - F_X(b_1, a_2) + F_X(a_1, a_2).$$

3.1.3 Hàm mật độ đồng thời

Véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$ có hàm mật độ của véc tơ ngẫu nhiên X được xác định bởi

$$f_X(x_1, x_2, \dots, x_n) = P\left(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\right).$$

- Trong trường hợp X thuộc dạng rời rạc (các biến thành phần $X_i, i = \overline{1, n}$ rời rạc), nếu F_X là hàm phân phối của X thì

$$F_X\left(x_1, x_2, \dots, x_n\right) = \sum_{u_1 < x_1} \sum_{u_2 < x_2} \cdots \sum_{u_n < x_n} f_X(u_1, u_2, \dots, u_n).$$

- Cho véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$ liên tục và có hàm phân phối là F_X. Nếu một hàm f_X khả tích trên R^n sao cho

$$F_X(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_X(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n$$

  thì f_X được gọi là hàm mật độ của véc tơ ngẫu nhiên X.

- Nếu véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$ liên tục có hàm phân phối F_X và hàm mật độ f_X thì

$$\forall (x_1, x_2, \dots, x_n) \in R^n, \qquad f_X(x_1, x_2, \dots, x_n) = \frac{\partial^n F_X(X_1, X_2, \dots, X_n)}{\partial x_1 \partial x_2 \dots \partial x_n}.$$

3.1.4 Hàm phân phối biên và hàm mật độ biên

Cho véc tơ ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$ có hàm phân phối F_X và hàm mật độ f_X.

- Hàm phân phối biến F_i của biến ngẫu nhiên $X_i, i \in \{1, 2, \dots, n\}$ được xác định trên R như sau

$$F_i(x_i) = \lim_{\substack{x_1 \to +\infty \\ \dots \\ x_{i-1} \to +\infty \\ x_{i+1} \to +\infty \\ \dots \\ x_n \to +\infty}} F_X(x_1, x_2, \dots, x_n)$$

- Hàm mật độ biên f_i của biến ngẫu nhiên $X_i, i \in \{1, 2, \dots, n\}$ được xác định trên R như sau

$$f_i(x_i) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(x_1, x_2, \dots, x_n) du_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

khi véc tơ ngẫu nhiên X liên tục, hoặc

$$f_i\left(x_i\right)=\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_n}f_X(x_1,x_2,\ldots,x_n)$$

khi X rời rạc.

3.1.5 Tính độc lập của các biến ngẫu nhiên

Các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ được gọi là độc lập nếu với mọi nhóm $X_{i_1}, X_{i_2}, \ldots, X_{i_k}, 1 \le k \le n$ của chúng, ta có

$$P\left(X_{i_1}<x_{i_1}, X_{i_2}<x_{i_2}, \ldots, X_{i_k}<x_{i_k}\right)=P\left(X_{i_1}<x_{i_1}\right)P\left(X_{i_2}<x_{i_2}\right)\ldots P\left(X_{i_k}<x_{i_k}\right)$$

với bất kỳ $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$.
Ta cũng có thể nói các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập nếu với bất kỳ $x_1, x_2, \ldots, x_n$ đẳng thức sau đây luôn đúng

$$P\left(X_1<x_1, X_2<x_2, \ldots, X_n<x_n\right)=P\left(X_1<x_1\right)P\left(X_2<x_2\right)\ldots P\left(X_n<x_n\right)$$

hay viết dưới dạng hàm phân phối xác suất

$$F_{X_1,X_2,\ldots,X_n}\left(x_1,x_2,\ldots,x_n\right)=F_{X_1}\left(x_1\right)F_{X_2}\left(x_2\right)\ldots F_{X_n}\left(x_n\right).$$

Đối với hàm mật độ, các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ được gọi là độc lập nếu và chỉ nếu với mọi $x_1, x_2, \ldots, x_k$ ta luôn có

$$P(X_1=x_1, X_2=x_2, \ldots X_n=x_n)=P(X_1=x_1)P(X_2=x_2)\ldots P(X_n=x_n)$$

hoặc f_X là hàm mật độ đồng thời và $f_1, f_2, \ldots, f_n$ là hàm mật độ biên thì

$$f_X(x_1,x_2,\ldots,x_n)=f_1(x_1)f_2(x_2)\ldots f_n(x_n).$$

Ngược lại nếu hàm phân phối đồng thời của các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ là $F_{X_1,X_2,\ldots,X_n}$ phân tích được dưới dạng

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=F_{X_1}\left(x_1\right)F_{X_2}\left(x_2\right)\ldots F_{X_n}\left(x_n\right)$$

và các hàm $F_{X_i}, i=1,2,\ldots,n$ thỏa $\lim\limits_{x\to+\infty} F_{X_i}(x)=1$ thì các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập và có hàm phân phối xác suất tương ứng là $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$.

3.2 Vector ngẫu nhiên hai chiều

3.2.1 Phân phối xác suất của vector ngẫu nhiên rời rạc hai chiều

1. Phân phối xác suất đồng thời của (X, Y)

$X \quad \backslash \quad Y$	y_1	y_2	...	y_j	...	y_n	tổng dòng
x_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1n}	$p_{1\bullet}$
x_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2n}	$p_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{in}	$p_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	...	p_{mj}	...	p_{mn}	$p_{m\bullet}$
Tổng cộng	$p_{\bullet 1}$	$p_{\bullet 2}$	...	$p_{\bullet j}$	...	$p_{\bullet n}$	1

trong đó $P(X = x_i; Y = y_i) = p_{ij}$ và $\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{m} p_{ij} = 1$.

2. Phân phối xác suất thành phần (phân phối lề)

a. Bảng phân phối xác suất của X

X	x_1	x_2	...	x_m
P	$p_{1\bullet}$	$p_{2\bullet}$	...	$p_{m\bullet}$

trong đó $p_{i\bullet} = p_{i1} + p_{i2} + ... + p_{in}$ (tổng dòng i của bảng phân phối xác suất đồng thời).

Kỳ vọng của X là

$$\mathrm{EX} = x_1 p_{1\bullet} + x_2 p_{2\bullet} + \cdots + x_m p_{m\bullet}$$

b. Bảng phân phối xác suất của Y

Y	y_1	y_2	...	y_n
P	$p_{\bullet 1}$	$p_{\bullet 2}$	...	$p_{\bullet n}$

trong đó $p_{\bullet j} = p_{1j} + p_{2j} + ... + p_{mj}$ (tổng cột j của bảng phân phối xác suất đồng thời).

Kỳ vọng của Y là

$$\mathrm{EY} = y_1 p_{\bullet 1} + y_2 p_{\bullet 2} + \cdots + y_n p_{\bullet n}$$

3. Phân phối xác suất có điều kiện

$$P\left(X = x_i \left| Y = y_j \right.\right) = \frac{P\left(X = x_i, Y = y_j\right)}{P\left(Y = y_j\right)} = \frac{p_{\text{ij}}}{p_{\bullet j}},\ i = \overline{1, m}$$

$$P\left(Y = y_i \left| X = x_i \right.\right) = \frac{P\left(X = x_i, Y = y_j\right)}{P\left(X = x_i\right)} = \frac{p_{\text{ij}}}{p_{i\bullet}},\ j = \overline{1, n}$$

a. Bảng phân phối xác suất của X với điều kiện $Y = y_i$

X	x_1	x_2	...	x_m
$P(X = x_i/Y = y_j)$	$\frac{p_{1j}}{p_{\bullet j}}$	$\frac{p_{2j}}{p_{\bullet j}}$	...	$\frac{p_{mj}}{p_{\bullet j}}$

Kỳ vọng của X với điều kiện $Y = y_j$ là

$$EX = \frac{1}{p_{\bullet j}}\left(x_1 p_{1j} + x_2 p_{2j} + \ldots + x_m p_{mj}\right)$$

b. Bảng phân phối xác suất của Y với điều kiện $X = x_i$

Y	y_1	y_2	...	y_n
$P(Y = y_j/X = x_i)$	$\frac{p_{i1}}{p_{i\bullet}}$	$\frac{p_{i2}}{p_{i\bullet}}$	...	$\frac{p_{in}}{p_{i\bullet}}$

Kỳ vọng của Y với điều kiện $X = x_i$ là

$$EY = \frac{1}{p_{i\bullet}}\left(y_1 p_{i1} + y_2 p_{i2} + \ldots + y_n p_{in}\right)$$

3.2.2 Phân phối xác suất của vector ngẫu nhiên liên tục hai chiều

1. Hàm mật độ đồng thời của (X, Y)

Hàm hai biến $f(x, y) \geq 0$ xác định trên $\mathbb{R}^2$ được gọi là hàm mật độ của vector ngẫu nhiên (X, Y) nếu:

$$\iint\limits_{\mathbb{R}\mathbb{R}} f(x,y)dxdy = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\mathrm{d}y = 1$$

Xác suất của vector (X, Y) trên tập $D \subset \mathbb{R}^2$ là

$$P\{(X,Y) \in D\} = \iint\limits_{\mathcal{D}} f(x,y)dxdy$$

2. Hàm mật độ thành phần

a. Hàm mật độ của X $f_X(x) = \int\limits_{-\infty}^{+\infty} (x,y)\,\mathrm{d}y$.

b. Hàm mật độ của Y $f_Y(y) = \int\limits_{-\infty}^{+\infty} (x,y)\,dx$.

Chú ý:

- Khi tìm hàm $f_X(x)$, ta lấy tích phân hàm $f(x,y)$ theo biến y và điều kiện x phải độc lập đối với y.
- Khi tìm hàm $f_Y(y)$, ta lấy tích phân hàm $f(x,y)$ theo biến x và điều kiện y phải độc lập đối với x

c. Trung bình thành phần

$$E\{f_X(x)\} = \int\limits_{-\infty}^{+\infty} x.f_X(x)\,d\mathrm{x},\ E\{f_Y(y)\} = \int\limits_{-\infty}^{+\infty} y.f_Y(y)\ dy$$

3. Hàm mật độ có điều kiện

a. Hàm mật độ có điều kiện của X khi biết $Y = x$

$$f_X(x\,|y) = \frac{f(x,y)}{f_Y(y)}$$

b. Hàm mật độ có điều kiện của Y khi biết $X = x$

$$f_Y(y\,|x) = \frac{f(x,y)}{f_X(x)}$$

3.2.3 Covarian - Hệ số tương quan

Cho X và Y là hai BNN trên cùng một không gian mẫu, có kỳ vọng lần lượt là μ_X và μ_Y . Nếu X và Y không độc lập, người ta muốn có một đại lượng đo mối liên hệ giữa chúng. Một mối liên hệ đơn giản là liên hệ tuyến tính. Để đánh giá điều này, người ta để ý đến BNN $(X-\mu_X)(Y-\mu_Y)$ và kỳ vọng của nó (nếu tồn tại).

1. Covarian

Covarian của hai biến ngẫu nhiên X, Y ký hiệu là $Cov(X,Y)$, được xác định bởi biểu thức

$$Cov\left(X,Y\right)=E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]=E\left(XY\right)-\mu_X.\mu_Y$$

2. Hệ số tương quan

Hệ số tương quan của hai biến ngẫu nhiên X, Y ký hiệu là $\rho(X,Y)$, được xác định bởi biểu thức

$$\rho\left(X,Y\right)=\frac{Cov\left(X,Y\right)}{\sigma_X\sigma_Y}=\frac{E\left(XY\right)-E\left(X\right)E\left(Y\right)}{\sigma_X\sigma_Y}$$

($-1\leq\rho(X,Y)\leq 1$ và $|\rho(X,Y)|=1\Leftrightarrow X,Y$ phụ thuộc tuyến tính)

3. Phương sai của hai biến ngẫu nhiên X, Y

$$\begin{aligned}D\left(X+Y\right)&=D\left(X\right)+D\left(Y\right)+2Cov\left(X,Y\right);\\ D\left(aX+bY\right)&=a^2D\left(X\right)+b^2D\left(Y\right)+2abCov\left(X,Y\right).\end{aligned}$$

4. Nếu X, Y độc lập thì

- $Cov(X,Y)=0$.
- $\rho(X,Y)=0$.
- $D(X+Y)=D(X)+D(Y)$.

(Nếu một trong hai biến BNN X, Y là hằng số thì $\rho(X,Y)=0$)

Một số ví dụ

Ví dụ 3.1 Chọn ngẫu nhiên 2 bi từ hộp có 3 bi đỏ, 2 bi vàng, 4 bi xanh. Gọi X, Y tương ứng là số bi đỏ và số bi vàng có trong 2 bi lấy ra.

a) Tìm phân phối xác suất đồng thời của X và Y .

b) Tính $P\{(X,Y) \in A\}$ với $P\{(x,y)/x+y \leq 1\}$.

c) Tìm các phân phối xác suất biên của X , của Y .

Ví dụ 3.2 Thống kê dân số của một vùng theo hai chỉ tiêu : Giới tính X ; học vấn Y được kết quả cho trong bảng

X | Y	Thất học: 0	Phổ thông: 1	Sau phổ thông: 2
Nam: 0	0,1	0,25	0,16
Nữ: 1	0,15	0,15	0,12

a) Lập bảng phân phối xác suất của học vấn ; giới tính.

b) Học vấn có độc lập với giới tính không?

c) Tìm xác suất để chọn ngẫu nhiên một người của vùng thì người đó không bị thất học.

d) Lập bảng phân phối xác suất học vấn của nữ ; tính trung bình học vấn của nữ.

e) Tìm tỉ lệ nữ có học vấn không vượt quá phổ thông.

Ví dụ 3.3 Tỉ lệ thời gian trong một ngày làm việc của một ngân hàng giành cho 2 loại dịch vụ ; ưu tiên, thông thường là X, Y. Giả sử (X, Y) có hàm mật độ xác suất:

$$f(x,y) = \begin{cases} k\left(x+y^2\right) & 0 \leq x \leq 1; 0 \leq y \leq 1 \\ 0 & x, y \notin [0,1] \end{cases}$$

a) Tính xác suất không có loại dịch vụ nào chiếm quá một phần tư thời gian của ngày làm việc.

b) Tính xác suất thời gian giành cho dịch vụ ưu tiên không vượt quá nửa thời gian giành cho dịch vụ thông thường.

c) Tìm hàm mật độ xác suất biên của X.

d) X và Y có độc lập với nhau không ?

Ví dụ 3.4 Tuổi thọ 2 bộ phận của một máy tính xách tay là X và Y có hàm mật độ xác suất đồng thời :

$$f(x,y) = \begin{cases} xe^{-x(1+y)} & \text{khi}\, x > 0, y > 0 \\ 0 & \text{khi}\, x \leq 0, y \leq 0\,. \end{cases}$$

a) Tìm hàm mật độ $f_X(x)$ của X .

b) Tìm mật độ điều kiện của Y khi $X = x$.

c) Tìm hàm hồi quy của Y theo X.

Ví dụ 3.5 Cho các biến ngẫu nhiên rời rạc X và Y có phân phối xác suất đồng thời:

$X \quad \vert \quad Y$	1	3
2	0,1	0,45
4	0,3	0,15

Tìm kỳ vọng của $h(X,Y) = XY^2$.
ĐS: 14,9.

Ví dụ 3.6 Cho biến ngẫu nhiên liên tục (X,Y) có hàm mât độ xác suất

$$f(x,y) = \begin{cases} \frac{x\left(1+3y^2\right)}{4} & \text{khi } 0 < x < 2,\ 0 < y < 1 \\ 0 & \text{khi } x \notin (0,2)\,, y \notin (0,1)\,. \end{cases}$$

Tìm kì vọng của $h(X,Y) = Y/X$.
ĐS: $\dfrac{5}{8}$.

Ví dụ 3.7 Thống kê về lãi suất cổ phiếu tính cho 100 (USD) khi đầu tư vào hai ngân hàng A và B trong 1 năm tương ứng là X (đơn vị %), Y (đơn vị %) cho kết quả trong bảng:

$X \quad \vert \quad Y$	-2	5	10
-1	0,1	0,15	0,1
4	0,05	0,2	0,1
8	0,1	0,15	0,05

a) Tính $Cov(X,Y)$

b) Tìm tỉ lệ đầu tư vào A và B để thu nhập hàng năm đều đặn nhất.

3.3 Bài tập có hướng dẫn

Bài 3.1. Cho véc tơ ngẫu nhiên $X = (X_1, X_2)$ có hàm mật độ đồng thời f_X được cho bởi

$$f_X(x_1, x_2) = \begin{cases} \frac{x_1+x_2}{21}, & x_1 \in \{1,2,3\},\ x_2 \in \{1,2\} \\ 0, & \text{nơi khác} \end{cases}$$

Tính $P(X_1 = 3), P(X_2 = 2)$ và hàm mật độ biên f_1, f_2 tương ứng của X_1, X_2.
Giải.

$$P(X_1 = 3) = f_X(3,1) + f_X(3,2) = \frac{3}{7}.$$

$$P(X_2 = 2) = f_X(2,1) + f_X(2,2) + f_X(3,2) = \frac{4}{7}.$$

Hàm mật độ biên f_1 của X_1 xác định bởi:

$$f_1(x_1) = \begin{cases} \displaystyle\sum_{x_2=1}^{2} \frac{x_1+x_2}{21} = \frac{2x_1+3}{21}, & x_1 \in \{1,2,3\}, \\ 0, & \text{nơi khác} \end{cases}$$

Tương tự

$$f_2(x_2) = \begin{cases} \displaystyle\sum_{x_1=1}^{3} \frac{x_1+x_2}{21} = \frac{3x_2+6}{21}, & x_2 \in \{1,2,\}, \\ 0, & \text{nơi khác} \end{cases}$$

Bài 3.2. Cho véc tơ ngẫu nhiên $X = (X_1, X_2)$ có hàm phân phối đồng thời F_X được cho bởi

$$F_X(x_1, x_2) = \begin{cases} 1 - e^{-x_1} - e^{-x_2} + e^{-(x_1+x_2)} & \text{với } (x_1, x_2) \in R_+^2 \\ 0, & \text{nơi khác} \end{cases}$$

(a) Tìm hàm phân phối biên của X_1.

(b) Tìm hàm mật độ đồng thời của X_1, X_2.

(c) Tìm hàm mật độ biên của X_1.

Giải.

(a) Hàm phân phối biên F_1 của X_1 được xác định với mọi $x \in R$ bởi:

$$F_1(x_1) = \lim_{x_2 \to +\infty} F_X(x_1, x_2) = \begin{cases} 1 - e^{-x_1} & \text{với } x_1 \geq 0 \\ 0, & \text{nơi khác} \end{cases}$$

(b) Hàm mật độ đồng thời của X_1, X_2 được xác định bởi:

$$f_X(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = \begin{cases} e^{-(x_1+x_2)} & \text{với } (x_1, x_2) \in R_+^2, \\ 0, & \text{nơi khác} \end{cases}$$

(c) Hàm mật độ biên f_1 của X_1 được xác định với mọi $x \in R$ bởi:

$$f_1(x_1) = \int_{-\infty}^{+\infty} f_X(x_1, x_2) dx_2 = \begin{cases} \int_0^{+\infty} e^{-(x_1+x_2)} dx_2 = e^{-x_1} & \text{với } x_1 \geq 0, \\ 0, & \text{với } x < 0. \end{cases}$$

Bài 3.3. Cho véc tơ ngẫu nhiên rời rạc $X = (X_1, X_2)$ có phân phối xác suất được cho trong bảng sau:

		X_2		
		1	2	3
X_1	0	0,15	0,13	0,16
	3	0,07	0,09	0,12
	5	0,05	0,10	0,13

a) Tìm luật phân phối biên của biến ngẫu nhiên X_1 và X_2?

b) Tính các xác suất $P(X_1 = 0; X_2 < 3), P(X_1 < 4; X_2 \leq 3)$?

Giải. a) Ta tính các xác suất sau:

$$\begin{aligned} f_1(0) = P(X_1 = 0) &= \sum_{x_2} P(X_1 = 0; X_2 = x_2) \\ &= P(X_1 = 0; X_2 = 1) + P(X_1 = 0; X_2 = 2) + P(X_1 = 0; X_2 = 3) \\ &= 0,15 + 0,13 + 0,16 = 0,44. \end{aligned}$$

Tương tự cũng dễ dàng tính được

$$f_1(3) = P(X_1 = 3) = 0,07 + 0,09 + 0,12 = 0,28$$

và

$$f_1(5) = P(X_1 = 5) = 0,05 + 0,10 + 0,13 = 0,28.$$

Luật phân phối biên của biến ngẫu nhiên X_1 được cho trong bảng sau:

$X_1 = x_1$	0	3	5
$f_1(x_1)$	0,44	0,28	0,28

Để tìm phân phối xác suất biên của X_2 ta tính các xác suất sau:

$$f_2(1) = P(X_2 = 1) = 0,15 + 0,07 + 0,05 = 0,27$$

$$f_2(2) = P(X_2 = 2) = 0,13 + 0,09 + 0,10 = 0,32$$

$$f_2(3) = P(X_2 = 3) = 0,16 + 0,12 + 0,13 = 0,41.$$

Luật phân phối biên của X_2 được cho trong bảng:

$X_2 = x_2$	1	2	3
$f_2(x_2)$	0,27	0,32	0,41

b) Tính các xác suất $P(X_1 = 0; X_2 < 3)$

$$\begin{aligned} P(X_1 = 0; X_2 < 3) &= \sum_{x_2<3} P(X_1 = 0; X_2 = x_2) \\ &= P(X_1 = 0; X_2 = 1) + P(X_1 = 0; X_2 = 2) \\ &= 0,15 + 0,13 = 0,28. \end{aligned}$$

Tính xác suất $P(X_1 < 4; X_2 \le 3)$

$$\begin{aligned} P(X_1 < 4; X_2 \le 3) &= \sum_{x_1<4} \sum_{x_2\le 3} P(X_1 = x_1; X_2 = x_2) \\ &= P(X_1 = 0; X_2 = 1) + P(X_1 = 0; X_2 = 2) + P(X_1 = 0; X_2 = 3) \\ &\quad + P(X_1 = 3; X_2 = 1) + P(X_1 = 3; X_2 = 2) + P(X_1 = 3; X_2 = 3) \\ &= 0,15 + 0,13 + 0,16 + 0,07 + 0,09 + 0,12 = 0,72. \end{aligned}$$

Bài 3.4. Cho bảng phân phối xác suất của véc tơ ngẫu nhiên $X = (X_1, X_2)$ sau

		X_2		
		1	2	3
X_1	1	0,15	0,20	0,15
	3	0,10	0,15	0,25

a) Xác định phân phối xác suất đồng thời của hai biến ngẫu nhiên X_1 và X_2.

b) Kiểm tra tính độc lập của X_1 và X_2.

c) Tính xác suất $P(X_1 = 1|X_2 = 3)$.

Giải. a) Hàm phân phối đồng thời của véc tơ X được cho bởi

$$F_X(x_1, x_2) = P(X_1 < x_1, X_2 < x_2) = \sum_{u_1 < x_1} \sum_{u_2 < x_2} P(X_1 = u_1, X_2 = u_2).$$

Kết quả tính được cho trong bảng sau:

		X_2			
		$x_2 < 1$	$1 \leq x_2 < 2$	$2 \leq x_2 < 3$	$x_2 \geq 3$
X_1	$x_1 < 1$	0	0	0	0
	$1 \leq x_1 < 3$	0	0,15	0,35	0,5
	$x_1 \geq 3$	0	0,25	0,50	1

b) Từ bảng phân phối xác suất của X ta có $P(X_1 = 1, X_2 = 2) = 0,20$. Mặt khác ta cũng tính được

$$P(X_1 = 1) = 0,15 + 0,20 + 0,15 = 0,5$$

và

$$P(X_2 = 2) = 0,20 + 0,15 = 0,35.$$

Từ đó ta có

$$P(X_1 = 1)P(X_2 = 2) = 0,5 \times 0,35 \neq P(X_1 = 1, X_2 = 2).$$

Vậy hai biến ngẫu nhiên X_1 và X_2 phụ thuộc.
c) Tính xác suất $P(X_1 = 1|X_2 = 3)$

$$P(X_1 = 1|X_2 = 3) = \frac{P(X_1 = 1, X_2 = 3)}{P(X_2 = 3)} = \frac{0,15}{0,4} = 0,375.$$

3.4 Bài tập đề nghị

Chương 4

Các định lý giới hạn của lý thuyết xác suất

4.1 Hàm đặc trưng

1. Định nghĩa
   Giả sử F là hàm phân phối của đại lượng ngẫu nhiên X, ta gọi hàm biến thực $\varphi(t), t \in R$ được xác định bằng hệ thức

$$\varphi(t) = \mathbb{E}(e^{itX}) = \int\limits_{-\infty}^{+\infty} e^{itx} dF(x) = \int\limits_{-\infty}^{+\infty} \cos tx dF(x) + i \int\limits_{-\infty}^{+\infty} \sin tx dF(x)$$

   là hàm đặc trưng của phân phối F.

2. Một vài tính chất quan trọng

   a) $|\varphi(t)| \leq \varphi(0) = 1, \quad |\varphi(t+h) - \varphi(t)| \leq 2\sqrt{1 - \mathrm{Re}\varphi(h)} \quad \forall t \in \mathbb{R}$, trong đó Rez là phần thực của z.

   b) $\varphi(t)$ liên tục đều trên $\mathbb{R}$.

   c) $\varphi(-t) = \overline{\varphi(t)}$.

   d) $\varphi(t)$ là hàm thực khi và chỉ khi X có phân phối đối xứng, nghĩa là X và $-X$ cùng phân phối.

   e) Nếu X và Y độc lập với nhau thì $\forall t \in \mathbb{R}$ ta có $\varphi_{X+Y}(t) = \varphi_X(t).\varphi_Y(t)$.
   Do đó nếu $X_1, X_2, \ldots X_n$ độc lập thì $\forall t \in \mathbb{R}$ ta có $\varphi_{\sum\limits_{i=1}^{n} X_i}(t) = \prod\limits_{i=1}^{n} \varphi_{X_i}(t)$.

f) Với mọi số thực a và b ta có:

$$\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at).$$

g) Nếu $\mathbb{E}|X|^n < \infty$ với $n \geq 1$ nào đó thì $\varphi(t)$. có đạo hàm đến cấp n tại mọi điểm và

$$\varphi^{(k)}(t) = \int\limits_{-\infty}^{+\infty} (ix)^k e^{itx} dF(x) = i^k \mathbb{E}\left(X^k e^{itX}\right).$$

4.2 Các dạng hội tụ trong xác suất

1. Hội tụ hầu chắc chắn
   Một dãy các biến ngẫu nhiên $(X_n)_{n\geq 1}$ hội tụ hầu chắc chắn đến biến ngẫu nhiên X nếu $N = \left\{w : \lim\limits_{n\to\infty} X_n(w) \neq X(w)\right\}$ có $P(N) = 0$. Ta kí hiệu $X_n \xrightarrow{a.s} X$.

2. Hội tụ trong L^p
   Một dãy biến ngẫu nhiên $(X_n)_{n\geq 1}$ hội tụ trong L^p (với $1 \leq p < \infty$) đến biến ngẫu nhiên X nếu $|X_n|, |X|$ đều thuộc L^p và $\lim\limits_{n\to\infty} \mathbb{E}\{|X_n - X|^p\} = 0$. Ta kí hiệu $X_n \xrightarrow{L^p} X$.

3. Hội tụ theo xác suất
   Một dãy biến ngẫu nhiên $(X_n)_{n\geq 1}$ hội tụ theo xác suất đến biến ngẫu nhiên X nếu với mọi $\varepsilon > 0$ ta có $\lim\limits_{n\to\infty} P\{w : |X_n(w) - X(w)| > \varepsilon\} = 0$. Ta kí hiệu $X_n \xrightarrow{P} X$.

4. Hội tụ theo phân phối
   Giả sử F_n, F lần lượt là hàm phân phối của biến ngẫu nhiên X_n, X. Khi đó người ta nói dãy biến ngẫu nhiên $(X_n)_{n\geq 1}$ hội tụ theo phân phối đến biến ngẫu nhiên X nếu $F_n(x) \to F(x)$ tại các điểm liên tục của hàm F. Ta kí hiệu $X_n \xrightarrow{d} X$.

4.3 Luật số lớn

Định lí 4.1. *(Luật số lớn Bernoulli)*

Với $\varepsilon > 0$ bất kì

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\} = 0$$

Định lí 4.2. *(Luật số lớn Chebyshev)*

Giả sử $\{X_n, n \geq 1\}$ *là một dãy các biến ngẫu nhiên độc lập, có phương sai hữu hạn bởi cùng một hằng số, tức là* $Var\,(X_n) = \sigma_n^2 \leq C,\ \forall n \geq 1.$. *Khi đó,* $\forall > 0,$

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \frac{1}{n}\sum_{i=1}^{n} E\,(X_i)\right| \geq \varepsilon\right\} = 0$$

Định lí 4.3. *(Luật số lớn Khintchine)*

Giả sử $\{X_n, n \geq 1\}$ *là một dãy các biến ngẫu nhiên độc lập, cùng phân phối với kì vọng* $E(X_n) = \mu\,\forall n \geq 1$. *Khi đó,*

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} = 0$$

khi $n \to \infty$.

Định lí 4.4. *(Luật số lớn Liapunov)*

Giả sử $\{X_n, n \geq 1\}$ *là một dãy các biến ngẫu nhiên độc lập với kì vọng hữu hạn* $E(X_k) = \mu_k$, *thỏa mãn điều kiện sau với* $0 < \delta \leq 1$,

$$\frac{1}{n^{1+\delta}}\sum_{k=1}^{n} E|X_k - \mu_k|^{l+\delta} \to 0, \text{ khi } n \to \infty.$$

Khi đó $\forall\varepsilon > 0$, *ta có*

$$\lim_{n\to\infty} P\left\{\left|S_n - \frac{E\,(\mu_1) + ... + E\,(\mu_n)}{n}\right| \geq \varepsilon\right\} = 0$$

Định lí 4.5. *(Luật mạnh số lớn Berol)*

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ *là một dãy các biến ngẫu nhiên độc lập, cùng phân phối với kì vọng* $X_1 = \mu$. *Khi đó,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \text{ khi } n \to \infty$$

hầu chắc chắn.

4.4 Định lý giới hạn trung tâm

Định lí 4.6. *(Định lí giới hạn địa phương Moivres-Laplace)*

Xét n phép thử Bernoulli với xác suất xuất hiện biến cố A bằng $P(A) = p, (0 < p < 1)$. Khi đó xác suất $P_n(k)$ để biến cố A xảy ra đúng k lần trong n phép thử thỏa mản

$$\sqrt{npq}P_n(k) : \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \to 1$$

khi $n \to \infty$ đều với tất cả k, x thuộc khoảng hữu hạn bất kì.

Định lí 4.7. *(Định lý giới hạn tích phân Moivres-Laplace)*

Giả sử S_n là số thành công trong n phép thử độc lập Bernoulli với xác suất thành công $p(0 < p < 1)$ trong mỗi phép thử. Khi đó

$$\lim_{n\to\infty} P\left\{\frac{S_n - np}{\sqrt{npq}} < x\right\} = \Phi(x),$$

với $x \in (-\infty; +\infty)$, ở đây

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

là hàm phân phối chuẩn.

Định lí 4.8. *(Sự suy rộng của định lý Moivres-Laplace)*

Giả sử $(X_n)_{n\geq 1}$ là một dãy các biến ngẫu nhiên độc lập, cùng phân phối với biến ngẫu nhiên X có kỳ vọng μ và phương sai $\sigma^2 < \infty$. Đặt $S_n = \sum\limits_{i=1}^{n} X_i$ và $\hat{S}_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, khi đó

$$\lim_{n\to\infty} P\left(\hat{S}_n < x\right) = \Phi(x) \quad \forall x \in \mathbb{R},$$

trong đó $\Phi(x) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$ là hàm phân phối chuẩn.

Định lí 4.9. *(Định lí Lindeberg)*

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là một dãy các đại lượng ngẫu nhiên độc lập, không nhất thiết cùng phân phối, với

$$F_k(x) = P\{X_k < x\},\ E(X_k) = 0 \text{ và } Var(X_k) = \sigma_k^2 < +\infty,\ k \geq 1.$$

Kí hiệu: $B_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$. *Khi đó* $\forall \varepsilon > 0$, *nếu*

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq \varepsilon B_n} x^2 dF_k(x) = 0$$

thì

$$\lim_{n\to\infty} \underset{-\infty<x<+\infty}{Sup} \left| P\left\{ \frac{S_n}{B_n} < x \right\} - \Phi(x) \right| = 0.$$

Định lí 4.10. *(Định lí Feller W.)*

Nếu

$$F_{\frac{S_n}{B_n}}(x) \xrightarrow{d} N(0,1)$$

và

$$\max_{i\leq k\leq n} \frac{\sigma_k}{B_n} \to 0 \text{ khi } n \to \infty$$

thì điều kiện Lindeberf thỏa mãn cho một dãy $X_1, X_2, \ldots, X_n$ *các biến ngẫu nhiên độc lập.*

Định lí 4.11. *(Định lí Lindeberg-Feller)*

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ *là một dãy các biến ngẫu nhiên độc lập với* $E(X_k) = 0$ *và* $Var(X_k) = \sigma_k^2 < +\infty$. *Khi đó,*

$$\lim_{n\to\infty} P\left\{ \frac{S_n}{B_n} < x \right\} = \Phi(x)$$

và

$$\frac{1}{B_n} \max_{1\leq k\leq n} \sigma_k \to 0 \text{ khi n} \to \infty$$

khi và chỉ khi $\forall \varepsilon > 0$ *và* $n \longrightarrow \infty$,

$$g_n(\varepsilon) = \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq \varepsilon B_n} x^2 dF_k(X) \to 0.$$

Định lí 4.12. *(Định lí Liapunov)*

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ *là một dãy các biến ngẫu nhiên độc lập. Nếu*

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|X_k|^{2+\delta} \to 0 \text{ khi } n \to \infty \text{ và } \delta > 0,$$

thì

$$\lim_{n\to\infty} \underset{x}{Sup} \left| P\left\{ \frac{S_n}{B_n} < x \right\} - \Phi(x) \right| = 0.$$

4.5 Định lý xấp xỉ Poisson

Định lí 4.13. *Cho $X_n \sim B(n; p_n)$. Nếu số phép thử $n \longrightarrow \infty$ và xác suất thành công $p_n \longrightarrow 0$ sao cho $np_n \longrightarrow \lambda$ thì $X_n \xrightarrow{d} Poisson(\lambda)$.*

Định lí 4.14. *Giả sử cho mảng tam giác $(X_{nj}, j = 1, 2, ..., n; n = 1, 2, ...)$ sao cho với mỗi n, dãy $X_{n1}, X_{n2}, ..., X_{nn}$ là các biến ngẫu nhiên Bernoulli độc lập, với $P(X_{nj} = 1) = 1 - P(X_{nj} = 0) = p_{nj}$ $(p_{nj} \in [0, 1]; j = 1, 2, \ldots, n; n = 1, 2, \ldots)$. Cho $Z_{\lambda_n} \sim Poisson(\lambda_n)$ với $\lambda_n = \sum\limits_{j=1}^{n} p_{nj}$. Đặt $S_n = \sum\limits_{j=1}^{n} X_{nj}$, khi đó ta có*

$$|P(S_n = k) - P(Z_{\lambda_n} = k)| \leq \sum_{j=1}^{n} p_{nj}^2 \qquad k = 0, 1, 2, \ldots$$

4.6 Định lý Renyi

Định lí 4.15. *(Renyi) Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên nhận giá trị không âm, độc lập, cùng phân phối và kỳ vọng $\mathbb{E}(X_1) = m$ với $0 < m < +\infty$. Đặt N là biến ngẫu nhiên có phân phối hình học tham số q, $q \in (0, 1)$. Khi đó*

$$\lim_{q \to 0^+} F_{qS_N}(x) = 1 - e^{-\frac{x}{m}}. \tag{4.1}$$

Trong đó $S_N = X_1 + X_2 + \cdots + X_N$ và $F_{qS_N}(x) = \mathbb{P}(qS_N \leq x)$.

Định lí 4.16. *Giả sử $S_N = X_1 + X_2 + \ldots + X_N$ là biến phức hợp hình học tham số q với các biến thành phần không âm và $E(X_k) = 1$. Khi đó, biến ngẫu nhiên pS_N có phân phối mũ với tham số $\lambda = 1$ khi và chỉ khi X_k có phân phối mũ với tham số $\lambda = 1$, tức là*

$$F_{pS_N}(x) = 1 - e^{-x};\ (x \geq 0) \Leftrightarrow F_{X_k}(x) = 1 - e^{-x}.$$

4.7 Bài tập có hướng dẫn

Bài 4.1. Chứng minh rằng, nếu $X \sim B(n, p)$ thì hàm đặc trưng có dạng

$$\varphi_X(t) = \left[p\left(e^{it} - 1\right) + 1\right]^n$$

Chứng minh

Ta có

$$\begin{aligned}\varphi_X(t) &= E\left(e^{itX}\right)\\ &= \sum_{k=0}^{n} e^{itk}.P(X=k)\\ &= \sum_{k=0}^{n} e^{itk}.C_n^k p^k q^{n-k}\\ &= \sum_{k=0}^{n} C_n^k \left(pe^{it}\right)^k q^{n-k}\\ &= \left[p\left(e^{it}-1\right)+1\right]^n.\end{aligned}$$

Bài 4.2. Chứng minh rằng, nếu $X \sim P(\lambda)$ thì hàm đặc trưng có dạng

$$\varphi_X(t) = e^{\lambda\left(e^{it}-1\right)}$$

Chứng minh

Ta có

$$\begin{aligned}\varphi_X(t) &= E\left(e^{itX}\right)\\ &= \sum_{k=0}^{n} e^{itk}.P(X=k)\\ &= \sum_{k=0}^{n} e^{itk}.e^{-\lambda}\frac{\lambda^k}{k!}\\ &= e^{-\lambda}\sum_{k=0}^{n} \frac{\left(\lambda e^{it}\right)^k}{k!}\\ &= e^{-\lambda}.e^{\lambda e^{it}}\\ &= e^{\lambda\left(e^{it}-1\right)}.\end{aligned}$$

Bài 4.3. Chứng minh rằng, nếu $X \sim N(0,1)$ thì hàm đặc trưng có dạng

$$\varphi_X(t) = e^{-\frac{t^2}{2}}.$$

Chứng minh

Ta có

$$\begin{aligned}\varphi_X(t) &= E\left(e^{itX}\right)\\ &= \int\limits_{-\infty}^{+\infty} e^{itx} f(x)\,dx\\ &= \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} e^{itx-\frac{x^2}{2}}\,dx\end{aligned}$$

Lấy đạo hàm theo t ta được:

$$\begin{aligned}\varphi'_X(t) &= \frac{i}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} xe^{itx-\frac{x^2}{2}}\,dx\\ &= \frac{i}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} (x-it+it)e^{itx-\frac{x^2}{2}}\,dx\\ &= \frac{i}{\sqrt{2\pi}}\left[-\int\limits_{-\infty}^{+\infty}(it-x)e^{itx-\frac{x^2}{2}}\,dx + it\int\limits_{-\infty}^{+\infty} e^{itx-\frac{x^2}{2}}\,dx\right]\\ &= \frac{-i}{\sqrt{2\pi}}e^{itx}.\frac{1}{e^{x^2/2}}\Bigg|_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} e^{itx-\frac{x^2}{2}}\,dx\\ &= 0 - t\varphi_X(t).\quad \left(\lim_{x\to\pm\infty}\frac{1}{e^{x^2/2}}=0\right)\end{aligned}$$

Như vậy ta có:

$$\varphi'_X(t) = -t\varphi_X(t)$$

$$\Leftrightarrow \varphi'_X(t) + t\varphi_X(t) = 0$$

$$\Leftrightarrow \varphi_X(t) = c.e^{-\int t dt} = ce^{-\frac{t^2}{2}}.$$

Vì

$$\varphi_X(0) = E\left(e^0\right) = 1$$

$$\Rightarrow 1 = c.e^0$$

$$\Rightarrow c = 1$$

Vậy

$$\varphi_X(t) = e^{-\frac{t^2}{2}}.$$

Bài 4.4. Chứng minh rằng, nếu $X \sim N\left(\mu, \sigma^2\right)$ thì hàm đặc trưng có dạng

$$\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}.$$

Chứng minh

Ta có: $X = \sigma.\dfrac{X-\mu}{\sigma} + \mu = \sigma Y + \mu$ với $Y \sim N(0,1)$.

Vậy

$$\varphi_X(t) = E\left(e^{itX}\right) = E\left(e^{it\sigma Y + it\mu}\right) = e^{it\mu}.E\left(e^{it\sigma Y}\right) = e^{it\mu}.e^{-\frac{(t\sigma)^2}{2}} = e^{it\mu - \frac{(t\sigma)^2}{2}}.$$

Bài 4.5. Chứng minh rằng, nếu $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$, $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$ với X_1 và X_2 là các đại lượng ngẫu nhiên độc lập, thì $X = X_1 + X_2 \sim N\left(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2\right)$.

Giải

Ta có

$$E(X_1) = \mu_1, \quad D(X_1) = \sigma_1^2;$$
$$E(X_2) = \mu_2, \quad D(X_2) = \sigma_2^2.$$

Khi đó hàm đặc trưng của X_1 và X_2 có dạng:

$$\varphi_{X_1}(t) = \exp\left\{i\mu_1 t - \tfrac{1}{2}\sigma_1^2 t^2\right\};$$
$$\varphi_{X_2}(t) = \exp\left\{i\mu_2 t - \tfrac{1}{2}\sigma_2^2 t^2\right\}.$$

Theo tính chất của hàm đặc trưng thì

$$\varphi_X(t) = \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t).\varphi_{X_2}(t) = \exp\left\{it(\mu_1+\mu_2) - \frac{1}{2}\left(\sigma_1^2+\sigma_2^2\right)t^2\right\}.$$

Đây là hàm đặc trưng của phân phối chuẩn với kỳ vọng $\mu = \mu_1 + \mu_2$ và phương sai $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Theo định lý duy nhất ta suy ra điều phải chứng minh.

Bài 4.6. Giả sử các đại lượng ngẫu nhiên X_1 và X_2 là độc lập và cùng có phân phối Poisson, với

$$P(X_1 = k) = \frac{\lambda_1^k e^{-\lambda}}{k!} \quad \text{và} \quad P(X_2 = k) = \frac{\lambda_2^k e^{-\lambda}}{k!}$$

Chứng minh rằng đại lượng ngẫu nhiên $X = X_1 + X_2$ sẽ có phân phối theo luật Poisson với tham số $\lambda = \lambda_1 + \lambda_2$.

Giải

Ta có

$$\varphi_{X_1}(t) = \exp\left\{\lambda_1\left(e^{it}-1\right)\right\};$$
$$\varphi_{X_2}(t) = \exp\left\{\lambda_2\left(e^{it}-1\right)\right\}.$$

Theo tính chất hàm đặc trưng, ta nhận được

$$\varphi_X(t) = \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t).\varphi_{X_2}(t) = \exp\left\{(\lambda_1+\lambda_2)\left(e^{it}-1\right)\right\}.$$

Đây là hàm đặc trưng của phân phối Poisson với kỳ vọng và phương sai là $\lambda_1+\lambda_2$. Theo định lý duy nhất ta suy ra điều phải chứng minh.

Bài 4.7. Chứng minh định lý sau đây

Giả sử $S_N = X_1+X_2+\ldots+X_N$ là biến phức hợp hình học tham số q với các biến thành phần không âm và $E(X_k)=1$. Khi đó, biến ngẫu nhiên pS_N có phân phối mũ với tham số $\lambda=1$ khi và chỉ khi X_k có phân phối mũ với tham số $\lambda=1$, tức là

$$F_{pS_N}(x) = 1-e^{-x};\ (x\geq 0) \Leftrightarrow F_{X_k}(x) = 1-e^{-x}.$$

Hướng dẫn giải

Với $\varphi_0(t)$ là hàm đặc trưng của phân phối mũ ($\lambda=1$) thì theo đề bài ta có:

$$\begin{aligned}
&\varphi_{pS_N}(t) = \varphi_0(t)\\
&\Leftrightarrow \tfrac{p\varphi(pt)}{1-(1-p)\varphi(pt)} = \tfrac{1}{1-it}\\
&\Leftrightarrow p\varphi(pt) - ipt\varphi(pt) = 1-\varphi(pt)+p\varphi(pt)\\
&\Leftrightarrow \varphi(pt) = \tfrac{1}{1-ipt}
\end{aligned}$$

Đặt $u = pt$ thì

$$\Leftrightarrow \varphi(u) = \frac{1}{1-iu}$$

hay

$$\Leftrightarrow \varphi(t) = \frac{1}{1-it}$$

Vậy $F_{pS_N}(x) = 1-e^{-x};\ (x\geq 0) \ \Leftrightarrow\ F_{X_k}(x) = 1-e^{-x}$. ■

Bài 4.8. Chứng minh định lý sau đây

Cho $X_1, X_2, \ldots$ là một dãy các biến ngẫu nhiên độc lập cùng phân phối, $X_i \sim N(0,1)$, $N_p \sim Geometric(q)\, q \in (0,1)$ và $Y \sim Exp(1)$. Khi đó, nếu $p \longrightarrow 0$ thì

$$p\left(X_1^2+X_2^2+\ldots+X_{N_p}^2\right) \xrightarrow{d} Y$$

Hướng dẫn giải

Gọi $\varphi(t), \varphi_0(t)$ lần lượt là hàm đặc trưng của X_j^2, Y.

Ta có $Y \sim Exp(1)$ nên $\varphi_0(t) = \frac{1}{1-it}$.

Ta có $X_j \sim N(0,1) \Rightarrow X_j^2 \sim \chi_1^2$, như vậy ta xác định được hàm đặc trưng của X_j^2 là

$$\varphi(t) = \frac{1}{\sqrt{1-2it}}$$

Do $N_p \sim Geometric(p)\ \ p \in (0,1)$, nên N_p có hàm sinh $g(t) = \frac{pt}{1-(1-p)t}$. Khi đó hàm đặc trưng của S_{N_p} là

$$\varphi_{S_{N_p}}(t) = g \circ \varphi(t) = \frac{p\varphi(t)}{1-(1-p)\varphi(t)}$$

Ta suy ra hàm đặc trưng của pS_{N_p} là

$$\varphi_{pS_{N_p}}(t) = \varphi_{S_{N_p}}(pt) = \frac{p\varphi(pt)}{1-(1-p)\varphi(pt)} = \frac{\sqrt{1-2ipt}+1-p}{2-2it-p}$$

Khi cho $p \to 0$, ta được:

$$\varphi_{pS_{N_p}}(t) \to \frac{1}{1-it} = \varphi_0(t) \quad \text{(đpcm)}.$$

Xét tổng ngẫu nhiên $S_N = X_1 + X_2 + \ldots + X_N$ là biến phức hợp hình học với $\mu = E(X_k) < +\infty$. Gọi $F_{\sqrt{p}S_N}(x)$ và $\varphi_{\sqrt{p}Z}(t)$ tương ứng là hàm phân phối và hàm đặc trưng của $\sqrt{p}S_N$; $F_{X_k}(x)$ và $\varphi(t)$ tương ứng là hàm phân phối và hàm đặc trưng của X_k; $\widehat{F_0}(x)$ và $\widehat{\varphi_0}(t)$ tương ứng là hàm phân phối và hàm đặc trưng của phân phối mũ hai phía. ■

Bài 4.9. Chứng minh định lý sau đây

Điều kiện cần và đủ để $\sqrt{p}S_N$ có phân phối mũ hai phía là các thành phần $X_k, k = 1, 2, \ldots$ có phân phối mũ hai phía, nghĩa là

$$F_{\sqrt{p}S_N}(x) = \widehat{F_0}(x) \Leftrightarrow F_{X_k}(x) = \widehat{F_0}(x).$$

Hướng dẫn giải

Ta có: $\hat{\varphi}_0(t) = \frac{1}{1+t^2}; \quad \forall t \in \mathbb{R}^1$.

Theo đề bài ta có:

$$\begin{aligned}
&\varphi_{\sqrt{p}S_N}(t) = \hat{\varphi}_0(t)\\
&\Leftrightarrow \tfrac{p\varphi(\sqrt{p}t)}{1-(1-p)\varphi(\sqrt{p}t)} = \tfrac{1}{1+t^2}\\
&\Leftrightarrow p\varphi(\sqrt{p}t) + pt^2\varphi(\sqrt{p}t) = 1 - \varphi(\sqrt{p}t) + p\varphi(\sqrt{p}t)\\
&\Leftrightarrow \varphi(\sqrt{p}t) = \tfrac{1}{1+pt^2}
\end{aligned}$$

Đặt $u = \sqrt{pt}$ thì

$$\Leftrightarrow \varphi(u) = \frac{1}{1+u^2} \qquad \forall u \in \mathbb{R}^1.$$

hay

$$\Leftrightarrow \varphi(t) = \frac{1}{1+t^2} \qquad \forall t \in {}^1.$$

Vậy định lí được được chứng minh. ■

Bài 4.10. Chứng minh định lý sau đây

Biến phức hợp hình học $S_N = X_1 + X_2 + \ldots + X_N$ tham số q có phân phối hình học tham số q khi và chỉ khi các biến thành phần suy biến với $E(X_k) = 1$.

Hướng dẫn giải

Vì N có phân phối hình học với tham số p nên hàm đặc trưng và hàm sinh của N lần lượt là:

$$\varphi_N(t) = \frac{pe^{it}}{1-(1-p)e^{it}};\ \ g\,(t) = \frac{pt}{1-(1-p)t}, \qquad t \in \mathbb{R}^1$$

Vậy ta có S_N có phân phối hình học tham số p khi và chỉ khi hàm đặc trưng $\varphi_{S_N}(t)$ có dạng như sau:

$$\varphi_{S_N}(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$$

Mặc khác ta có

$$\varphi_{S_N}(t) = g \circ \varphi\,(t) = \frac{p\varphi(t)}{1-(1-p)\varphi(t)}$$

với $\varphi(t)$ là hàm đặc trưng của biến ngẫu nhiên X_k.

Nên

$$\frac{p\varphi(t)}{1-(1-p)\varphi(t)} = \frac{pe^{it}}{1-(1-p)e^{it}} \Leftrightarrow \varphi(t) = e^{it} \quad t \in {}^1$$

tức là các thành phần X_k suy biến với $E(X_k) = 1$. ■

Bài 4.11. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là dãy các biến ngẫu nhiên độc lập và cùng tuân theo phân phối mũ với tham số λ. Gọi $N \sim NB(r,p)$ và độc lập với các biến ngẫu nhiên X_n, $n \geq 1$. Khi đó tổng ngẫu nhiên

$$S_N = \sum_{i=1}^{N} X_i \sim G(r, \lambda p)$$

.

<u>*Hướng dẫn giải*</u>

\+ Ta có hàm sinh của biến ngẫu nhiên N là

$$\psi_N(t) = \left[\frac{pt}{1-(1-p)\,t}\right]^r$$

\+ Ta có hàm đặc trưng của X là

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

\+ Khi đó ta tính được hàm đặc trưng của S_N là

$$\varphi_{S_N}(t) = \psi \circ \varphi_X(t) = \left[\frac{p.\frac{\lambda}{\lambda-it}}{1-(1-p)\,.\frac{\lambda}{\lambda-it}}\right]^r = \left(\frac{\lambda p}{\lambda p - it}\right)^r.$$

Vậy $S_N \sim Gamma(r, \lambda p)$. ■

Bài 4.12. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là dãy các biến ngẫu nhiên độc lập và cùng phân phối với biến ngẫu nhiên X, với $X \sim L(0,\lambda)$. Gọi $N \sim Geometric(p)$ và độc lập với các biến ngẫu nhiên X_n, $n \geq 1$. Khi đó tổng ngẫu nhiên

$$\sqrt{p}S_N = \sqrt{p}\sum_{i=1}^{N} X_i \sim L(0,\lambda)$$

.

<u>*Hướng dẫn giải*</u>

\+ Ta có hàm sinh của biến ngẫu nhiên N là

$$\psi_N(t) = \frac{pt}{1-(1-p)\,t}$$

+ Ta có hàm đặc trưng của X là

$$\varphi_X(t) = \frac{\lambda^2}{t^2 + \lambda^2}$$

+ Khi đó ta tính được hàm đặc trưng của S_N là

$$\varphi_{S_N}(t) = \psi \circ \varphi(t) = \frac{p\varphi(t)}{1 - (1-p)\varphi(t)} = \frac{p\lambda^2}{t^2 + p\lambda^2}.$$

+ Vậy hàm đặc trừng của $\sqrt{p}S_N$ là

$$\varphi_{\sqrt{p}S_N}(t) = \varphi_{S_N}(\sqrt{p}t) = \frac{\lambda^2}{t^2 + \lambda^2}.$$

Hay $\sqrt{p}S_N \sim L(0, \lambda)$. ■

Bài 4.13. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots, X_n, \ldots$ là dãy các biến ngẫu nhiên độc lập và cùng phân phối với biến ngẫu nhiên X, với $X \sim L(0, \sqrt{p}a, \sigma)$. Gọi $N \sim \mathit{Geometric}(p)$ và độc lập với các biến ngẫu nhiên X_n, $n \geq 1$. Khi đó tổng ngẫu nhiên

$$\sqrt{p}S_N = \sqrt{p}\sum_{i=1}^{N} X_i \sim L(0, a, \sigma)$$

.

Hướng dẫn giải

+ Ta có hàm sinh của biến ngẫu nhiên N là

$$\psi_N(t) = \frac{pt}{1 - (1-p)t}$$

+ Ta có hàm đặc trưng của X là

$$\varphi_X(t) = \frac{1}{1 - i\sqrt{p}at + \frac{\sigma^2 t^2}{2}}$$

+ Khi đó ta tính được hàm đặc trưng của S_N là

$$\begin{aligned}
\varphi_{S_N}(t) &= \psi_N \circ \varphi_X(t) \\
&= \frac{p\varphi_X(t)}{1-(1-p)\varphi_X(t)} \\
&= \frac{\dfrac{p}{1-i\sqrt{p}at+\dfrac{\sigma^2t^2}{2}}}{1-(1-p)\dfrac{1}{1-i\sqrt{p}at+\dfrac{\sigma^2t^2}{2}}} \\
&= \frac{p}{p-i\sqrt{p}at+\dfrac{\sigma^2t^2}{2}}
\end{aligned}$$

+ Vậy hàm đặc trừng của $\sqrt{p}S_N$ là

$$\varphi_{\sqrt{p}S_N}(t) = \varphi_{S_N}(\sqrt{p}t) = \frac{1}{1-iat+\frac{\sigma^2t^2}{2}}.$$

Hay $\sqrt{p}S_N \sim L(0, a, \sigma)$. ■

Bài 4.14. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots X_n, \ldots$ là dãy các biến ngẫu nhiên nhận giá trị không âm, độc lập, cùng phân phối với biến ngẫu nhiên X và kỳ vọng hữu hạn $\mathbb{E}(X) = m < +\infty$. Đặt $N \sim NB(r, p)$ và độc lập với tất cả các $X_i,\ i \geq 1$. Khi đó với $p \to 0^+$ thì

$$\frac{p}{r}S_N \xrightarrow{d} Z.$$

Trong đó $S_N = \sum_{i=1}^{N} X_i$ và $Z \sim G(r, \frac{r}{m})$.

<u>*Hướng dẫn giải*</u>

Gọi $\varphi_X(t)$ là hàm đặc trưng của biến ngẫu nhiên X.

Do $N \sim NB(r, p)$ nên ta có hàm sinh của N là

$$\psi_X(t) = \left[\frac{pt}{1-(1-p)t}\right]^r$$

Từ đây ta suy ra hàm đặc trưng của S_N được xác định như sau:

$$\varphi_{S_N}(t) = \psi \circ \varphi_X(t) = \left[\frac{p\varphi_X(t)}{1-(1-p)\varphi_X(t)}\right]^r$$

Theo đó hàm đặc trưng của $\frac{p}{r}S_N$ là

$$\varphi_{\frac{p}{r}S_N}(t) = \varphi_{S_N}\left(\frac{p}{r}t\right) = \left[\frac{p\varphi_X\left(\frac{p}{r}t\right)}{1-(1-p)\varphi_X\left(\frac{p}{r}t\right)}\right]^r$$

Theo định lý Taylor, tồn tại c ở giữa 0 và pt:

$$\varphi_X\left(\frac{p}{r}t\right) = \varphi_X(0) + \frac{p}{r}t\varphi'(c) = 1 + \frac{p}{r}t\varphi'(c)$$

Do đó

$$\begin{aligned}
\varphi_{\frac{p}{r}S_N}(t) &= \left[\frac{p\varphi_X\left(\frac{p}{r}t\right)}{1-(1-p)\varphi_X\left(\frac{p}{r}t\right)}\right]^r \\
&= \left[\frac{p\left(1+\frac{p}{r}t\varphi'(c)\right)}{1-(1-p)\left(1+\frac{p}{r}t\varphi'(c)\right)}\right]^r \\
&= \left[\frac{p\left(1+\frac{p}{r}t\varphi'(c)\right)}{1-1-\frac{p}{r}t\varphi'(c)+p+\frac{p^2}{r}t\varphi'(c)}\right]^r \\
&= \left[\frac{1+\frac{p}{r}t\varphi'(c)}{1-\frac{t}{r}\varphi'(c)+\frac{p}{r}t\varphi'(c)}\right]^r.
\end{aligned}$$

Cho $p \to 0^+$ thì $pt \to 0^+$ tức là $c \to 0$. Khi đó

$$\lim_{p\to 0^+} \varphi_{\frac{p}{r}S_N}(t) = \left[\frac{1}{1-\frac{t}{r}\varphi'(0)}\right]^r = \left[\frac{1}{1-\frac{t}{r}iE(X)}\right]^r = \left[\frac{1}{1-ti\frac{m}{r}}\right]^r = \left[\frac{\frac{r}{m}}{\frac{r}{m}-it}\right]^r.$$

Vậy khi $p \to 0^+$ thì

$$\frac{p}{r}S_N \sim G(r, \frac{r}{m}). \qquad \blacksquare$$

Bài 4.15. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots X_n, \ldots$ là dãy các biến ngẫu nhiên độc lập, cùng phân phối, $X_j \sim N(0,1), N \sim NB(r,p)$ và $Z \sim G(r,r)$. Khi đó, nếu $p \to 0$ thì

$$\frac{p}{r}(X_1^2 + X_2^2 + ... + X_N^2) \xrightarrow{d} Z$$

Hướng dẫn giải

Gọi $\varphi(t), \varphi_Z(t)$ lần lượt là hàm đặc trưng của X_j^2, Z. Ta có $Z \sim G(r,r)$ nên

$$\varphi_Z\left(t\right) = \left(\frac{r}{r - it}\right)^r.$$

Ta có $X_j \sim N(0,1)$ nên suy ra $X_j^2 \sim \chi_1^2$, như vậy ta xác định được hàm đặc trưng của X_j^2 là

$$\varphi\left(t\right) = \frac{1}{\sqrt{1 - 2it}}.$$

Do $N \sim NB(r,p)$ nên N có hàm sinh

$$\psi_N\left(t\right) = \left[\frac{pt}{1 - \left(1 - p\right)t}\right]^r$$

Khi đó hàm đặc trưng của S_N là

$$\varphi_{S_N}\left(t\right) = \psi_N \circ \varphi\left(t\right) = \left[\frac{p\varphi\left(t\right)}{1 - \left(1 - p\right)\varphi\left(t\right)}\right]^r$$

Ta suy ra hàm đặc trưng của $\frac{p}{r}S_N$ là

$$\varphi_{\frac{p}{r}S_N}\left(t\right) = \varphi_{S_N}\left(\frac{p}{r}t\right) = \left[\frac{p\varphi\left(\frac{p}{r}t\right)}{1 - \left(1 - p\right)\varphi\left(\frac{p}{r}t\right)}\right]^r = \left(\frac{\sqrt{1 - 2i\frac{p}{r}t} + 1 - p}{2 - 2i\frac{t}{r} - p}\right)^r.$$

Khi cho $p \to 0$, ta được

$$\varphi_{\frac{p}{r}S_N}\left(t\right) \to \left(\frac{r}{r - it}\right)^r = \varphi_Z\left(t\right). \quad \blacksquare$$

Bài 4.16. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots X_n, \ldots$ là dãy các biến ngẫu nhiên độc lập, cùng phân phối với biến ngẫu nhiên X, với $\mathbb{E}(X) = 0$ và $D(X) = \sigma^2 < \infty$. Đặt N là biến ngẫu nhiên có phân phối hình học tham số p và độc lập với tất cả các $X_i,\ i \geq 1$. Khi đó, với $p \to 0^+$ thì

$$\sqrt{p}S_N \xrightarrow{d} Z$$

Trong đó $S_N = \sum\limits_{i=1}^{N} X_i$ và $Z \sim L(0, \frac{\sqrt{2}}{\sigma})$.

<u>*Hướng dẫn giải*</u>

Gọi $\varphi_X(t)$ là hàm đặc trưng của biến ngẫu nhiên X.

Do $N \sim Geometric(p)$ nên ta có hàm sinh của N là

$$\psi(t) = \frac{pt}{1-(1-p)t}.$$

Từ đây ta suy ra hàm đặc trưng của S_N được xác định như sau:

$$\varphi_{S_N}(t) = \psi \circ \varphi(t) = \frac{p\varphi(t)}{1-(1-p)\varphi(t)}.$$

Theo đó hàm đặc trưng của $\sqrt{p}S_N$ là

$$\varphi_{\sqrt{p}S_N}(t) = \varphi_{S_N}(\sqrt{p}t) = \frac{p\varphi\left(\sqrt{p}t\right)}{1-(1-p)\varphi\left(\sqrt{p}t\right)}.$$

Theo định lý Taylor, tồn tại c ở giữa 0 và $\sqrt{p}t$:

$$\begin{aligned}
\varphi\left(\sqrt{p}t\right) &= \varphi(0) + \sqrt{p}t\varphi'(0) + \frac{pt^2}{2}\varphi''(c) \\
&= 1 + \sqrt{p}tiE(X) + \frac{pt^2}{2}\varphi''(c) \\
&= 1 + \frac{pt^2\varphi''(c)}{2}.
\end{aligned}$$

Do đó

$$\begin{aligned}
\varphi_{\sqrt{p}S_N}(t) &= \frac{p\varphi\left(\sqrt{p}t\right)}{1-(1-p)\varphi\left(\sqrt{p}t\right)} \\
&= \frac{p\left[1+\dfrac{pt^2\varphi''(c)}{2}\right]}{1-(1-p)\left[1+\dfrac{pt^2\varphi''(c)}{2}\right]} \\
&= \frac{1+\dfrac{pt^2\varphi''(c)}{2}}{1-\dfrac{t^2\varphi''(c)}{2}+\dfrac{pt^2\varphi''(c)}{2}}.
\end{aligned}$$

Cho $p \to 0^+$ thì $\sqrt{p}t \to 0^+$ tức là $c \to 0$. Khi đó

$$\lim_{p\to 0} \varphi_{\sqrt{p}S_N}(t) = \frac{1}{1 - \dfrac{t^2\varphi''(0)}{2}} = \frac{1}{1 + \dfrac{t^2 E\left(X^2\right)}{2}} = \frac{1}{1 + \dfrac{t^2\sigma^2}{2}}.$$

Vậy khi $p \to 0^+$ thì

$$\sqrt{p}S_N \xrightarrow{d} Z \sim L(0, \frac{\sqrt{2}}{\sigma}). \qquad \blacksquare$$

Bài 4.17. Chứng minh định lý sau đây

Giả sử $X_1, X_2, \ldots X_n, \ldots$ là dãy các biến ngẫu nhiên độc lập, cùng phân phối với biến ngẫu nhiên X, với $\mathbb{E}(X) = \sqrt{p}a$ và $E(X^2) = \sigma^2 < \infty$. Đặt N là biến ngẫu nhiên có phân phối hình học tham số p và độc lập với tất cả các $X_i,\ i \geq 1$. Khi đó, với $p \to 0^+$ thì

$$\sqrt{p}S_N \xrightarrow{d} Z$$

Trong đó $S_N = \displaystyle\sum_{i=1}^{N} X_i$ và $Z \sim L(0, a, \sigma)$.

<u>*Hướng dẫn giải*</u>

Gọi $\varphi_X(t)$ là hàm đặc trưng của biến ngẫu nhiên X.

Do $N \sim Geometric(p)$ nên ta có hàm sinh của N là

$$\psi(t) = \frac{pt}{1 - (1-p)t}.$$

Từ đây ta suy ra hàm đặc trưng của S_N được xác định như sau:

$$\varphi_{S_N}(t) = \psi \circ \varphi(t) = \frac{p\varphi(t)}{1 - (1-p)\varphi(t)}.$$

Theo đó hàm đặc trưng của $\sqrt{p}S_N$ là

$$\varphi_{\sqrt{p}S_N}(t) = \varphi_{S_N}(\sqrt{p}t) = \frac{p\varphi\left(\sqrt{p}t\right)}{1 - (1-p)\varphi\left(\sqrt{p}t\right)}.$$

Theo định lý Taylor, tồn tại c ở giữa 0 và $\sqrt{p}t$:

$$\begin{aligned}
\varphi\left(\sqrt{p}t\right) &= \varphi(0) + \sqrt{p}t\varphi'(0) + \frac{pt^2}{2}\varphi''(c) \\
&= 1 + \sqrt{p}tiE(X) + \frac{pt^2}{2}\varphi''(c) \\
&= 1 + piat + \frac{pt^2}{2}\varphi''(c).
\end{aligned}$$

Do đó

$$\begin{aligned}
\varphi_{\sqrt{p}S_N}(t) &= \frac{p\varphi\left(\sqrt{p}t\right)}{1-(1-p)\varphi\left(\sqrt{p}t\right)} \\
&= \frac{p\left[1+piat+\frac{pt^2}{2}\varphi''(c)\right]}{1-(1-p)\left[1+piat+\frac{pt^2}{2}\varphi''(c)\right]} \\
&= \frac{1+piat+\frac{pt^2}{2}\varphi''(c)}{1-iat-\frac{t^2}{2}\varphi''(c)+piat+\frac{pt^2}{2}\varphi''(c)}.
\end{aligned}$$

Cho $p \to 0^+$ thì $\sqrt{p}t \to 0^+$ tức là $c \to 0$. Khi đó

$$\lim_{p\to 0^+} \varphi_{\sqrt{p}S_N} = \frac{1}{1-iat-\frac{t^2}{2}\varphi''(0)} = \frac{1}{1-iat+\frac{\sigma^2t^2}{2}}.$$

Vậy khi $p \to 0^+$ thì

$$\sqrt{p}S_N \xrightarrow{d} Z \sim L(0,a,\sigma). \qquad \blacksquare$$

Bài 4.18. Chứng minh định lý sau đây

Giả sử đối với các biến ngẫu nhiên X và N_n, kỳ vọng và phương sai đều tồn tại hữu hạn. Khi đó
i. $\mathbb{E}\left(S_{N_n}\right) = \mathbb{E}\left(N_n\right).\mathbb{E}\left(X\right)$,
ii. $D\left(S_{N_n}\right) = \mathbb{E}\left(N_n\right).D\left(X\right) + \left[\mathbb{E}\left(X\right)\right]^2.D\left(N_n\right)$.

Hướng dẫn giải

i) $E\left(S_N\right) = \sum\limits_{m=0}^{\infty} P\left(N=m\right).E\left(S_m\right) = \left(\sum\limits_{m=0}^{\infty} P\left(N=m\right).m\right) E\left(X\right) = E\left(N\right).E\left(X\right)$.
ii) Ta có

$$\begin{aligned}
E\left(X^2\right) &= E\left(E\left(S_N^2/N\right)\right) \\
&= \sum_{m=1}^{\infty} P\left(N=m\right) E\left(S_m^2/N=m\right) \\
&= \sum_{m=1}^{\infty} P\left(N=m\right) E\left(S_m^2\right) \\
&= \sum_{m=1}^{\infty} P\left(N=m\right) \left(\sum_{j=1}^{m} EX_j^2 + \sum_{\substack{i\neq j \\ i,j=1}}^{m} E\left(X_i\right).E\left(X_j\right)\right) \\
&= \sum_{m=1}^{\infty} P\left(N=m\right) \left[mE\left(X^2\right) + \left(m^2-m\right)\left(E\left(X\right)\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{m=1}^{\infty} P(N=m)\left[mD(X) + m^2(E(X))^2\right] \\
&= E(N)\,D(X) + E\left(N^2\right)(E(X))^2.
\end{aligned}
$$

Do đó

$$
\begin{aligned}
D(S_N) &= E\left(X^2\right) - (E(X))^2 \\
&= E(N)\,D(X) + E\left(N^2\right)(E(X))^2 - (E(N)\,E(X))^2 \\
&= E(N)\,D(X) + (E(X))^2 D(N). \qquad \blacksquare
\end{aligned}
$$

Bài 4.19. Chứng minh định lý sau đây

Giả sử $g(t) = \mathbb{E}\left(t^{N_n}\right)$ là hàm sinh của biến ngẫu nhiên N và biến ngẫu nhiên X có hàm sinh $f(t) = \mathbb{E}\left(t^X\right)$ cùng hàm đặc trưng $\varphi(t) = \mathbb{E}\left(e^{itX}\right)$. Khi đó
i. S_{N_n} có hàm sinh $h(t)$ được định nghĩa bởi $h(t) = g \circ f(t)$.
ii. S_{N_n} có hàm đặc trưng $\psi(t)$ được định nghĩa bởi $\psi(t) = g \circ \varphi(t)$.

<u>Hướng dẫn giải</u>

i) Đặt $p_k = P(N=k), k = 0, 1, \ldots$ Ta có hàm sinh của S_N là

$$
\begin{aligned}
h(t) = E\left(t^{S_N}\right) &= E\left(E\left(t^{S_N}/N\right)\right) = \sum_{k=0}^{\infty} p_k E\left(t^{S_N}/N=k\right) \\
&= \sum_{k=0}^{\infty} p_k E\left(t^{S_k}\right) = \sum_{k=0}^{\infty} p_k\left[E\left(t^X\right)\right]^k = \sum_{k=0}^{\infty} p_k [f(t)]^k \\
&= E\left([f(t)]^N\right) = g \circ f(t).
\end{aligned}
$$

ii) Đặt $p_k = P(N=k), k = 0, 1, \ldots$ Ta có hàm đặc trưng của S_N là

$$
\begin{aligned}
\psi(t) = E\left(e^{itS_N}\right) &= E\left(E\left(e^{itS_N}/N\right)\right) = \sum_{k=0}^{\infty} p_k E\left(e^{itS_N}/N=k\right) \\
&= \sum_{k=0}^{\infty} p_k E\left(e^{itS_k}\right) = \sum_{k=0}^{\infty} p_k\left[E\left(e^{itX}\right)\right]^k = \sum_{k=0}^{\infty} p_k [\varphi(t)]^k \\
&= E\left([\varphi(t)]^N\right) = g \circ \varphi(t). \qquad \blacksquare
\end{aligned}
$$

Bài 4.20. Chứng minh định lý sau đây

Giả sử $X \sim Becnoulli\,(p)$, $p \in (0,1)$. Khi đó

i. Nếu $N_n \sim Poisson\,(\lambda)\,, \lambda > 0$, thì $S_{N_n} \sim Poisson\,(\lambda p)$.
ii. Nếu $N_n \sim Binomial\,(n, q)\,,\ q \in (0, 1)$ thì $S_{N_n} \sim Binomial\,(n, pq)$.
iii. Nếu $N_n \sim Geometry\,(q)\,,\, p \in (0, 1)$ thì $S_{N_n} \sim Geometry\left(\dfrac{q}{p + q - pq}\right)$.

<u>*Hướng dẫn giải*</u>

i) Ta có hàm sinh của N là $g(t) = e^{\lambda(t-1)}$ và hàm đặc trưng của X là $\varphi(t) = 1 + p(e^{it} - 1)$. Khi đó hàm đặc trưng của S_N là

$$\psi\,(t) = g \circ \varphi\,(t) = g\,(\varphi\,(t)) = e^{\lambda p\left(e^{it} - 1\right)}.$$

Vậy $S_N \sim Poisson\,(\lambda p)$.
ii) Ta có hàm sinh của N là $g\,(t) = [1 + q\,(t - 1)]^n$ và hàm đặc trưng của X là $\varphi(t) = 1 + p(e^{it} - 1)$. Khi đó hàm đặc trưng của S_N là

$$\psi\,(t) = g \circ \varphi\,(t) = g\,(\varphi\,(t)) = \left[1 + pq\left(e^{it} - 1\right)\right]^n.$$

Vậy $S_N \sim Binomial\,(n, pq)$.
iii) Ta có hàm sinh của N là $g\,(t) = \dfrac{q}{1 - (1 - q)\,t}$ và hàm đặc trưng của X là $\varphi(t) = 1 + p(e^{it} - 1)$. Khi đó hàm đặc trưng của S_N là

$$\begin{aligned}
\psi\,(t) &= g \circ \varphi\,(t) = g\,(\varphi\,(t)) \\
&= \frac{q}{1 - (1 - q)\left[1 + p\left(e^{it} - 1\right)\right]} \\
&= \frac{q}{p + q - pq - (p - pq)\,e^{it}} \\
&= \frac{\dfrac{q}{p + q - pq}}{1 - \left(1 - \dfrac{q}{p + q - pq}\right) e^{it}}.
\end{aligned}$$

Vậy $S_N \sim Geometry\left(\dfrac{q}{p + q - pq}\right)$. ■

Bài 4.21. Chứng minh định lý sau đây

Giả sử với mỗi $n = 1, 2, \ldots, \{X_{nk}, k = 1, 2, \ldots n\}$ là một dãy biến ngẫu nhiên độc lập, cùng phân phối Bernoulli với tham số $p_n \in [0, 1]$ và $P_n \longrightarrow 0, np_n \longrightarrow \lambda(\lambda > 0)$ khi $n \longrightarrow \infty$. Khi đó, với $n \longrightarrow \infty$,

$$S_n = \sum_{k=1}^{n} X_{nk} \xrightarrow{d} Poisson\,(\lambda)\,.$$

Hướng dẫn giải

Giả sử $N_n \sim Poisson(n)$, khi đó theo định lý 1.3 thì $S_{N_n} \sim Poisson(np_n)$. Như vậy ta có:

$$P\left(S_{N_n} = k\right) = \frac{e^{-np_n}\left(np_n\right)^k}{k!} \to \frac{e^{-\lambda}\lambda^k}{k!} \quad (n \to \infty)\,.$$

Hay

$$S_{N_n} \xrightarrow{d} Poisson(\lambda) \tag{4.2}$$

Gọi $\varphi_{S_{N_n}}(t)\,, \varphi_{S_n}(t)$ lần lượt là hàm đặc trưng của S_{N_n} và S_n. Ta có:

$$\begin{aligned}
\left|\varphi_{S_{N_n}}(t) - \varphi_{S_n}(t)\right| &= \left|e^{np_n\left(e^{it}-1\right)} - \left[1 - p_n\left(1 - e^{it}\right)\right]^n\right| \\
&\le n\left|e^{p_n\left(e^{it}-1\right)} - 1 - p_n\left(e^{it} - 1\right)\right| \\
&\le \tfrac{n\left|p_n\left(e^{it}-1\right)\right|^2}{2} \\
&= np_n\tfrac{p_n\left|e^{it}-1\right|^2}{2} \to 0 \quad (n \to 0)\,.
\end{aligned}$$

Như vậy $\lim\limits_{n\to\infty}\left|\varphi_{S_{N_n}}(t) - \varphi_{S_n}(t)\right| = 0$. Hay

$$S_n \xrightarrow{d} S_{N_n} \tag{4.3}$$

Từ (1.1) và (1.2) ta suy ra $S_n \xrightarrow{d} Poisson(\lambda)$. ■

Bài 4.22. Giả sử X không âm, khi đó với giá trị tùy ý $a > 0$ ta có bất đẳng thức Markov

$$P\left(X \ge a\right) \le \frac{\mathbb{E}\left[X\right]}{a}.$$

. Hãy chứng minh bất đẳng thức Markov?

Hướng dẫn giải

Ta xác định biến ngẫu nhiên Y như sau:

$$Y = \begin{cases} a & khi \;\; X \ge a, \\ 0 & khi \;\; X < a. \end{cases}$$

Theo giả thuyết $X \ge 0$ và từ cách xác định trên suy ra: $X \ge Y$
Do đó

$$E\left(X\right) \ge E\left(Y\right) \Leftrightarrow E\left(X\right) \ge a.P\left(X \ge a\right) \Leftrightarrow P\left(X \ge a\right) \le \frac{E\left(X\right)}{a}. \qquad ■$$

Bài 4.23. Giả sử X là biến ngẫu nhiên có kỳ vọng μ và phương sai $\sigma^2 < \infty$. Khi đó với giá trị tùy ý $k > 0$, ta có bất đẳng thức Chebyshev

$$P\left(|X-\mu| \geq k\right) \leq \frac{\sigma^2}{k^2}.$$

Hãy chứng minh bất đẳng thức Chebyshev?

Hướng dẫn giải

Cách 1: Theo bất đẳng thức Markov ta có

$$P\left(|X-\mu| \geq k\sigma\right) = P\left(\frac{(X-\mu)^2}{\sigma^2} \geq k^2\right) \leq \frac{E\left(\frac{(X-\mu)^2}{\sigma^2}\right)}{k^2} = \frac{1}{k^2} \quad \text{(đpcm)}.$$

Cách 2:

$$\begin{aligned} D(X) &= \int\limits_R |x-\mu|^2 dF_X(x) \\ &= \int\limits_{|x-\mu|<k} |x-\mu|^2 dF_X(x) + \int\limits_{|x-\mu|\geq k} |x-\mu|^2 dF_X(x) \\ &\geq \int\limits_{|x-\mu|\geq k} |x-\mu|^2 dF_X(x) \end{aligned}$$

Từ đó suy ra

$$\forall k > 0,\ \frac{1}{k^2} D(X) \geq \frac{1}{k^2} \int\limits_{|x-\mu|\geq k} (x-\mu)^2 dF_X(x) \geq \int\limits_{|x-\mu|\geq k} dF_X(x) = P\left(|X-\mu| \geq k\right). \quad \blacksquare$$

Bài 4.24. Chứng minh luật số lớn Chebyshev sau

Giả sử $\{X_n, n \geq 1\}$ *là một dãy các biến ngẫu nhiên độc lập, có phương sai hữu hạn bởi cùng một hằng số, tức là* $Var(X_n) = \sigma_n^2 \leq C,\ \forall n \geq 1.$. *Khi đó,* $\forall > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \geq \varepsilon\right\} = 0$$

Hướng dẫn giải

Ta có

$$E\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] = \mu$$

$$D\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^2}\left[D(X_1) + D(X_2) + ... + D(X_n)\right] = \frac{\sigma^2}{n}.$$

Theo bất đẳng thức Chebyshev, với giá trị $k > 0$ tùy ý ta có:

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \mu\right| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}.$$

Do cách đặt $k > 0$ tùy ý, ta đặt

$$\varepsilon = k\frac{\sigma}{\sqrt{n}} \Leftrightarrow k = \frac{\varepsilon\sqrt{n}}{\sigma} \Leftrightarrow \frac{1}{k^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Vậy

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Cho $n \to \infty$ ta được điều phải chứng minh. ■

Bài 4.25. Chứng minh luật số lớn Khintchine sau

Giả sử $\{X_n, n \geq 1\}$ là một dãy các biến ngẫu nhiên độc lập, cùng phân phối với kì vọng $E(X_n) = \mu \, \forall n \geq 1$. Khi đó,

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right\} = 0$$

khi $n \to \infty$.

Hướng dẫn giải

+ Ta cần chứng minh

$$f_{\frac{S_n}{n}}(t) \to f_\mu(t) \quad \text{as} \quad n \to \infty$$

$$\Leftrightarrow f_{\frac{S_n}{n}}(t) \to e^{it\mu} \quad \text{as} \quad n \to \infty$$

+ Ta có

$$f_{\frac{S_n}{n}}(t) = \left[f_{\frac{X_j}{n}}(t)\right]^n = \left[f_{X_j}\left(\frac{t}{n}\right)\right]^n$$

+ Đặt

$$l(t) = \ln f_{X_j}(t)$$

+ Khi đó

$$\begin{aligned} f_{\frac{S_n}{n}}(t) &= \left[f_{X_j}\left(\tfrac{t}{n}\right)\right]^n \\ &= \exp\left[nl\left(\tfrac{t}{n}\right)\right] \\ &= \exp\left[\frac{l\left(\frac{t}{n}\right) - l(0)}{\frac{t}{n}}.t\right] \\ &\to \exp(i\mu t) \quad \textit{as} \quad n \to \infty. \end{aligned}$$

■

Bài 4.26. Chứng minh định lý giới hạn trung tâm (suy rộng của định lý Moivres-Laplace) sau đây:

Giả sử $(X_n)_{n\geq 1}$ là một dãy các biến ngẫu nhiên độc lập, cùng phân phối với biến ngẫu nhiên X có kỳ vọng μ và phương sai $\sigma^2 < \infty$. Đặt $S_n = \sum\limits_{i=1}^{n} X_i$ và $\hat{S}_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, khi đó

$$\lim_{n\to\infty} P\left(\hat{S}_n < x\right) = \Phi(x) \quad \forall x \in \mathbb{R},$$

trong đó $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ là hàm phân phối chuẩn.

Hướng dẫn giải

Chúng ta có

$$S_n = \frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu).$$

Biến ngẫu nhiên $(X-\mu)$ có kỳ vọng 0 và phương sai σ^2. Gọi $\varphi(t)$ là hàm đặc trưng của $(X-\mu)$ ta có $\varphi'(0) = 0, \quad \varphi''(0) = -\sigma^2$.
Do $X_1, X_2, \ldots, X_n$ độc lập nên hàm đặc trưng của S_n được xác định là

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

Theo định lý Taylor, có một số c ở giữa 0 và $\frac{t}{\sigma\sqrt{n}}$ sao cho:

$$\begin{aligned}
\varphi\left(\tfrac{t}{\sigma\sqrt{n}}\right) &= \varphi(0) + \varphi'(0)\frac{t}{\sigma\sqrt{n}} + \varphi''(c)\frac{t^2}{2n\sigma^2} \\
&= 1 + \varphi''(c)\frac{t^2}{2\sigma^2 n} \\
&= 1 - \frac{t^2}{2n} + \frac{\left[\varphi''(c) + \sigma^2\right] t^2}{2n\sigma^2}.
\end{aligned}$$

Ta suy ra

$$\varphi_n(t) = \left[1 - \frac{t^2}{2n} + \frac{\left[\varphi''(c) + \sigma^2\right] t^2}{2n\sigma^2}\right]^n.$$

Do $\varphi''(t)$ liên tục tại 0 và khi $n \longrightarrow \infty$ thì $\frac{1}{\sigma\sqrt{n}} \to 0$ dẫn đến $c \longrightarrow 0$ nên ta có

$$\lim_{n\to\infty} \frac{\left[\varphi''(c) + \sigma^2\right] t^2}{2n\sigma^2} = 0.$$

Vậy
$$\lim_{n\to\infty}\varphi_n(t)=e^{-\frac{t^2}{2}} \text{ hay } S_n\xrightarrow{d}X^*\sim N(0,1). \qquad \blacksquare$$

Bài 4.27. Chứng minh định lý xấp xỉ Poisson sau đây

Cho $X_n\sim B(n;p_n)$. Nếu số phép thử $n\longrightarrow\infty$ và xác suất thành công $p_n\longrightarrow 0$ sao cho $np_n\longrightarrow\lambda$ thì $X_n\xrightarrow{d}Poisson(\lambda)$.

<u>*Hướng dẫn giải*</u>

Với $q_n=1-p_n$ ta có

$$\begin{aligned} C_n^k p_n^k q_n^{n-k} &= \frac{n!}{k!\,(n-k)!}p_n^k q_n^{n-k}\\ &= \frac{n(n-1)\ldots.(n-k+1)}{k!}p_n^k q_n^{n-k}\\ &= \frac{1}{k!}\frac{n(n-1)\ldots.(n-k+1)}{n^k}(np)^k\left[(1-p_n)^{-\frac{1}{p_n}}\right]^{-p_n(n-k)}. \end{aligned}$$

Cho $n\longrightarrow\infty, p_n\longrightarrow 0$, ta có:

- $\dfrac{n(n-1)\ldots.(n-k+1)}{n^k}=1.\left(1-\dfrac{1}{n}\right).\left(1-\dfrac{2}{n}\right)\ldots.\left(1-\dfrac{k-1}{n}\right)\to 1,$
- $(np_n)^k\to\lambda^k,$
- $-p_n(n-k)\to-\lambda,$
- $(1-p_n)^{-\frac{1}{p_n}}\to e.$

Vậy khi $n\longrightarrow\infty$ và $p_n\longrightarrow 0$ thỏa mãn $np_n\longrightarrow\lambda$ thì
$$C_n^k p_n^k q_n^{n-k}\to e^{-\lambda}\frac{\lambda^k}{k!} \text{ hay } X_n\xrightarrow{d}Poisson(\lambda). \qquad \blacksquare$$

Bài 4.28. Chứng minh định lý sau đây

Nếu X là biến ngẫu nhiên nhận giá trị nguyên với hàm đặc trưng $\varphi_X(t)$ thì
$$\mathrm{P}[X=k]=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-itk}\varphi_X(t)\,dt, \qquad k=0,\pm1,\pm2,\ldots$$

<u>*Hướng dẫn giải*</u>

Ta có

$$\begin{aligned}
\int\limits_{-\pi}^{\pi} e^{-itk}\varphi_X(t)\,dt &= \int\limits_{-\pi}^{\pi} e^{-itk} E e^{itX} dt \\
&= \int\limits_{-\pi}^{\pi} e^{-itk} \sum_{j=0}^{\infty} e^{itj} \mathrm{P}\left[X=j\right] dt \\
&= \int\limits_{-\pi}^{\pi} \left[\mathrm{P}\left[X=k\right] + \sum_{j\neq k} e^{it(j-k)} \mathrm{P}\left[X=j\right] \right] dt \\
&= 2\pi \mathrm{P}\left[X=k\right] + \int\limits_{-\pi}^{\pi} \sum_{j\neq k} e^{it(j-k)} \mathrm{P}\left[X=j\right] dt \\
&= 2\pi \mathrm{P}\left[X=k\right] + \sum_{j\neq k} \int\limits_{-\pi}^{\pi} e^{it(j-k)} \mathrm{P}\left[X=j\right] dt \\
&= 2\pi \mathrm{P}\left[X=k\right] \\
&\quad + \sum_{j\neq k} \mathrm{P}\left[X=j\right] \int\limits_{-\pi}^{\pi} [\cos t\,(j-k) + i \sin t\,(j-k)] dt.
\end{aligned} \tag{4.4}$$

Mặt khác

$$\begin{aligned}
&\int\limits_{-\pi}^{\pi} [\cos t\,(j-k) + i \sin t\,(j-k)] dt \\
&= \left[\frac{1}{j-k} \sin t\,(j-k) - \frac{i}{j-k} \cos t\,(j-k) \right] \Bigg|_{-\pi}^{\pi} \\
&= -\frac{i}{j-k} \cos \pi\,(j-k) + \frac{i}{j-k} \cos \pi\,(j-k) \\
&= 0.
\end{aligned} \tag{4.5}$$

Từ (4.1) và (4.2) suy ra bổ đề đã chứng minh xong. ■

Bài 4.29. Chứng minh định lý sau đây

Giả sử cho mảng tam giác $(X_{nj}, j = 1, 2, ..., n; n = 1, 2, ...)$ *sao cho với mỗi* n*, dãy* $X_{n1}, X_{n2}, ..., X_{nn}$ *là các biến ngẫu nhiên Bernoulli độc lập, với* $P(X_{nj} = 1) = 1 - P(X_{nj} = 0) = p_{nj}$ $(p_{nj} \in [0, 1]; j = 1, 2, \ldots, n; n = 1, 2, \ldots)$*. Cho* $Z_{\lambda_n} \sim$

Poisson (λ_n) *với* $\lambda_n = \sum\limits_{j=1}^{n} p_{nj}$. *Đặt* $S_n = \sum\limits_{j=1}^{n} X_{nj}$, *khi đó ta có*

$$|P(S_n = k) - P(Z_{\lambda_n} = k)| \leq \sum_{j=1}^{n} p_{nj}^2 \qquad k = 0, 1, 2, \ldots$$

<u>*Hướng dẫn giải*</u>

Ta lần lượt tính các hàm đặc trưng của $X_{nj},\ S_n,\ Z_{\lambda_n}$

$$\begin{aligned}
\varphi_{X_{nj}}(t) &= \mathbb{E}e^{itX_{nj}} = p_{kj}e^{it} + 1 - p_{nj} = p_{nj}\left(e^{it} - 1\right) + 1 \\
\varphi_{S_n}(t) &= \mathbb{E}e^{itS_n} = \prod_{j=1}^{n} \mathbb{E}e^{itX_{nj}} = \prod_{j=1}^{n} \left[p_{nj}\left(e^{it} - 1\right) + 1\right] \\
\psi(t) &= \mathbb{E}e^{itZ_{\lambda_n}} = e^{\lambda_n\left(e^{it} - 1\right)} = \prod_{j=1}^{n} e^{p_{nj}\left(e^{it} - 1\right)}.
\end{aligned}$$

Ta có

$$\begin{aligned}
|\varphi_{S_n}(t) - \psi(t)| &= \left|\prod_{j=1}^{n}\left[p_{nj}\left(e^{it} - 1\right) + 1\right] - \prod_{j=1}^{n} e^{p_{nj}\left(e^{it} - 1\right)}\right| \\
&\leq \sum_{j=1}^{n}\left|p_{nj}\left(e^{it} - 1\right) + 1 - e^{p_{nj}\left(e^{it} - 1\right)}\right| \\
&\leq \frac{1}{2}\sum_{j=1}^{n} p_{nj}^2\left|e^{it} - 1\right|^2 \\
&= \sum_{j=1}^{n} p_{nj}^2 (1 - \cos t).
\end{aligned}$$

Theo bài tập trên ta có

$$\begin{aligned}
|P(S_n = k) - P(Z_{\lambda_n} = k)| &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\left(\varphi_{S_n}(t) - \psi(t)\right) dt\right| \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |\varphi_{S_n}(t) - \psi(t)|\, dt \\
&\leq \frac{1}{\pi}\int_{0}^{\pi}\sum_{j=0}^{n} p_{nj}^2 (1 - \cos t)\, dt \\
&= \sum_{j=0}^{n} p_{nj}^2.
\end{aligned}$$

Vậy

$$|P(S_n = k) - P(Z_{\lambda_n} = k)| \leq \sum_{j=1}^{n} p_{nj}^2. \qquad \blacksquare$$

Bài 4.30. Chứng minh định lý Renyi sau đây

Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên nhận giá trị không âm, độc lập, cùng phân phối và kỳ vọng $\mathbb{E}(X_1) = m$ với $0 < m < +\infty$. Đặt N là biến ngẫu nhiên có phân phối hình học tham số q, $q \in (0,1)$. Khi đó

$$\lim_{q \to 0^+} F_{qS_N}(x) = 1 - e^{-\frac{x}{m}}. \tag{4.6}$$

Trong đó $S_N = X_1 + X_2 + \cdots + X_N$ và $F_{qS_N}(x) = \mathbb{P}(qS_N \leq x)$.

Hướng dẫn giải

Đặt φ_{X_1} là hàm đặc trưng của biến ngẫu nhiên X_1, khi đó hàm đặc trưng của qX_1 là φ_{qX_1} được xác định bởi

$$\forall t \in \mathbb{R}, \qquad \varphi_{qX_1}(t) = \varphi_{X_1}(qt).$$

Theo giả thiết ta có $\mathbb{E}(qX_1) = qm < +\infty$ nên $\varphi_{qX_1}(t)$ tồn tại đạo hàm cấp một liên tục. Từ đó ta có khai triển Taylor của hàm φ_{qX_1} tại 0 dạng

$$\begin{aligned} \varphi_{qX_1}(t) &= \varphi_{X_1}(0) + qt\varphi'_{X_1}(\eta) \\ &= 1 + qt\varphi'_{X_1}(\eta), \end{aligned}$$

trong đó η nằm giữa 0 và qt. Mặt khác $\{X_i, i \geq 1\}$ là dãy các biến ngẫu nhiên độc lập cùng phân phối, N là biến ngẫu nhiên tuân theo phân phối hình học tham số q độc lập với tất cả các X_i và có hàm sinh ψ_N được xác định

$$\psi_N(s) = \frac{qs}{1 - (1-q)s}.$$

Từ đó ta xác định được hàm đặc trưng φ_{qS_N} của qS_N dạng

$$\begin{aligned} \varphi_{qS_N}(t) &= \psi_N\left[\varphi_{qX_1}(t)\right] \\ &= \frac{q\left[1 + qt\varphi'_{X_1}(\eta)\right]}{1 - (1-q)\left[1 + qt\varphi'_{X_1}(\eta)\right]} \\ &= \frac{1 + qt\varphi'_{X_1}(\eta)}{1 + qt\varphi'_{X_1}(\eta) - t\varphi'_{X_1}(\eta)}. \end{aligned}$$

Với số thực η được xác định như trên ta có, khi $q \to 0$ thì $\eta \to 0$. Và vì φ'_{X_1} liên tục nên ta có

$$\lim_{q\to 0} \varphi'_{X_1}(\eta) = \varphi'_{X_1}(0) = im.$$

Từ đó ta có

$$\begin{aligned}
\lim_{q\to 0^+} \varphi_{qS_N}(t) &= \lim_{q\to 0^+} \frac{1 + qt\varphi'_{X_1}(\eta)}{1 + qt\varphi'_{X_1}(\eta) - t\varphi'_{X_1}(\eta)}. \\
&= \frac{1}{1 - itm} \\
&= \frac{\frac{1}{m}}{\frac{1}{m} - it}.
\end{aligned}$$

Từ đó suy ra $\lim\limits_{q\to 0^+} F_{qS_N}(x) = 1 - e^{-\frac{x}{m}}$. ■

Bài 4.31. Chứng minh bất đẳng thức Jensen

Nếu $g : \mathbb{R} \to \mathbb{R}$ là hàm Borel, lồi và $X \in L_1$ thì

$$g(EX) \leq E(gX)$$

Hướng dẫn giải

Vì g là hàm lồi nên g liên tục và có đạo hàm trái, đạo hàm phải tại mọi điểm. Khi đó đồ thị của g có tiếp tuyến tại mọi điểm và tiếp tuyến có dạng $y = a + bx$. Xét điểm $\mu = E(X)$ ta có:

$$\begin{aligned}
E[g(X)] &\geq E(a + bx) \\
&= a + bE(x) \\
&= a + b\mu \\
&= g(\mu) \\
&= g[E(X)].
\end{aligned}$$ ■

Bài 4.32. Chứng minh bất đẳng thức Lyapunov

Giả sử X là biến ngẫu nhiên. Nếu $0 < p_1 < p_2 < \infty$ thì

$$(E|X|^{p_1})^{1/p_1} \leq (E|X|^{p_2})^{1/p_2}$$

Hướng dẫn giải

Đặt$r = \frac{p_2}{p_1} > 1$. Đặt $Y = |X|^{p_1}$ và sử dụng bất đẳng thức Jensen với hàm $g(y) = |y|^r$ ta được

$$(E[|Y|])^r \leq E[|Y|^r]$$

Hay

$$(E[|X|^{p_1}])^{p_2/p_1} \leq E[|X|^{p_2}]$$

Lấy căn bậc p_2 hai vế ta được:

$$(E|X|^{p_1})^{1/p_1} \leq (E|X|^{p_2})^{1/p_2}. \qquad \blacksquare$$

Bài 4.33. Tung đồng tiền 1000 lần. Ước lượng trên xác suất của độ lệch tần suất xuất hiện mặt sấp so với xác suất xuất hiện mặt đó nhỏ hơn 0,1.

Giải

áp dụng luật số lớn Bernoulli với $n = 1000; p = q = \frac{1}{2}; \varepsilon = 0,1$.

Ta có

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2} \Leftrightarrow P\left(\left|\frac{m}{1000} - \frac{1}{2}\right| < 0,1\right) > 1 - \frac{1}{1000.0,01} = \frac{39}{40}.$$

Vì bất đẳng thức $\left|\frac{m}{1000} - \frac{1}{2}\right| < 0,1$ tương đương với bất đẳng thức kép $400 < m < 600$, nên ta có thể nói rằng xác suất số lần rơi mặt sấp trong khoảng (400;600) lớn hơn 39/40.

Bài 4.34. Trong thùng có 1000 bi trắng và 2000 bi đen. Rút có hoàn lại 300 bi. Ước lượng trên xác suất để số m bi trắng rút ra thỏa mãn bất đẳng thức kép $80 < m < 120$.

Giải

Bất đẳng thức kép đã cho có thể viết dưới dạng

$$-20 < m - 100 < 20 \Leftrightarrow -\frac{1}{15} < \frac{m}{300} - \frac{1}{3} < \frac{1}{15}.$$

Vậy chỉ cần ước lượng xác suất của bất đẳng thức

$$\left|\frac{m}{300} - \frac{1}{3}\right| < \frac{1}{15}$$

do đó $\varepsilon = \dfrac{1}{15}$.

và

$$P\left(\left|\frac{m}{300} - \frac{1}{3}\right| < \frac{1}{15}\right) > 1 - \frac{pq}{n\varepsilon^2} = \frac{\frac{1}{3}\cdot\frac{2}{3}}{300.\frac{1}{225}} = \frac{5}{6}.$$

Bài 4.35. Giả sử xác suất để biến cố A xuất hiện trong một dãy n phép thử bằng n. Tính xác suất để biến cố A xuất hiện từ α đến β lần?

Giải

Hiển nhiên rằng, từ $\left(a < \dfrac{x - np}{\sqrt{npq}} < b\right)$.

Suy ra $\left(np + a\sqrt{npq} < x < np + b\sqrt{npq}\right)$.

Đặt

$$np + a\sqrt{npq} = \alpha \Leftrightarrow a = \frac{\alpha - np}{\sqrt{npq}};$$
$$np + b\sqrt{npq} = \beta \Leftrightarrow b = \frac{\beta - np}{\sqrt{npq}}.$$

áp dụng định lý Moivre-Laplac, ta nhận được

$$P(\alpha < X < \beta) = \frac{1}{2}\left[\Phi_0\left(\frac{\beta - np}{\sqrt{2npq}}\right) - \Phi_0\left(\frac{\alpha - np}{\sqrt{2npq}}\right)\right].$$

Bài 4.36. Xác suất biến cố A trong phép thử bằng 0,7. Hỏi phải lặp lại phép thử bao nhiêu lần để với xác suất 0,9 có thể tin tưởng rằng tần số xuất hiện biến cố A sẽ lệch so với xác suất không lớn hơn 0,05?

Giải

Từ đầu bài suy ra $\left|\frac{X}{n} - 0,7\right| < 0,05$.

$$\left|\frac{X}{n} - 0,7\right| < 0,05 \Leftrightarrow 0,65n < X < 0,75n.$$

Khi đó, ta có

$$\begin{aligned} P(0,65n < X < 0,75n) &= \frac{1}{2}\left[\Phi\left(\frac{0,75n - 0,7n}{\sqrt{2n(1/2)(1/2)}}\right) - \Phi\left(\frac{0,65n - 0,7n}{\sqrt{2n(1/2)(1/2)}}\right)\right] \\ &= \Phi\left(\frac{\sqrt{2n}}{20}\right). \end{aligned}$$

Từ phương trình $\Phi\left(\frac{\sqrt{2n}}{20}\right) = 0,9$, tra bảng tích phân Laplace ta có

$$\frac{\sqrt{2n}}{20} = 1,282 \Leftrightarrow n = 328,7048.$$

Bài 4.37. Gieo 10000 hạt đậu tương với xác suất nảy mầm của mỗi hạt là 0,5. Tính gần đúng xác suất để số hạt nảy mầm trong khoảng 1000 đến 2000.

Giải

Gọi k là số hạt nảy mầm, theo giả thiết ta có $n = 10000; p = 0,5$. Khi đó

$$\begin{aligned} P\left(1000 < k < 2000\right) &= \frac{1}{2}\left\{\Phi\left(\frac{2000 - 10^4.0,5}{\sqrt{10^4.0,5.0,5}}\right) - \Phi\left(\frac{1000 - 10^4.0,5}{\sqrt{10^4.0,5.0,5}}\right)\right\} \\ &= \frac{1}{2}\left\{\Phi\left(-60\right) - \Phi\left(-80\right)\right\} \end{aligned}$$

Theo tính chất đối xứng của hàm $\Phi(x)$, ta có:

$$\Phi(60) = 1 - \Phi(80)$$

nên tra bảng Laplace, ta có $\Phi(60) = \Phi(80) \approx 1$.

dẫn tới $P(1000 < k < 2000) \approx 0$.

Bài 4.38. Gieo 5000 hạt giống với xác suất để mỗi hạt bị lép là . Tính gần đúng xác suất để trong 5000 hạt có đúng 6 hạt lép.

Giải

Theo định lý xấp xỉ Poisson, với p khá bé và n khá lớn, ta có công thức xấp xỉ

$$P\left(X = k\right) \approx \frac{\lambda^k e^{-\lambda}}{k!};\ \ k = 0, 1, 2, ...$$

Với giả thiết $k = 6; n = 5000; p = 0,001$ ta có

$$\lambda = np = 5000.0,001 = 5$$

Vậy xác suất cần tìm là $p \approx \dfrac{5^6 e^{-5}}{6!}$.

Bài 4.39. Người ta lấy ngẫu nhiên 400 hạt lúa từ một kho lúa rất lớn, có tỉ lệ hạt lép là 0,2 để kiểm tra. Tính xác suất để trong đó có từ 80 đến 100 hạt lép.

Giải

Gọi X là biến ngẫu nhiên chỉ số hạt lép trong mẫu thì $X \sim B(n;p)$ với $n = 400$ và $p = 0,2$.

Vì $n > 30, np = 80 > 5$ và $n(1-p) = 320 > 5$ nên chúng ta có thể áp dụng cách tính xấp xỉ bằng phân phối chuẩn $N(80, 64)$.

Xác suất để có từ 80 đến 100 hạt lép:

$$\begin{aligned} P\left(80 \le X \le 100\right) &\approx \Phi_0\left(\frac{100-80}{\sqrt{80.0,8}}\right) - \Phi_0\left(\frac{80-80}{\sqrt{80.0,8}}\right) \\ &= \Phi_0\left(2,5\right) - \Phi_0\left(0\right) = 0,4938 \end{aligned}$$

Bài 4.40. Giả sử xác suất để máy M sản xuất ra một phế phẩm là 1/250. Khi xuất xưởng, người ta xếp các sản phẩm đó vào từng kiện gồm 500 sản phẩm. Tính tỉ lệ các kiện hàng chứa không quá một phế phẩm.

Giải

Gọi X là BNN chỉ số phế phẩm trong một kiện hàng thì $X \sim B(n,p)$ với $n = 500$ và $p = \dfrac{1}{250}$. Khi đó, xác suất phải tính là

$$P\left(X \le 1\right) = \left(\frac{249}{250}\right)^{500} + 500\left(\frac{1}{250}\right)\left(\frac{249}{250}\right)^{499} = 0,40546.$$

Vì $n = 500$ khá lớn và $np = 2 < 5$ nên chúng ta có thể tính xấp xỉ xác suất trên bằng phân phối Poisson(2):

$$P\left(X = 0\right) + P\left(X = 1\right) \approx \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} = 0,406006.$$

Bài 4.41. Một lô hàng có 5000 sản phẩm, trong đó có 250 phế phẩm. Lấy ngẫu nhiên 10 sản phẩm từ lô hàng này. Tính xác suất để trong 10 sản phẩm lấy ra có 2 phế phẩm.

Giải

Gọi X là số phế phẩm lấy ra. Khi đó $X \sim H(5000, 250, 10)$.

Vì $n = 10$ rất nhỏ so với $N = 5000$ nên ta có thể xem $X \sim B(n,p)$ với $p = \dfrac{250}{5000} = 0,05$.

Vậy

$$P\left(X = 2\right) = C_{10}^2.(0,05)^2.(0,95)^8 \approx 0,074635.$$

Bài 4.42. Cho biết xác suất trúng đích trong mỗi lần bắn của một xạ thủ là 0,99. Xạ thủ bắn 500 phát, tính xác suất để có ít nhất 2 phát bắn trượt.

Giải

Gọi X là số lần xạ thủ bắn trượt. Ta cần tính $P(X \geq 2)$. Theo giả thiết thì $X \sim B(500; 0,01)$.

Vì $p = 0,01$ rất nhỏ so với $n = 500$ và $np = 500.0,01 = 5$ nên ta có thể xấp xỉ $X \sim P(5)$. Khi đó ta có

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-5}\frac{5^0}{0!} - e^{-5}\frac{5^1}{1!} \\ &\approx 0,95957 \end{aligned}$$

Bài 4.43. Theo thống kê của một công ty sản xuất xe máy thì có 60% khách mua sản phẩm của họ là phụ nữ. Chọn ngẫu nhiên 200 người mua xe của công ty này. Tính xác suất để trong 200 người này có ít nhất 140 phụ nữ.

Giải

Gọi X là số phụ nữ trong 200 người được chọn. Khi đó $X \sim B(200; 0,6)$, ta cần tính $P(140 \leq X \leq 200)$.

Vì $p = 0,6$ khá nhỏ so với $n = 200$ và $np = 200.0,6 = 120 > 5$ nên ta có thể xấp xỉ $X \sim N(\mu, \sigma^2)$ với $\mu = np = 120$ và $\sigma^2 = npq = 200.0,6.0,4 = 48$.

Ta có

$$\begin{aligned} P(140 \leq X \leq 200) &\approx \Phi_0\left(\frac{200-\mu}{\sigma}\right) - \Phi_0\left(\frac{140-\mu}{\sigma}\right) \\ &= \Phi_0\left(\frac{200-120}{\sqrt{48}}\right) - \Phi_0\left(\frac{140-120}{\sqrt{48}}\right) \\ &= \Phi_0(11,55) - \Phi_0(2,89) = 0,0019. \end{aligned}$$

Bài 4.44. Từ một lô hàng có 1000 sản phẩm, trong đó có 10 phế phẩm. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Tính xác suất để có ít nhất 2 phế phẩm.

Giải

Gọi X là số phế phẩm chọn được. Khi đó $X \sim H(1000, 10, 20)$.

Do $n = 20$ khá nhỏ so với $N = 1000$ nên ta có thể xấp xỉ $X \sim B(n, p)$ trong đó

$n = 20, p = \frac{10}{1000} = 0,01.$

Xác suất để có ít nhất 2 phế phẩm

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - C_{20}^{0}.(0,01)^{0}.(0,99)^{20} - C_{20}^{1}.(0,01)^{1}.(0,99)^{19} \\ &\approx 0,0169. \end{aligned}$$

Bài 4.45. Mỗi chuyến xe chở được 1000 chai bia. Xác suất để một chai bia bị vỡ khi vận chuyển là 0,001. Tìm xác suất khi vận chuyển có:

a) 2 chai vỡ, số chai vỡ không ít hơn 2.

b) Số chai vỡ trung bình khi vận chuyển.

Giải

a)Gọi X là số chai vỡ khi vận chuyển 1000 chai. Khi đó $X \sim B(1000; 0,001)$.

Vì n khá lớn, p khá bé và $np = 1$ nên ta có thể xem $X \sim P(\lambda), \lambda = 1$. Ta có

$$P(X = 2) \approx e^{-1}.\frac{(-1)^2}{2!} = 0,1839;$$
$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \approx 0,2642.$$

b) Số chai vỡ trung bình: $E(X) = 1$.

Bài 4.46. Một nhà máy có 3 phân xưởng cùng sản xuất ra một loại sản phẩm. Tỉ lệ sản phẩm loại B của các phân xưởng tương ứng là: 10%, 20%, 30%. Từ một lô hàng có 10000 sản phẩm gồm 3000 sản phẩm của phân xưởng 1; 4000 sản phẩm của phân xưởng 2 và 3000 sản phẩm của phân xưởng 3. Người ta chọn ngẫu nhiên ra 100 sản phẩm để kiểm tra. Nếu thấy có không quá 24 sản phẩm loại B trong 100 sản phẩm kiểm tra thì mua lô hàng đó. Tính xác suất lô hàng được mua.

Giải

Số sản phẩm loại B của các phân xưởng tương ứng là 300; 800; 900. Gọi X là số sản phẩm loại B trong 100 sản phẩm kiểm tra, khi đó ta có $X \sim H(10000; M; 100)$ với $M = 300 + 800 + 900 = 2000$.

Vì $N = 10000$ rất lớn so với $n = 100$.

Suy ra $X \sim B(100; p)$ với $p = \dfrac{2000}{10000} = 0,2,\ np = 20,\ \sqrt{npq} = 4.$

Xác suất lô hàng được mua:

$$\begin{aligned} P(X \le 24) &= P(0 \le X < 25) \\ &= \Phi_0\left(\frac{25 - 20 - 0,5}{4}\right) - \Phi_0\left(\frac{0 - 20 - 0,5}{4}\right) \\ &= 0,89065. \end{aligned}$$

Bài 4.47. Xác suất để một máy sản xuất ra một phế phẩm là 0,1%. Cho máy sản xuất 1000 sản phẩm. Tính xác suất có đúng 2 phế phẩm.

Giải

Gọi X là số phế phẩm có khi sản xuất 1000 sản phẩm. Khi đó $X \sim B(1000; 0,001)$.

Vì $n = 1000$ khá lớn và $p = 0,001$ khá bé nên ta có thể xem $X \sim P(np)$ với $np = 1$. Ta có

$$P(X = 2) \approx e^{-1}\frac{(-1)^2}{2!} = 0,1839.$$

Bài 4.48. Một nhân viên thuế chọn ngẫu nhiên một số tờ khai từ nhóm những tờ khai thuế đặc thù để kiểm tra. Kinh nghiệm cho thấy tỷ lệ tờ khai không thích hợp là 30%.

a) Nếu chọn 6 tờ khai từ nhóm 100 tờ khai thuế đặc thù mà có hơn 1 tờ khai không thích hợp thì nhóm sẽ bị kiểm tra toàn bộ. Tính xác suất nhóm bị kiểm tra toàn bộ.

b) Số tờ khai tối thiểu không thích hợp là bao nhiêu khi kiểm tra 18 trong nhóm 400 tờ khai thuế đặc thù để xác suất nhóm này bị kiểm tra lại toàn bộ là 0,31.

Giải

a) Gọi X là số tờ khai không thích hợp trong 6 tờ được kiểm tra. Khi đó, $X \sim H(100; M; 6)$ với $M = p.N = 0,3.100 = 30$.

Xác suất nhóm bị kiểm tra toàn bộ:

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{C_{70}^6}{C_{100}^6} - \frac{C_{30}^1 . C_{70}^5}{C_{100}^6} \approx 0,5854.$$

b) Gọi Y là số tờ khai không thích hợp khi kiểm tra 18 tờ. Khi đó $Y \sim H(400; M; 18)$ với $M = 0,3.400 = 120$.

Gọi a là số tờ khai không thích hợp tối thiểu thì $P(a \le Y \le 18) = 0,31$.

Vì $N = 400$ khá lớn so với $n = 18$ ($n < 0,05.N$), suy ra $Y \sim B(18; 0,3)$.

Phân phối nhị thức này lại được xấp xỉ bởi phân phối chuẩn $N(np; npq)$ với $np = 5,4$ và $npq = 3,78$. Ta có

$$\begin{aligned} P\left(a \le Y \le 18\right) &= P\left(a \le Y < 19\right) \\ &\approx \Phi_0\left(\frac{19 - 5,4 - 0,5}{\sqrt{3,78}}\right) - \Phi_0\left(\frac{a - 5,4 - 0,5}{\sqrt{3,78}}\right) \\ &= 0,5 - \Phi_0\left(\frac{a - 5,9}{1,9442}\right) = 0,31 \end{aligned}$$

Suy ra

$$\begin{aligned} &\Phi_0\left(\frac{a - 5,9}{1,9442}\right) = 0,19 = \Phi_0\left(0,49\right) \\ &\Leftrightarrow \frac{a - 5,9}{1,9442} = 0,49 \\ &\Leftrightarrow a = 6,85 \end{aligned}$$

Vậy a tối thiểu là 7.

4.8 Bài tập đề nghị

Bài 4.49. Giả sử đại lượng ngẫu nhiên X có hàm mật độ phân phối

$$F_X\left(x\right) = \begin{cases} a, & \text{nếu } -a \le x \le b \\ 0, & \text{nếu } x < -b,\ x > b \end{cases}$$

Hãy tìm hàm đặc trưng $\varphi_X\left(t\right)$.

ĐS: $\dfrac{sinbt}{bt}; b > 0$.

Bài 4.50. Giả sử $X \sim P\left(\lambda\right),\ \lambda > 0$, tức

$$P\left(X = k\right) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 1, 2, ...$$

a) Tìm hàm đặc trưng $\varphi_X(t)$, từ đó suy ra $E(X)$ và $D(X)$.

b) Chứng minh rằng nếu $X_i \sim P(\lambda_i)$, $\lambda_i > 0\ \forall i = \overline{1,n}$ và các X_i là độc lập, thì

$$(X_1 + X_2 + ... + X_n) \sim P(\lambda_1 + \lambda_2 + ... + \lambda_n).$$

ĐS:

$$\varphi_X(t) = e^{-\lambda(1-e^{it})};\ E(X) = D(X) = \lambda.$$

Bài 4.51.

a) Cho $X \sim H(1000, 600, 5)$. Tính $P(X \geq 1)$.

b) Cho $X \sim B(250; 0,008)$. Tính $P(X > 2)$.

c) Cho $X \sim H(10000, 4000, 200)$. Tính $P(X \leq 100)$.

Bài 4.52. Gieo một con xúc xắc 10000 lần. Ước lượng xác suất độ lệch giữa tần suất xuất hiện mặt 6 so với xác suất xuất hiện mặt 6 nhỏ hơn 0,01.

Bài 4.53. Trong một hộp có 100 bi trắng và 100 bi đen. Rút có hoàn lại 50 viên. Ước lượng trên xác suất để số bi trắng trong các bi rút ra thỏa mãn bất đẳng thức kép $15 < m < 35$.

ĐS:

$$P\left(\left|\frac{m}{50} - \frac{1}{2}\right| < \frac{1}{5}\right) \geq \frac{7}{8}.$$

Bài 4.54. Với xác suất nào để với 100 lần tung đồng tiền mặt sấp xuất hiện từ 40 đến 60 lần?

ĐS: 0,954.

Bài 4.55. Trong một hộp có 80 bi trắng và 20 bi đen. Hỏi phải rút (có hoàn lại) bao nhiêu bi từ hộp, để với xác suất 0,95 có thể tin tưởng rằng tần số xuất hiện bi trắng sẽ lệch so với xác suất nhỏ hơn 0,1?

ĐS: 61.

Bài 4.56. Xác suất một lọ thuốc bị vỡ khi vận chuyển từ nơi sản xuất đến nơi tiêu thụ là 0.01.

a) Dùng định lý xấp xỉ Poisson tính xác suất có đúng 5 lọ thuốc bị vỡ khi vận chuyển 300 lọ thuốc.

b) Phải vận chuyển bao nhiêu lọ thuốc cùng loại để xác suất có ít nhất một lọ thuốc bị vỡ không bé hơn 0.999.

Bài 4.57. Gieo 90 lần con xúc xắc cân đối đồng chất. Gọi X là số điểm sau 30 lần gieo con xúc xắc

a) Tính $E(X)$ và $D(X)$.

b) Tính xác suất $P(X > 360)$

Bài 4.58. Trong kho có 100 thùng sản phẩm, trong mỗi thùng có 90 chính phẩm và 10 phế phẩm. Mỗi thùng hàng lấy ra (có hoàn lại) kiểm tra 5 sản phẩm. Tính xác suất để tổng số phế phẩm thuộc khoảng $(40, 55)$.

Bài 4.59. Gieo 3200 lần một đồng tiền có xác suất xuất hiện mặt sấp là 0.519. Gọi X là số lần sấp xuất hiện sau 3200 lần gieo.

a) Tính số lần sấp có khả năng suất hiện nhất và xác suất tương ứng.

b) Tính xác suất $P(1600 + 5\sqrt{2} < X < 1600 + 10\sqrt{2})$.

Bài 4.60. Gieo ngẫu nhiên một con xúc xắc cân đối và đồng chất 2010 lần.

a) Gọi X là số mặt "lục" xuất hiện. Tính xác suất $P(X = 335)$.

b) Tính xác suất $P(X \in (335, 360))$.

c) Gọi Y là tổng số điểm sau 2010 lần gieo. Hãy tính $E(2Y + 5)$ và $D(4Y - 9)$.

Bài 4.61. Gieo ngẫu nhiên một đồng xu cân đối và đồng chất 2012 lần.

a) Gọi X là số mặt "sấp" xuất hiện. Tính xác suất $P(X = 1006)$.

b) Tính xác suất $P(X \in (1006, 1060))$.

c) Gọi Y là tổng số điểm sau 2012 lần gieo đồng xu. Hãy tính $E(2Y + 1)$ và $D(2Y - 3)$.

Bài 4.62. Gieo ngẫu nhiên một khối tứ diện đều cân đối và đống chất 2012 lần. Gán số điểm 1, 2, 3 và 4 cho sự xuất hiện các mặt thứ 1, 2, 3, và 4. Gọi X là tổng số điểm sau 2012 lần gieo.

a) Tính $E(X)$ và $D(X)$.

b) Tính $P(X \geq 503)$

Bài 4.63. Một khách sạn nhận đặt chổ của 325 khách hàng cho 300 phòng vào ngày 30 tháng 4 vì theo kinh nghiệm của những năm trước cho thấy tỷ lệ khách đặt chổ nhưng không đến là 10%. Tính xác suất:

a) Có 300 khách đến vào ngày 30 tháng 4 để nhận phòng.

b) Tất cả các khách đến vào ngày 30 tháng 4 đều nhận được phòng.

ĐS: $0,0281; 93,06\%$.

Bài 4.64. Trong một lô hàng có 800 sản phẩm loại I và 200 sản phẩm loại II. Lấy ngẫu nhiên đồng thời 3 sản phẩm từ lô hàng này. Gọi X là số sản phẩm loại I lấy được.

a) Tìm quy luật phân phối xác suất của X.

b) Tính $E(X)$ và $D(X)$.

c) Tính xác suất để lấy được ít nhất 1 sản phẩm loại I.

ĐS: $X \sim H(1000, 800, 3); 2,4; 0,47904; 0,9921$.

Bài 4.65. Một kiện hàng có 1000 sản phẩm, trong đó có 200 sản phẩm loại II. Một người lấy ngẫu nhiên từ kiện hàng ra 10 sản phẩm để kiểm tra. Tính xác suất để trong 10 sản phẩm lấy ra

a) Có 6 sản phẩm loại II.

b) Không quá 3 sản phẩm loại II.

c) Nhiều hơn 2 sản phẩm loại II.

ĐS: 0,0055; 0,8791; 0,3222.

Bài 4.66. Bắn 400 phát đạn vào mục tiêu một cách độc lập. Xác suất trúng đích ở mỗi lần bắn là như nhau và bằng 0,85. Tính xác suất để có từ 315 đến 335 lần bắn trúng đích.

ĐS: 0,2418.

Bài 4.67. Một nhà máy sản xuất 100000 sản phẩm, trong đó có 30000 sản phẩm loại II còn lại là sản phẩm loại I. KCS đến kiểm tra và lần lượt lấy ngẫu nhiên ra 500 sản phẩm để thử (theo phương thức hoàn lại).

a) Tính trung bình số sản phẩm loại II mà KCS phát hiện ra.

b) Tính xác suất để số sản phẩm loại II mà KCS phát hiện ra từ 145 đến 155 sản phẩm.

c) Tính xác suất để số sản phẩm loại II mà KCS phát hiện ra ít hơn 151 sản phẩm.

ĐS: 150; 0,3758; 0,5398.

Bài 4.68. Một xe tải vận chuyển 1000 chai rượu vào kho. Xác suất để mỗi chai bị vỡ trong khi vận chuyển là 0,0035. Tính xác suất để sau khi vận chuyển, có 6 chai rượu bị vỡ; có từ 2 đến 8 chai rượu bị vỡ. (giả sử rằng sự kiện các chai rượu bị vỡ là độc lập nhau, do chất lượng của mỗi chai).

Bài 4.69. Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên có trung bình 50g, độ lệch tiêu chuẩn 10g. Các sản phẩm được đóng thành hộp, mỗi hộp 100 sản phẩm. Hộp có trọng lượng trên 4,85kg là đạt tiêu chuẩn. Tính tỉ lệ hộp đạt tiêu chuẩn.

ĐS: $93,32\%$.

Bài 4.70. Với xác suất nào để với 100 lần tung đồng tiền mặt sấp xuất hiện từ 40 đến 60 lần?

Bài 4.71. Trong một hộp có 80 bi trắng và 20 bi đen. Hỏi phải rút (có hoàn lại) bao nhiêu bi từ hộp, để với xác suất 0,95 có thể tin tưởng rằng tần số xuất hiện bi trắng sẽ lệch so với xác suất nhỏ hơn 0,1?

Bài 4.72. Chứng minh rằng, nếu $f(t)$ là hàm đặc trưng, thì $|f(t)|^2$ cũng là hàm đặc trưng.

Bài 4.73. Có hai hộp cùng chứa một số lượng bi như nhau là 10 viên được đánh số từ 1 tới 10. Tiến hành 100 phép thử, trong đó mỗi phép thử là rút từ mỗi hộp 1 bi (có hoàn lại). Gọi S là tổng các số viết trên các viên bi rút ra từ hai hộp. Ước lượng trên xác suất để tổng $\sum\limits_{i=1}^{100} x_i$ rơi vào khoảng $(800, 1400)$.

Bài 4.74. Gieo một con xúc xắc 10000 lần. Ước lượng xác suất độ lệch giữa tần suất xuất hiện mặt 6 so với xác suất xuất hiện mặt 6 nhỏ hơn 0,01.

Bài 4.75. Trong một hộp có 100 bi trắng và 100 bi đen. Rút có hoàn lại 50 viên. Ước lượng trên xác suất để số bi trắng trong các bi rút ra thõa mãn bất đẳng thức kép $15 < m < 35$.

PHẦN II. THỐNG KÊ TOÁN HỌC

Tổng quan về thống kê toán học

1. Thống kê là gì?

- Thống kê là hệ thống các phương pháp dùng để thu thập, xử lý và phân tích các con số của những hiện tượng số lớn để tìm hiểu bản chất và tính quy luật vốn có của chúng trong điều kiện thời gian và không gian cụ thể.
- Thống kê là một ngành khoa học lớn, mang tính liên ngành và có phạm vi ứng dụng rất rộng lớn. Thống kê là một khoa học, một công nghệ cung cấp cho ta những công cụ hữu ích để thu thập dữ liệu, hiểu dữ liệu, tạo dữ liệu, xử lý dữ liệu nhằm rút ra từ dữ liệu những thông tin tri thức hữu ích.
- Thống kê nằm giữa trừu tượng và cụ thể, giữa lý thuyết và ứng dụng. Nó mang hương vị toán học nhưng không đơn giản là một ngành của toán học. Các bài toán cốt lõi của nó pha trộn với các bài toán của nhiều lĩnh vực nhằm đi sâu tìm hiểu bản chất của trí tuệ và tư duy.

2. Công việc của một nhà thống kê

- Thu thập và xử lý số liệu.
- Điều tra chọn mẫu.
- Nghiên cứu mối quan hệ giữa các hiện tượng.
- Dự đoán.
- Nghiên cứu các hiện tượng trong điều kiện không chắc chắn.
- Ra quyết định trong điều kiện không chắc chắn.

3. Các loại thống kê

- Thống kê mô tả: bao gồm các phương pháp thu thập số liệu, mô tả và trình bày số liệu, tính toán các tham số đặc trưng như: trung bình, phương sai, độ lệch chuẩn, mode, trung vị,...
- Thống kê suy diễn: bao gồm các phương pháp như ước lượng, kiểm định, phân tích mối quan hệ, dự đoán,... trên cơ sở các thông tin thu thập từ mẫu.

4. Các trường phái thống kê

+ Thống kê có hai yếu tố chính: dữ liệu X và quy luật θ sinh ra dữ liệu X. Chất kết dính giữa dữ liệu X và quy luật θ sẽ được biểu diễn bằng một mô hình toán học, mô tả cơ chế sinh dữ liệu X nếu ta đã biết quy luật θ .

+ Tuy nhiên, kể cả khi biết chính xác quy luật θ ta có thể vẫn không thể có được X một cách chắc chắn, do sự can thiệp của nhân tố ngẫu nhiên từ nhiều khía cạnh. Nghĩa là dữ liệu X là một biến ngẫu nhiên và dữ liệu mà ta quan sát được chỉ là một thể hiện có thể của X. Thành thử ta phải dùng ngôn ngữ xác suất trong việc biểu diễn mô hình toán học này. Mô hình toán học mô tả cơ chế sinh dữ liệu X từ quy luật θ phải là một mô hình xác suất. Dựa trên mô hình này nhà thống kê căn cứ trên dữ liệu X quan sát được sẽ cố gắng đưa ra một phán đoán, một ước lượng về quy luật X . Vì thế thống kê luôn đi song hành cùng lý thuyết xác suất, lĩnh vực toán học nghiên cứu các mô hình toán học về sự ngẫu nhiên và các phương pháp tính toán cái ngẫu nhiên. Ngôn ngữ xác suất đóng vai trò nền tảng trong các suy luận thống kê.

+ Trong khoa học thống kê có hai trường phái song hành, cạnh tranh với nhau. Đó là **trường phái tần suất** (frequentist school) và **trường phái Bayes** (Bayesian school). Sự bất đồng căn bản nhất giữa hai trường phái như sau. Với trường phái tần suất, quy luật θ nằm trong tập hợp Θ , mà ta có thể chưa biết giá trị chính xác của nó, nhưng dứt khoát rằng θ là xác định, không phải là biến ngẫu nhiên. Với trường phái Bayes, quy luật θ là một biến ngẫu nhiên lấy giá trị trong tập hợp Θ nào đấy. Đối với trường phái tần suất, người ta cần tìm một ước lượng, một suy đoán tốt nhất về θ trên cơ sở những dữ liệu quan sát được. Còn đối với trường phái Bayes, vì θ là một biến ngẫu nhiên, câu hỏi xác định giá trị của θ là vô nghĩa; thay vào đó, cần phải tìm được phân bố của θ trên cơ sở dữ liệu quan sát được. Hai cách tiếp cận này có sự khác biệt lớn mang tính triết lý khoa học sâu sắc.

+ Phần lớn các phương pháp thống kê đang được sử dụng ngày nay được phát triển từ trường phái tần suất. trường phái tần suất về cơ bản hiện đang "thắng thế" với những tên tuổi lớn như Ronald Fisher, Karl Pearson, Jerzy Neyman, Abraham Wald,... Trong những tên tuổi đó phải kể đến Karl Pearson, tác giả của phương pháp "thống kê chi-square". Thống kê chi-square được thừa nhận là một trong số 20 phát minh quan trọng nhất của thế kỉ 20. Tuy nhiên những năm gần đây trường phái Bayes đang trên đà chinh phục nhiều người, số người theo trường phái Bayes có xu hướng ngày càng tăng.

5. Các loại thang đo trong thống kê

- Stanley Stevens (1946) đã đề xuất 4 thang đo cơ bản: định danh, thứ bậc, khoảng và tỷ lệ. Thang đo định danh và thứ bậc gọi chung là thang đo định tính; thang đo khoảng và tỷ lệ gọi chung là thang đo định lượng.
- Việc sử dụng thang đo thường phụ thuộc vào phương pháp hoặc công cụ đo hơn là thuộc tính.
- Theo thứ tự, thang đo sau có chất lượng đo lường cao hơn thang đo trước, đồng thời xây dựng thang đo phức tạp hơn. Việc lựa chọn loại thang đo tùy thuộc vào đặc điểm của hiện tượng và mục đích nghiên cứu.

6. Dữ liệu trong thống kê

- Các loại dữ liệu: sơ cấp và thứ cấp; định tính và định lượng.
- Các phương pháp thu thập dữ liệu: thực nghiệm, quan sát, phỏng vấn, từ nguồn có sẵn.

7. Một số kí hiệu dùng trong tính toán thống kê

Kí hiệu	Tên gọi, ý nghĩa
μ	trung bình tổng thể
$\overline{x}$	trung bình mẫu
σ^2	phương sai tổng thể
$\widehat{s}^2$	phương sai mẫu (máy fx-570ES là phiếm $(x\sigma n)^2$)
s^2	phương sai mẫu hiệu chỉnh (máy fx-570ES là phiếm $(x\sigma n-1)^2$)
p	tỉ lệ tổng thể (tỉ lệ cần xác định)
p_0	tỉ lệ ban đầu có được do các thông tin trước đó
f	tỉ lệ mẫu (nếu không có mẫu thăm dò thì lấy $f = 0.5$)
n	kích thước mẫu
n_A	số phần tử có tính chất A trong một mẫu cụ thể.
γ	độ tin cậy ($\gamma = 95\%$): xác suất để tham số rơi vào khoảng ước lượng
α	mức ý nghĩa ($\alpha = 5\%$): xác suất của việc mắc sai lầm loại một
ε	độ chính xác (sai số)
2ε	bề rộng khoảng ước lượng
H_0	giả thiết về sự ổn định (không có sự khác biệt có ý nghĩa về mặt tk)
H_1	đối thiết
W_α	miền bác bỏ (là tập hợp các giá trị trên đường thẳng thực)

Chương 5

Lí thuyết mẫu

5.1 Tổng thể và mẫu

- Tổng thể là toàn bộ các đối tượng cần được nghiên cứu.
- Tập hợp các phần tử được lấy ra từ tổng thể để nghiên cứu gọi là mẫu.
- Số phần tử của mẫu gọi là cỡ mẫu.

5.2 Phương pháp chọn mẫu

- Chọn mẫu ngẫu nhiên: đơn giản, hệ thống, cả khối, phân tầng, nhiều giai đoạn,...
- Chọn mẫu phi ngẫu nhiên: thuận tiện, phán đoán, định ngạch,...

5.3 Khái niệm mẫu ngẫu nhiên và mẫu cụ thể

- Giả sử mẫu có kích thước n được chọn theo phương pháp ngẫu nhiên từ tổng thể có biến quan sát X. Mỗi lần chọn một phần tử của mẫu chính là thực hiện một phép thử độc lập rút ngẫu nhiên một giá trị của X từ tập giá trị của nó. Khi đó thành phần thứ i trong mẫu là biến ngẫu nhiên X_i có cùng luật phân phối của X.
- Mẫu ngẫu nhiên kích thước n là tập hợp của n biến ngẫu nhiên độc lập $X_1, X_2, ..., X_n$ được lập từ biến ngẫu nhiên X và cùng luật phân phối với X. Ký hiệu $W = (X_1, X_2, ..., X_n)$. Như vậy mẫu ngẫu nhiên kích thước n là

một vectơ ngẫu nhiên n chiều. Giả sử một giá trị của nó là $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$. Ta gọi $(x_1, x_2, ..., x_n)$ là một mẫu cụ thể kích thước n.

5.4 Hàm phân phối thực nghiệm

Giả sử $(X_1; X_2; ...; X_n)$ là một mẫu ngẫu nhiên được xây dựng từ đại lượng ngẫu nhiên X với hàm phân phối xác suất $F_X(x)$.

1. **Định nghĩa:** Hàm phân phối thực nghiệm ngẫu nhiên tương ứng với mẫu $(X_1; X_2; ...; X_n)$, kí hiệu là $F_n(x)$, xác định bởi công thức sau

$$F_n(x) = \begin{cases} 0 & \text{nếu} \quad x \leq min(X_1, X_2, ..., X_n), \\ \dfrac{k}{n} & \text{nếu} \quad \text{có k phần tử trong mẫu} < x, \\ 1 & \text{nếu} \quad x > max(X_1, X_2, ..., X_n). \end{cases}$$

   Xuất phát từ các mẫu khác nhau ta nhận được các hàm phân phối thực nghiệm khác nhau. Đồ thị của chúng đều là các hàm bậc thang. Các đường bậc thang khác nhau đều có chung một tính chất là: khi cỡ mẫu tăng vô hạn thì các hàm phân phối thực nghiệm tiệm cận đến hàm phân phối lý thuyết. Điều này thể hiện qua định lý sau

2. **Định lý Glivenko:** Nếu $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên từ một họ có hàm phân phối $F_X(x)$, thì hàm phân phối thực nghiệm $F_n(x)$ tương ứng với mẫu $(X_1; X_2; ...; X_n)$ thỏa mãn hệ thức

$$P\left(\lim_{n\to\infty} \sup_{-\infty<x<+\infty} |F_n(x) - F_X(x)| = 0\right) = 1$$

5.5 Tham số đặc trưng của mẫu

1. Tham số đặc trưng của mẫu ngẫu nhiên

   a. Trung bình mẫu: $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

   b. Phương sai mẫu: $\widehat{S}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

   c. Phương sai mẫu hiệu chỉnh: $S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

   d. Độ lệch chuẩn mẫu hiệu chỉnh: $S = \sqrt{S^2}$.

e. Tỉ lệ mẫu:

$$F_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad X_i = \left[\begin{array}{l} 0 \\ 1 \end{array}\right.$$

Chú ý: Giá trị của $\overline{X}$ là biến ngẫu nhiên, khác với giá trị của trung bình số học là một hằng số. Cách xác định trung bình số học là theo quan điểm đồng khả năng, còn cách xác định của kỳ vọng là theo xác suất, tổng quát hơn trung bình số học.

2. Tham số đặc trưng của mẫu cụ thể

Mẫu cụ thể $W_n = (x_1, x_2, ..., x_n)$, $x_1 < x_2 < ... < x_k$ và $n_1 + n_2 + ... + n_k = n$.

⋆ Bảng phân phối tần số thực nghiệm

x_i	x_1	x_2	x_3	.....	x_k
n_i	n_1	n_2	n_3	.....	n_k

⋆ Bảng phân phối tần suất thực nghiệm

x_i	x_1	x_2	x_3	.....	x_k
f_i	f_1	f_2	f_3	.....	f_k

Trong đó $f_i = \dfrac{n_i}{n}$.

a. **Trung bình mẫu**: $\overline{x} = \dfrac{1}{n}\sum\limits_{i=1}^{k} n_i x_i$.

b. **Phương sai mẫu**: $\widehat{s}^2 = \dfrac{1}{n}\left[\sum\limits_{i=1}^{k} n_i x_i^2 - n(\overline{x})^2\right]$.

c. **Phương sai mẫu hiệu chỉnh**: $s^2 = \dfrac{1}{n-1}\left[\sum\limits_{i=1}^{k} n_i x_i^2 - n(\overline{x})^2\right]$ hay $s^2 = \dfrac{n}{n-1}\widehat{s}^2$.

d. **Độ lệch chuẩn mẫu hiệu chỉnh**: $s = \sqrt{s^2}$.

e. **Tỉ lệ mẫu**: $f = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i, \quad x_i = \left[\begin{array}{l} 0 \\ 1 \end{array}\right.$ hay $f = \dfrac{n_A}{n}$.

f. **Sai số chuẩn**: $se = \dfrac{s}{\sqrt{n}}$.

g. **Mode**: Mode của chuỗi dữ liệu là giá trị của dữ liệu có nhiều phần tử nhận được giá trị đó nhất.

h. **Midian**: Midian hay còn gọi là trung vị, là giá trị nằm giữa các giá trị của chuỗi dữ liệu. Nghĩa là có 50% dữ liệu nằm phía trên và 50% dữ liệu nằm phía dưới trung vị. Nếu một chuỗi dữ liệu có n giá trị được sắp xếp từ nhỏ đến lớn, hoặc từ lớn đến nhỏ (giá trị giống nhau được xếp liền nhau) thì trung vị là số thứ $\frac{n+1}{2}$.

5.6 Kiểm tra và trình bày dữ liệu mẫu

1. Kiểm tra dữ liệu
   Khi đã thu được số liệu, không phải chúng ta lấy tất cả chúng đi xử lý mà trước tiên ta phải kiểm tra số liệu này để loại bỏ các giá trị bất thường. Khi loại bỏ được giá trị bất thường thì tập mẫu chọn ra sẽ đại diện cho tổng thể tốt hơn.

2. Dùng biểu bảng trong trình bày số liệu
   Biểu bảng thống kê là sự sắp xếp có hệ thống số liệu về các chỉ tiêu thống kê trên các hàng và cột.

3. Dùng đồ thị
   Thông thường các đồ thị sau được sử dụng phổ biến: biểu đồ tròn, biểu đồ đường, biểu đồ cột, đồ thị phân tán,...

Một số ví dụ

Ví dụ 5.1 Giả sử ta có 4 tham số đặc trưng điểm trung bình (thang điểm 4) của sinh viên nam và nữ tại một trường đại học được cho bởi bảng số liệu sau:

Tham số thống kê	Điểm trung bình của nam	Điểm trung bình của nữ
Trung bình	2.49	2.9
Độ lệch chuẩn	0.52	0.42
Giá trị lớn nhất	0.83	1.1
Giá trị nhỏ nhất	3.74	3.91

Từ bảng này ta có thể rút ra được các nhận xét sau: Điểm trung bình của nam (2.49) thấp hơn nữ (2.90) tương đối nhiều. Điểm trung bình lớn nhất của nam là 3.74 cũng thấp hơn nữ là 3.91. Điểm trung bình thấp nhất của nam là 0.83 cũng thấp hơn nữ là 1.10.độ lệch chuẩn về điểm trung bình của nam là 0.52, cao hơn nữ là 0.42 cho chúngta thấy kết quả học tập của nữ ổn định hơn nam. Như vậy tất cả tham số thống kê phản ảnh kết quả học tập của nữ tốt hơn của nam.

Ví dụ 5.2 Cho bảng phân phối thực nghiệm

x_i	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

Tính trung bình, phương sai, độ lệch tiêu chuẩn của mẫu.

Ví dụ 5.3 Lượng xăng hao phí của một ô tô đi từ A đến B sau 30 lần chạy, kết quả cho trong bảng:

Lượng xăng hao phí (lít)	9,6 - 9,8	9,8 - 10	10 - 10,2	10,2 - 10,4	10,4 - 10,6
Số lần tương ứng	3	5	10	8	4

Tính $\overline{x}, s^2$.

5.7 Bài tập có hướng dẫn

Bài 5.1 Chiều cao (cm) của một loại cây công nghiệp là BNN tuân theo luật phân phối chuẩn với trung bình là 75 và độ lệch chuẩn là 10. Người ta đo ngẫu nhiên 25 cây loại trên, tính xác suất để chiều cao trung bình của 25 cây đó nằm trong khoảng từ 71cm đến 79cm.

Giải

Cho một mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ kích thước n được lập từ biến ngẫu nhiên mang đặc trưng của tổng thể X. Nếu $X \sim \mathcal{N}(\mu, \sigma^2)$ thì trung bình mẫu $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ và do đó ta có chuẩn hóa sau

$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0, 1).$$

Sử dụng kết quả trên, gọi X là biến ngẫu nhiên cho thông tin chiều cao của loại cây công nghiệp nói trên. Theo giả thiết $X \sim \mathcal{N}(75, 10^2)$ nên trung bình mẫu 25 cây $\overline{X} \sim \mathcal{N}(75, \frac{100}{25})$ và dạng chuẩn hóa của $\overline{X}$ là $\dfrac{\overline{X}-75}{10}\sqrt{25} = \dfrac{\overline{X}-75}{2} \sim \mathcal{N}(0,1)$. Xác suất cần tìm là

$$P\left(71 < \overline{X} < 79\right) = P\left(\frac{71-75}{2} < \frac{\overline{X}-75}{2} < \frac{79-75}{2}\right) = \Phi_0(2) - \Phi_0(-2) = 0.9544.$$

<u>Bài 5.2</u> Giả sử độ tăng theo phần trăm lương hàng năm của mỗi công nhân viên chức trong công ty Alpha tuân theo luật phân phối chuẩn với trung bình 12,2% và độ lệch chuẩn 3,6%. Một mẫu ngẫu nhiên gồm 9 phần tử được chọn từ tổng thể ấy. Tìm xác suất để trung bình mẫu nhỏ hơn 10%.

Giải

Gọi X là biến ngẫu nhiên cho thông tin độ tăng theo phần trăm lương hàng năm của mỗi công nhân viên chức trong công ty Alpha. Theo giả thiết $X \sim \mathcal{N}(12,2; 3,6^2)$, với mẫu ngẫu nhiên có kích thước $n = 9$ ta có chuẩn hóa của trung bình mẫu

$$\frac{\overline{X} - 12,2}{3,6}\sqrt{9} \sim \mathcal{N}(0,1).$$

Từ đó ta tính xác suất

$$P\left(\overline{X} < 10\right) = P\left(\frac{\overline{X}-12,2}{3,6}.3 < \frac{10-12,2}{3,6}.3\right) = \Phi(-1,833).$$

Tra bảng tích phân Laplace $\Phi_0(1,833) \approx 0,4664$ suy ra

$$\Phi(-1,833) = \frac{1}{2} + \Phi_0(-1,833) = \frac{1}{2} - \Phi_0(1,833) = 0,0336.$$

Vậy $P(\overline{X} < 10) = 0,0336$.

<u>Bài 5.3</u> Giả sử $\overline{X}$ là trung bình của một mẫu kích thước n được thành lập từ tổng thể phân phối chuẩn với kỳ vọng μ và phương sai là 100. Hãy xác định n sao cho:

a) $P\left(\mu - 10 < \overline{X} < \mu + 10\right) = 0,9544$,

b) $P\left(\mu - 2 < \overline{X} < \mu + 2\right) = 0,9544$.

Giải

a. Với kích thước mẫu là n ta chuẩn hóa trung bình mẫu $\dfrac{\overline{X}-\mu}{10}\sqrt{n} \sim \mathcal{N}(0,1)$. Từ giả thiết ta có

$$P\left(\mu - 10 < \overline{X} < \mu + 10\right) = P\left(-\sqrt{n} < \frac{\overline{X}-\mu}{10}\sqrt{n} < \sqrt{n}\right) = \Phi_0(\sqrt{n}) - \Phi_0(-\sqrt{n}).$$

Tra bảng tích phân Laplace ta được $2\Phi_0(\sqrt{n}) = 0,9544$ ứng với $n = 4$.
b. Tương tự câu a ta có

$$P\left(\mu - 2 < \overline{X} < \mu + 2\right) = P\left(\frac{-\sqrt{n}}{5} < \frac{\overline{X} - \mu}{10}\sqrt{n} < \frac{\sqrt{n}}{5}\right) = \Phi_0(\frac{\sqrt{n}}{5}) - \Phi_0(\frac{-\sqrt{n}}{5}).$$

Tra bảng Laplace $2\Phi_0(\frac{\sqrt{n}}{5}) = 0,9544$ ứng với $\frac{\sqrt{n}}{5} = 2$ hay $n = 100$.

<u>Bài 5.4</u> Thống kê lượng đường cát trắng được bán ra mỗi ngày tại một cửa hàng theo đơn vị kg cho trong bảng sau

25	27,5	22	25	18	16	20	21,5	16	25
18	17,5	21,5	30	18	25	19,5	20	18,5	21

Hãy tính các đặc trưng của mẫu bao gồm: trung bình mẫu, phương sai mẫu, phương sai mẫu điều chỉnh?

Giải

Với mẫu trên ta có kích thước mẫu là $n = 20$. Ta lần lượt tính các đặc trưng của mẫu trên.
Giá trị trung bình mẫu được tính như sau

$$\overline{x} = \sum_{i=1}^{n} x_i = \frac{1}{20}(25 + 27,5 + 22 + \cdots + 21) = \frac{425}{20} = 21,25.$$

Vậy trung bình mẫu là 21,25kg.
Tính giá trị phương sai mẫu s^2

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\ &= \frac{1}{20}\left(25^2 + 27,5^2 + 25^2 + \cdots + 21^2\right) - 21,25^2 \\ &= \frac{1}{20}.9318,5 - 21,25^2 = 14,3625. \end{aligned}$$

Vậy giá trị phương sai mẫu tính được là $s^2 = 14,3625$. Từ giá trị phương sai mẫu, tính được giá trị phương sai mẫu điều chỉnh

$$\hat{s}^2 = \frac{n}{n-1}s^2 = \frac{20}{20-1}.14,3625 \approx 15,118.$$

<u>Bài 5.5</u> Lấy 18 mẫu thép của một nhà máy sản xuất thép để kiểm tra chất lượng. Kết quả kiểm tra về sức chịu lực R (đơn vị kg/cm2) như sau:

12,1	10,5	13,4	10,0	11,3	12,8	13,2	12,4	12,5
11,2	12,6	12,5	11,7	12,8	11,6	14,1	13,2	13,7

Tính trung bình mẫu, độ lệch chuẩn mẫu và độ lệch chuẩn mẫu có điều chỉnh?

Giải

Ta có khích thước mẫu là $n = 18$. Giá trị trung bình mẫu được tính như sau

$$\overline{x} = \sum_{i=1}^{n} x_i = \frac{1}{18}(12,1 + 10,5 + 13,4 + \cdots + 13,7) = \frac{221,6}{18} \approx 12,311.$$

Vậy trung bình $\overline{x} = 12,311$. Tính giá trị độ lệch chuẩn mẫu

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\ &= \frac{1}{18}\left(12,1^2 + 10,5^2 + 13,4^2 + \cdots + 13,7^2\right) - \left(\frac{221,6}{18}\right)^2 \\ &= \frac{1}{18}.2748,48 - \left(\frac{221,6}{18}\right)^2 \approx 1,130. \end{aligned}$$

Vậy độ lệch chuẩn mẫu là $s = \sqrt{1,13} \approx 1,063$.
Tính phương sai mẫu có điều chỉnh

$$\hat{s}^2 = \frac{n}{n-1}s^2 \Leftrightarrow \hat{s} = s\sqrt{\frac{n}{n-1}} = 1,063\sqrt{\frac{18}{18-1}} = 1,094.$$

Vậy độ lệch chuẩn mẫu có điều chỉnh là 1,094.

5.8 Bài tập đề nghị

Bài 5.6 Một mẫu ngẫu nhiên kích thước $n = 64$ được rút từ tổng thể phân phối chuẩn với kỳ vọng là 50 và độ lệch chẩn là 4. Tính xác suất để trung bình mẫu nằm trong khoảng từ 48,5 đến 51,5.
Bài 5.7 Giả sử S^2 là phương sai của một mẫu kích thước 6 được rút ra từ tổng thể phân phối chuẩn $\mathcal{N}(\mu; 12)$. Tính xác suất $P(2,3 < S^2 < 22,2)$.
Bài 5.8 Giả sử độ tăng theo phần trăm lương hàng năm của mỗi công nhân viên chức trong công ty Alpha tuân theo luật phân phối chuẩn với trung bình 12,2% và độ lệch chuẩn 3,6%. Một mẫu ngẫu nhiên gồm 9 phần tử được chọn từ tổng thể ấy. Tìm xác suất để trung bình mẫu nhỏ hơn 10%.
Bài 5.9 Tính các đặc trưng của mẫu được cho trong bảng số liệu sau:

SĐĐH (x_i)	0	1	2	3	4	5	Tổng
Số ngày (n_i)	5	8	6	3	2	1	n=25

<u>**Bài 5.10**</u> Để nghiên cứu về thâm niên công tác (tính tròn năm) của nhân viên ở một công ty lớn, người ta khảo sát thâm niên của 100 nhân viên được chọn ngẫu nhiên trong công ty. Kết quả như sau:

Thâm niên	5-7	8-10	11-13	14-16	17-19
Số nhân viên	8	21	36	25	10

Hãy tính giá trị trung bình mẫu và giá trị độ lệch chuẩn mẫu.
<u>**Bài 5.11**</u> Để nghiên cứu nhu cầu (kg/tháng) của một loại hàng ở một khu vực, người ta tiến hành khảo sát 500 hộ gia đình. Kết quả cho ở bảng dưới đây:

Nhu Cầu	10-12	12-14	14-16	16-18	18-20	20-22	22-24	24-26	26-28
Số hộ	25	32	47	50	80	85	95	66	20

Tính trung bình mẫu, phương sai mẫu và phương sai mẫu điều chỉnh.
<u>**Bài 5.12**</u> Để nghiên cứu chiều cao (CC) của thanh niên lứa tuổi từ 18 đến 22 tuổi ở thành phố HCM, người ta đo trên một mẫu gồm một số thanh niên (STN) được chọn ngẫu nhiên ở thành phố HCM. Kết quả như sau (đơn vị cm):

CC	[154,158)	[158,162)	[162,166)	[166,170)	[170,174)	[174,178)	[178,182)
STN	10	16	29	37	15	10	4

Tính giá trị trung bình mẫu và giá trị độ lệch chuẩn mẫu.
<u>**Bài 5.13**</u> Khảo sát về năng suất (tạ/ha) của một loại cây trồng, người ta thu được kết quả như sau:

Năng suất (tạ/ha)	40	45	50	55	60	65
Số điểm thu hoạch	2	5	4	6	7	4

Tính trung bình mẫu, phương sai mẫu, phương sai hiệu chỉnh và độ lệch chuẩn mẫu.
<u>**Bài 5.14**</u> Khảo sát lượng nước tiêu thụ X (m3/tháng) của một số hộ gia đình ở thành phố A, người ta có kết quả sau:

Lượng nước	2-4	4-6	6-8	8-10	10-12	12-15
Số hộ	5	7	19	22	6	11

Tính trung bình mẫu, phương sai mẫu, phương sai hiệu chỉnh và độ lệch chuẩn mẫu.

Bài 5.15 Quan sát thời gian (phút) cần thiết để sản xuất một chi tiết máy, người ta thu được các số liệu ở bảng sau:

Thời gian	20	25	30	35	40	45	50
Số chi tiết	12	17	8	15	14	11	9

Tính trung bình mẫu, phương sai mẫu hiệu chỉnh.

Chương 6

Lí thuyết ước lượng tham số

6.1 Giới thiệu tổng quan

- Giả sử đặc trưng của tổng thể cần nghiên cứu được biểu diễn bởi một biến ngẫu nhiên X, xác định trên không gian mẫu Ω . Có thể nói gọn là "tổng thể X". Tổng thể X có các giá trị cần biết như kỳ vọng, phương sai,... được gọi là các tham số của tổng thể (gọi tắt là tham số). Vì chúng ta không nghiên cứu trên toàn bộ tổng thể, nên các tham số này chưa được biết một cách chính xác, mà chỉ được ước tính nhờ các quan sát trên mẫu.

- **Ước lượng** là phỏng đoán một giá trị chưa biết của tổng thể dựa vào quan sát trên mẫu lấy ra từ tổng thể đó.

- Một trong những bài toán quan trọng của thống kê toán là ước lượng giá trị của một hoặc nhiều tham số tổng thể. Việc ước lượng các giá trị tham số của X dựa trên một mẫu $(X_1, X_2, ..., X_n)$ gọi là bài toán **ước lượng tham số**. Lời giải đáp cho vấn đề này có thể có dạng một giá trị duy nhất, gọi là **ước lượng điểm**, hoặc có dạng một khoảng, gọi là **ước lượng khoảng**.

- **Thống kê:** Cho mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ của tổng thể X, một hàm của biến ngẫu nhiên $T = T(X_1, X_2, ..., X_n)$ được gọi là một thống kê của X.

6.2 Ước lượng điểm

1. Định nghĩa: Bài toán tìm một thống kê $\widehat{\theta}(X_1, X_2, ..., X_n)$ để thay thế tham số θ chưa biết được gọi là bài toán ước lượng điểm của θ .

2. Các tiêu chuẩn ước lượng điểm: ước lượng không chệch, ước lượng vững, ước lượng hiệu quả,...

   a. Ước lượng không chệch: thống kê $\widehat{\theta}$ được gọi là ước lượng không chệch của tham số θ nếu $E(\theta) = \theta$. Ví dụ: $\overline{X}, S^2$ lần lượt là ước lượng không chệch của μ, σ^2 .

   b. Ước lượng vững: thống kê $\widehat{\theta}$ được gọi là ước lượng vững của tham số θ nếu với mọi ε cho trước ta có $\lim\limits_{n \to \infty} P\left(\left|\widehat{\theta} - \theta\right| < \varepsilon\right) = 1$. Ví dụ: $\overline{X}, S^2$ lần lượt là ước lượng vững của μ, σ^2 ; $\widehat{S}^2$ cũng là ước lượng vững của σ^2 .

   c. Ước lượng hiệu quả: thống kê $\widehat{\theta}$ được gọi là ước lượng hiệu quả của tham số θ nếu nó là một ước lượng không chệch và có phương sai nhỏ nhất. Ví dụ: tỉ lệ mẫu F là ước lượng hiệu quả của tỉ lệ tổng thể p; $\overline{X}$ là ước lượng hiệu quả của μ nếu X có phân phối chuẩn.

3. Các phương pháp ước lượng điểm: sử dụng các đặc trưng mẫu, ước lượng hợp lý cực đại, phương pháp Moment, phương pháp cực tiểu χ^2 , ước lượng minimax, các ước lượng Bayes,...

6.3 Ước lượng khoảng

✠ Mặc dù có nhiều tiêu chuẩn cho bài toán ước lượng điểm, nhưng ước lượng điểm dù tốt nhất cũng chỉ cho ta biết một giá trị trong tập vô hạn các giá trị, nên ta không biết được mức độ chính xác của ước lượng. Do đó, không đánh giá được mức độ sai lầm khi ta dùng $\widehat{\theta}$ thay cho θ . Để khắc phục các hạn chế đó, người ta sử dụng ước lượng khoảng tin cậy cho tham số θ . Cụ thể bài toán ước lượng khoảng cho tham số được đặt ra như sau:

Giả sử θ là một tham số của biến ngẫu nhiên X cần biết. Ước lượng khoảng tham số θ là xác định khoảng $(\theta_1; \theta_2)$ sao cho xác suất để $\theta \in (\theta_1; \theta_2)$ bằng một độ tin cậy cho trước.

✠ Thông thường 3 tham số cần được ước lượng trong thống kê là trung bình, tỷ lệ và phương sai.

✠ Trong thực tế ta chỉ ước lượng khoảng tham số với khoảng tin cậy đối xứng. Khi đó tham số θ cần ước lượng thuộc khoảng $(\theta_1; \theta_2) = (\theta_0 - \varepsilon; \theta_0 - \varepsilon)$, trong đó:

- θ_0 là ước lượng điểm của tham số ,

- ε là độ chính xác hay sai số của ước lượng.

✠ Bài toán ước lượng khoảng tổng quát
Cho xác suất $1-\alpha$. Từ mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ của tổng thể X có phân phối xác suất $f(x,\theta)$ (θ chưa biết) tìm các thống kê θ_1, θ_2 sao cho tham số θ nằm trong khoảng (θ_1, θ_2) với mức xác suất $1-\alpha$.

$$P(\theta_1 < \theta < \theta_2) = 1-\alpha$$

Các bước xây dựng khoảng ước lượng

Bước 1 Tìm một thống kê $G = G(X_1, X_2, ..., X_n)$ sao cho phân phối xác suất của nó được xác định, không phụ thuộc vào tham số.

Bước 2 Cho độ tin cậy $1-\alpha$; lấy các cặp số α_1, α_2 đều không âm sao cho $\alpha_1 + \alpha_2 = \alpha$. (thông thường đối với bài toán ước lượng khoảng đối xứng ta chọn $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$).

Ta tìm các giá trị tới hạn g_1; g_2 của G thỏa mãn:
$P(G > g_2) = \alpha_2$,
$P(G > g_1) = 1 - \alpha_1$.
Khi đó:

$$P(g_1 < G < g_2) = P(G < g_2) - P(G < g_1) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha.$$

Bước 3 Biến đổi tương đương biến cố

$$(g_1 < G < g_2) = (\theta_1 < \theta < \theta_2)$$

Vậy $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$.

6.3.1 Ước lượng khoảng tham số trung bình

Giả sử đại lượng ngẫu nhiên X có tham số trung bình $E(X) = \mu$ chưa biết. Với một mức ý nghĩa α cho trước, ta cần tìm khoảng (x_1, x_2) chứa m sao cho $P(x_1 < \mu < x_2) = 1 - \alpha$.

Chọn một mẫu cụ thể gồm n phần tử, ta được các giá trị cụ thể của X là $\{x_1, x_2, ..., x_n\}$. Gọi $\overline{x}$ là trung bình mẫu, s là phương sai mẫu điều chỉnh, khi đó bài toán ước lượng khoảng tham số trung bình được chia thành 3 trường hợp với các bước thực hiện cụ thể như sau:

1. Trường hợp 1: $\begin{cases} Var(X) = \sigma^2 & \text{đã biết} \\ n \geq 30 & \text{hoặc } (n < 30 \text{ và } X \text{ có phân phối chuẩn}) \end{cases}$

- Khoảng ước lượng cần tìm là $(x_1, x_2) = (\overline{x} - \varepsilon, \overline{x} + \varepsilon)$.
- Độ chính xác $\varepsilon = z_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ với $\Phi_0\left(z_{\frac{1-\alpha}{2}}\right) = \frac{1-\alpha}{2} = \frac{\gamma}{2}$.
  (tra bảng 2 - bảng ứng với $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x \exp\left(-\frac{t^2}{2}\right) dt$)
- Kích thước mẫu $n = \left[\frac{z^2_{\frac{1-\alpha}{2}} . \sigma^2}{\varepsilon^2}\right] + 1$.
- Khoảng tin cậy tối đa $\mu \leq \overline{x} + z_{0,5-\alpha} \frac{\sigma}{\sqrt{n}}$
  với $\Phi_0(z_{0,5-\alpha}) = 0,5 - \alpha = \gamma - 0,5$.
- Khoảng tin cậy tối thiểu $\mu \geq \overline{x} - z_{0,5-\alpha} \frac{\sigma}{\sqrt{n}}$.

Gợi ý chứng minh

Theo định lý giới hạn trung tâm ta có

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Ta đi tìm z thỏa mãn $P(-z < Z < z) = \gamma$ (1). Ta có

$$\begin{aligned} &P(-z < Z < z) = \gamma \\ &\Leftrightarrow \Phi_0\left(\tfrac{z-0}{1}\right) - \Phi_0\left(\tfrac{-z-0}{1}\right) = \gamma \\ &\Leftrightarrow \Phi_0(z) = \tfrac{\gamma}{2}. \end{aligned}$$

với $\Phi_0(z) = \frac{1}{\sqrt{2\pi}} \int\limits_0^z e^{-\frac{x^2}{2}} dx$

Đặt $\Phi_0(z_\alpha) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^{z_\alpha} e^{-\frac{x^2}{2}} dx = \alpha$. Khi đó ta suy ra $z = z_{\frac{\gamma}{2}}$ thỏa $\Phi_0\left(z_{\frac{\gamma}{2}}\right) = \frac{\gamma}{2}$

Như vậy từ (1) ta nhận được

$$P\left(-z_{\frac{\gamma}{2}} < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\gamma}{2}}\right) = \gamma \Leftrightarrow P\left(\overline{X} - z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}}\right) = \gamma$$

Kết luận: khoảng ước lượng tham số trung bình với độ tin cậy γ là

$$\left(\overline{X} - \varepsilon; \overline{X} + \varepsilon\right) \text{ với } \varepsilon = z_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

2. Trường hợp 2: $\begin{cases} Var(X) = \sigma^2 & \text{chưa biết} \\ n \geq 30 \end{cases}$

- Khoảng ước lượng cần tìm là $(x_1, x_2) = (\overline{x} - \varepsilon, \overline{x} + \varepsilon)$.
- Độ chính xác $\varepsilon = z_{\frac{1-\alpha}{2}} \frac{s}{\sqrt{n}}$
- Kích thước mẫu $n = \left[\frac{z^2_{\frac{1-\alpha}{2}} . s^2}{\varepsilon^2} \right] + 1$.
- Khoảng tin cậy tối đa $\mu \leq \overline{x} + z_{0,5-\alpha} \frac{s}{\sqrt{n}}$
  với $\Phi_0 \left(z_{0,5-\alpha} \right) = 0,5 - \alpha = \gamma - 0,5$.
- Khoảng tin cậy tối thiểu $\mu \geq \overline{x} - z_{0,5-\alpha} \frac{s}{\sqrt{n}}$.

<u>Gợi ý chứng minh</u>

Theo định lý giới hạn trung, ta có

$$Z = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0, 1)$$

ta chứng minh tương tự trường 1.

3. Trường hợp 3: $\begin{cases} Var(X) = \sigma^2 & \text{chưa biết} \\ n < 30 & \text{và } X \text{ có phân phối chuẩn} \end{cases}$

- Khoảng ước lượng cần tìm là $(x_1, x_2) = (\overline{x} - \varepsilon, \overline{x} + \varepsilon)$.
- Độ chính xác $\varepsilon = t_{\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}}$ với $t_{\frac{\alpha}{2}}(n-1)$ là phân vị Student với $n-1$ bậc tự do mức xác suất $\frac{\alpha}{2}$.
  (tra bảng 4 - bảng ứng với $P\left(T > t_\alpha (n-1)\right) = \alpha$)
- Kích thước mẫu $n = \left[\frac{t^2_{\frac{\alpha}{2}} (n-1) . s^2}{\varepsilon^2} \right] + 1$
- Khoảng tin cậy tối đa $\mu \leq \overline{x} + t_\alpha (n-1) \frac{s}{\sqrt{n}}$.
- Khoảng tin cậy tối thiểu $\mu \geq \overline{x} - t_\alpha (n-1) \frac{s}{\sqrt{n}}$.

<u>Gợi ý chứng minh</u>

Ta đi xác định phân phối của $T = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$?

Ta có

$$\begin{cases} Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \\ V = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2 (n-1) \end{cases} \Rightarrow \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{Z\sqrt{n-1}}{\sqrt{V}} \sim t(n-1).$$

Vậy $T = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim t\,(n-1)\,.$
Ta đi tìm t thỏa $P(-t < T < t) = \gamma\,(2)$.
Ta có $P\,(-t < T < t) = \gamma \Leftrightarrow 1 - 2P\,(T > t) = \gamma \Leftrightarrow P\,(T > t) = \dfrac{1-\gamma}{2}$.
Đặt $P(T > t_\alpha) = \alpha$. Khi đó ta suy ra $t = t_{\frac{1-\gamma}{2}}$ thỏa $P\left(T > t_{\frac{1-\gamma}{2}}\right) = \dfrac{1-\gamma}{2}$.
Như vậy từ (2) ta nhận được

$$P\left(-t_{\frac{1-\gamma}{2}}\,(n-1) < \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} < t_{\frac{1-\gamma}{2}}\,(n-1)\right) = \gamma$$

$$\Leftrightarrow P\left(-t_{\frac{1-\gamma}{2}}\,(n-1)\,\frac{s}{\sqrt{n}} < \mu < t_{\frac{1-\gamma}{2}}\,(n-1)\,\frac{s}{\sqrt{n}}\right) = \gamma.$$

Kết luận: khoảng ước lượng tham số trung bình với độ tin cậy γ là

$$\left(\overline{X} - \varepsilon; \overline{X} + \varepsilon\right) \text{ với } \varepsilon = t_{\frac{\alpha}{2}}\,(n-1)\,\frac{s}{\sqrt{n}}$$

Chú ý: thực ra ở trường hợp thứ 3 khi mà **phương sai tổng thể chưa biết** và X **có phân phối chuẩn** thì ta hoàn toàn có đủ cơ sở để giải quyết bài toán theo "dạng T" mà không cần quan tâm đến kích thước mẫu n. Tuy nhiên, khi $n \geq 30$ thì theo định lý giới hạn trung tâm ta sẽ giải quyết bài toán theo "dạng Z" như trường hợp 2 sẽ thuận lợi hơn.

6.3.2 Ước lượng khoảng tham số tỉ lệ

Giả sử tổng thể có hai loại phần tử: loại phần tử có tính chất A và loại phần tử không có tính chất A với tỷ lệ phần tử có tính chất A là p chưa biết. Với một độ tin cậy $1-\alpha$ cho trước, ta cần tìm khoảng (f_1, f_2) chứa p sao cho $P(f_1 < p < f_2) = 1-\alpha$.

Chọn một mẫu gồm n phần tử. Gọi f là tỉ lệ phần tử có tính chất A của mẫu. Việc ước lượng tham số p được tính toán như sau:

- Khoảng ước lượng cần tìm là $(f_1, f_2) = (f-\varepsilon, f+\varepsilon)$.
- Độ chính xác $\varepsilon = z_{\frac{1-\alpha}{2}}\sqrt{\dfrac{f(1-f)}{n}}$.
- Kích thước mẫu $n = \left[\dfrac{z^2_{\frac{1-\alpha}{2}} f\,(1-f)}{\varepsilon^2}\right] + 1$.
- Khoảng tin cậy tối đa $0 \leq p \leq f + z_{0{,}5-\alpha}\sqrt{\dfrac{f\,(1-f)}{n}}$

$*$ Khoảng tin cậy tối thiểu $f - z_{0,5-\alpha}\sqrt{\dfrac{f\,(1-f)}{n}} \le p \le 1$

<u>Gợi ý chứng minh</u>

Giả sử $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên của tổng thể X có tỉ lệ p và F là tỉ lệ mẫu của X.

+ Theo định lý giới hạn trung tâm, ta có:

$$G = \frac{F-p}{\sqrt{\dfrac{p\,(1-p)}{n}}} \xrightarrow{d} Z \sim N\,(0,1) \qquad (*)$$

+ Từ (*) với n đủ lớn thì phân phối của G sẽ xấp xỉ phân phối chuẩn tắc và với độ tin cậy $1-\alpha$, ta tìm các giá trị tới hạn chuẩn $z_{\frac{1-\alpha}{2}}$, $-z_{\frac{1-\alpha}{2}}$ thỏa mãn:

$P(G > z_{\frac{1-\alpha}{2}}) = \frac{\alpha}{2}$,
$P(G > -z_{\frac{1-\alpha}{2}}) = 1 - \frac{\alpha}{2}$.

Khi đó

$$P\left(-z_{\frac{1-\alpha}{2}} < G < z_{\frac{1-\alpha}{2}}\right) = P\left(G < z_{\frac{1-\alpha}{2}}\right) - P\left(G < -z_{\frac{1-\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

+ Ta có

$$P\left(-z_{\frac{1-\alpha}{2}} < G < z_{\frac{1-\alpha}{2}}\right) = 1 - \alpha$$

$$\Leftrightarrow \quad P\left(-z_{\frac{1=\alpha}{2}} < \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\frac{1-\alpha}{2}}\right) = 1 - \alpha$$

Khi n đủ lớn ta thay p (ở mẫu số) bằng ước lượng điểm của nó là F, ta có

$$P\left(F - z_{\frac{1-\alpha}{2}}\sqrt{\frac{F\,(1-F)}{n}} < p < F + z_{\frac{1-\alpha}{2}}\sqrt{\frac{F\,(1-F)}{n}}\right) = 1 - \alpha$$

Với mẫu cụ thể $(x_1, x_2, ..., x_n)$; f là tỉ lệ mẫu cụ thể; khi $\begin{cases} n \ge 30 \\ nf \ge 5 \\ n\,(1-f) \ge 5 \end{cases}$ ta có khoảng tin cậy đối xứng cho p là

$$P\left(f - z_{\frac{1-\alpha}{2}}\sqrt{\frac{p\,(1-p)}{n}} < p < f + z_{\frac{1-\alpha}{2}}\sqrt{\frac{p\,(1-p)}{n}}\right) = 1 - \alpha$$

6.3.3 Ước lượng khoảng tham số phương sai

a) Bài toán

Giả sử biến ngẫu nhiên X có phân phối chuẩn với phương sai σ^2 chưa biết. Từ tổng thể ta chọn ra một mẫu ngẫu nhiên kích thước n. Với độ tin cậy $1-\alpha$ cho trước, ta cần xác định khoảng (σ_1^2, σ_2^2) sao cho $P(\sigma_1^2 < \sigma^2 < \sigma_2^2) = 1-\alpha$.

b) Các bước thực hiện

i. Trường hợp 1: biết trung bình μ
Khoảng ước lượng của phương sai cần tìm là

$$\left(\frac{\sum\limits_{i=1}^{n} (x_i - \mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)} ; \frac{\sum\limits_{i=1}^{n} (x_i - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \right)$$

Trong đó $\chi^2_{\frac{\alpha}{2}}(n), \chi^2_{1-\frac{\alpha}{2}}(n)$ là phân vị Khi bình phương bậc tự do n mức xác suất α và $1 - \dfrac{\alpha}{2}$, được xác định từ bảng phụ lục 5.

ii. Trường hợp 2: chưa biết trung bình μ
Khoảng ước lượng của phương sai cần tìm là

$$\left(\frac{(n-1)\,s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} ; \frac{(n-1)\,s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \right)$$

Một số ví dụ

Ví dụ 6.1 Khảo sát điểm thi ngoại ngữ đầu vào (thang điểm 150) ở một trường đại học, người ta chọn một mẫu 100 sinh viên được kết quả cụ thể sau:

Điểm thi	Số sinh viên
70-80	10
80-90	15
90-100	25
100-110	30
110-120	15
120-130	5

a) Ước lượng điểm thi trung bình của sinh viên với độ tin cậy 90

b) Trong ước lượng trên nếu độ chính xác của ước lượng trên là 2,5 điểm, thì độ tin cậy của ước lượng này là bao nhiêu?

c) Giả sử những sinh viên đạt điểm từ 110 trở lên thì được xếp nhóm I, và điểm thi có phân phối chuẩn. Ước lượng điểm thitrung bình của sinh viên loại I với độ tin cậy 95%.

ĐS: a) (96,82; 101,18); b) 94%; c) (115,42; 119,58).

Ví dụ 6.2 . Chiều cao của thanh niên tuân theo luật phân phối chuẩn với độ lệch chuẩn là 5 cm. Cần phải lấy một mẫu bằng bao nhiêu để với độ tin cậy 95%, sai số của ước lượng cho chiều cao trung bình không vượt quá 0,6 cm.
ĐS: tối thiểu 267 thanh niên.

Ví dụ 6.3 Tại một vùng rừng nguyên sinh, người ta đeo vòng cho 1000 con chim. Sau một thời gian bắt lại 200 con thì thấy 40 con có đeo vòng. Bạn hãy ước lượng số chim trong rừng này với độ tin cậy 99%.
ĐS: (3664; 7868).

Ví dụ 6.4 Để ước lượng tỷ lệ trẻ em bị mắc bệnh sâu răng ở một huyện vùng nông thôn với mức ý nghĩa 5%, sai số không vượt quá 2% thì cần khám bao nhiêu đứa trẻ, biết rằng tỷ lệ sâu răng thực nghiệm là 0,8.
ĐS: tối thiểu 1537 đứa trẻ.

Ví dụ 6.5 . Mức qui định độ lệch chuẩn trọng lượng của sản phẩm là 5mg. Kiểm tra ngẫu nhiên 20 sản phẩm, ta có phương sai điều chỉnh là 10. Bằng cách ước lượng phương sai, hãy cho biết độ lệch chuẩn trọng lượng sản phẩm này có cao hơn mức qui định không.
ĐS: (2,04; 4,62).

Ví dụ 6.6 Để đánh giá chất lượng sản phẩm của một nhà máy, người ta kiểm tra ngẫu nhiên 600 sản phẩm của nhà máy này sau ca sản xuất, thấy có 50 phế phẩm. Với độ tin cậy 95%.

a) Hãy ước lượng tỉ lệ phế phẩm của nhà máy sau ca sản xuất.

b) Hãy ước lượng số phế phẩm của nhà máy sau ca sản xuất ; biết rằng mỗi ca nhà máy sản xuất được 5500 sản phẩm.

ĐS: a) (0,0612; 0,1054); b) (337; 579).

Ví dụ 6.7 Sử dụng ví dụ 6.6.

a) Nếu sử dụng mẫu điều tra để ước lượng tỉ lệ phế phẩm trong một ca sản xuất của nhà máy đạt độ chính xác là 2% thì đảm bảo độ tin cậy là bao nhiêu ?

b) Nếu muốn ước lượng tỉ lệ phế phẩm của nhà máy trong một ca sản xuất đạt độ tin cậy là 96% và độ chính xác là 1,8% thì cần điều tra thêm ít nhất bao nhiêu sản phẩm nữa.

ĐS: a) 0,92328; b) 401.

Ví dụ 6.8 Sử dụng ví dụ 6.6. với độ tin cậy 95% hãy ước lượng tỉ lệ phế phẩm tối đa của một ca sản xuất của nhà máy. ĐS: 10,19%.

Ví dụ 6.9 Chủ một kho cung cấp sơn muốn ước lượng lượng sơn chứa trong một thùng được sản xuất từ một công nghệ dây chuyền quốc gia. Biết rằng theo tiêu chuẩn của dây chuyền công nghệ đó, độ lệch tiêu chuẩn của lượng sơn là 0,08 thùng. Điều tra một mẫu 50 thùng tính được lượng sơn trung bình chứa trong một thùng là 0,97 (thùng). Với độ tin cậy 99%, hãy ước lượng:

a) Lượng sơn trung bình chứa trong một thùng.

b) Lượng sơn trung bình tối thiểu chứa trong một thùng.

ĐS: a) (0,9408; 0,9992); b) 0,9436.

Ví dụ 6.10 Giá bán của một loại thiết bị (đơn vị USD) trên thị trường là biến ngẫu nhiên có phân phối chuẩn. một người định mua một thiết bị loại này, khảo sát ngẫu nhiên tại 18 cửa hàng tính được giá trung bình của thiết bị là 137,75 USD và độ lệch tiêu chuẩn 7,89 USD. Với độ tin cậy 90%, hãy ước lượng giá bán trung bình của thiết bị.
ĐS: (134,4772; 141,0228).

Ví dụ 6.11 Mức hao phí nguyên liệu X để sản xuất một sản phẩm của nhà máy (tính bằng kilôgam) tuân theo quy luật chuẩn. Mẫu điều tra về mức hao phí nguyên liệu để sản xuất 25 sản phẩm loại này cho kết quả trong bảng :

Mức hao phí nguyên liệu hao phí (kg)	19,5	20,0	20,5
Số sản phẩm	5	18	2

Với độ tin cậy 95%.

1) Khi không biết kì vọng $E(X)$, hãy:

  a) Ước lượng phương sai mức hao phí nguyên liệu để sản xuất một sản phẩm của nhà máy.

b) Ước lượng tối đa cho phương sai của mức hao phí nguyên liệu để sản xuất một sản phẩm của nhà máy.

2) Khi biết kì vọng $E(X) = 20$, hãy ước lượng phương sai mức hao phí nguyên liệu để sản xuất một sản phẩm của nhà máy.

6.4 Bài tập có hướng dẫn

Ước lượng trung bình tổng thể

Bài 6.1 Giả sử đại lượng ngẫu nhiên $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với μ chưa biết, còn σ^2 đã biết. Hãy xây dựng khoảng ước lượng cho μ với độ tin cậy $\gamma \in (0,1)$ cho trước.
Lưu ý: Ta sử dụng bảng tích phân Laplace

$$\Phi_0(z_\gamma) = \frac{1}{\sqrt{2\pi}} \int_0^{z_\gamma} \exp\left(-\frac{x^2}{2}\right) dx = \gamma.$$

Giải

Để ước lượng μ ta lấy ngẫu nhiên mẫu $(X_1, X_2, \ldots, X_n)$ và xét thống kê $\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n}\right) \sim \mathcal{N}(0,1)$. Với độ tin cậy $\gamma \in (0,1)$ cho trước ta có thể tìm được số c thỏa

$$P\left(-c < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < c\right) = \gamma \tag{6.1}$$

hay

$$P\left(\overline{X} - c\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + c\frac{\sigma}{\sqrt{n}}\right) = \gamma.$$

Mặt khác (6.1) tương đương với

$$\frac{1}{\sqrt{2\pi}} \int_{-c}^{c} \exp\left(-\frac{x^2}{2}\right) dx = \gamma \Leftrightarrow \frac{1}{\sqrt{2\pi}} \int_0^{c} \exp\left(-\frac{x^2}{2}\right) dx = \frac{\gamma}{2}.$$

Theo bảng tích phân Laplace ta xác định được $c = z_{\frac{\gamma}{2}}$. Từ đó ta xác định được khoảng ước lượng cho trung bình μ với độ tin cậy γ là

$$\left(\overline{X} - z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}\right).$$

Bài 6.2 Một mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ kích thước n sinh ra từ đại lượng ngẫu nhiên $X \sim \mathcal{N}(\mu, \sigma_0^2)$, với σ_0 đã biết. Xét bài toán ước lượng trung bình tổng thể μ với độ tin cậy γ ở bài tập 1.

a) Giả sử bán kính của khoảng ước lượng cho μ (hay gọi là độ chính xác của ước lượng) với độ tin cậy γ là ϵ_0 cho trước, hãy xác định kích thước mẫu ngẫu nhiên cần quan sát?

b) Cho trước bán kính ước lượng cho μ là ϵ_0 và kích thước mẫu ngẫu nhiên quan sát là n_0, hãy tìm độ tin cậy γ cho khoảng ước lượng đó?

Giải

Sử dụng các thông tin trong bài tập 1, từ khoảng ước lượng trung bình μ với độ tin cậy γ ta xác định bán kính của ước lượng này là

$$\frac{1}{2}\left[\overline{X} + z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}} - \left(\overline{X} - z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}\right)\right] = z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}.$$

a. Theo giả thiết ta có

$$z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}} = \epsilon_0.$$

Từ đó xác định được kích thước mẫu là

$$n = \frac{z^2_{\frac{\gamma}{2}}\sigma^2}{\epsilon_0^2}.$$

b. Tương tự câu a ta cũng có

$$z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n_0}} = \epsilon_0.$$

Ta xác định độ tin cậy thông qua $z_{\frac{\gamma}{2}}$ cho bởi

$$z_{\frac{\gamma}{2}} = \frac{\epsilon_0\sqrt{n_0}}{\sigma}.$$

<u>Bài 6.3</u> Giả sử đại lượng ngẫu nhiên $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với μ và σ^2 chưa biết. Hãy xây dựng khoảng ước lượng cho μ với độ tin cậy $\gamma \in (0,1)$ cho trước.
Lưu ý: Ta sử dụng bảng giá trị $t_\gamma^{(k)}$ cho bởi

$$\int_{t_\gamma(k)}^{+\infty} f_k(x)dx = \gamma,$$

trong đó f_k là hàm mật độ của phân phối Student với bậc tự do là k.

Giải

Chọn mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ từ đại lượng ngẫu nhiên X và xét thống kê $\left(\frac{\overline{X} - \mu}{\hat{S}_n}\sqrt{n}\right) \sim t(n-1)$. Với độ tin cậy $\gamma \in (0,1)$ cho trước ta có thể tìm được số c thỏa

$$P\left(-c < \frac{\overline{X} - \mu}{\hat{S}_n}\sqrt{n} < c\right) = \gamma \tag{6.2}$$

hay

$$P\left(\overline{X} - c\frac{\hat{S}_n}{\sqrt{n}} < \mu < \overline{X} + c\frac{\hat{S}_n}{\sqrt{n}}\right) = \gamma.$$

Mặt khác (6.2) tương đương với

$$\int_{-c}^{c} f_{n-1}(x)dx = \gamma \Leftrightarrow \int_{c}^{+\infty} f_{n-1}(x)dx = \frac{1-\gamma}{2}.$$

Do đó $c = t_{\frac{1-\gamma}{2}}(n-1)$. Từ đó ta xác định được khoảng ước lượng cho trung bình μ với độ tin cậy γ là

$$\left(\overline{X} - t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{S}_n}{\sqrt{n}}, \overline{X} + t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{S}_n}{\sqrt{n}}\right).$$

Bài 6.4 Giả sử đại lượng ngẫu nhiên $X \sim B_1(p)$, $p \in (0,1)$. Hãy xây dựng khoảng ước lượng đối xứng cho p với độ tin cậy là γ.

Giải

Tiến hành chọn mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$. Từ mẫu này ta xác định được k thành công và tần suất $f_n = \dfrac{k}{n}$ là một ước lượng điểm cho p. Khi đó $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = f_n$, theo định lý giới hạn trung tâm Moivres - Laplace $\dfrac{f_n - p}{\sqrt{p(1-p)}}\sqrt{n}$ có hàm phân phối hội tụ về phân phối chuẩn tắc $\mathcal{N}(0,1)$.

Do đó, chọn kích thước mẫu đủ lớn ta có thống kê $\dfrac{f_n - p}{\sqrt{p(1-p)}}\sqrt{n} \sim \mathcal{N}(0,1)$. Với độ tin cậy $\gamma \in (0,1)$ cho trước ta có thể tìm được số c thỏa

$$P\left(-c < \frac{f_n - p}{\sqrt{p(1-p)}}\sqrt{n} < c\right) = \gamma, \tag{6.3}$$

hay

$$P\left(f_n - c\sqrt{\frac{p(1-p)}{n}} < p < f_n + c\sqrt{\frac{p(1-p)}{n}}\right) = \gamma.$$

Thay ước lượng điểm của p là f_n vào hai cận trên, dưới ta được

$$P\left(f_n - c\sqrt{\frac{f_n(1-f_n)}{n}} < p < f_n + c\sqrt{\frac{f_n(1-f_n)}{n}}\right) = \gamma.$$

Dựa vào bảng tích phân Laplace, từ (6.3) xác định $c = z_{\frac{\gamma}{2}}$. Do đó khoảng ước lượng đối xứng cho p với độ tin cậy γ là

$$\left(f_n - z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}, f_n + z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}\right).$$

Bài 6.5 Một mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ kích thước n lập từ đại lượng ngẫu nhiên $X \sim B_1(p)$. Xét bài toán khoảng ước lượng cho p với độ tin cậy γ ở bài tập ?.

a) Cho trước bán kính ước lượng p là ϵ_0 và kích thước mẫu ngẫu nhiên quan sát là n_0, hãy tìm độ tin cậy γ cho khoảng ước lượng đó?

b) Giả sử bán kính của khoảng ước lượng cho p với độ tin cậy γ là ϵ_0 cho trước, hãy xác định kích thước mẫu ngẫu nhiên cần quan sát trong hai trường hợp biết tần suất mẫu và không biết tần suất mẫu?

Lưu ý: Tần suất mẫu có thể có được thông qua việc lấy mẫu thăm dò lần đầu.

Giải

Theo khoảng ước lượng cho p với độ tin cậy γ được xây dựng ở bài tập ? ta có bán kính của ước lượng là

$$\epsilon = z_{\frac{\gamma}{2}} \sqrt{\frac{f_n(1-f_n)}{n}}.$$

a. Với bán kính ước lượng là ϵ_0, kích thước mẫu n_0 cho trước, độ tin cậy γ được xác định qua $z_{\frac{\gamma}{2}}$ như sau

$$z_{\frac{\gamma}{2}} = \epsilon_0 \sqrt{\frac{n_0}{f_{n_0}(1-f_{n_0})}}.$$

b. Với bán kính ước lượng là ϵ_0 cho trước, trong trường hợp có giả thiết tần suất mẫu f_n đã biết ta có kích thước n_0 được xác định như sau

$$n_0 = z^2_{\frac{\gamma}{2}} \frac{f_{n_0}(1-f_{n_0})}{\epsilon_0^2}.$$

Trong trường hợp chưa biết tần suất mẫu quan sát, sử dụng bất đẳng thức Cauchy ta có đánh giá sau

$$\epsilon = z_{\frac{\gamma}{2}} \sqrt{\frac{f_n(1-f_n)}{n}} \le z_{\frac{\gamma}{2}} \frac{1}{2\sqrt{n}}.$$

Từ đây ta chấp nhận giá trị $n_0 = \dfrac{z^2_{\frac{\gamma}{2}}}{4\epsilon_0^2}$ là một đánh giá thô để bán kính của ước lượng cho p không vượt qua ϵ_0.

Bài 6.6 Giả sử khối lượng của một nam sinh viên năm thứ nhất trường đại học A tuân theo luật phân phối chuẩn với độ lệch chuẩn 3kg. Chọn ngẫu nhiên 25 nam sinh viên năm thứ nhất, người ta tính được khối lượng trung bình là 52kg.

a) Hãy tìm khoảng tin cậy 95% cho khối lượng trung bình của một nam sinh viên năm thứ nhất trường đại học A.

b) Với mẫu trên, nếu muốn bề rộng của khoảng ước lượng trung bình tổng thể là 1,8kg thì độ tin cậy là bao nhiêu?

Giải

a. Gọi X là biến ngẫu nhiên cho thông tin khối lượng của mỗi nam sinh viên năm thứ nhất ở trường Đại học A, $X \sim \mathcal{N}(\mu, 3^2)$. Bài toán ước lượng trung bình μ khi biết phương sai.

Khoảng tin cậy $\gamma = 95\%$ cho μ là

$$\left(\overline{x} - z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}\right),$$

thay giá trị cụ thể ta được

$$\left(52 - z_{\frac{0,95}{2}}\frac{3}{\sqrt{25}}, 52 + z_{\frac{0,95}{2}}\frac{3}{\sqrt{25}}\right).$$

Tra bảng tích phân Laplace $z_{0,475} = 1,96$. Khoảng ước lượng cần tìm là $(50,824; 53,176)$.

b. Bề rộng của khoảng ước lượng trung bình tổng thể được cho bởi

$$\overline{x} + z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}} - \left(\overline{x} - z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}\right) = 2z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}.$$

Theo giả thiết ta có

$$2z_{\frac{\gamma}{2}}\frac{3}{\sqrt{25}} = 1,8 \Leftrightarrow z_{\frac{\gamma}{2}} = 1,5.$$

Tra bảng tích phân Laplace ta có $\dfrac{\gamma}{2} = 0,4332$ hay $\gamma = 86,64\%$.

Bài 6.7 Biết rằng chiều cao của các thanh niên cùng một lứa tuổi tuân theo luật phân phối chuẩn. Khảo sát ngẫu nhiên chiều cao của 121 thanh niên cùng lứa tuổi đó, người ta tính được chiều cao trung bình là 162cm và độ lệch chuẩn có điều chỉnh là 14cm. Hãy ước lượng chiều cao trung bình của mỗi thanh niên ở lứa tuổi trên bằng khoảng tin cậy 95%.

Giải

Gọi X là biến ngẫu nhiên chỉ chiều cao của mỗi thanh niên, X có phân phối chuẩn với kì vọng μ và phương sai σ^2 đều chưa biết. Bài toán ước lượng trung bình μ khi không biết phương sai.

Ta đã biết giá trị trung bình mẫu $\overline{x} = 162$ và độ lệch chuẩn mẫu điều chỉnh là $\hat{s} = 14$. Từ đó xác định được khoảng ước lượng với độ tin cậy 95% cho μ là

$$\left(\overline{x} - t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{s}}{\sqrt{n}}, \overline{x} + t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{s}}{\sqrt{n}}\right)$$

hay

$$\left(162 - t_{\frac{1-0,95}{2}}(121-1).\frac{14}{\sqrt{121}}; 162 + t_{\frac{1-0,95}{2}}(121-1).\frac{14}{\sqrt{121}}\right).$$

Thực hiện tính toán cụ thể với $t_{0,025}(120) = 1,980$ ta thu được khoảng ước lượng chiều cao trung bình của mỗi thanh niên bằng khoảng tin cậy 95% là

$$(159,480; 164,520).$$

Chú ý: Trong trường hợp này kích thước mẫu khá lớn ($n = 121 > 30$) nên ta cũng có thể dùng ước lượng

$$\left(\overline{x} - z_{\frac{\gamma}{2}}\frac{\hat{s}}{\sqrt{n}}, \overline{x} + z_{\frac{\gamma}{2}}\frac{\hat{s}}{\sqrt{n}}\right)$$

hay

$$\left(162 - z_{\frac{0,95}{2}}.\frac{14}{\sqrt{121}}; 162 + z_{\frac{0,95}{2}}.\frac{14}{\sqrt{121}}\right).$$

Tra bảng tích phân Laplace $z_{0,475} = 1,96$. Từ đó xác định được khoảng ước lượng cho μ là

$$(159,505; 164,495).$$

<u>Bài 6.8</u> Một nông trại, người ta nhập về nuôi thí nghiệm một giống heo mới. Qua thời gian nuôi 4 tháng, cân thử 100 con ta có bảng số liệu:

Trọng lượng (kg)	Số con
70-80	10
80-90	15
90-100	25
100-110	30
110-120	15
120-130	5

a) Ước lượng trọng lượng trung bình của những con heo giống mới sau 4 tháng nuôi với độ tin cậy 95%.

b) Trong ước lượng trên nếu độ chính xác của ước lượng trên là 2.5kg, thì độ tin cậy của ước lượng này là bao nhiêu?

c) Giả sử những con heo có trọng lượng từ 110kg trở lên thì được xếp loại I, và trọng lượng của nó có phân phối chuẩn. Ước lượng trọng lượng trung bình những con heo loại I với độ tin cậy 95%.

Giải

Gọi X biến ngẫu nhiên chỉ trọng lượng mỗi con heo giống mới mà nông trại mới nhập về sau thời gian 4 tháng nuôi, X có trung bình μ, phương sai σ^2 và phân

phối xác suất chưa biết.
a. Bài toán ước lượng trung bình μ trong trường hợp phương sai σ^2 chưa biết. Với mẫu quan sát có kích thước khá lớn ($n = 100 > 30$) thì có thể bỏ qua giả thiết về tổng thể có phấn phối chuẩn.
Với dữ liệu mẫu trên ta tính được giá trị trung bình mẫu $\overline{x} = 99$ và phương sai mẫu có điều chỉnh $\hat{s} = 13,257$. Khoảng ước lượng trọng lượng trung bình với độ tin cậy 95% được cho bởi

$$\left(\overline{x} - z_{\frac{\gamma}{2}}\frac{\hat{s}}{\sqrt{n}}, \overline{x} + z_{\frac{\gamma}{2}}\frac{\hat{s}}{\sqrt{n}}\right)$$

hay

$$\left(99 - z_{\frac{0.95}{2}}\frac{13,257}{\sqrt{100}}; 99 + z_{\frac{0.95}{2}}\frac{13,257}{\sqrt{100}}\right).$$

Tra bảng tích phân Laplace ta được $z_{0,475} = 1,960$. Tính toán ta được khoảng ước lượng trên là

$$(96,402; 101,598).$$

b. Độ chính xác 2,5kg của ước lượng với độ tin cậy γ được cho bởi

$$2,5 = z_{\frac{\gamma}{2}}\frac{13,257}{\sqrt{100}} \Leftrightarrow z_{\frac{\gamma}{2}} = 1,886.$$

Tra bảng tích phân Laplace ta tìm được giá trị gần bằng $\frac{\gamma}{2} = 0,4706$ hay $\gamma = 94,12\%$. c.Kích thước mẫu của heo loại I là $n = 20$, trọng lượng tuân theo phân phối chuẩn, phương sai chưa biết bài toán yêu cầu ước lượng trọng lượng trung bình với độ tin cậy 95%. Do đó, ta dùng khoảng ước lượng cho trung bình là

$$\left(\overline{x} - t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{s}}{\sqrt{n}}, \overline{x} + t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{s}}{\sqrt{n}}\right).$$

Từ mẫu của heo loại I ta tính được giá trị trung bình mẫu $\overline{x} = 117,5$, phương sai mẫu điều chỉnh là $\hat{s} = 4,443$ và tra bảng giá trị Student $t_{0,025}(19) = 2.093$. Từ đây xác định khoảng ước lượng trọng lượng trung bình những con heo loại I với độ tin cậy 95%

$$\left(117,5 - t_{\frac{1-0,95}{2}}(20-1)\frac{4,443}{\sqrt{20}}, 117,5 + t_{\frac{1-0,95}{2}}(20-1)\frac{4,443}{\sqrt{20}}\right)$$

hay

$$\left(117,5 - t_{0,025}(19)\frac{4,443}{\sqrt{20}}, 117,5 + t_{0,025}(19)\frac{4,443}{\sqrt{20}}\right) = (115,421; 119,580).$$

<u>Bài 6.9</u> Biết chiều cao của những cây bạch đàn trồng một năm tuổi là đại lượng ngẫu nhiên có phân phối chuẩn với độ lệch tiêu chuẩn là σ. Khảo sát chiều cao của 25 cây ta có số liệu sau:

Chiều cao (m)	0.8	0.9	1.0	1.1	1.2
Số cây	2	4	8	6	5

a) Ước lượng chiều cao trung bình của cây bạch đàn với độ tin cậy 95%, cho biết $\sigma = 0.122$.

b) Ước lượng chiều cao trung bình của cây bạch đàn với độ tin cậy 99% trong trường hợp không biết giá trị của σ.

Giải

Gọi X là biến ngẫu nhiên chỉ chiều cao của những cây bạch đàn trồng một năm tuổi. Với dữ liệu mẫu trên ta tính được trung bình mẫu $\overline{x} = 1,032$ và phương sai mẫu có điều chỉnh $\hat{s} = 0,122$ và mẫu có kích thước $n = 25$.

a. Theo giả thiết $X \sim \mathcal{N}(\mu; 0,122^2)$, ta cần ước lượng trung bình μ. Độ tin cậy $\gamma = 0,95$ tra bảng tích phân Laplace ta có $z_{\frac{0,95}{2}} = z_{0,475} = 1,96$. Khoảng chiều cao trung bình của những cây bạch đàn với độ tin cậy 95% là

$$\left(\overline{x} - z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\frac{\gamma}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

hay

$$\left(1,032 - z_{\frac{0,95}{2}}\frac{0,122}{\sqrt{25}}; 1,032 + z_{\frac{0,95}{2}}\frac{0,122}{\sqrt{25}}\right) = (0,984; 1,080).$$

b. Với giả thiết $X \sim \mathcal{N}(\mu, \sigma^2)$, trong đó σ chưa biết và kích thước mẫu $n < 30$ nên ta sử dụng khoảng ước lượng cho trung bình là

$$\left(\overline{x} - t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{s}}{\sqrt{n}}, \overline{x} + t_{\frac{1-\gamma}{2}}(n-1)\frac{\hat{s}}{\sqrt{n}}\right).$$

Tra bảng giá trị Student, với độ tin cậy $\gamma = 0,99$ ta có $t_{\frac{1-0,99}{2}(24)} = t_{0,005}(24) = 2,797$. Từ đó tính khoảng ước lượng trên

$$\left(1,032 - t_{\frac{1-0,99}{2}}(25-1)\frac{0,122}{\sqrt{25}}, 1,032 + t_{\frac{1-0,99}{2}}(25-1)\frac{0,122}{\sqrt{25}}\right) = (0,959; 1,105).$$

Bài 6.10 Tại một cổng ATM của ngân hàng A, số tiền được khách hàng rút ra mỗi ngày là một đại lượng ngẫu nhiên. Thống kê trong 120 ngày, nhân viên ngân hàng A tính được lượng tiền rút ra trung bình mỗi ngày là 25 triệu đồng và độ lệch chuẩn có điều chỉnh là 2,5 triệu đồng. Hãy ước lượng số tiền được rút ra trung bình mỗi ngày tại cổng ATM của ngân hàng A đó với độ tin cậy 98,5%?

Giải

Gọi X là đại lượng ngẫu nhiên chỉ số tiền được rút ra mỗi ngày tại cổng ATM

của ngân hàng A. Nhận xét thấy X có phân phối xác suất chưa biết và chưa có thông tin về phương sai, bài toán yêu cầu ước lượng trung bình μ với độ tin cậy $\gamma = 98,5\%$ từ mẫu quan sát có kích thước 120.
Từ những nhận xét trên ta sử dụng khoảng ước lượng cho μ sau

$$\left(\overline{x} - z_{\frac{\gamma}{2}}\frac{\hat{s}}{\sqrt{n}}, \overline{x} + z_{\frac{\gamma}{2}}\frac{\hat{s}}{\sqrt{n}}\right).$$

Từ các đặt trưng của mẫu $\overline{x} = 25, \hat{s} = 2,5$ và tra bảng tích phân Laplace $z_{0,4925} = 2,43$ ta tính được khoảng ước lượng trên

$$\left(25 - z_{\frac{0.985}{2}}\frac{2,5}{\sqrt{120}}, 25 + z_{\frac{0,985}{2}}\frac{2,5}{\sqrt{120}}\right) = (24,445; 25,555).$$

Vậy khoảng ước lượng số tiền được rút ra trung bình mỗi ngày là $(24,445; 25,555)$, đơn vị triệu đồng

Ước lượng tỉ lệ tổng thể

Bài 6.11 Trong một đợt điều tra nha khoa, khám ngẫu nhiên 100 trẻ em ở một địa phương, người ta thấy có 36 trẻ bị sâu răng. Hãy tìm khoảng tin cậy 99% cho tỉ lệ trẻ bị sâu răng ở địa phương đó.

Giải

Gọi p là tỉ lệ trẻ bị sâu răng ở địa phương đó, ta cần tìm khoảng tin cậy 99% cho p.
Tính giá trị tần suất mẫu

$$f_n = \frac{36}{100} = 0,36.$$

Với $\gamma = 0.99$ tra bảng tích phân Laplace $z_{\frac{0,99}{2}} = z_{0,495} = 2,58$. Khoảng tin cậy 99% cho p là

$$\left(f_n - z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}, f_n + z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}\right)$$

hay

$$\left(0,36 - 2,58\sqrt{\frac{0,36(1-0,36)}{100}}; 0,36 + 2,58\sqrt{\frac{0,36(1-0,36)}{100}}\right) = (0,236; 0,484).$$

Vậy khoảng tin cậy 99% cho p là $(0,236; 0,484)$.
Bài 6.12 Tại một vùng rừng nguyên sinh, người ta đeo vòng cho 1000 con chim. Sau một thời gian bắt lại 200 con thì thấy 40 con có đeo vòng. Hãy ước lượng số

chim trong rừng với độ tin cậy 99%.

Giải

Gọi p là tỉ lệ chim có đeo vòng trong vùng rừng nguyên sinh, ta ước lượng p với độ tin cậy 99%. Tính giá trị tần suất mẫu

$$f_n = \frac{40}{200} = 0,2.$$

Với độ tin cậy 99%, $z_{\frac{0,99}{2}} = z_{0,495} = 2,58$. Khoảng tin cậy 99% cho p là

$$\left(f_n - z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}, f_n + z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}\right).$$

Tính giá trị $z_{\frac{\gamma}{2}}\sqrt{\dfrac{f_n(1-f_n)}{n}} = z_{0,495}\sqrt{\frac{0,2(1-0,2)}{200}} = 0,073$. Suy ra khoảng ước lượng

$$(0,2 - 0,073; 0,2 + 0,073) = (0,127; 0,273).$$

Từ đây suy ra khoảng ước lượng cho số chim trong rừng với độ tin cậy 99% là

$$\left(\frac{1000}{0,273}; \frac{1000}{0,127}\right) = (3663; 7874).$$

<u>Bài 13</u> Để đánh giá chất lượng một lô hàng, nhân viên kiểm tra đã lấy ngẫu nhiên 100 sản phẩm. Kết quả kiểm tra cho thấy trong 100 sản phẩm có 8 sản phẩm không đạt yêu cầu.

a) Hãy ước lượng tỉ lệ sản phẩm không đạt yêu cầu của lô hàng đó với độ tin cậy 95%.

b) Nếu nhân viên kiểm tra đưa ra khoảng ước lượng tỉ lệ sản phẩm không đạt yêu cầu là (7%;12%) thì độ tin cậy ước lượng này là bao nhiêu?

c) Với độ tin cậy 95%, muốn bán kính ước lượng cho tỉ lệ sản phẩm không đạt yêu cầu không vượt qua 2% thì cần kiểm tra ít nhất bao nhiêu sản phẩm? Cho biết rằng tỉ lệ sản phẩm không đạt yêu cầu trên mẫu thăm dò là 8%.

d) Cùng lô hàng trên, ở độ tin cậy 95% muốn bán kính ước lượng cho tỉ lệ sản phẩm không đạt yêu cầu không vượt qua 2% thì cần kiểm tra ít nhất bao nhiêu?

Giải

Gọi p là tỉ lệ sản phẩm không đạt yêu cầu của lô hàng trên.
a. Tính giá trị tần suất mẫu

$$f_n = \frac{8}{100} = 0,08.$$

Với độ tin cậy 95%, tra bảng Laplace $z_{\frac{0,95}{2}} = z_{0,475} = 1,96$. Khoảng ước lượng cho tỉ lệ sản phẩm không đạt yêu cầu với độ tin cậy 95% là

$$\left(f_n - z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}, f_n + z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}}\right).$$

Tính giá trị $z_{\frac{\gamma}{2}}\sqrt{\dfrac{f_n(1-f_n)}{n}} = z_{0,475}\sqrt{\frac{0,08(1-0,08)}{100}} = 0,053$. Từ đó tính được khoảng ước lượng là

$$(0,08 - 0,053; 0,08 + 0,053) = (2,7\%; 13,3\%)\,.$$

b. Do khoảng ước lượng đối xứng nên từ khoảng ước lượng ta tìm được bán kính

$$\epsilon = \frac{12\% - 7\%}{2} = 5\%.$$

Từ đó suy ra

$$\epsilon = z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}},$$

hay

$$z_{\frac{\gamma}{2}} = \epsilon\sqrt{\frac{n}{f_n(1-f_n)}} = 0,05\sqrt{\frac{100}{0,08(1-0,08)}} = 1,843 \approx 1,84.$$

Tra bảng tích phân Laplace ta được $\dfrac{\gamma}{2} = 0,4671$ tương đương với $\gamma = 93,42\%$.

c. Theo giả thiết ta có

$$\epsilon = z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}} \le 0,02$$

tương đương với

$$n \ge z^2_{\frac{\gamma}{2}}.\frac{f_n(1-f_n)}{0,02^2} \Leftrightarrow n \ge z^2_{\frac{0,95}{2}}.\frac{0,08(1-0,08)}{0,02^2} = 706,854.$$

Vậy số sản phẩm tối thiểu cần kiểm tra là 707 sản phẩm.

d. Khi không có mẫu thăm dò để tính tỉ lệ mẫu sản phẩm không đạt yêu cầu ta buộc sử dụng đánh giá thô sau

$$\epsilon = z_{\frac{\gamma}{2}}\sqrt{\frac{f_n(1-f_n)}{n}} \le z_{\frac{\gamma}{2}}\frac{1}{2\sqrt{n}} \le \epsilon_0 = 0,02.$$

Suy ra

$$n \ge z^2_{\frac{\gamma}{2}}\frac{1}{4\epsilon_0^2} = 1,96^2\frac{1}{4.0,02^2} = 2401.$$

Vậy cần kiểm tra ít nhất 2401 sản phẩm.

6.5 Bài tập đề nghị

<u>Bài 6.14</u> Cân thử 100 sản phẩm của một nhà máy, ta có trọng lượng trung bình là 500g, độ lệch chuẩn mẫu hiệu chỉnh s=150g. Hãy ước lượng trọng lượng trung bình của một sản phẩm của nhà máy đó với độ tin cậy 95%.
<u>Bài 6.15</u> Để xác định trọng lượng trung bình của các bao bột mì bán ở một cửa hàng. Cân thử 25 bao bột mì của cửa hàng đó ta được trọng lượng trung bình của các bao bột đó là 49,2 kg; độ lệch chuẩn có hiệu chỉnh của các bao bột đó là 0,5kg. Haxy ước lượng trọng lượng trung bình của các bao bột của cửa hàng đó với độ tin cậy 95%. Biết rằng trọng lượng của các bao bột mì ở cửa hàng đó là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn.
<u>Bài 6.16</u> Để điều tra nhu cầu tiêu dùng 1 loại sản phẩm ở một khu dân cư, người ta theo dõi 100 hộ gia đình của khu này thấy có 60 hộ gia đình có nhu cầu về loại sản phẩm này. Với độ tin cây 95%, hãy ước lượng tỉ lệ hộ gia đình này có nhu cầu về loại sản phẩm này.
<u>Bài 6.17</u> Điểm trung bình môn toán của 100 thí sinh vào Trường Đại học Tài Chính Marketing là 5,3 với $s^2 = 1$.

a) Ước lượng điểm trung bình môn toán của các thí sinh thi vào trường với độ tin cậy 96%.

b) Với độ chính xác 0,25 điểm thì bảo đảm độ tin cậy là bao nhiêu?

<u>Bài 6.18</u> Tuổi thọ của một loại bóng đèn được biết theo quy luật chuẩn với độ lệch chuẩn 100 giờ.

a) Chọn ngẫu nhiên 100 bóng để thử nghiệm, thấy mỗi bóng tuổi thọ trung bình là 1000 giời. Hãy ước lượng tuổi thọ trung bình của bóng đèn xí nghiệp A sản xuất với độ tin cậy 95%.

b) Với độ chính xác là 15 giờ. Hãy xác định độ tin cậy.

c) Với độ chính xác là 25 giờ và độ tin cậy là 95% thì cần thử nghiệm bao nhiêu bóng.

<u>Bài 6.19</u> Để ước lượng tỷ lệ sản phẩm xấu của một kho đồ hộp, người ta kiểm tra ngẫu nhiên 100 hộp thấy có 11 hộp xấu.

a) Ước lượng tỷ lệ sản phẩm xấu của một kho đồ hộp với độ tin cậy 94%.

b) Với sai số cho phép $\alpha = 3\%$, hãy ước lượng độ tin cậy.

<u>Bài 6.20</u> Một lô bút bi của xí nghiệp A sản xuất ra gồm 1000 hộp, mỗi hộp 10 cây. Kiểm tra ngẫu nhiên 50 hộp, thấy có 45 cây bút bi bị hư hỏng.

a) Tìm khoảng tin cậy 95% cho tỉ lệ bút bị hỏng và số bút bị hỏng của lô hàng.

b) Với mẫu trên, nếu muốn ước lượng tỉ lệ bút bị hỏng với độ chính xác 1.5% thì độ tin cậy đạt được là bao nhiêu?

Bài 6.21 Quan sát ở một mẫu, người ta có kết quả về chiều cao X(m) của loại cây công nghiệp ở một nông trường như sau:

x_i	3	4	5	6	7	8
Số cây	2	8	23	32	23	12

a) Hãy ước lượng chiều cao trung bình của loại cây đó bằng khoảng tin cậy 90%.

b) Để ước lượng chiều cao trung bình của loại cây đó ở độ tin cậy 95%, với sai số không quá 2 dm thì cần phải quan sát thêm bao nhiêu cây nữa?

c) Những cây cao từ 7m trở lên gọi là cây loại A. Hãy tìm khoảng tin cậy 99,44% cho tỉ lệ cây loại A của nông trường.

Bài 6.22 Người ta muốn ước lượng tỉ lệ viên thuốc bị sức mẻ trong một lô thuốc rất nhiều viên.

a) Nếu muốn sai số cho phép không quá 1% ở độ tin cậy 95% thì phải quan sát ít nhất mấy viên?

b) Quan sát ngẫu nhiên 200 viên, thấy có 20 viên bị sức mẻ. Hãy tìm khoảng tin cậy 95% cho tỉ lệ tổng thể. Nếu muốn sai số cho phép không quá 1% ở độ tin cậy 95% thì phải quan sát ít nhất mấy viên?

Bài 6.23 Để nghiên cứu sản lượng sữa hàng ngày (SLSHN) của một đàn bò, người ta điều tra ngẫu nhiên trên 100 con bò của nông trường và có kết quả sau:

SLSHN (kg)	9	10	12	14	15
Số con bò	10	24	42	16	8

a) Ước lượng sản lượng sữa trung bình mỗi ngày của một con bò bằng khoảng tin cậy 97%.

b Với độ tin cậy 97%, có thể nói sản lượng sữa trung bình hàng ngày của một con bò nhiều nhất bằng bao nhiêu?

c) Tìm khoảng tin cậy 90% cho tỉ lệ bò cho SLSHN trên 11kg.

d) Muốn sai số khi ước lượng sản lượng sữa trung bình mỗi ngày không vượt quá 0,5kg và sai số khi ước lượng tỉ lệ bò cho SLSHN trên 11kg không vượt quá 12%, với cùng độ tin cậy 98%, thì cần điều tra bao nhiêu con bò?

Bài 6.24 Để đánh giá trữ lượng cá trong một hồ lớn, người ta bắt 2000 con cá từ hồ đó, đánh dấu rồi thả lại xuống hồ. Vài ngày sau, họ đánh bắt lại 400 con thì thấy có 80 con có đánh dấu.

a) Hãy ước lượng trữ lượng cá trong hồ bằng khoảng tin cậy 95%.

b) Nếu muốn sai số của ước lượng giảm đi một nửa thì lần sau phải đánh bắt bao nhiêu con cá?

Chương 7

Lí thuyết kiểm định giả thiết thống kê

7.1 Các khái niệm

✠ Những thông tin từ mẫu của một tổng thể được dùng để suy đoán về các đặc trưng của tổng thể đó, chẳng hạn ước lượng các tham số của một tổng thể mà chúng ta gặp ở chương VII. Trong chương này, chúng ta bàn đến một dạng suy đoán khác , liên quan đến các giả thiết thống kê và các phép kiểm định để có quyết định chấp nhận hay bác bỏ các giả thiết đó.

Định nghĩa 1.1. *Một giả thiết thống kê là một khẳng định về phân phối của một hoặc nhiều biến ngẫu nhiên. Nếu giả thiết thống kê xác định hoàn toàn một phân phối , thì nó được gọi là một giả thiết thống kê đơn; trường hợp ngược lại nó được gọi là một giả thiết thống kê hợp.*

✠ Trong quá trình đi đến một quyết định, chúng ta thường dựa vào một quy luật hay một kinh nghiệm nào đó để đặt ra một giả thiết thống kê; sau đó, xây dựng những thủ tục, theo đó, những giả thiết đã đặt ra được chấp nhận hay bị bác bỏ. Những thủ tục đó được gọi là những phép kiểm định giả thiết thống kê.

✠ Phép kiểm định thường là phép so sánh giữa hai hay nhiều giá trị. Giả thiết được đặt ra thường được gọi là "giả thiết không", kí hiệu H_0 . Chữ "không" ở đây có nghĩa là không có sự khác biệt có ý nghĩa về mặt thống kê giữa các giá trị cần so sánh. Khi bác bỏ H_0 chúng ta sẽ chập nhận một giả thiết H_1 khác, được gọi là "giả thiết đối" của H_0 . Phương pháp của thủ tục là dùng kết quả của mẫu để chứng minh một giả thiết.

✠ Xét mẫu kích thước n $(X_1, X_2, ..., X_n)$ được thành lập từ tổng thể X. Người ta chia $Im(X_1, X_2, ..., X_n)$ (còn gọi là không gian mẫu) thành hai tập con, lần lượt được ký hiệu là W_α và $W = Im(X_1, X_2, ..., X_n) - W_\alpha$. Khi mẫu cụ thể $(x_1, x_2, ..., x_n)$ được thực hiện, nếu $(x_1, x_2, ..., x_n) \in W_\alpha$ thì giả thiết H_0 bị bác bỏ; nếu $(x_1, x_2, ..., x_n) \in W$ thì giả thiết H_0 được chấp nhận. Tập hợp W_α được gọi là miền bác bỏ của phép kiểm định.

Định nghĩa 1.2. Khi ta có một tiêu chuẩn kiểm định G, với một mức ý nghĩa α cho trước, ta thiết lập một miền W_α sao cho

$$P\left(G \in W_\alpha / H\text{đúng}\right) = \alpha$$

W_α được gọi là miền bác bỏ (miền bác bỏ là tập các giá trị trên đường thẳng thực).

Định nghĩa 1.3. *Một phép kiểm định một giả thiết thống kê là một qui tắc, theo đó, dựa vào một mẫu cụ thể được thực hiện, chúng ta có thể chấp nhận hay bác bỏ giả thiết đang xét.*

Định nghĩa 1.4. ta định nghĩa 2 loại sai lầm sau

- Sai lầm loại 1: Bác bỏ giả thiết trong khi thực tế giả thiết đúng (bác bỏ giả thiết đúng).
- Sai lầm loại 2: Chấp nhận giả thiết trong khi thực tế giả thiết sai (chấp nhận giả thiết sai).

Chú ý: khi ta nói "chấp nhận" điều đó không có nghĩa là giả thiết H_0 đúng mà chỉ có nghĩa là với số liệu của mẫu ta chưa đủ cơ sở để bác bỏ H_0 .

7.2 Kiểm định tham số

✠ Nguyên tắc chung của kiểm định giả thiết thống kê là dựa trên **nguyên lý xác suất nhỏ**: khi thực hiện một phép thử, một sự kiện có xác suất xuất hiện đủ bé thì coi như không xuất hiện. Như vậy, chúng ta quyết định bác bỏ giả thiết H_0 nếu xác suất xuất hiện của một sự kiện quan sát được, tính trong điều kiện H_0 đúng là nhỏ.

✠ Để thực hiện bài toán kiểm định tham số θ với giả thiết $H_0: \theta = \theta_0$ và một trong đối thiết $H_1: \left[\begin{array}{l}\theta > \theta_0 \\ \theta < \theta_0 \\ \theta \neq \theta_0\end{array}\right.$ ta làm theo các bước sau:

Bước 1: Chọn giả thiết H_0 và đối thiết H_1.

Bước 2: Tìm miền bác bỏ W_α.

Bước 3: Tính giá trị quan sát từ số liệu mẫu cụ thể.

Bước 4: Trả lời (nếu giá trị quan sát thuộc miền bác bỏ thì ta bác bỏ giả thiết, chấp nhận đối thiết và ngược lại.)

7.2.1 Kiểm định trung bình

Các bước thực hành:

Bước 1. Chọn giả thiết $H_0: \mu = \mu_0, H_1: \left[\begin{array}{l}\mu > \mu_0 \\ \mu < \mu_0 \\ \mu \neq \mu_0\end{array}\right.$

Bước 2. Tìm miền bác bỏ

i) Miền bác bỏ dạng Z như sau
Nếu $\mu > \mu_0$ thì $W_\alpha = (z_{0.5-\alpha}; +\infty)$,
Nếu $\mu < \mu_0$ thì $W_\alpha = (-\infty; -z_{0.5-\alpha})$,
Nếu $\mu \neq \mu_0$ thì $W_\alpha = \left(-\infty, -z_{\frac{1-\alpha}{2}}\right) \cup \left(z_{\frac{1-\alpha}{2}}, +\infty\right)$.

ii) Miền bác bỏ dạng T như sau (không biết σ^2 và $n < 30$): Nếu $\mu > \mu_0$ thì $W_\alpha = (t_\alpha(n-1), +\infty)$,
Nếu $\mu < \mu_0$ thì $W_\alpha = (-\infty, -t_\alpha(n-1))$,
Nếu $\mu \neq \mu_0$ thì $W_\alpha = \left(-\infty, -t_{\frac{\alpha}{2}}(n-1)\right) \cup \left(t_{\frac{\alpha}{2}}(n-1), +\infty\right)$.

Bước 3. Tính giá trị quan sát

$$z_{qs} = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$$

Lưu ý nếu không có σ ta thay bằng s.

Bước 4. Kết luận

+ Nếu $z_{qs} \in W_\alpha$: bác bỏ giả thiết, chấp nhận đối thiết,

+ Nếu $z_{qs} \notin W_\alpha$: chấp nhận giả thiết, bác bỏ đối thiết.

7.2.2 Kiểm định tỷ lệ

Sau khi kiểm tra điều kiện $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$ được thỏa mãn, ta tiến hành các bước thực hành sau:

Bước 1. Chọn giả thiết $H_0 : p = p_0, H_1 : \left[\begin{array}{l} p > p_0 \\ p < p_0 \\ p \neq p_0 \end{array}\right.$

Bước 2. Tìm miền bác bỏ
Miền bác bỏ dạng Z như sau
Nếu $p > p_0$ thì $W_\alpha = (z_{0.5-\alpha}; +\infty)$,
Nếu $p < p_0$ thì $W_\alpha = (-\infty; -z_{0.5-\alpha})$,
Nếu $p \neq p_0$ thì $W_\alpha = \left(-\infty, -z_{\frac{1-\alpha}{2}}\right) \cup \left(z_{\frac{1-\alpha}{2}}, +\infty\right)$.

Bước 3. Tính giá trị quan sát

$$z_{qs} = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$$

Bước 4. Kết luận

+ Nếu $z_{qs} \in W_\alpha$: bác bỏ giả thiết, chấp nhận đối thiết,

+ Nếu $z_{qs} \notin W_\alpha$: chấp nhận giả thiết, bác bỏ đối thiết.

7.2.3 Kiểm định phương sai

Các bước thực hành

Bước 1. Chọn giả thiết $H_0 : \sigma^2 = \sigma_0^2, H_1 : \left[\begin{array}{l} \sigma^2 > \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \\ \sigma^2 \neq \sigma_0^2 \end{array}\right.$

Bước 2. Miền bác bỏ

+ Trường hợp biết trung bình μ ta có miền bác bỏ dạng Khi-bình phương với bậc tự do n, mức ý nghĩa α:

Nếu $\sigma^2 > \sigma_0^2$ thì $W_\alpha = \left(\chi_\alpha^2\left(n\right), +\infty\right)$,
Nếu $\sigma^2 < \sigma_0^2$ thì $W_\alpha = \left(0, \chi_{1-\alpha}^2\left(n\right)\right)$,
Nếu $\sigma^2 \neq \sigma_0^2$ thì $W_\alpha = \left(0, \chi_{1-\frac{\alpha}{2}}^2\left(n\right)\right) \cup \left(\chi_{1-\frac{\alpha}{2}}^2\left(n\right), +\infty\right)$.
Trong đó n là số phần tử mẫu chọn được, $\chi_\alpha^2(n)$ là phân vị Student bậc tự do n mức xác suất α, được xác định từ bảng 5.

+ Trường hợp chưa biết trung bình μ ta có miền bác bỏ dạng Khi-bình phương với bậc tự do $n-1$, mức ý nghĩa α:

Nếu $\sigma^2 > \sigma_0^2$ thì $W_\alpha = \left(\chi_\alpha^2\left(n-1\right), +\infty\right)$,
Nếu $\sigma^2 < \sigma_0^2$ thì $W_\alpha = \left(0, \chi_{1-\alpha}^2\left(n-1\right)\right)$,
Nếu $\sigma^2 \neq \sigma_0^2$ thì $W_\alpha = \left(0, \chi_{1-\frac{\alpha}{2}}^2\left(n-1\right)\right) \cup \left(\chi_{1-\frac{\alpha}{2}}^2\left(n-1\right), +\infty\right)$.

Bước 3. Giá trị quan sát

+ Trường hợp biết trung bình μ,

$$z_{qs} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$$

+ Trường hợp chưa biết trung bình μ,

$$z_{qs} = \frac{\left(n-1\right)s^2}{\sigma_0^2}$$

Bước 4. Kết luận

+ Nếu $z_{qs} \in W_\alpha$: bác bỏ giả thiết, chấp nhận đối thiết,
+ Nếu $z_{qs} \notin W_\alpha$: chấp nhận giả thiết, bác bỏ đối thiết.

7.3 So sánh hai tham số

7.3.1 So sánh hai trung bình

Các bước thực hành:

Bước 1. Chọn giả thiết $H_0 : \mu_1 = \mu_2, H_1 : \left[\begin{array}{l} \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \\ \mu_1 \neq \mu_2 \end{array}\right.$

Bước 2. Tìm miền bác bỏ

i) Miền bác bỏ dạng Z như sau
Nếu $\mu_1 > \mu_2$ thì $W_\alpha = (z_{0.5-\alpha}; +\infty)$,
Nếu $\mu_1 < \mu_2$ thì $W_\alpha = (-\infty; -z_{0.5-\alpha})$,
Nếu $\mu_1 \neq \mu_2$ thì $W_\alpha = \left(-\infty, -z_{\frac{1-\alpha}{2}}\right) \cup \left(z_{\frac{1-\alpha}{2}}, +\infty\right)$.

ii) Miền bác bỏ dạng T như sau (không biết σ^2 và $n_1 + n_2 - 2 < 30$): Nếu $\mu_1 > \mu_2$ thì $W_\alpha = (t_\alpha(n_1 + n_2 - 2), +\infty)$,
Nếu $\mu_1 < \mu_2$ thì $W_\alpha = (-\infty, -t_\alpha(n_1 + n_2 - 2))$,
Nếu $\mu_1 \neq \mu_2$ thì $W_\alpha = \left(-\infty, -t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)\right) \cup \left(t_{\frac{\alpha}{2}}(n_1 + n_2 - 2), +\infty\right)$.

Bước 3. Tính giá trị quan sát

+ Dạng Z: $z_{qs} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

Lưu ý nếu không có σ_1^2, σ_2^2 ta thay bằng s_1^2, s_2^2.

+ Dạng T: $z_{qs} = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$.

Trong đó $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.

Bước 4. Kết luận

+ Nếu $z_{qs} \in W_\alpha$: bác bỏ giả thiết, chấp nhận đối thiết,
+ Nếu $z_{qs} \notin W_\alpha$: chấp nhận giả thiết, bác bỏ đối thiết.

7.3.2 So sánh hai tỷ lệ

Các bước thực hành

Bước 1. Chọn giả thiết $H_0 : p_1 = p_2, H_1 : \left[\begin{array}{l} p_1 > p_2 \\ p_1 < p_2 \\ p_1 \neq p_2 \end{array}\right.$

Bước 2. Tìm miền bác bỏ
Miền bác bỏ dạng Z như sau

Nếu $p_1 > p_2$ thì $W_\alpha = (z_{0.5-\alpha}; +\infty)$,
Nếu $p_1 < p_2$ thì $W_\alpha = (-\infty; -z_{0.5-\alpha})$,
Nếu $p_1 \neq p_2$ thì $W_\alpha = \left(-\infty, -z_{\frac{1-\alpha}{2}}\right) \cup \left(z_{\frac{1-\alpha}{2}}, +\infty\right)$.

Bước 3. Tính giá trị quan sát

$$z_{qs} = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

trong đó $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$.

Bước 4. Kết luận

+ Nếu $z_{qs} \in W_\alpha$: bác bỏ giả thiết, chấp nhận đối thiết,
+ Nếu $z_{qs} \notin W_\alpha$: chấp nhận giả thiết, bác bỏ đối thiết.

7.3.3 So sánh cặp

Trong các trường hợp trên, khi so sánh hai trung bình của hai biến ngẫu nhiên X, Y chúng ta đều xét trong trường hợp chúng độc lập. Bây giờ, ta giả sử X, Y là một cặp gồm hai biến ngẫu nhiên phụ thuộc nhau, khi đó ta không sử dụng được các trường hợp về mẫu độc lập ở trên để so sánh hai trung bình. Trường hợp này phương pháp so sánh cặp thường được sử dụng.

a) Bài Toán
Trên mỗi n đối tượng, ta xét biến ngẫu nhiên X qua hai giai đoạn trước và sau, được số liệu cụ thể

Giai đoạn trước	x_1	x_2	...	x_n
Giai đoạn sau	y_1	y_2	...	y_n

Với một độ tin cậy $1-\alpha$ cho trước ta cần kiểm tra dấu hiệu X qua hai giai đoạn như thế nào (có khác nhau, lớn hơn, hay nhỏ hơn).

b) Các bước thực hiện
Đặt $d_i = x_i - y_i, i = 1, 2, ..., n$ lúc này bài toán đặt ra chính là bài toán kiểm định trung bình với $\mu_0 = 0$, do đó các bước thực hiện được cụ thể như sau:

Bước 1. Chọn giả thiết $H_0 : \mu_d = 0, H_1 : \left[\begin{array}{l} \mu_d > 0 \\ \mu_d < 0 \\ \mu_d \neq 0 \end{array}\right.$

Bước 2. Tìm miền bác bỏ
Miền bác bỏ dạng Z hoặc dạng T với bậc tự do $n-1$ tùy theo $n \geq 30$ hay $n < 30$.

Bước 3. Tính giá trị quan sát

$$z_{qs} = \frac{\overline{d}\sqrt{n}}{s_d}$$

Trong đó $\overline{d}$ và s_d lần lượt là trung bình và độ lệch chuẩn mẫu điều chỉnh của tập dữ liệu $\{d_1, d_2, ..., d_n\}$.

Bước 4. Kết luận

+ Nếu $z_{qs} \in W_\alpha$: bác bỏ giả thiết, chấp nhận đối thiết,
+ Nếu $z_{qs} \notin W_\alpha$: chấp nhận giả thiết, bác bỏ đối thiết.

7.4 Kiểm định phi tham số

Kiểm định tham số đã trình bày có đặc điểm chung là phải dựa trên một số điều kiện liên quan đến tổng thể (giả thiết có phân phối chuẩn hoặc cỡ mẫu lớn). Tuy nhiên trong thực tế nhiều trường hợp không đưa ra bất kì giả thiết nào liên quan đến tham số của tổng thể hoặc dạng phân phối xác suất của tổng thể, vì thế ta không áp dụng dạng kiểm định tham số được. Kiểm định phi tham số được áp dụng trong các trường hợp này. Kiểm định phi tham số phù hợp cho dữ liệu định danh và thứ bậc. Nó cũng hữu dụng khi mẫu có những giá trị quan sát bất thường. Cần khẳng định rằng kiểm định phi tham số không mạnh như kiểm định tham số (khi mà các điều kiện để thực hiện kiểm định tham số được thỏa mãn) vì nó đã bác bỏ qua những thông tin có giá trị về tham số.

7.4.1 Kiểm định sự độc lập

a) **Bài toán**
Giả sử ta cần nghiên cứu đồng thời hai dấu hiệu X và Y, với X có h dấu hiệu thành phần $\{x_1, x_2, ..., x_h\}$; Y có c dấu hiệu thành phần $\{y_1, y_2, ..., y_c\}$. Chọn một mẫu gồm n phần tử , quan sát đồng thời hai dấu hiệu X và Y ta có bảng số liệu sau:

$X \quad \vert \quad Y$	y_1	y_2	....	y_c
x_1	n_{11}	n_{12}	....	n_{1c}
x_2	n_{21}	n_{22}	....	n_{2c}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_h	n_{h1}	n_{h2}	....	n_{hc}

trong đó n_{ij} là tần số thực tế ứng với các phần tử đồng thời mang dấu hiệu (x_i, y_j).

Với một mức ý nghĩa α cho trước, ta cần kiểm tra X và Y có độc lập nhau hay không?

b) **Các bước thực hiện**

Ta có thể thực hiện theo các bước sau:

+ *Bước 1:* Chọn giả thiết H_0 và đối thiết H_1:

  H_0: X và Y độc lập.
  H_1: X và Y phụ thuộc.

+ *Bước 2:* Tìm miền bác bỏ:

$$W_\alpha = \left(\chi^2_\alpha\left(\left(h-1\right)\left(c-1\right)\right); +\infty\right)$$

  trong đó $\chi^2_\alpha\left(\left(h-1\right)\left(c-1\right)\right)$ là phân vị Khi bình phương bậc tự do $(h-1)(c-1)$, mức xác suất α.

+ *Bước 3:* Tính giá trị quan sát:

$$\chi^2_{qs} = \sum_{i=1}^{h}\sum_{j=1}^{c}\frac{\left(n_{ij} - \widehat{n_{ij}}\right)^2}{\widehat{n_{ij}}}$$

  Trong đó
  n_{ij}: Tần số thực tế ứng với các phần tử đồng thời mang dấu hiệu được xác định từ mẫu chọn ra.
  $\widehat{n_{ij}}$: Tần số lý thuyết ứng với các phần tử đồng thời mang dấu hiệu (x_i, y_j) . Tần số lý thuyết được xác định bởi

$$\widehat{n_{ij}} = \frac{h_i c_j}{n}$$

  Với
  h_i : Tổng số liệu của hàng thứ i,
  c_j : Tổng số liệu của cột j.

+ *Bước 4:* Trả lời.

7.4.2 Kiểm định luật phân phối

Kiểm định giả thiết các quan sát tuân theo một quy luật nào đó: nhị thức, Poisson, siêu bội, chuẩn,... Giả sử tổng thể X có luật phân phối xác suất $F_X(x)$ chưa biết. Với mức ý nghĩa α cho trước, từ mẫu quan sát $(X_1, X_2, ..., X_n)$ kiểm định giả thiết $H_0 : F_X(x) = F_X^*(x)$ và đối thiết $H_1 : F_X(x) \neq F_X^*(x)$, trong đó $F_X^*(x)$ là luật phân phối xác suất đã biết. Với mẫu n quan sát lấy từ biến X được chia thành k nhóm $\{E_1, E_2, ..., E_k\}$ với số lượng cụ thể như sau:

Các nhóm E_k	E_1	E_2	...	E_k
Tần số	n_1	n_2	...	n_k

Trong đó

$$\sum_{i=1}^{k} n_i = n, n_i \geq 5,$$

E_i có thể là một giá trị, khoảng giá trị hay tập hợp một số giá trị. Các bước thực hành như bài toán kiểm định tổng quát, trong đó:
- Giả thiết: X có phân phối chuẩn (nhị thức, Poisson, siêu bội,...).
- Miền bác bỏ: $W_\alpha = \left(\chi^2_\alpha (k - r - 1); +\infty\right)$.
Trong đó r là số lượng tham số trong dạng kiểm định mà ta xét đến. Chẳng hạn

$$X \in P(a) \Rightarrow r = 1$$
$$X \in N\left(\mu, \sigma^2\right) \Rightarrow r = 2$$
$$X \in H(N, M, n) \Rightarrow r = 3$$

- Giá trị quan sát:

$$f_{qs} = \sum_{i=1}^{k} \frac{(n_i - \widehat{n_i})^2}{\widehat{n_i}}$$

Trong đó $p_i = P(E_i)$, $i = 1, 2, ..., k$, và tần số lý thiết $\widehat{n_i} = np_i$.

Một số ví dụ

Ví dụ 7.1 Giả sử một huyện năm trước có tỷ lệ trẻ em bị suy dinh dưỡng là 10%, năm nay huyện thực hiện nhiều chính sách nhằm làm giảm tỷ lệ này xuống. chọn 400 đứa trẻ, kiểm tra ta thấy có 32 đứa trẻ vẫn còn bị suy dinh dưỡng. với mức ý nghĩa 1% hãy cho kết luận về việc giảm tỷ lệ trẻ em suy dinh dưỡng của huyện này.
ĐS: $g = -1,33 \notin W_\alpha = (-\infty; -2,326)$.

Ví dụ 7.2 Kiểm tra 100 đứa trẻ của vùng I phát hiện 42 đứa trẻ bị sâu răng, vùng II có 92 đứa trẻ bị sâu răng khi kiểm tra 200 đứa trẻ. Với mức ý nghĩa 5% có thể xem tỷ lệ trẻ bị sâu răng ở 2 vùng bằng nhau được không?
ĐS: $g = -0,66 \notin W_\alpha = (-\infty; -1,96) \cup (1,96; +\infty)$.

Ví dụ 7.3 Một trường cấp 3 báo cáo học sinh giỏi, khá, trung bình và yếu kém của trường mình lần lượt là 15%, 40%, 35% và 10%. Kiểm tra ngẫu nhiên các học sinh của trường này ta lần lượt có số liệu về học sinh giỏi, khá, trung bình và yếu kém lần lượt là 20, 75, 75, và 30. Hỏi báo cáo của trường về tỷ lệ các nhóm học sinh có đúng không với mức ý nghĩa 5%.
ĐS: $g = 9 \in W_\alpha = (7,815; +\infty)$.

Ví dụ 7.4 Khảo sát chiều cao (đơn vị cm) của các em bé 10 tuổi ở một tỉnh, người ta chọn một mẫu 75 đứa bé và được số liệu:

Chiều cao	≤ 130	$(130-135]$	$(135-140]$	$(140-145]$	> 145
Số em	5	15	30	20	5

Với độ tin cậy 95% có thể khẳng định chiều cao trung bình học sinh của tỉnh này lớn hơn 137 cm được không?
ĐS: $g = 0,953 \notin W_\alpha = (1,645; +\infty)$.

Ví dụ 7.5 Đo chiều cao (đơn vị cm) của 24 trẻ em 2 tuổi tại 1 huyện ta có số liệu:
84,4; 89,9; 89,0; 91,9; 87,0; 78,5; 84,5; 86,3; 80,6; 80,0; 81,3; 86,8; 83,4; 89,8; 85,4; 80,6; 85,0; 82,5; 80,7; 84,3; 95,4; 85,0; 85,5; 81,6
Biết chiều cao của trẻ em hai tháng tuổi chung của đất nước là biến ngẫu nhiên có phân phối chuẩn $N(86,5; 9,67)$. Hỏi với mức ý nghĩa 1% có sự khác biệt đáng kể về chiều cao trung bình của trẻ em huyện này so với chiều cao trung bình chung của đất nước không?
ĐS:$g = -2,426 \notin W_\alpha = (-\infty; -2,576) \cup (2,576; +\infty)$.

Ví dụ 7.6 Một trại chăn nuôi gà đã nuôi thí nghiệm bằng khẩu phần thức ăn có bổ sung kháng sinh. Kiểm tra 81 con gà ta có số liệu:

Trọng lượng (kg)	3,8	3,9	4,0	4,1	4,2	4,3	4,4	4,5	4,6	4,7
Số gà	5	7	9	12	15	10	9	6	5	3

a) Trại chăn nuôi báo cáo trọng lượng trung bình của những con gà nuôi thí nghiệm sau 8 tuần nuôi là 4,3 kg thì có đúng không với độ tin cậy 95%?

b) Giả sử những con gà có trọng lượng lớn hơn 4,3 kg được xếp loại I và trọng lượng của nó có phân phối chuẩn. với mức ý nghĩa 5%, chúng ta có thể kết luận trọng lượng trung bình của những con gà loại I lớn hơn 4,5 kg được không?

ĐS: $g = -3,273 \in W_\alpha = (-\infty; -1,96) \cup (1,96; +\infty)$.

Ví dụ 7.7 So sánh mức thu nhập theo tuần giữa nam và nữ tại một công ty liên doanh ta có số liệu mẫu như sau:

- Nữ: chọn một mẫu 40 người, tính được thu nhập trung bình .
- Nam: chọn một mẫu 50 người, tính được thu nhập trung bình .

Biết rằng phương sai thu nhập theo tuần của nữ là 80 và của nam là 100. Với mức ý nghĩa 1%, có thể kết luận thu nhập trung bình của nữ thấp hơn nam được không?
ĐS: $g = -5 \in W_\alpha = (-\infty; -2,326)$.

Ví dụ 7.8 Khảo sát chiều cao (đơn vị cm) của học sinh nữ tại hai trường phổ thông trung học huyện A và huyện B ta có số liệu:

Chiều cao	Số học sinh nữ của huyện A	Số học sinh nữ của huyện B
(150 – 152]	3	5
(152 – 154]	5	10
(154 – 156]	7	14
(156 – 158]	15	18
(158 – 160]	26	22
(160 – 162]	25	11
(162 – 164]	12	9
(164 – 166]	13	5
(166 – 168]	10	4
(168 – 170]	5	2

a) Với mức ý nghĩa 1% có thể xem chiều cao trung bình học sinh trung học nữ của huyện A cao hơn huyện B được không?

b) Những học sinh có chiều cao từ 154 cm trở xuống được xem là nhóm thấp. giả sử chiều cao học sinh nhóm thấp ở hai huyện là biến ngẫu nhiên có phân phối chuẩn có phương sai xấp xỉ bằng nhau. Một người nói chiều cao trung bình học sinh nhóm thấp của hai huyện là như nhau thì có đúng không với độ tin cậy là 95%.

ĐS: a) $g = 3,713 \in W_\alpha = (2,326; +\infty)$;
b) $g = -0,18 \notin W_\alpha = (-\infty; -2,08) \cup (2,08; +\infty)$.

Ví dụ 7.9 Đo hiều ngang xương cằm (cm) của bé gái lúc 5 tuổi và lúc 6 tuổi ta có số liệu:
5 tuổi: 7,33 7,49 7,27 7,93 7,56 7,81 7,46 6,94 7,49 7,44 7,95 7,47 7,04 7,10 7,64.
6 tuổi: 7,53 7,70 7,46 8,21 7,81 8,01 7,72 7,13 7,68 7,66 8,11 7,66 7,20 7,25 7,79.
Với mức ý nghĩa 1% hãy xác định xem giá trị trung bình của chiều ngang xương cằm có thay đổi theo độ tuổi không?
ĐS: $g = -19,72 \in W_\alpha = (-\infty; -2,977) \cup (2,977; +\infty)$.

Ví dụ 7.10 Nghiên cứu 200 cặp vợ chồng có thời gian kết hôn trên 5 năm để tìm hiểu có mối quan hệ giữa thời gian tìm hiểu trước hôn nhân và tình trạng hiện tại của cuộc hôn nhân hay không. Có 3 mức độ thời gian tìm hiểu trước hôn nhân là ngắn, trung bình và dài. Cũng có 3 mức độ tình trạng hôn nhân hiện tại là hạnh phúc, không hạnh phúc và ly dị. Phỏng vấn 200 người ta có số liệu sau:

	Ngắn	Trung bình	Dài
Hạnh phúc	38	58	54
Không hạnh phúc	12	14	4
Ly dị	10	8	2

Với độ tin cậy 95%, đánh giá xem có mối quan hệ giữa thời gian tìm hiểu trước hôn nhân và tình trạng hiện tại của cuộc hôn nhân hay không?
ĐS: $g = 12,4 \in W_\alpha = (9,488; +\infty)$.

Ví dụ 7.11 Tiến hành đo ngẫu nhiên chiều cao của 100 cây bạch đàn trong khu rừng trồng bạch đàn của một lâm trường ta được các kết quả sau:

Khoảng chiều cao	Số cây
$(-\infty; 8,425)$	7
$(8,425; 8,475)$	7
$(8,475; 8,525)$	9
$(8,525; 8,575)$	7
$(8,575; 8,625)$	7
$(8,625; 8,675)$	7
$(8,675; 8,725)$	7
$(8,725; 8,775)$	7
$(8,775; 8,825)$	7
$(8,825; +\infty)$	7

Hãykiểm tra giả thiết cho rằng chiều cao cây bạch đàn ở khu rừng trên tuân theo luật phân phối chuẩn với mức ý nghĩa 5%.
ĐS: $g = 2,405686 \notin W_\alpha = (14,0671; +\infty)$.

Ví dụ 7.12 ở một nước, một đảng chính trị tuyên bố rằng 45% cử tri sẽ bỏ phiếu bầu cho ông A là ứng cử viên của họ. Chọn ngẫu nhiên 200 người hỏi ý kiến có 80 người sẽ bầu cho ông A. với mức ý nghĩa 5% hãy cho nhận xét về tuyên bố trên.
ĐS: Với mức ý nghĩa 5% chưa có cơ sở bác bỏ tuyên bố trên.

Ví dụ 7.13 Một công ty tuyên bố rằng 40% người tiêu dùng ưa thích sản phẩm của công ty. Để xem xét tuyên bố đó, người ta điều tra 400 người tiêu dùng thấy có 120 người ưa thích sản phẩm của công ty. Với mức ý nghĩa 5%, hãy xem tỉ lệ trong tuyên bố của công ty có cao hơn thực tế không ?
ĐS: Với mức ý nghĩa 5%, tỉ lệ trong tuyên bố của công ty là cao hơn thực tế.

Ví dụ 7.14 Trọng lượng của một hộp sản phẩm do một máy tự động đóng theo quy định là 6 kg. sau một thời gian máy sản xuất, người ta tiến hành kiểm tra 121 hộp sản phẩm tính được trọng lượng trung bình $\overline{X} = 5,975kg$ và phương sai mẫu là 5,7596. Với mức ý nghĩa 5% hãy cho kết luận về sự hoạt động của máy.
ĐS: Với mức ý nghĩa 5% chưa có cơ sở để cho rằng máy hoạt động không bình thường.

Ví dụ 7.15 Trọng lượng của một bao gạo là biến ngẫu nhiên có phân phối chuẩn trung bình 50kg. Nghi ngờ máy đóng bao gạo làm việc không bình thường

làm cho trọng lượng của bao gạo có xu hướng giảm sút. Người ta cân thử 25 bao tính được $\overline{X} = 49,27kg$ và độ lệch tiêu chuẩn mẫu $s = 0,49kg$. Với mức ý nghĩa 1% hãy cho kết luận về nghi ngờ trên.
ĐS: Với mức ý nghĩa 1%, nghi ngờ trên là có cơ sở.

Ví dụ 7.16 Người ta muốn kiểm định giả thuyết $H_0 : \mu = 370$ và $H_1 : \mu \neq 370$ để kiểm tra một kết luận trong một tài liệu cho rằng : "tuổi thọ trung bình của loại bóng đèn A là 370 giờ với độ lệch tiêu chuẩn 20 giờ", biết rằng tuổi thọ trung bình của bóng đèn không vượt quá 380 giờ. Nếu muốn mức ý nghĩa của kiểm định $\alpha = 0,01$ và xác suất sai lầm loại II là β không vượt quá 0,05, thì cần điều tra tối thiểu bao nhiêu bóng đèn A ?
ĐS: vậy cần điều tra tối thiểu 72 bóng đèn.

Ví dụ 7.17 Chủ hãng sản xuất một loại thiết bị đo cho biết sai số đo của thiết bị loại này có độ lệch tiêu chuẩn bằng 5mm. kiểm tra một mẫu 19 thiết bị loại này tính được phương sai mẫu $s^2 = 33$. Với mức ý nghĩa 5%, hãy cho nhận xét về ý kiến trên của chủ hãng. Biết sai số đo của thiết bị có phân phối chuẩn.
ĐS: Với mức ý nghĩa 5%, chưa có cơ sở bác bỏ ý kiến của chủ hãng.

Ví dụ 7.18 Nếu độ biến động về đường kính của các sản phẩm được sản xuất bởi một dây chuyền tự động vượt quá 0,2 (mm) thì dây chuyền phải dừng lại để điều chỉnh. Kiểm tra ngẫu nhiên 12 sản phẩm của dây chuyền đo được độ lệch tiêu chuẩn của đường kính $s = 0,3$. Với mức ý nghĩa 5%, hãy xem dây chuyền có phải dừng lại để điều chỉnh không, biết đường kính của các sản phẩm có phân phối chuẩn.
ĐS: Với mức ý nghĩa 5%, dây chuyền cần điều chỉnh vì độ biến động của đường kính sản phẩm lớn hơn mức cho phép.

Ví dụ 7.19 Mức hao phí nguyên liệu X để sản xuất một sản phẩm của nhà máy tuân theo quy luật chuẩn có trung bình 20 kg. Một mẫu điều tra mức hao phí nguyên liệu để sản xuất 25 sản phẩm loại này được kết quả cho trong bảng :

Mức hao phí nguyên liệu hao phí (kg)	19,5	20,0	20,5
Số sản phẩm	5	18	2

Với mức ý nghĩa 5% có thể cho rằng phương sai của mức hao phí nguyên liệu là 0,06 kg^2 hay không.
ĐS: với mức ý nghĩa 5% có thể cho rằng phương sai của mức hao phí nguyên liệu là 0,06 kg^2.

Ví dụ 7.20 Kiểm tra chất lượng sản phẩm về một loại hàng do hai nhà máy A và B sản xuất cho kết quả : trong 500 sản phẩm của A có 50 phế phẩm và trong 400 sản phẩm của B có 60 phế phẩm. với mức ý nghĩa 5%, hãy xem

a) Chất lượng sản phẩm của A và B có khác nhau không ?

b) Chất lượng sản phẩm của A có tốt hơn B không ?

ĐS: a) khác nhau; b) A tốt hơn B.

Ví dụ 7.21 Giám đốc một hãng sản xuất thép muốn xác định xem có sự khác nhau về năng suất giữa ca ngày và ca tối hay không. Điều tra một mẫu 100 công nhân của ca ngày sản xuất được trung bình 74,3 (đơn vị : phần/1 giờ/ người) với độ lệch tiêu chuẩn 16; một mẫu khác 100 công nhân của ca tối sản xuất được trung bình 69,7 (đơn vị : phần/1 giờ/ người) với độ lệch tiêu chuẩn 18. Hãy xem có sự khác nhau về năng suất giữa hai ca không, với mức ý nghĩa 5%.
ĐS: với mức ý nghĩa 5% chưa có bằng chứng về sự khác nhau về năng suất trung bình của hai ca.

Ví dụ 7.22 Giám đốc của công ty thực phẩm muốn xác định liệu một kiểu đóng gói mới có làm tăng lượng hàng bán được hay không. Chọn ngẫu nhiên 30 quầy tương đương nhau gồm 15 quầy bán hàng được đóng theo gói mới và 15 quầy khác bán hàng được đóng theo gói cũ, tính được trong thời gian nghiên cứu lượng hàng bán được tại mỗi quầy hàng đóng gói mới có trung bình $\overline{X} = 130$ hộp với $s_1 = 10$; hàng đóng gói cũ có trung bình : $\overline{Y} = 117$ hộp với $s_2 = 12$. Với mức ý nghĩa 5% hãy xem kiểu đóng gói mới có làm tăng lượng hàng bán được không. Giả sử lượng hàng bán được theo hai kiểu đóng gói có phân phối chuẩn và cùng phương sai.
ĐS: với mức ý nghĩa 5% kiểu đóng gói mới làm tăng lượng hàng trung bình bán được.

Ví dụ 7.23 Điều tra doanh số bán hàng tại các cửa hàng ở hai vùng A và B của một công ty trong thời gian T, kết quả cho như sau: (đơn vị : triệu đồng)

Vùng A	Vùng B
$\overline{X} = 121$	$\overline{Y} = 89,17$
$s_1^2 = 76,57$	$s_2^2 = 17,37$
$m = 16$	$n = 10$

Với mức ý nghĩa 10%, hãy xem có sự khác nhau về doanh số bán hàng trung bình của mỗi cửa hàng ở hai vùng không. Giả sử doanh số bán hàng của mỗi

cửa hàng có phân phối chuẩn.
ĐS: với mức ý nghĩa 10% doanh số bán hàng trung bình của mỗi cửa hàng tại hai vùng khác nhau và vùng A cao hơn B.

Ví dụ 7.24 Để khảo sát giá bán của hai cửa hiệu thực phẩm lớn A và B, người ta điều tra ngẫu nhiên giá bán của 12 mặt hàng(MH) thông dụng nhất như sau:

MH	1	2	3	4	5	6	7	8	9	10	11	12
A	0,89	0,59	1,29	1,50	2,49	0,65	0,99	1,99	2,25	0,50	1,99	1,79
B	0,95	0,55	1,49	1,69	2,39	0,79	0,99	1,79	2,39	0,59	2,19	1,99

Với mức ý nghĩa 2% hãy xem có sự khác nhau về giá bán ở hai cửa hiệu hay không.
ĐS: Với mức ý nghĩa 2%, chưa có cơ sở để cho rằng giá bán trung bình của hai cửa hiệu là khác nhau.

Ví dụ 7.25 Người ta dùng phương sai hay độ lệch tiêu chuẩn làm độ đo đánh giá sự rủi ro của cổ phiếu. điều tra ngẫu nhiên giá cổ phiếu của công ty A trong 25 ngày tính được $s_1^2 = 6,52$; của công ty B trong 22 ngày tính được $s_2^2 = 3,47$. Với mức ý nghĩa 5% có thể cho rằng độ rủi ro cổ phiếu của công ty A cao hơn của công ty B không ?
ĐS: Với mức ý nghĩa 5%, chưa có cơ sở để cho rằng cổ phiếu của A nhiều rủi ro hơn cổ phiếu của B.

Ví dụ 7.26 Ở một nước, để xác định xem sự ủng hộ của những công dân nước đó về một sắc thuế mới có thụ thuộc vào mức thu nhập của họ hay không , người ta tiến hành điều tra ngẫu nhiên 1000 người và phân loại theo các mức thu nhập của họ cho trong bảng sau :

Sắc sau thuế	Mức thu nhập			
	Thấp	Trung bình	Cao	Tổng
Ủng hộ	182	213	203	598
Phản đối	154	138	110	402
Tổng	336	351	313	1000

Với mức ý nghĩa 5%, hãy cho kết luận về việc xác định trên.
ĐS: bác bỏ H_0, chấp nhận H_1.

Ví dụ 7.27 Sản phẩm sản xuất ra trên một dây chuyền tự động được đóng gói một cách ngẫu nhiên cứ 3 sản phẩm trong một hộp. mẫu kiểm tra ngẫu nhiên 100 hộp sản phẩm loại này cho kết quả :

Số sản phẩm loại I	0	1	2	3
Số hộp	0	10	40	50

Với mức ý nghĩa 5%, có thể xem số sản phẩm loại I có trong một hộp là biến ngẫu nhiên có phân phối thị thức không?
ĐS: Chấp nhận H_0.

Ví dụ 7.28 Để tìm hiểu số thiết bị bị hỏng trong một tháng của một hệ thống máy, người ta theo dõi 50 tháng liền và được số thiết bị bị hỏng cho trong bảng:

x_i	0	1	2	3	4	6	8
n_i	10	4	12	8	7	6	3

Với mức ý nghĩa 5%, có thể cho rằng số thiết bị bị hỏng X tuân theo quy luật Poisson $P(\lambda)$ hay không ?
ĐS: Với mức ý nghĩa 5%, chưa có cơ sở cho rằng số thiết bị bị hỏng X tuân theo quy luật Poisson.

Ví dụ 7.29 Đo chỉ tiêu X (đơn vị là gam) của một loại sản phẩm kết quả cho trong bảng

Lớp	18-20	20-22	22-24	24-26	26-28	28-30
n_i	3	4	14	33	27	19

Với mức ý nghĩa 5% có thể cho rằng có phân phối chuẩn không ?
ĐS: Chấp nhận H_0.

Ví dụ 7.30 Một báo cáo cho rằng "sự phân chia thị trường năm ngoái về tỉ lệ khách hàng sử dụng sản phẩm của ba công ty A, B, C tương ứng là $p_A = 30\%, p_B = 50\%, p_C = 20\%$ đã bị thay đổi trong năm nay khi công ty C cho ra một loại sản phẩm mới" . Hãy cho nhận xét về báo cáo đó với mức ý nghĩa 5%. Biết rằng mẫu điều tra thị trường năm nay của các công ty cho kết quả : trong 200 khách hàng, số ngừoi dùng sản phẩm của các công ty A, B, C tương ứng là 48 ; 98 ; 54 (người).
ĐS: Bác bỏ H_0.

7.5 Bài tập có hướng dẫn

Kiểm định trung bình tổng thể

Trường hợp 1: Biết phương sai

<u>Bài 7.1</u> Đo chiều cao (đơn vị cm) của 24 trẻ em 2 tuổi tại 1 huyện ta có số liệu:

84,4	89,9	89,0	91,9	87,0	785	84,5	86,3	80,6	80,0	81,3	86,8
83,4	89,8	85,4	80,6	85,0	82,5	80,7	84,3	95,4	85,0	85,5	81,6

Biết chiều cao của trẻ em hai tháng tuổi chung của đất nước là đại lượng ngẫu nhiên có phân phối chuẩn $\mathcal{N}(86,5;96,7)$. Hỏi với mức ý nghĩa 1% có sự khác biệt đáng kể về chiều cao trung bình của trẻ em huyện này so với chiều cao trung bình của đất nước không?

Giải

Gọi X là biến ngẫu nhiên cho thông tin chiều cao của trẻ em hai tháng tuổi tại huyện này $X \sim \mathcal{N}(\mu;\sigma^2)$.

Đây là bài toán kiểm định trung bình μ với $n = 24 < 30, \sigma^2 = 96,7$.

Kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 86,5; \overline{H}_0 : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_1 = \frac{(\overline{X} - \mu_0)\sqrt{24}}{\sigma} \sim \mathcal{N}(0,1).$$

Miền bác bỏ:

$$W_{\frac{\alpha}{2}} = (-\infty; -z_{\frac{1-0,01}{2}}) \cup (z_{\frac{1-0,01}{2}}; +\infty) = (-\infty; -2,58) \cup (2,58; +\infty)$$

Với mẫu trên, ta tính được trung bình mẫu $\overline{x} = 84,975$.

Từ đó tính giá trị của thống kê K_1

$$k_1 = \frac{(84,975 - 86,5)\sqrt{24}}{\sqrt{96.7}} = -0,76.$$

Do $k_1 = -0,76 \notin W_{\frac{\alpha}{2}}$ nên ta chấp nhận giả thiết, nghĩa là chiều cao trung bình của trẻ em huyện này không khác so với cả nước.

<u>Bài 7.2</u> Khối lượng của một loại sản phẩm do một nhà máy sản xuất là một đại lượng ngẫu nhiên tuân theo luật phân phối chuẩn $\mathcal{N}\left(500,(8,5)^2\right)$. Sau một thời gian sản xuất, do phản ánh của khách hàng, ban lãnh đạo của nhà máy nghi ngờ rằng khối lượng trung bình của loại sản phẩm này có xu hướng giảm, nên tiến hành cân thử 25 sản phẩm và thu được kết quả sau:

Khối lượng(g)	480	485	490	495	500	510
Số sản phẩm	2	3	8	5	3	4

Với mức ý nghĩa $\alpha = 0,05$ hãy cho kết luận về điều nghi ngờ trên.

Giải

Gọi X là biến ngẫu nhiên chỉ khối lượng sản phẩm do nhà máy sản suất $X \sim \mathcal{N}(\mu, \sigma^2)$.

Đây là bài toán kiểm định trung bình μ với phương sai đã biết $\sigma = 8,5$.

Kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 500; \overline{H}_0 : \mu < \mu_0;$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_1 = \frac{(\overline{X} - \mu_0)\sqrt{25}}{\sigma} \sim \mathcal{N}(0, 1).$$

Với mức ý nghĩa $\alpha = 0,05$, ta có miền bác bỏ

$$W_\alpha = (-\infty, -z_{\frac{1}{2}-0.05}) = (-\infty, -1.65).$$

Với mẫu trên ta có $\overline{x} = 494$.

Khi đó

$$k_1 = \frac{(494 - 500)\sqrt{25}}{8,5} = -3,529 \in W_\alpha,$$

nên giả thiết H_0 bị bác bỏ, nghĩa là điều nghi ngờ trên là có cơ sở.

Bài 7.3 Nếu máy móc làm việc bình thường thì trọng lượng của sản phẩm có kỳ vọng toán là 100g và độ lệch tiêu chuẩn $\sigma = 1$. Qua một thời gian sản xuất, người ta nghi ngờ trọng lượng trung bình của loại sản phẩm này đã thay đổi. Cân thử 100 sản phẩm và tính được $x = 100,3g$. Với mức ý nghĩa $\alpha = 0,05$, hãy kết luận điều nghi ngờ trên có đúng hay không?

Giải

Gọi X là biến ngẫu nhiên chỉ trọng lượng của sản phẩm có kì vọng μ và phương sai $\sigma^2 = 1$.

Đây là bài toán kiểm định trung bình μ với phương sai $\sigma^2 = 1$. Chưa biết phân phối của X nhưng kích thước mẫu lớn $n = 100 > 30$.

Kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 100; H_0 : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_1 = \frac{(\overline{X} - \mu_0)\sqrt{100}}{\sigma} \sim \mathcal{N}(0, 1).$$

Với mức ý nghĩa $\alpha = 0.05$, miền bác bỏ là

$$W_{\frac{\alpha}{2}} = (-\infty, -z_{\frac{1-0,05}{2}}) \cup (z_{\frac{1-0,05}{2}}, +\infty) = (-\infty, -1.96) \cup (1.96, +\infty).$$

Với mẫu trên ta có $\overline{x} = 100,3$ và

$$k_1 = \frac{(100,3 - 100)\sqrt{100}}{1} = 3 \in W_{\frac{\alpha}{2}},$$

nên ta bác bỏ giả thiết, nghĩa là điều nghi ngờ trên là có cơ sở.
<u>Bài 7.4</u> Mì chính được đóng gói 453 gam một gói trên máy tự động. Có thể coi trọng lượng các gói mì chính tuân theo luật phân phối chuẩn với độ lệch chuẩn 36 gam. Kiểm tra ngẫu nhiên 81 gói thấy trọng lượng trung bình là 458 gam. Với mức ý nghĩa $\alpha = 0.05$; có thể kết luận trọng lượng các gói mì chính có xu hướng tăng không?

Giải

Gọi X là biến ngẫu nhiên chỉ khối lượng của mỗi gói mì chính $X \sim \mathcal{N}(\mu, \sigma^2)$.
Đây là bài toán kiểm định trung bình μ với phương sai $\sigma^2 = 36^2$.
Ta kiểm định gải thiết

$$H_0 : \mu = \mu_0 = 453; \overline{H}_0 : \mu > \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_1 = \frac{(\overline{X} - \mu_0)\sqrt{81}}{\sigma} \sim \mathcal{N}(0, 1).$$

Với $\alpha = 0,05$, miền bác bỏ là:

$$W_\alpha = (z_{\frac{1}{2}-0,05}, +\infty) = (1,65, +\infty).$$

Với mẫu trên ta có

$$\overline{x} = 458, n = 81, \sigma = 36.$$

Khi đó

$$k_1 = \frac{(458 - 453)\sqrt{81}}{36} = 1,25 \notin W_\alpha,$$

nên ta chấp nhận giả thiết.
Vậy trọng lượng mì chính có xu hướng tăng là không đúng.
Trường hợp 2: Không biết phương sai
<u>Bài 7.5</u> Công ty A muốn sản xuất loại bóng đèn có tuổi thọ trung bình 1600 giờ. Nếu thời gian dùng ngắn hơn 1600 giờ thì công ty sẽ mất khách hàng; nếu thời gian dùng dài hơn thì chi phí sẽ tăng lên. Để biết xem qui trình sản xuất có tốt không, công ty chọn một mẫu ngẫu nhiên gồm 64 bóng đèn đốt thử và thấy tuổi

thọ trung bình của chúng là 1570 giờ với độ lệch chuẩn là 121 giờ. Hãy cho kết luận về qui trình sản xuất ở mức ý nghĩa 5%.

Giải

Gọi X là biến ngẫu nhiên chỉ tuổi thọ của loại bóng đèn do công ty A sản suất. Đây là bài toán kiểm định trung bình hai phía với

$$n = 64, \overline{x} = 1570, \hat{s} = 121$$

và phương sai chưa biết.
Kiểm định giả thiết:

$$H_0 : \mu = \mu_0 = 1600; \overline{H_0} : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{64}}{\hat{S}} \sim t(63).$$

Miền bác bỏ

$$W_{\frac{\alpha}{2}} = \Big(-\infty; -t_{\frac{0,05}{2}}(63)\Big) \cup \Big(t_{\frac{0,05}{2}}(63); +\infty\Big) = (-\infty; -1.96 \cup (1.96; +\infty).$$

Với ý nghĩa $\alpha = 5\%$, thì

$$k_2 = \frac{(1570 - 1600).\sqrt{64}}{121} = -1,983 \in W_{\frac{\alpha}{2}}$$

nên ta bác bỏ giả thiết.
Vậy quy trình sản suất không tốt.
<u>Bài 7.6</u> Trở lại công ty A trong bài tập trên, công ty tuyên bố rằng tuổi thọ trung bình của bóng đèn do họ sản xuất là không dưới 1600 giờ. Với mẫu trên, bạn hãy cho kết luận về lời tuyên bố của công ty, ở mức ý nghĩa 4%.

Giải

Gọi X là biến ngẫu nhiên chỉ tuổi thọ của loại bóng đèn do công ty A sản suất. Đây là bài toán kiểm định trung bình phía trái với $n = 64, \overline{x} = 1570, \hat{s} = 121$ và phương sai chưa biết.
Chọn giả thiết:

$$H_0 : \mu = \mu_0 = 1600; \overline{H_0} : \mu < \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{64}}{S} \sim t(63).$$

Với ý nghĩa $\alpha = 4\%$,miền bác bỏ

$$W_\alpha = (-\infty; -t_{0,04}(63)) = (-\infty; -1,76).$$

Tính giá trị của thống kê K_2

$$k_2 = \frac{(1570 - 1600).\sqrt{64}}{121} = -1,983 \in W_\alpha$$

nên ta bác bỏ giả thiết và chấp nhận đối thiết.
Vậy lời tuyên bố của công ty là chưa hợp lý.
Bài 7.7 Khảo sát chiều cao của các em bé 10 tuổi ở một tỉnh, người ta chọn một mẫu 75 đứa bé và được số liệu:

Chiều cao	<30	130-135	135-140	140-145	>145
Số em	5	15	30	20	5

Với độ tin cậy 95% có thể khẳng định chiều cao trung bình học sinh của tỉnh này lớn hơn 137cm được không?

Giải

Gọi X là biến ngẫu nhiên chỉ chiều cao của mỗi học sinh ở tỉnh đó, X có phân phối xác suất chưa biết và kích thước mẫu khá lớn $n = 75$.
Chọn cặp giả thiết

$$H_0 : \mu = \mu_0 = 137; \overline{H}_0 : \mu > \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{75}}{\hat{S}} \sim t(74).$$

Với độ tin cậy 95%, ta có miền bác bỏ

$$W_\alpha = (t_{0,05}, +\infty) = (1,645, +\infty).$$

Với mẫu trên ta tính được $\overline{x} = 131,167; \hat{s} = 27,492$. Từ đó tính giá trị của thống kê K_2.

$$k_2 = \frac{(131,167 - 137)\sqrt{75}}{27,492} = -1,837 \notin W_\alpha,$$

nên ta chấp nhận giả thiết, nghĩa là chiều cao của học sinh không lớn hơn 137cm.
Bài 7.8 Giám đốc một xí nghiệp cho biết lương trung bình của một công nhân thuộc xí nghiệp hiện nay là 600 ngàn đồng/tuần. Chọn ngẫu nhiên 36 công nhân thấy lương trung bình là 520 ngàn đồng/tuần, với độ lệch chuẩn $\sigma = 40$ ngàn đồng/tuần. Lời báo cáo của giám đốc có tin cậy được không, với mức ý nghĩa $\alpha = 5\%$.

Giải

Gọi X là biến ngẫu nhiên chỉ mức lương của mỗi công nhân, X có phân phối chưa biết và cỡ mẫu là $n = 36 > 30$.

Đây là bài toán kiểm định trung bình với phương sai chưa biết.
Cặp giả thiết kiểm định

$$H_0 : \mu = \mu_0 = 600; \overline{H}_0 : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{36}}{\hat{S}} \sim t(35).$$

Với mức ý nghĩa $\alpha = 0,05$, ta có miền bác bỏ:

$$W_\alpha = (-\infty, -t_{0,05}(35)) \cup (t_{0,05}(35), +\infty) = (-\infty, -1.96) \cup (1.96, +\infty).$$

Với mẫu trên ta có $\overline{x} = 520, \hat{s} = 40$. Từ đó tính giá trị của thống kê

$$k_2 = \frac{(520 - 600).6}{40} = -12 \in W_\alpha,$$

nên ta bác bỏ H_0, nghĩa là lời tuyên bố của giám đốc không thể tin được.
<u>Bài 7.9</u> Một nhà máy sản xuất săm lốp ô tô tuyên bố rằng tuổi thọ trung bình một chiếc lốp ô tô của họ là 30000 dặm. Cơ quan giám định chất lượng muốn kiểm tra tính chính xác của lời tuyên bố này đã lấy ngẫu nhiên 100 chiếc lốp để thực nghiêm và tìm được trung bình mẫu 29000 dặm với độ lệch chuẩn là 5000 dặm.

a) Với $\alpha = 0,05$, cơ quan giám định có bác bỏ được lời tuyên bố trên không?

b) Với $\alpha = 2\%$, cơ quan giám định có bác bỏ được lời tuyên bố trên không?

Giải

Gọi X là biến ngẫu nhiên chỉ tuổi thọ của lốp ô tô.
a) Kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 30000; \overline{H}_0 : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{100}}{\hat{S}} \sim t(99).$$

Với mức ý nghĩa $\alpha = 0,05$, miền bác bỏ

$$W_{\frac{\alpha}{2}} = \left(-\infty, -t_{\frac{0,05}{2}}(99)\right) \cup \left(t_{\frac{0,05}{2}}(99), +\infty\right) = (-\infty, -1,96) \cup (1.96, +\infty),$$

với mẫu trên ta có $\overline{x} = 29000; \hat{s} = 5000$ và giá trị của thống kê K_2

$$k_2 = \frac{(29000 - 30000).\sqrt{100}}{5000} = -2 \in W_{\frac{\alpha}{2}},$$

nên ta bác bỏ H_0, nghĩa là tuyên bố chưa chính xác.
b)Kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 30000; \overline{H}_0 : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{100}}{\hat{S}} \sim t(99).$$

Với mức ý nghĩa $\alpha = 0,02$, miền bác bỏ

$$W_{\frac{\alpha}{2}} = \Big(-\infty, -t_{\frac{0.02}{2}}(99)\Big) \cup \Big(t_{\frac{0.02}{2}}(99), +\infty\Big) = (-\infty, -2.33) \cup (2.33, +\infty).$$

Theo mẫu trên $\overline{x} = 29000$, $\hat{s} = 5000$ và tính giá trị của thống kê

$$k_2 = \frac{(29000 - 30000).\sqrt{100}}{5000} = -2 \notin W_{\frac{\alpha}{2}},$$

nên ta không thể bác bỏ H_0, nghĩa là tuyên bố chính xác..
<u>Bài 7.10</u> Mức quy định cho mỗi gói bánh được đóng gói tự động là 225g. Kiểm tra ngẫu nhiên 81 gói, thấy khối lượng trung bình là 210g, với độ lệch chuẩn mẫu là 36g. Với mức ý nghĩa $\alpha = 0.01$, hãy cho kết luận về tình hình sản suất.

Giải

Đây là bài toán kiểm định trung bình với phương sai chưa biết.
Gọi X là biến ngẫu nhiên chỉ khối lượng của mỗi gói bánh.
Chúng ta kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 225; \overline{H}_0 : \mu \neq \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{81}}{\hat{S}} \sim t(80).$$

Với mức ý nghĩa $\alpha = 0,01$, miền bác bỏ là

$$W_{\frac{\alpha}{2}} = \Big(-\infty, -t_{\frac{0,01}{2}}(80)\Big) \cup \Big(t_{\frac{0,01}{2}}(80), +\infty\Big) = (-\infty, -2.58) \cup (2.58, +\infty).$$

Với mẫu trên ta có $n = 81, \overline{x} = 210, \hat{s} = 36$. Khi đó

$$k_2 = \frac{(225 - 210)\sqrt{81}}{36} = 3,75 \in W_{\frac{\alpha}{2}},$$

nên ta bác bỏ giả thiết H_0.
Vậy trọng lượng đóng gói thay đổi.

Bài 7.11 Được biết nhịp đập trung bình của nam thanh niên là 72 lần/phút. Kiểm tra 64 thanh niên làm việc trong hầm lò, thấy nhịp mạch trung bình của họ là 74 lần/ phút với độ lệch chuẩn mẫu là 9 lần/phút. Hãy xét xem làm việc trong hầm lò có làm tăng nhịp mạch hay không, hãy kết luận với mức ý nghĩa $\alpha = 0,01$.

Giải

Đây là bài toán kiểm định trung bình với phương sai chưa biết.
Gọi X là biến ngẫu nhiên chỉ nhịp mạch của nam thanh niên làm việc trong hầm lò.
Chúng ta kiểm định cặp giả thiết

$$H_0 : \mu = \mu_0 = 72; \overline{H}_0 : \mu > \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{64}}{\hat{S}} \sim t(63).$$

Với $\alpha = 0,01$, miền bác bỏ là:

$$W_\alpha = (t_{0,01}(63), +\infty) = (2,33, +\infty).$$

Theo mẫu trên ta có $\overline{x} = 74; \hat{s} = 9, n = 81$. Khi đó

$$k_2 = \frac{(74 - 72)\sqrt{81}}{9} = 2 \notin W_\alpha,$$

nên ta chấp nhận giả thiết.
Vậy nhịp mạch thanh niên làm việc trong hầm lò không có xu hướng tăng.

Bài 7.12 Trong điều kiện chăn nuôi bình thường, lượng sữa trung bình của một con bò là 14kg/ ngày. Nghi ngờ điều kiện chăn nuôi kém làm cho lượng sữa giảm, người ta tiến hành điều tra ngẫu nhiên 50 con và tính được lượng sữa trung bình của một con trong một ngày là 12.5kg và độ lệch chuẩn mẫu là 2.5. Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận điều nghi ngờ nói trên.

Giải

Đây là bài toán kiểm định trung bình với phương sai chưa biết và $n = 50 > 30$.
Gọi X là biến ngẫu nhiên chỉ lượng sữa của mỗi con bò.
Kiểm định giả thiết

$$H_0 : \mu = \mu_0 = 14; \overline{H}_0 : \mu < \mu_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_2 = \frac{(\overline{X} - \mu_0)\sqrt{50}}{\hat{S}} \sim t(49).$$

Với $\alpha = 0,05$, miền bác bỏ là

$$W_\alpha = (-\infty, -t_{0,05}(49)) = (-\infty, -1,65).$$

Theo mẫu trên ta có $\overline{x} = 12,5; n = 50; \hat{s} = 2,5$ nên

$$k_2 = \frac{(12,5 - 14)\sqrt{50}}{2,5} = -4,243 \in W_\alpha,$$

nên ta bác bỏ giả thiết, nghĩa là lượng sữa có xu hướng giảm.
Vậy điều nghi ngờ trên là có cơ sở.

Kiểm định tỉ lệ tổng thể

Bài 7.13 Tại một địa phương, bệnh B có tỉ lệ 34%. Sau một đợt điều trị, kiểm tra lại trên 100 người, thấy có 24 người bị bệnh B. Hỏi đợt điều trị có thực sự làm giảm tỉ lệ bệnh B ở địa phương trên không? (kết luận ở mức ý nghĩa 5%).

Giải

Đây là bài toán kiểm định tỉ lệ, mẫu có kích thước $n = 100$.
Gọi p là tỉ lệ bệnh B ở địa phương sau đợt điều trị.
Kiểm định giả thiết:

$$H_0 : p = p_0 = 0,34; \overline{H_0} : p < p_0 = 0,34.$$

Tần suất bệnh B của mẫu:

$$f_n = \frac{24}{100} = 0,24.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{100}}{\sqrt{p_0(1-p_0)}} \sim N(0,1).$$

Với mức ý nghĩa $\alpha = 5\%$, miền bác bỏ là

$$W_\alpha = (-\infty; -z_{\frac{1}{2}-0,05}) = (-\infty; -1,65).$$

Giá trị thống kê K_3

$$k_3 = \frac{(0,24 - 0,34)\sqrt{100}}{\sqrt{0,34.0,66}} = -2,111 \notin W_\alpha$$

nên ta bác bỏ giả thiết, nghĩa là sau đợt điều trị thực sự có làm giảm tỉ lệ bệnh B tại địa phương.

Bài 7.14 Giả sử một huyện năm trước có tỷ lệ trẻ em bị suy dinh dưỡng là 10%, năm nay huyện thực hiện nhiều chính sách nhằm làm giảm tỷ lệ này xuống. Chọn 400 đứa trẻ, kiểm tra thấy có 32 đứa trẻ vẫn còn bị suy dinh dưỡng. Với mức ý nghĩa 1% hãy cho kết luận về việc giảm tỷ lệ trẻ em suy dinh dưỡng của huyện này.

Giải

Đây là bài toán kiểm định tỷ lệ với kích thước mẫu $n = 400$.
Gọi p là tỷ lệ trẻ em bị suy dinh dưỡng sau khi thực hiện chính sách.
Kiểm định cặp giả thiết:

$$H_0 : p = p_0 = 0,1; \overline{H}_0 : p < p_0.$$

Tần suất trẻ bị suy dinh dưỡng trên mẫu: $f_n = \dfrac{32}{400} = 0,08$.
Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{400}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với mức $\alpha = 0,01$, ta có miền bác bỏ:

$$W_\alpha = (-\infty, -z_{\frac{1}{2}-0,01}) = (-\infty, -2,33)$$

Tính giá trị thống kê K_3

$$k_3 = \frac{(0,08 - 0,1)\sqrt{400}}{\sqrt{0,1.0,9}} = -1,333 \notin W_\alpha,$$

nên ta chấp nhận giả thiết H_0.
Vậy tỉ lệ trẻ em suy dinh dưỡng vẫn chưa giảm.
<u>Bài 7.15</u> Theo một nguồn tin thì tỉ lệ hộ dân thích xem dân ca trên tivi là 80%. Thăm dò 36 hộ dân thấy có 25 hộ thích xem dân ca. Với mức ý nghĩa là 5%. Kiểm định xem nguồn tin này có đáng tin cậy không?

Giải

Đây là bài toán kiểm định tỷ lệ với mẫu có kích thước $n = 36$.
Gọi p là tỷ lệ hộ dân thích xem dân ca trên tivi.
Cặp giả thiết

$$H_0 : p = p_0 = 0,8; \overline{H}_0 : p \neq p_0.$$

Giá trị tần suất dân thích xem dân ca trên mẫu $f_n = 0,694$.
Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{36}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,05$, ta có miền bác bỏ là

$$W_{\frac{\alpha}{2}} = \left(-\infty, -z_{\frac{1-0,05}{2}}\right) \cup \left(z_{\frac{1-0,05}{2}}, +\infty\right) = (-\infty, -1.96) \cup (1.96, +\infty).$$

Với mẫu trên ta có

$$k_3 = \frac{(0,694 - 0,8).\sqrt{36}}{\sqrt{0,8.0,2}} = -1,59 \notin W_{\frac{\alpha}{2}},$$

nên ta chấp nhận giả thiết, nghĩa là nguồn tin này đáng tin cậy.
<u>Bài 7.16</u> Tỷ lệ phế phẩm của một nhà máy trước đây là 5%. Năm nay nhà máy áp dụng một biện pháp kĩ thuất mới. Để nghiên cứu tác dụng của biện pháp kĩ thuật mới, người ta lấy một mẫu gồm 800 sản phẩm để kiểm tra và thấy có 24 phế phẩm.

a) Với $\alpha = 1\%$ hãy cho kết luận về biện pháp kĩ thuật mới này?

b) Nếu nhà máy báo cáo tỷ lệ phế phẩm sau khi áp dụng biện pháp kỹ thuật mới là 2% thì có chấp nhận được không?($\alpha = 0.05$).

Giải

Gọi p là tỷ lệ phế phẩm sau khi áp dụng kỹ thuật mới.
a) Kiểm định giả thiết

$$H_0 : p = p_0 = 0 = 0,05; \overline{H}_0 : p < p_0.$$

Giá trị tần suất trên mẫu $f_n = 0,03$.
Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{800}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,01$, ta có miền bác bỏ

$$W_\alpha = (-\infty, -z_{\frac{1}{2}-0,01}) = (-\infty, -2,33).$$

Với mẫu trên ta có

$$k_3 = \frac{(0,03 - 0,05)\sqrt{800}}{\sqrt{0,05.0,95}} = -2,596 \in W_\alpha,$$

nên ta bác bỏ H_0, nghĩa là biện pháp kỹ thuật mới làm giảm tỷ lệ phế phẩm.
b) Kiểm định giả thiết

$$H_0 : p = p_0 = 0,02; \overline{H}_0 : p \neq p_0.$$

Giá trị tần suất trên mẫu $f_n = 0,03$.
Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{800}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,05$, ta có miền bác bỏ

$$W_{\alpha/2} = (-\infty, -z_{\frac{1-0,05}{2}}) \cup (z_{\frac{1-0,05}{2}}, +\infty) = (-\infty, -1.96) \cup (1.96, +\infty).$$

Với mẫu trên ta có

$$k_3 = \frac{(0,03-0,02)\sqrt{800}}{\sqrt{0,02.0,98}} = 2,02 \in W_{\frac{\alpha}{2}},$$

nên ta bác bỏ H_0, nghĩa là báo cáo trên không được chấp nhận.
<u>Bài 7.17</u> Một báo cáo nói rằng 18% gia đình ở thành phố HCM có máy tính cá nhân ở nhà. Người ta chọn ngẫu nhiên 80 gia đình trong thành phố có trẻ em đang học và thấy có 22 gia đình có máy tính. Với mức ý nghĩa 2%, hãy kiểm định xem liệu trong các gia đình có trẻ em đang đi học, tỷ lệ gia đình có máy tính cao hơn tỷ lệ chung hay không?

Giải

Đây là bài toán kiểm định tỷ lệ với mẫu có kích thước $n = 80$.
Gọi p là tỷ lệ gia đình có trẻ em có máy tính.
Tỉ lệ mẫu $f_n = \dfrac{22}{80} = 0,275$.
Kiểm định giả thiết

$$H_0 : p = p_0 = 0,18; \overline{H}_0 : p > p_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{80}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với mức $\alpha = 0,02$, miền bác bỏ là

$$W_\alpha = \left(z_{\frac{1}{2}-0,02}, +\infty\right) = (2,06, +\infty).$$

Với mẫu trên ta có $f_n = 0,275$,

$$k_3 = \frac{(0,275-0,18)\sqrt{80}}{\sqrt{0,18.0,82}} = 2,212 \in W_\alpha,$$

nên ta bác bỏ H_0, nghĩa là tỷ lệ gia đình có trẻ em có máy tính cao hơn tỷ lệ chung.
<u>Bài 7.18</u> Máy sản suất tự động với tỷ lệ chính phẩm là 98%. Sau một thời gian hoạt động, người ta nghi ngờ tỉ lệ trên đã giảm. Kiểm tra ngẫu nhiên 500 sản phẩm thấy có 28 phế phẩm. Với $\alpha = 0,05$ hãy kiểm tra chất lượng làm việc của máy có còn được như trước không?

Giải

Đây là bài toán kiểm định tỉ lệ, kích thước mẫu là $n = 500$.

Gọi p là tỉ lệ chính phẩm do máy sản suất.
Kiểm định giả thiết

$$H_0 : p = p_0 = 0,98; \overline{H}_0 : p < p_0.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{500}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với $\alpha = 0.05$, miền bác bỏ là

$$W_\alpha = (-\infty, -z_{\frac{1-0,05}{2}}) = (-\infty, -1.96).$$

Với mẫu trên ta có tỉ lệ mẫu $f_n = \dfrac{500-28}{500} = 0,944$ và

$$k_3 = \frac{(0,944 - 0,98)\sqrt{500}}{\sqrt{0,98.0,02}} = -5,750 \in W_\alpha,$$

nên ta bác bỏ giả thiết. Vậy chất lượng làm việc của máy đã giảm.
Bài 7.19 Hàm lượng dầu lý thuyết của một loại cây bằng 0,5 (sau một thời gian chăm bón). Lấy ngẫu nhiên một mẫu gồm 100 cây, thấy có tỉ lệ dầu $f_n = 0,7$. Kiểm định giả thiết $H_0 : p = 0,5$ với đối thiết $H_1 : p \neq 0,5$; với mức ý nghĩa $\alpha = 0,01$.

Giải

Kiểm định giả thiết $H_0 : p = p_0 = 0,5$ với đối thiết $H_1 : p \neq 0,5$. Nếu H_0 đúng thì biến ngẫu nhiên

$$K_3 = \frac{(f_n - p_0)\sqrt{100}}{\sqrt{p_0(1-p_0)}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,01$, miền bác bỏ là:

$$W_\alpha = (-\infty, -z_{\frac{1-0,01}{2}}) \cup (z_{\frac{1-0,01}{2}}, +\infty) = (-\infty, -2,58) \cup (2,58, +\infty).$$

Với mẫu trên ta có $f_n = 0,7; n = 100$ và

$$k_3 = \frac{(0,7 - 0,5)\sqrt{100}}{\sqrt{0,5.0,5}} = 4 \in W_\alpha,$$

nên ta bác bỏ giả thiết.
Vậy hàm lượng dầu của cây khác 0,5.
Bài 7.20 Tỉ lệ mắc bệnh sốt rét ở một huyện miền núi là 0,07, trong lần kiểm tra sức khỏe ngẫu nhiên 350 người thấy có 30 người mang vi trùng sốt rét. Với mức ý nghĩa 0,05 có thể khẳng định tỉ lệ mắc bệnh sốt rét trong vùng có thay đổi hay

không?

Giải

Gọi p là tỉ lệ người mắc bệnh sốt rét.
Với mẫu trên ta có tỉ lệ người mắc bệnh sốt rét trên mẫu là $f_n = \frac{30}{350}$ Ta kiểm định giả thiết

$$H_0 : p = 0,07; \overline{H}_0 : p \neq 0,07.$$

Với mức ý nghĩa $\alpha = 0,05$, miền bác bỏ là:

$$W_{\frac{\alpha}{2}} = \left(-\infty, -z_{\frac{1-0,05}{2}}\right) \cup \left(z_{\frac{1-0,05}{2}}, +\infty\right) = (-\infty, -1,96) \cup (1,96, +\infty).$$

Tính giá trị của thống kê K_3

$$k_3 = \frac{(\frac{30}{350} - 0,07)\sqrt{350}}{\sqrt{0,07.0,93}} = 1,152 \notin W_{\frac{\alpha}{2}},$$

nên ta chấp nhận giả thiết.
Vậy tỉ lệ mắc bệnh sốt rét trong vùng không thay đổi.
Bài 7.21 Tỉ lệ bệnh nhân khỏi bệnh T khi điều trị bằng thuốc A là 85%. Thí nghiệm dùng loại thuốc B để chữa bệnh thì trong số 900 người mắc bệnh T có 810 người được chữa khỏi bệnh. Như vậy, với mức ý nghĩa $\alpha = 0,05$ có thể kết luận là thuốc B hiệu quả hơn loại thuốc A hay không?

Giải

Đây là bài toán kiểm định tỉ lệ, kích thước mẫu ngẫu nhiên là $n = 900$.
Gọi p là tỉ lệ bệnh nhân khỏi bệnh khi dùng thuốc B.
Kiểm định giả thiết

$$H_0 : p = 0,85; \overline{H}_0 : p > 0,85.$$

Với $\alpha = 0.05$, miền bác bỏ là:

$$W_\alpha = (z_{\frac{1}{2}-0,05}, +\infty) = (1,65, +\infty).$$

Với mẫu trên ta có $f_n = \frac{810}{900} = 0,9$ và

$$u = \frac{(0,9 - 0,85)\sqrt{900}}{\sqrt{0,85.0,15}} = 4,201 \in W_\alpha,$$

nên ta bác bỏ giả thiết.
Vậy thuốc B hiệu quả hơn thuốc A

Kiểm định giả thiết về sự bằng nhau của trung bình hai tổng thể

Dạng 1: Biết phương sai hai tổng thể
Bài 7.22 Người ta cho hai nhóm học sinh, theo thứ tự, đại diện cho hai trường A

và B, làm một bài kiểm tra. Nhóm thứ nhất gồm 40 học sinh, có điểm trung bình 7,4; nhóm thứ hai gồm 50 học sinh, có điểm trung bình 7,8. Dựa vào mẫu trên, có thể kết luận rằng: điểm trung bình của trường B tốt hơn điểm trung bình của trường A không? (kết luận ở mức ý nghĩa 4%). Biết rằng điểm số của mỗi học sinh của hai trường A và B có phân phối chuẩn với độ lệch chuẩn, theo thứ tự, là 0,8 và 0,7.

Giải

Gọi X, Y tương ứng là biến ngẫu nhiên chỉ điểm số của mỗi học sinh trường A và trường B.
Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với

$$n = 40; \overline{x} = 7,4; \sigma_X = 0,8;$$

$$m = 50; \overline{y} = 7,8; \sigma_Y = 0,7.$$

Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y,$$

$$\overline{H}_0 : \mu_X < \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_6 = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}(0, 1).$$

Miền bác bỏ

$$W_\alpha = (-\infty; -z_{\frac{1}{2}-0,04}) = (-\infty, -1,76).$$

Với mẫu trên, chúng ta có

$$u = \frac{7,4 - 7,8}{\sqrt{\frac{0,8^2}{40} + \frac{0,7^2}{50}}} = -2,490 \in W_\alpha,$$

nên ta bác bỏ H_0, nghĩa là điểm trung bình của học sinh trường B cao hơn học sinh trường A.

Bài 7.23 So sánh mức lương thu nhập theo tuần ở một công ty liên doanh giữa nam và nữ, ta có số liệu mẫu như sau: Nữ: chọn một mẫu 40 người, tính được thu nhập trung bình là 130\$, Nam: chọn một mẫu 50 người, tính được thu nhập trung bình 140\$. Biết rằng phương sai thu thập theo tuần của nữ là 80 và của nam là 100. Với mức ý nghĩa 5%, có thể kết luận thu nhập trung bình của nữ thấp hơn nam được không?

Giải

Gọi X, Y lần lượt là biến ngẫu nhiên chỉ thu nhập của nữ và nam. Đây là bài toán

kiểm định sự bằng nhau trung bình hai tổng thể với $\sigma_1^2 = 80; \sigma_2^2 = 100$.
Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y, \overline{H}_0 : \mu_X < \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_6 = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{40} + \frac{\sigma_Y^2}{50}}} \sim \mathcal{N}(0, 1).$$

Với mức ý nghĩa $\alpha = 0,05$, miền bác bỏ là

$$W_\alpha = (-\infty, -z_{\frac{1}{2}-0,05}) = (-\infty, -1,6449).$$

Với mẫu trên ta có:

$$\overline{x} = 130; \overline{y} = 140$$

Khi đó

$$k_6 = \frac{130 - 140}{\sqrt{\frac{80}{40} + \frac{100}{50}}} = -5 \in W_\alpha,$$

nên ta bác bỏ H_0, nghĩa là thu nhập của nam lớn hơn thu nhập của nữ.
Bài 7.24 Trọng lượng sản phẩm do hai nhà máy sản xuất là các đại lượng ngẫu nhiên phân phối theo quy luật chuẩn và có độ lệch tiêu chuẩn là $\sigma = 1kg$. Với mức ý nghĩa 5% có thể xem trọng lượng trung bình của sản phẩm do hai nhà máy sản xuất ra là như nhau hay không? Nếu cân thử 25 sản phẩm của nhà máy A ta tính được $\overline{x} = 50kg$; Cân 20 sản phẩm của nhà máy B thì tính được $\overline{y} = 50,6kg$.

Giải

Gọi X, Y là biến ngẫu nhiên chỉ trọng lượng sản phẩm do nhà máy A và nhà máy B sản suất.
Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với phương sai $\sigma_X^2 = \sigma_Y^2 = 1$
Chúng ta kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y \text{và } \overline{H}_0 : \mu_X \neq \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_6 = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{25} + \frac{\sigma_Y^2}{20}}} \sim \mathcal{N}(0, 1),$$

với mức ý nghĩa $\alpha = 0,05$, miền bác bỏ là

$$W_{\alpha/2} = (-\infty, -z_{\frac{1-0,05}{2}}) \cup (z_{\frac{1-0,05}{2}}, +\infty) = (-\infty, -1,96) \cup (1,96, +\infty).$$

Với mẫu trên ta có $\overline{x} = 50; \overline{y} = 50,6$ và

$$k_6 = \frac{50 - 50,6}{\sqrt{\frac{1}{25} + \frac{1}{20}}} = -2 \in W_{\alpha/2},$$

nên ta bác bỏ H_0, nghĩa là trọng lượng trung bình sản phẩm do hai nhà máy sản suất là khác nhau.

Bài 7. 25 Sau 30 lần quan sát thí nghiệm, người ta nhận thấy thời gian làm việc trung bình của một linh kiện điện tử là 60 giờ. Trong cùng một điều kiện, người ta tiến hành quan sát 20 lần khi chi tiết đó chưa được cải tiến, thì thời gian làm việc trung bình của chi tiết là 55 giờ. Biết rằng $\sigma_1 = 7$ và $\sigma_2 = 5$, hỏi $\alpha = 0,001$; sự cải tiến có thực sự có tác dụng hay không?

Giải

Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với phương sai $\sigma_1 = 7$ và $\sigma_2 = 5$.

Gọi X, Y lần lượt là thời gian làm việc của linh kiện điện tử trước và sau cải tiến. Khi đó

$$X \sim \mathcal{N}(\mu_X, 5^2); Y \sim \mathcal{N}(\mu_Y, 7^2).$$

Kiểm định gỉa thiết

$$H_0 : \mu_X = \mu_Y; \overline{H}_0 : \mu_X < \mu_Y.$$

Với mức ý nghĩa $\alpha = 0,001$, miền bác bỏ là:

$$W_\alpha = (-\infty, -z_{\frac{1}{2}-0,001}) = (-\infty, -3,1).$$

Với mẫu trên ta có

$$\overline{x} = 55; n = 20; \sigma_2 = 5$$

$$\overline{y} = 60; n = 30; \sigma_1 = 7.$$

Khi đó

$$k_6 = \frac{55 - 60}{\sqrt{\frac{5^2}{20} + \frac{7^2}{30}}} = -2,945 \notin W_\alpha,$$

nên ta chấp nhận giả thiết.

Vậy sự cải tiến thực sự không có tác dụng.

Dạng 2: Không biết phương sai

Bài 7.26 Trong một công ty sản xuất bóng đèn, người ta muốn kiểm tra sự làm việc của hai phân xưởng A và B. Một mẫu gồm $n = 8$ bóng đèn của phân xưởng A cho tuổi thọ trung bình 4000 giờ với độ lệch chuẩn 200 giờ. Một mẫu gồm $m = 10$ bóng đèn của phân xưởng B cho tuổi thọ trung bình 4300 giờ với độ lệch chuẩn 250 giờ. Biết rằng tuổi thọ của mỗi bóng đèn của hai phân xưởng A và B, theo thứ tự, tuân theo luật phân phối chuẩn có cùng phương sai. Hãy cho kết luận về sự khác nhau giữa tuổi thọ trung bình của hai loại bóng đèn trên ở mức ý nghĩa 1%.

Giải

Gọi X, Y lần lượt là biến ngẫu nhiên chỉ tuổi thọ của bóng đèn do phân xưởng A và phân xưởng B sản suất.

Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với kích thước $n = 8, m = 20$.

Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y,$$

$$\overline{H}_0 : \mu_X \neq \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_7 = \frac{\overline{X} - \overline{Y}}{\sqrt{\hat{S}^2(\frac{1}{10} + \frac{1}{8})}} \sim t(16)$$

Miền bác bỏ

$$W_{\alpha/2} = (-\infty; -t_{\frac{0,01}{2}}(16)) \cup (t_{\frac{0,01}{2}}(16); +\infty) = (-\infty, -2,9208) \cup (2,9208, +\infty).$$

Với mẫu trên ta có:

$$\hat{s}^2 = \frac{9S_X^2 + 7s_Y^2}{10 + 8 - 2} = 49843,75;$$

$$k_7 = \frac{\overline{x} - \overline{y}}{\sqrt{s^2(\frac{1}{10} + \frac{1}{8})}} = \frac{4000 - 4300}{\sqrt{49843,75 \times 0,225}} = -2,8329 \notin W_\alpha.$$

nên với mức ý nghĩa $\alpha = 0,01$ ta không thể bác bỏ H_0.

Vậy chúng ta kết luận rằng: với mức ý nghĩa 1% sự khác nhau về tuổi thọ trung bình của hai loại bóng đèn trên là không có ý nghĩa về mặt thống kê.

Bài 7.27 Khảo sát chiều cao của học sinh nữ tại hai trường phổ thông trung học huyện A và huyện B ta có số liệu:

Chiều cao (cm)	Số học sinh trường A	Số học sinh trường B
150-152	3	5
152-154	5	10
154-156	7	14
156-158	15	18
158-160	26	22
160-162	25	11
162-164	12	9
164-166	13	5
166-168	10	4
168-170	5	2

a) Với mức ý nghĩa 1% có thể xem chiều cao trung bình học sinh trung học nữ của huyện A cao hơn huyện B được không?

b) Những học sinh có chiều cao nhỏ hơn 154cm được xem là thấp. Giả sử chiều cao học sinh nhóm thấp ở hai trường là đại lượng ngẫu nhiên có phân phối chuẩn có phương sai xấp xỉ bằng nhau. Một người nói chiều cao trung bình học sinh thấp của hai trường là như nhau thì có đúng không với độ tin cậy 95%.

Giải

a) Gọi X, Y lần lượt là chiều cao của học sinh nữ tại trường A và B.
Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với phương sai chưa biết và $n = 121; m = 100$.
Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y; \overline{H}_0 : \mu_X > \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_7 = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\hat{S}_X^2}{n} + \frac{\hat{S}_Y^2}{m}}} \sim t(219).$$

Miền bác bỏ

$$W_\alpha = (t_{0,01}(219); +\infty) = (2,326, +\infty).$$

Với mẫu trên ta có

$$n = 121; \overline{x} = 160.603; \hat{s}_X = 4.216$$

$$m = 100; \overline{y} = 158,48; \hat{s}_Y = 5,777$$

và

$$k_7 = \frac{160,603 - 158,48}{\sqrt{\frac{4,216^2}{121} + \frac{5,777^2}{100}}} = 3,062 \in W_\alpha,$$

nên ta bác bỏ giả thiết.
Vậy chiều cao trung bình của học sinh nữ trường A cao hơn học sinh trường B.
b) Gọi X, Y lần lượt là chiều cao của học sinh nữ tại trường A và B nỏ hơn 154cm.
Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với phương sai chưa biết và $n = 121; m = 100$.
Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y; \overline{H}_0 : \mu_X \neq \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_7 = \frac{\overline{X} - \overline{Y}}{\sqrt{\hat{S}^2(\frac{1}{121} + \frac{1}{100})}} \sim t(219).$$

Với độ tin cậy 95% ta có, miền bác bỏ

$$W_{\alpha/2} = (-\infty, -t_{\frac{0,05}{2}}(219)) \cup (t_{\frac{0,05}{2}}(219), +\infty) = (-\infty, -2,08) \cup (2,08, +\infty).$$

Với mẫu trên ta có

$$n = 8; \overline{x} = 152,25; \hat{s}_X = 1,035$$

$$m = 15; \overline{y} = 152,333; \hat{s}_Y = 0,352$$

và

$$\hat{s}^2 = \frac{7\hat{s}_X^2 + 14\hat{s}_Y^2}{8 + 15 - 2} = 0,44$$

$$k_7 = \frac{152,25 - 152,333}{\sqrt{0,44(\frac{1}{8} + \frac{1}{15})}} = -0,125 \notin W_{\alpha/2},$$

nên ta chấp nhận giả thiết.
Vậy chiều cao trung bình của học sinh nữ có chiều cao nhỏ hơn 154cm của hai trường là như nhau.
Bài 7.28 Bắn thử nghiệm 100 viên đạn tên lửa loại mới thấy khoảng cách tầm bắn trung là 68.9 km, trong khi nếu bắn 90 quả đạn tên lửa loại cũ thì tầm bắn trung bình là 61,52 km. Cho biết $S_1^2 = 49,7$ và $S_2^2 = 58,8$. Hỏi sự cải tiến có thực sự hiệu quả với độ tin cậy 0,99.

Giải

Gọi X, Y lần lượt là biến ngẫu nhiên chỉ tầm bắn của tên lửa loại cũ và loại mới.
Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với phương sai

chưa biết và $n = 100; m = 90$.
Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y; \overline{H}_0 : \mu_X < \mu_Y.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_7 = \frac{\overline{X} - \overline{Y}}{\sqrt{\hat{S}^2(\frac{1}{100} + \frac{1}{90})}} \sim t(188).$$

Với độ tin cậy 0.99, ta có miền bác bỏ là:

$$W_\alpha = (-\infty, -t_{0,01}(188)) = (-\infty, -2,33).$$

Với mẫu trên ta có

$$\overline{x} = 61,52; n = 90; S_2^2 = 58,8;$$

$$\overline{y} = 68,9; m = 100; S_1^2 = 49,7$$

và

$$\hat{s}^2 = \frac{89 \times 58,8 + 99 \times 49,7}{90 + 100 - 2} = 54,008;$$

$$k_7 = \frac{61,52 - 68,9}{\sqrt{54,008(\frac{1}{90} + \frac{1}{100})}} = -6,911 \in W_\alpha,$$

nên ta bác bỏ giả thiết. Vậy sự cải tiến thực sự có hiệu quả.
Bài 7.29 Một giống lúa được trồng trên hai thửa ruộng với chế độ phân bón khác nhau. Khi thu hoạch, ta có kết quả như sau: Thửa ruộng thứ nhất lấy 1000 bông lúa thì đếm được số hạt trung bình của mỗi bông lúa là $\overline{X} = 70; S_X = 10$ còn thửa thứ hai lấy mẫu 500 bông lúa với số hạt trung bình của mỗi bông là $\overline{Y} = 72; S_Y = 20$. Với mức ý nghĩa $\alpha = 0,05$; hỏi $\overline{X}$ và $\overline{Y}$ có khác nhau không?

Giải

Đây là bài toán kiểm định sự bằng nhau trung bình hai tổng thể với phương sai chưa biết và $n = 1000; m = 500$.
Kiểm định giả thiết

$$H_0 : \mu_X = \mu_Y; \overline{H}_0 : \mu_X \neq \mu_Y.$$

Với $\alpha = 0,05$, miền bác bỏ là:

$$W_{\alpha/2} = (-\infty, -t_{\frac{0,05}{2}}(1498)) \cup (t_{\frac{0,05}{2}}(1498); +\infty) = (-\infty, -1,96) \cup (1,96, +\infty).$$

Với mẫu trên ta có

$$\overline{x} = 70; n = 1000; S_X = 10;$$

$$\overline{y} = 72; m = 500; S_Y = 20$$

và

$$\hat{s}^2 = \frac{999.100 + 499.20^2}{1000 + 500 - 2} = 199,933;$$

$$k_7 = \frac{70 - 72}{\sqrt{199,933(\frac{1}{1000} + \frac{1}{500})}} = -2,582 \in W_{\alpha/2},$$

nên ta bác bỏ gỉa thiết.
Vậy số hạt lúa trên một bông lúa của hai thửa ruộng khác nhau.

Kiểm định giả thiết về sự bằng nhau của tỉ lệ hai tổng thể

Bài 7.30 Để so sánh về tỉ lệ một loại bệnh B đối với trẻ sơ sinh trai và trẻ sơ sinh gái, người ta quan sát 100 bé trai thấy có 20 cháu mắc bệnh B; quan sát 120 bé gái thấy có 30 cháu mắc bệnh B. Hỏi tỉ lệ nhiễm bệnh B đối với bé trai và gái có như nhau không? (hãy kết luận ở mức ý nghĩa 1%).

Giải

Gọi p_1, p_2 lần lượt là tỉ lệ bệnh B của bé trai và bé gái.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2; \overline{H}_0 : p_1 \neq p_2.$$

Tỉ lệ chung

$$f = \frac{20 + 30}{220} = \frac{5}{22}.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_5 = \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1 - f)(\frac{1}{100} + \frac{1}{120})}} \sim \mathcal{N}(0, 1).$$

Miền bác bỏ

$$W_{\alpha/2} = (-\infty; -z_{\frac{1-0,01}{2}}) \cup (z_{\frac{1-0,01}{2}}; +\infty) = (-\infty; -2,58) \cup (2,58; +\infty).$$

Với mẫu trên ta có

$$f_1 = \frac{20}{100} = 0,2; f_2 = \frac{30}{120} = 0,25;$$

$$k_5 = \frac{0,2 - 0,25}{\sqrt{\frac{5}{22} \cdot \frac{17}{22} \cdot (\frac{1}{100} + \frac{1}{120})}} = -0,881 \notin W_{\alpha/2},$$

nên ta không thể bác bỏ H_0, nghĩa là tỉ lệ nhiễm bệnh B đối với bé trai và bé gái là như nhau.

Bài 7.31 Kiểm tra 100 đứa trẻ của vùng I phát hiện 42 đứa trẻ bị sâu răng. Ở vùng II có 92 đứa trẻ bị sâu răng khi kiểm tra 200 đứa trẻ. Với mức ý nghĩa 5% có thể xem tỷ lệ bị sâu răng ở hai vùng bằng nhau được không?

Giải

Gọi p_1, p_2 lần lượt là tỉ lệ trẻ em bị sâu răng ở vùng I và vùng II.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2; \overline{H}_0 : p_1 \neq p_2.$$

Tỉ lệ chung là

$$f = \frac{42+92}{100+200} = 0,447.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_5 = \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1-f)(\frac{1}{100} + \frac{1}{200})}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,05$, ta có miền bác bỏ

$$W_{\alpha/2} = (-\infty; -z_{\frac{1-0,05}{2}}) \cup (z_{\frac{1-0,05}{2}}; +\infty) = (-\infty; -1,96) \cup (1,96; +\infty).$$

Với mẫu cụ thể ta có:

$$f_1 = 0,42; f_2 = 0,46;$$

$$k_5 = \frac{0,42 - 0,46}{\sqrt{0,447.0,553.(\frac{1}{100} + \frac{1}{200})}} = -0,721 \notin W_{\alpha/2},$$

nên ta không thể bác bỏ H_0, nghĩa là tỷ lệ sâu răng của hai vùng là như nhau.

Bài 7.32 Điều tra ngẫu nhiên 200 người có hút thuốc lá, thấy có 28 người bị lao phổi; 170 người không hút thuốc lá, thấy có 12 người bị lao phổi. Tỉ lệ lao phổi giữa những người có và không hút thuốc lá có khác nhau không? (kết luân ở mức ý nghĩa $\alpha = 1\%$).

Giải

Gọi p_1, p_2 lần lượt là tỷ lệ người hút thuốc lá và người không hút thuốc lá.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2; \overline{H}_0 : p_1 \neq p_2.$$

Tỷ lệ chung là

$$f = \frac{28+12}{200+170} = \frac{4}{37}.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_5 = \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1-f)(\frac{1}{200} + \frac{1}{170})}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,01$, miền bác bỏ là

$$W_{\alpha/2} = (-\infty; -z_{\frac{1-0,01}{2}}) \cup (z_{\frac{1-0,01}{2}}; +\infty) = (-\infty; -2,58) \cup (2,58; +\infty),$$

với mẫu trên ta có $f_1 = 0,14; f_2 = 0,071;$

$$k_5 = \frac{0,14 - 0,071}{\sqrt{\frac{4}{37} . \frac{33}{37} . (\frac{1}{200} + \frac{1}{170})}} = 2,13 \notin W_{\alpha/2},$$

nên ta chấp nhận giả thiết.
Vậy tỷ lệ người hút thuốc và không hút thuốc là như nhau.
<u>Bài 7.33</u> Điều tra ngẫu nhiên 180 em học sinh ở trường trung học A, thấy có 27 em hút thuốt lá; điều tra ngẫu nhiên 150 em học sinh ở trường trung học B, thấy có 30 em hút thuốc lá. Với mức ý nghĩa 5%, có thể nói rằng tỉ lệ học sinh hút thuốc lá ở trường B cao hơn ở trường A không?

Giải

Gọi p_1, p_2 lần lượt là tỷ lệ học sinh hút thuốc trường A và trường B.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2 \text{và } \overline{H}_0 : p_1 < p_2.$$

Tỷ lệ chung là

$$f = \frac{27 + 30}{180 + 150} = 0,173$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_5 = \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1-f)(\frac{1}{180} + \frac{1}{150})}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,05$, miền bác bỏ là

$$W_\alpha = (-\infty; -z_{\frac{1}{2}-0,05}) = (-\infty; -1,65).$$

Theo mẫu trên ta có $f_1 = 0,15; f_2 = 0,2$ và

$$k_5 = \frac{0,15 - 0,2}{\sqrt{0,173.0,827.(\frac{1}{180} + \frac{1}{150})}} = -1,196 \notin W_\alpha,$$

nên ta chấp nhận giả thiết.
Vậy tỷ lệ hút thuốc lá ở hai trường là như nhau.
<u>Bài 7.34</u> Một cuộc điều tra về tỉ lệ sinh viên ra trường có việc làm trong năm đầu tiên của hai trường đại học lớn A và B cho biết: điều tra ngẫu nhiên 100 sinh viên mới ra trường của trường đại học A có 20 sinh viên có việc làm, trong 120 sinh

viên của trường đại học B có 30 sinh viên có việc làm. Hỏi tỉ lệ sinh viên có việc làm trong năm đầu tiên khi ra trường của hai trường A và B có như nhau không?($\alpha = 1\%$).

Giải

Gọi p_1, p_2 là tỷ lệ sinh viên trường A và trường B có việc làm năm đầu tiên.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2; \overline{H}_0 : p_1 \neq p_2.$$

Tỉ lệ chung là:

$$f = \frac{20 + 30}{100 + 120} = \frac{5}{22}$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_5 = \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1-f)\left(\frac{1}{100} + \frac{1}{120}\right)}} \sim \mathcal{N}(0, 1).$$

Với mức ý nghĩa $\alpha = 0,01$, miền bác bỏ là

$$W_{\alpha/2} = (-\infty; -z_{\frac{1-0,01}{2}}) \cup (z_{\frac{1-0,01}{2}}; +\infty) = (-\infty; -2,58) \cup (2,58; +\infty).$$

Với mẫu trên ta có

$$f_1 = \frac{20}{100} = 0,2; f_2 = \frac{30}{120} = 0,25$$

và

$$k_5 = \frac{0,2 - 0,25}{\sqrt{\frac{5}{22} \cdot \frac{17}{22} \cdot (\frac{1}{100} + \frac{1}{120})}} = -0,881 \notin W_{\alpha/2},$$

nên ta chấp nhận giả thiết, nghĩa là tỷ lệ sinh viên ra trường có việc làm là như nhau.

Bài 7.35 Theo dõi số tai nạn lao động trong xí nghiệp của một thành phố, ta thấy xí nghiệp I có 500 công nhân thì có 28 người bị tai nạn trong 6 tháng đầu năm. Trong cùng thời gian đó xí nghiệp II có 20 người bị tai nạn trong tổng số 400 công nhân. Với mức ý nghĩa 0,01; hỏi có sự khác nhau đáng kể về chất lượng công tác phòng hộ lao động ở hai xí nghiệp không?

Giải

Gọi p_1, p_2 lần lượt là tỉ lệ công nhân bị tai nạn lao động của xí nghiệp I và xí nghiệp II.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2; \overline{H}_0 : p_1 \neq p_2.$$

Tỉ lệ chung

$$f = \frac{28 + 20}{500 + 400} = \frac{4}{75}.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_= \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1-f)(\frac{1}{500} + \frac{1}{400})}} \sim \mathcal{N}(0,1).$$

Với mức ý nghĩa $\alpha = 0,01$, miền bác bỏ là:

$$W_{\alpha/2} = (-\infty; -z_{\frac{1-0,01}{1}}; +\infty) \cup (z_{\frac{1-0,01}{1}}; +\infty) = (-\infty; -2,58) \cup (2,58; +\infty).$$

Với mẫu trên ta có

$$f_1 = \frac{28}{500} = 0,056; f_2 = \frac{20}{400} = 0,05;$$

nên Khi đó

$$k_5 = \frac{0,056 - 0,05}{\sqrt{\frac{4}{75}.\frac{71}{75}.(\frac{1}{500} + \frac{1}{400})}} = 3,981 \in W_{\alpha/2},$$

nên ta bác bỏ giả thiết.
Vậy về chất lượng công tác phòng hộ lao động của hai xí nghiệp khác nhau.
<u>Bài 7.36</u> Dùng thuốc A điều trị căn bệnh X cho 200 bệnh nhân thì có 150 người đã khỏi bệnh; tương tự khi dùng thuốc B để điều trị cho 100 bệnh nhân mắc bệnh X thì có 70 người khỏi bệnh. Với số liệu thu được có thể nói hiệu quả của hai loại thuốc A và B là như nhau không? (Với mức ý nghĩa 5%)

Giải

Gọi p_1, p_2 lần lượt là tỉ lệ bệnh nhân khỏi bệnh sau khi điều trị bằng thuốc A và B.
Đây là bài toán kiểm sự bằng nhau tỉ lệ của hai tổng thể.
Kiểm định giả thiết

$$H_0 : p_1 = p_2; \overline{H}_0 : p_1 \neq p_2.$$

Tỉ lệ chung

$$f = \frac{150 + 70}{200 + 100} = \frac{11}{15}.$$

Nếu H_0 đúng thì biến ngẫu nhiên

$$K_5 = \frac{f_{n_1} - f_{n_2}}{\sqrt{f(1-f)(\frac{1}{200} + \frac{1}{100})}}.$$

Với $\alpha = 0,05$, miền bác bỏ:

$$W_{\alpha/2} = (-\infty; -z_{\frac{1-0,05}{2}}) \cup (z_{\frac{1-0,05}{2}}; +\infty) = (-\infty; -1,96) \cup (1,96; +\infty).$$

Với mẫu trên, ta có

$$f_1 = 0,75; f_2 = 0,7.$$

Khi đó

$$k_5 = \frac{0,75 - 0,7}{\sqrt{\frac{11}{15} . \frac{4}{15} . (\frac{1}{200} + \frac{1}{100})}} = 0,923 \notin W_{\alpha/2},$$

nên ta chấp nhận giả thiết.
Vậy hiệu quả của hai loại thuốc như nhau.

7.6 Bài tập đề nghị

Bài 7.37 Theo dõi giá cổ phiếu (đơn vị ngàn đồng) của hai công ty A và B trong vòng 31 ngày, người ta tính được: Đối với công ty A, giá cổ phiếu trung bình là 37,58 với độ lệch chuẩn 1,5; đối với công ty B, giá cổ phiếu trung bình là 38,24 với độ lệch chuẩn 2,2. ở mức ý nghĩa 5%, có sự khác biệt thực về giá cổ phiếu trung bình của hai công ty A và B không? Biết rằng giá cổ phiếu cảu hai công ty A và B là các biến ngẫu nhiên tuân theo luật phân bố có cùng phương sai.
Bài 7.38 Để so sánh thời gian cắt trung bình của một máy tiện loại cũ với một máy tiện loại mới, người ta cho mỗi máy cắt thử 10 lần và đo thời gian cắt (tính bằng giây). Kết quả thu được như sau:

Máy loại cũ	58	58	56	38	70	38	42	75	68	67
Máy loại mới	57	55	63	24	67	43	33	68	56	54.

Biết rằng thời gian cắt của máy loại cũ và loại mới là các biến ngẫu nhiên tuân theo luật phân phối chuẩn có độ lệch chuẩn theo thứ tự là 13,5 giây và 14,5 giây. Với mức ý nghĩa 3%, có thể cho rằng máy loại mới tốt hơn (có thời gian cắt trung bình ít hơn) máy loại cũ hay không?
Bài 7.39 Một cuộc điều tra của Hội phụ nữ để đánh giá về một dư luận xã hội cho rằng lương của phụ nữ thấp hơn lương của nam giới. Một mẫu ngẫu nhiên gồm 4 người đàn ông có lương trung bình là 78,0 (ngàn đồng), với độ lệch chuẩn mẫu là 24,4; một mẫu ngẫu nhiên độc lập khác độc lập với mẫu trên gồm 4 phụ nữ có lương trung bình là 63,5 (ngàn đồng), với độ lệch chuẩn là 20,2. Giả sử rằng lương của cả nam và nữ giới đều là các biến ngẫu nhiên tuân theo luật phân phối chuẩn có cùng phương sai. Hãy cho kết luận về cuộc điều tra trên ở mức ý nghĩa 10%.
Bài 7.40 Một mẫu gồm 300 cử tri ở khu vực A và một mẫu gồm 200 của tri ở khu vực B cho thấy có 56% và 48% theo thứ tự ủng hộ úng cử viên X. ở mức ý nghĩa 5%, hãy kiểm định giả thiết.

a) Có sự khác biệt giữa hai khu vực về sự ủng hộ ứng cử viên X.

b) Ứng cử viên X được ủng hộ hơn ở khu vực A.

Bài 7.41 Một công ty bào chế một loại thuốc trị dị ứng tuyên bố rằng thuốc của họ có hiệu quả không dưới 90% trong việc làm giảm cơn dị ứng trong vòng 8 giờ. Một mẫu gồm 200 người bị dị ứng sử dụng loại thuốc trên, có 160 người giảm cơn dị ứng. Hãy xác định xem lời tuyên bố của công ty có giá trị không?

Bài 7.42 Một máy sản suất tự động có tỉ lệ sản suất ra sản phẩm loại A lúc đầu là 48%. Máy được cải riến và sau một thời gian áp dụng, người ta kiểm tra 40 hộp, mỗi hộp 10 sản phẩm và ghi lại số sản phẩm loại A trong mỗi hộp (SSPLA/h) như sau:

SSPLA/h	1	2	3	4	5	6	7	8	9	10
Số hộp	2	0	4	6	8	10	4	5	1	0

a) Hãy ước lượng tỉ lệ sản phẩm loại A sau khi máy được cải tiến bằng khoảng tin cậy 95,44%.

b) Hãy cho kết luận về hiệu qảu của việc cải tiến máy ở mức ý nghĩa $\alpha = 0,05$.

Bài 7.43 Người ta muốn nghiên cứu tác dụng của việc cho sinh viên đi thực tế xem sự tiếp thu kiến thức có tốt hơn không bằng cách so sánh điểm thi của nhóm sinh viên không đi thực tế (SVKTT) với nhóm sinh viên có đi thực tế (SVCTT). Kết quả như sau:

Điểm	0	1	2	3	4	5	6	7	8	9	10
SVCTT	0	0	3	9	7	5	17	10	11	4	1
SVKTT	3	3	6	11	7	13	10	12	4	1	3

Gọi X và Y lần lượt là biến ngẫu nhiên biểu thị điểm số của sinh viên có đi thực tế và sinh viên không đi thực tế.
Điểm thi của nhóm sinh viên có đi thực tế có thực sự tốt hơn không? (kết luận ở mức ý nghĩa $\alpha = 0,01$).

Bài 7.44 Một trường đại học muốn đánh giá tác dụng của hội nghị học tốt đến thời gian tự học hàng tuần (viết tắt là TGTH) X của sinh viên X. Sau hội nghị, điều tra về TGTH của một số sinh viên được chọn ngẫu nhiên của trường. Kết quả như sau:

TGTH(giờ)	3	4	5	6	7	8	10	11	12
Số sinh viên	7	10	30	35	25	16	10	8	3

Trước hội nghị, TGTH trung bình của một sinh viên đại học là 6,2 giờ. Hội nghị học tốt vừa qua có làm tăng TGHT trung bình của sinh đại học đó không? (kết luận ở mức ý nghĩa 1%).

Bài 7.45 Sản phẩm của một xí nghiệp đúc cho phép số khuyết tật trung bình cho một sản phẩm là 3. Sau một đợt cải tiến kỹ thuật, người ta lấy ngẫu nhiên 30 sản phẩm để kiểm tra số khuyết tật trên mỗi sản phẩm (SKTTMSP). Kết quả thu được như sau:

SKTTMSP	0	1	2	3	4	5	6
Số sản phẩm	7	4	4	6	8	6	1

a) Hãy cho kết luận về hiệu quả của đợt cải tiến kỹ thuật đối với số khuyết tật trung bình của mỗi sản phẩm ở mức ý nghĩa $\alpha = 10\%$.

b) Sản phẩm không quá 2 khuyết tật được gọi là sản phẩm A. Tỉ lệ sản phẩm loại A trước đợt cải tiến kỹ thuật là 40%. Đợt cải tiến kỹ thuật có thực sự làm tăng tỉ lệ sản phẩm loại A không? (kết luận ở mức ý nghĩa 5%).

Bài 7.46 Một hãng bào chế thuốc đang thử nghiệm hai loại thuốc gây mê A và B mới. Việc thử nghiệm được tiến hành trên hai nhóm thú vật khác nhau. Nhóm thứ nhất gồm 100 con dùng thuốc A thì có 71 con bị mê; nhóm thứ hai, khi dùng thuốc B, có 58 con bị mê trong 90 con. Hãng bào chế muốn kiểm định xem tác dụng của hai loại thuốc trên có khác nhau không ở mức ý nghĩa 5%. Hãy cho biết kết luận?

Bài 7.47 Với ý muốn làm tăng chỉ số mỡ sữa của loại giống bò A, một trại chăn nuôi cho lai bò giống A với một loại bò giống B. Đo chỉ số mỡ sữa của 130 con bò lai giống được chọn ngẫu nhiên trong đàn bò của trại, người ta có kết quả sau:

Chỉ số mỡ sữa	Số bò lai
[3 ; 3,6)	2
[3,6 ; 4,2)	8
[4,2 ; 4,8)	35
[4,8 ; 5,4)	43
[5,4 ; 6)	22
[6 ; 6,6)	15
[6,6 ; 7,2)	5

Biết rằng chỉ số mỡ sữa trung bình của giống bò A thuần chủng là 4,95. Hãy cho kết luận về hiệu quả của việc lại giống ở mức ý nghĩa 1%.

<u>Bài 7.48</u> Điều tra về một nguyên nhân gây ung thư phổi:
thăm dò trong 200 người có hút thuốc lá, thấy có 28 người bị K phổi; trong 170 người không hút thuốc lá, có 12 người K phổi. Hai tỉ lệ này có khác nhau không? Kết luận rằng thuốc lá là nguyên nhân gây K phổi có đúng không? (ở mức ý nghĩa $\alpha = 5\%$).

Chương 8

Hồi quy và Tương quan

8.1 Giới thiệu về phân tích hồi quy và phân tích tương quan

8.1.1 Phân tích hồi quy

✠ Thuật ngữ hồi qui được nhà nghiên cứu Francis Galton sử dụng lần đầu tiên vào cuối thế kỉ XIX trong một nghiên cứu nhằm tìm hiểu tại sao có sự ổn định trong chiều cao trung bình của dân số.

✠ Phân tích hồi qui là nghiên cứu mối liên hệ phụ thuộc của một biến (gọi là biến phụ thuộc) vào một hay nhiều biến khác (gọi là các biến độc lập), với ý tưởng ước lượng và dự đoán giá trị trung bình (tổng thể) của biến phụ thuộc trên cơ sở các giá trị biết trước (trong mẫu) của các biến độc lập.

✠ Trong phân tích hồi qui, ta không quan tâm chủ yếu tới đại lượng hệ số tương quan, thay vào đó, ta cố gắng ước tính một hàm số toán học (tức là phương trình hồi qui) rồi trên phương trình đó ước lượng giá trị của biến phụ thuộc dựa vào các giá trị xác định của biến độc lập trên cơ sở có mối liên hệ nhân quả giữa các biến.

8.1.2 Phân tích tương quan

✠ Mục tiêu chủ yếu của phân tích tương quan là tính độ mạnh hay mức độ liên hệ tuyến tính giữa các biến ngẫu nhiên.

✠ Trong phân tích tương quan ta tính toán một chỉ số mà chỉ số này cho chúng ta một hình ảnh về sự biến chuyển cùng với nhau của hai biến số. Trong thuyết tương quan chúng ta không cần để ý đến sự liên hệ nhân quả, chỉ số cho biết sự tương quan giữa X và Y được ước tính bất kể là:

– X có ảnh hưởng lên Y hay ngược lại,

– Cả Y và X đều có ảnh hưởng qua lại đến nhau,

– Không có biến số nào ảnh hưởng trực tiếp lên biến số kia, nhưng cả hai đều biến chuyển cùng nhau vì một biến số thứ ba có tác động lên cả hai.

✠ Tuy tương quan là một kỹ thuật kém sức mạnh hơn hồi qui, nhưng hai vấn đề này có một mối liên hệ toán học rất chặt chẽ và thường thì tương quan được sử dụng như công cụ bổ trợ hữu ích cho hồi qui.

8.2 Hệ số tương quan

8.2.1 Công thức

Hệ số tương quan của hai biến ngẫu nhiên X và Y, kí hiệu ρ_{xy} được xác định bởi công thức:

$$\rho_{xy} = \frac{Cov\left(X, Y\right)}{\sqrt{V\left(X\right)}.\sqrt{V\left(Y\right)}} \tag{8.1}$$

Trong đó $V(X), V(Y)$ lần lượt là phương sai của X và Y; $Cov(X, Y)$ là hiệp phương sai giữa hai biến X và Y và được xác định bởi công thức sau:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X).E(Y).$$

8.2.2 Ý nghĩa

Hệ số tương quan của hai biến là đại lượng thể hiện chiều hướng và độ mạnh hay yếu của mối quan hệ tuyến tính giữa hai biến đó $|\rho_{xy}|$ càng gần 1 thì mối quan hệ tuyến tính càng chặt và ρ_{xy} càng gần 0 thì mối quan hệ tuyến tính càng yếu đi.

8.2.3 Tính chất

Hệ số tương quan mẫu có các tính chất sau:

- $-1 \geq \rho_{xy} \geq 1$

- $\rho_{xy} < 0$: X và Y có mối liên hệ tuyến tính nghịch ($\rho = -1$ thể hiện một mối liên hệ tuyến tính nghịch hoàn hảo).
- $\rho_{xy} > 0$: X và Y có mối liên hệ tuyến tính thuận ($\rho = 1$ thể hiện một mối liên hệ tuyến tính thuận hoàn hảo).
- $\rho = 0$: X và Y không có mối liên hệ tuyến tính.

8.2.4 Hệ số tương quan mẫu

Trong thực tế, chúng ta không biết được chính xác $Cov(X, Y), V(X), V(Y)$ để tính hệ số tương quan ρ_{xy} vì khi đó ta phảI biết luật phân phối xác suất của nó. Do đó ta phảI ước lượng các tham số của tổng thể trong công thức (8.1) bởi các tham số mẫu đặt trưng. Giả sử từ tổng thể ta chọn ra một mẫu gồm n phần tử. Quan sát hai biến ngẫu nhiên X và Y trên n phần tử này ta có số liệu cụ thể: $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Khi đó hệ số tương quan xác định bởi công thức (8.1) sẽ được ước lượng bằng hệ số tương quan mẫu (kí hiệu: R)

$$R = \frac{S_{xy}}{\sqrt{S_x}\sqrt{S_y}} = \frac{\overline{xy} - \overline{x}.\overline{y}}{\sqrt{S_x}\sqrt{S_y}} \tag{8.2}$$

Ví dụ 8.1. Thời gian sinh trưởng của một loại cây trồng là đại lượng ngẫu nhiên X (đơn vị là ngày) và chiều cao tương ứng Y (đơn vị là cm) được cho bảng sau:

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Hãy xác định hệ số tương quan giữa X và Y.

ĐS: 0.977

8.3 Hồi quy tuyến tính

8.3.1 Hồi quy tuyến tính đơn

1. **Mô hình** Mục đích của phân tích hồi qui là mô hình hóa mối liên hệ giữa các đại lượng bằng một mô hình toán học tối ưu nhất. Giả sử hai đại lượng X và Y có mối quan hệ như sau:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{8.3}$$

Với

β_0 là hệ số chặn,
β_1 là hệ số gốc,
ε là thành phần ngẫu nhiên, không chệch giữa Y và $E(Y \setminus X)$, có thể bằng 0, hoặc lớn hơn 0, hoặc nhỏ hơn 0 khi các giá trị nằm ngay, hoặc phía trên, hoặc phía dưới đường hồi qui.

Giả sử $E(\varepsilon) = 0; V(\varepsilon) = \sigma^2$ và $\varepsilon \sim N(0, \sigma^2)$ là những biến ngẫu nhiên không tương quan nhau. Mô hình (8.3) được gọi là mô hình hồi qui tuyến tính đơn.

2. **Mô hình hồi quy mẫu**

Giả sử ta có n cặp dữ liệu mẫu $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$; $y = \widehat{\beta_0} + \widehat{\beta_1}x$ là đường hồi qui tuyến tính ước lượng của (8.3) từ mẫu. Đặt

$$\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x_1, \varepsilon_i = y_i - \widehat{y_i}, \varepsilon^2 = \sum_{i=1}^{n} \varepsilon_i^2, i = 1, 2, \ldots, n. \tag{8.4}$$

xem $\widehat{\beta_0}$ và $\widehat{\beta_1}$ là các biến, lý luận để có ε^2 đạt giá trị nhỏ nhất (phương pháp bình phương tối thiểu), ta có được kết quả như sau:

$$\begin{cases} \widehat{\beta_1} = \dfrac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \\ \widehat{\beta_0} = \dfrac{\left(\sum\limits_{i=1}^{n} y_i\right) - \widehat{\beta_1}\left(\sum\limits_{i=1}^{n} x_i\right)}{n} \end{cases} \tag{8.5}$$

Khi đó ta có được đường lối hồi qui tuyến tính mẫu là $y = \widehat{\beta_0} + \widehat{\beta_1}x$

Chú ý:

i) $\widehat{\beta_0}$ và $\widehat{\beta_1}$ là ước lượng không chênh lệch của β_0, β_1.

ii) Đặt

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i; S_{xy}^2 = \sum_{i=1}^{n} y_i\left(x_i - \overline{x}\right) = n\left(\overline{xy} - \overline{x}.\overline{y}\right)$$

$$S_x^2 = \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - \tfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2;$$

$$S_y^2 = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

Khi đó hai hệ số $\widehat{\beta_0}$ và $\widehat{\beta_1}$ trở thành

$$\begin{cases} \widehat{\beta_1} = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{S_x^2} \\ \widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x} \end{cases} \tag{8.6}$$

Ví dụ 6.2. Một nghiên cứu ghi lại hai chỉ số X và Y của 10 đối tượng như sau:

X	70	65	90	95	110	115	120	140	155	150
Y	80	100	120	140	160	180	200	220	240	260

Giả sử X và Y có mối quan hệ tuyến tính, tìm mối quan hệ của Y theo X.

ĐS: $y = 24,4545 + 0,5091x$

8.3.2 Hồi quy tuyến tính bội

1. **Mô hình**

   Trong mô hình hồi qui tuyến tính đơn, chúng ta chỉ đơn thuần xây dựng mối quan hệ tuyến tính của hai biến X và Y. Nhưng trong thực tế, chúng ta thường gặp không chỉ biến X ảnh hưởng đến Y mà còn các biến khác cũng ảnh hưởng đến Y. Khi đó để dự báo Y được tốt chúng ta cần xây dựng mô hình hồi qui bội của biến Y qua tất cả các biến độc lập.

   Giả sử Y phụ thuộc vào k biến độc lập $X_1, X_2, ..., X_k$ bởi mô hình:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \tag{8.7}$$

   Trong đó

   β_0 là hệ số tung độ gốc,
   β_i là hệ số độ dốc của Y theo biến X_i khi các biến $X_1, X_2, ..., X_{i-1}, X_{i_1}, ..., X_k$ không đổi,
   ε là thành phần ngẫu nhiên với $E(\varepsilon) = 0$ và $Var(\varepsilon) = \sigma^2$

   Nếu ε là biến ngẫu nhiên không tương quan thì (8.7) được gọi là mô hình hồi qui tuyến tính bội.

2. **Mô hình hồi quy mẫu**

Giả sử chúng ta có n quan sát, mỗi quan sát có $k+1$ giá trị $(y_i, x_{1i}, x_{2i}, \ldots, x_{ki})$, $i = 1, 2, \ldots, n$. Khi đó $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i, i = 1, 2, \ldots, n$.

Mô hình trên được viết dưới dạng ma trận như sau:

$$y = X\beta + \varepsilon$$

Trong đó

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_k \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{bmatrix}$$

Gọi $\widehat{\beta_0}, \widehat{\beta_1}, \ldots, \widehat{\beta_k}$ là các hệ số ước lượng của mô hình (8.7). Bằng phương pháp bình phương tối thiểu, các hệ số này được xác định bởi hệ phương trình tuyến tính sau:

$$\begin{bmatrix} n & \sum\limits_{i=1}^{n} x_{1i} & \sum\limits_{i=1}^{n} x_{2i} & \ldots & \sum\limits_{i=1}^{n} x_{ki} \\ \sum\limits_{i=1}^{n} x_{1i} & \sum\limits_{i=1}^{n} {x_{1i}}^2 & \sum\limits_{i=1}^{n} x_{1i}x_{2i} & \ldots & \sum\limits_{i=1}^{n} x_{1i}x_{ki} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \sum\limits_{i=1}^{n} x_{ki} & \sum\limits_{i=1}^{n} x_{ki}x_{1i} & \sum\limits_{i=1}^{n} x_{ki}x_{2i} & \ldots & \sum\limits_{i=1}^{n} {x_{ki}}^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \vdots \\ \widehat{\beta_k} \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_{1i}y_i \\ \ldots \\ \sum\limits_{i=1}^{n} x_{ki}y_i \end{bmatrix}$$

Hệ phương trình trên có nghiệm là

$$\widehat{\beta} = \left(X^T X\right)^{-1} X^T y \tag{8.8}$$

Ví dụ 8.3 Nghiên cứu về mối liên hệ giữa độ tuổi, tỷ trọng cơ thể (được tính bằng cách lấy trọng lượng chia cho chiều cao bình phương) và cholesterol trong máu của 18 người, ta có kết quả đo lường như sau:

STT	Độ tuổi (age)	BMI (bmi)	Cholesterol (chos)
1	46	25,4	3,5
2	20	20,6	1,9
3	52	26,2	4
4	30	22,6	2,6
5	57	25,4	4,5
6	25	23,1	3
7	28	22,7	2,9
8	36	24,9	3,8
9	22	19,8	2,1
10	43	25,3	3,8
11	57	23,2	4,1
12	33	21,8	3
13	22	20,9	2,5
14	63	26,7	4,6
15	40	26,4	3,2
16	48	21,2	4,2
17	28	21,2	2,3
18	49	22,8	4

Tìm đường hồi qui tuyến tính thể hiện mối quan hệ giữa cholesterol với độ tuổi và tỷ trọng cơ thể.
ĐS: $Y = 0,45545759 + 0,05405192x_1 + 0,03336380x_2$.

8.4 Bài tập có hướng dẫn

Bài 8.1. Giả sử các giá trị quan sát được trên một mẫu của vectơ ngẫu nhiên (X, Y) tuân theo luật phân phối chuẩn hai chiều được cho trong bảng sau:

x_i	1	3	4	6	8	9	11	14
y_i	1	2	4	4	5	7	8	9

a) Tìm giá trị hệ số tương quan mẫu.

b) Viết phương trình đường thẳng hồi quy của Y theo X. Hãy dự báo giá trị của Y theo X lấy giá trị 12.

Giải

a) Giá trị hệ số tương quan mẫu:

$$\overline{p} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)\, s_X s_Y} = \frac{364 - 8.7.5}{7.4,342.2,828} = 0,977.$$

b) Phương trình đường thẳng hồi quy của Y theo X là:

$$y = 0,6364x + 0,5455.$$

Khi X lấy giá trị 12 thì dự báo Y có giá trị là:

$$y_0 = 0,6364.12 + 0,5455 = 8,1823$$

Bài 8.2. Theo dõi mật độ X (số cây/$100m^2$) và sản lượng Y của một loại cây, ta được kết quả:

X	5,8	6,8	7,7	6,5	6,6	5,6	5,7	6,6	5,7	9,4
Y	44	56	62	68	66	52	34	68	65	90

a) Tìm giá trị hệ số tương quan mẫu.

b) Viết phương trình đường thẳng hồi quy của Y theo X.

Giải

a) Giá trị hệ số tương quan mẫu:

$$\overline{p} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)\, s_X s_Y} = 0,932.$$

b) Phương trình đường thẳng hồi quy của Y theo X là:

$$y = 0,93x + 12,93.$$

Bài 8.3. Số vi khuẩn Y (đơn vị: triệu con) sinh sản sau X giờ được ghi lại trong bảng sau qua một thí nghiệm

X	0	1	2	3	4	5	6	7	8
Y	30	32	35	40	48	52	58	62	69

a) Tìm giá trị hệ số tương quan mẫu.

b) Viết phương trình đường thẳng hồi quy của Y theo X.

c) Dự báo số vi khuẩn có sau 10 giờ.

Giải

a) Giá trị hệ số tương quan mẫu:

$$\overline{p} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)\, s_X s_Y} = 0,9937.$$

b) Phương trình đường thẳng hồi quy của Y theo X là:

$$y = 5x + 27,2.$$

c) Dự báo số vi khuẩn có sau 10 giờ là 77,2 triệu con.

Bài 8.4. Kiểm tra hai môn toán và vật lý trong một nhóm 10 học sinh được chọn ngẫu nhiên từ một lớp chuyên vật lý, ta có bảng kết quả sau, với Y (điểm vật lý) và X (điểm toán):

X	7	6	7	10	4	5	7	8	8	9
Y	6	7	7	9	5	3	8	9	6	7

a) Tìm giá trị hệ số tương quan mẫu.

b) Viết phương trình đường thẳng hồi quy của X theo Y.

Giải

a) Giá trị hệ số tương quan mẫu:

$$\overline{p} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)\, s_X s_Y} = 0,722.$$

b) Phương trình đường thẳng hồi quy của Y theo X là:

$$x = 0,708y + 3,36.$$

Bài 8.5. Tính hệ số tương quan và phương trình đường hồi quy của y theo x của mẫu sau:

x_i	1	2	2	3	3	4
y_i	3	4	5	5	6	7
m_i	3	2	1	2	1	1

Giải

a) Giá trị hệ số tương quan mẫu:

$$\overline{p} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)\, s_X s_Y} = 0,95.$$

b) Phương trình đường thẳng hồi quy của Y theo X là:

$$y = 1,25x + 1,75.$$

8.5 Bài tập đề nghị

Trong mỗi bài tập dưới đây, giả sử rằng vectơ ngẫu nhiên đang xét tuân theo luật phân phối chuẩn hai chiều.

Bài 8.6. Xem vectơ ngẫu nhiên (X, Y) mà một mẫu ngẫu nhiên gồm 8 cặp được chọn ra như sau:

x_i	1	2	3	4	5	6	7	8
y_i	4	8	12	16	20	24	28	32

Hãy tính giá trị hệ số tương quan mẫu của X và Y đồng thời cho nhận xét.

Bài 8.7. Một cơ sở sản xuất đã ghi lại số tiền đã chi cho việc nghiên cứu phát triển và lợi nhuận hàng năm của cơ sở trong 6 năm vừa qua như sau (đơn vị 106 VNĐ):

Chi nghiên cứu	5	11	4	5	3	2
Lợi nhuận	31	40	30	34	25	20

a) Vẽ biểu đồ phân tán cho dữ liệu trong bảng trên.

b) Hãy tính giá trị hệ số tương quan mẫu giữa chi nghiên cứu và lợi nhuận.

c) Chi nghiên cứu và lợi nhuận thực sự có tương quan không? (kết luận ở mức ý nghĩa 2%).

d) Viết phương trình đường hồi quy tuyến tính mẫu của lợi nhuận theo chi phí nghiên cứu.

Bài 8.8. Đo chiều cao Y (cm) và chiều dài chi dưới X (cm) của một nhóm thanh niên cùng lứa tuổi, người ta thu được số liệu sau:

x_i	160	161,5	163	165	167	168	171	172
y_i	78	79	80	81	82	83	84	85

a) Tìm khoảng tin cậy 90% cho chiều cao trung bình của những thanh niên lứa tuổi trên.

b) Tính giá trị hệ số tương quan mẫu của X và Y.

c) Ở mức ý nghĩa 5%, hãy cho nhận xét về tài liệu cho rằng hệ số tương quan của X và Y là 0,9.

d) Viết phương trình đường hồi quy mẫu của Y theo X.

Bài 8.9. Nhịp tim tối đa sau khi gắng sức Y tương ứng với tuổi X của người đo được qua kết quả thực nghiệm sau:

x_i	30	48	41	38	29	39	46	41	42	24
y_i	186	183	171	177	191	177	175	176	171	196

a) X và Y có thực sự tương quan không?

b) Nếu hệ số tương quan của X và Y được cho là 0,8 thì kết quả thực nghiệm trên có phù hợp với giá trị đã cho không? (kết luận ở mức ý nghĩa 5%).

Bài 8.10. Một giảng viên dạy môn xác suất thống kê yêu cầu mỗi sinh viên phải làm một đồ án phân tích dữ liệu và dự kỳ thi hết môn. Sau đó, một mẫu gồm 10 sinh viên được chọn ngẫu nhiên, điểm số được ghi lại như sau:

Điểm thi	81	62	74	78	93	69	72	83	90	84
Điểm đồ án	76	71	69	76	87	62	80	75	92	79

a) Tìm khoảng tin cậy 95% cho điểm trung bình của một sinh viên.

b) Với mức ý nghĩa 5%, hãy đánh giá về sự tương quan tuyến tính giữa hai loại điểm trên.

Bài 8.11. Để thực hiện một công trình nghiên cứu về mối quan hệ giữa chiều cao Y(m) và đường kính X(cm) của một loại cây, người ta quan sát một mẫu ngẫu nhiên và có kết quả sau:

x_i	28	28	24	30	60	30	32	42	43	49
y_i	5	6	5	6	10	5	7	8	9	10

a) Hãy tính giá trị hệ số tương quan mẫu cảu X , Y và cho nhận xét.

b) Hãy viết phương trình đường thẳng hồi qui của Y theo X. Hãy dự báo chiều cao của cây có đường kính 45cm.

Bài 8.12. Nghiên cứu lượng phân bón X (kg) được dùng để bón cho ruộng trong một vụ; Y (kg/1000m2) là năng suất lúa. Thống kê ở 30 hộ gia đình, kết quả như sau:

Số hộ	3	5	2	6	4	3	5	2
x_i	40	40	50	50	50	60	60	60
y_i	270	280	280	290	300	300	310	320

a) Tính giá trị hệ số tương quan mẫu của X và Y.

b) Kiểm định giả thiết cho rằng hệ số tương quan của X và Y bằng 0,9 ở mức ý nghĩa 5%.

Bài 8.13. Chiều dài xương đùi X(cm) và chiều cao Y(cm) của những người đàn ông độ tuổi 20-30 là các biến ngẫu nhiên tuân theo luật phân phối chuẩn. Đo chiều dài xương đùi và chiều cao của 10 người đàn ông, được chọn ngẫu nhiên, ở độ tuổi trên. Kết quả được cho trong bảng sau:

x_i	44	46	47	47	48	49	50	50	51	52
y_i	155	159	163	166	169	172	174	176	176	179

a) Tìm khoảng tin cậy 96% cho chiều cao trung bình của những người đàn ông độ tuổi 20-30.

b) Tính giá trị hệ số tương quan mẫu của X và Y. Hãy cho nhận xét về mức độ tương quan giữa X và Y.

c) Một tài liệu y khoa cho rằng hệ số tương quan của X và Y là 0,9. Hãy cho nhận xét về tài liệu trên ở mức ý nghĩa 5%.

d) Viết phương trình đường hồi qui tuyến tính mẫu của Y theo X. Hãy dự báo xem nếu giá trị của X giảm bớt 1cm thì giá trị tương ứng của Y biến thiên như thế nào? Tại sao?

Bài 8.14. Hãy chứng minh rằng nếu X_1 và X_2 là hai đại lượng ngẫu nhiên độc lập thì hệ số tương quan giữa X_1 và X_2 bằng 0. Điều ngược lại có đúng không?

Bài 8.15. Một nghiên cứu được tiến hành ở mỹ được xác định mối quan hệ giữa chiều cao một người và cỡ giày của họ. Nhà nghiên cứu đã thu được số liệu sau:

Chiều cao (inches)	66	63	67	71	62	65	72	68	60	66
Số giày	9	7	8.5	10	6	5	12	10.5	5.5	8

Giả sử mối quan hệ giữa chiều cao và số giày là tuyến tính. Tính hệ số tương quan của hai đại lượng này và thiết lập đường hồi qui tuyến tính giữa chúng. **Bài 8.16** Đo chiều cao và đường kính của một loại cây, ta được kết quả cho bởi bảng sau:

Y | X	6	8	10	12	14
30	2	17	9	3	
35		10	17	9	
40		3	24	16	13
45			6	24	12
50			2	11	22

Trong đó X là đường kính (cm) và Y là chiều cao (m).

a) Xác định hệ số tương quan của X và Y .

b) Tìm phương trình hồi qui tuyến tính của Y theo X .

Bài 8.17. Số liệu về thu nhập (USD) của công nhân tại một nhà máy (Y) theo thâm niên (X) và giới tính ($Z = 1$: nam; $Z = 0$: nữ) được cho bởi bảng số liệu sau:

Y	X	Z
230	12	1
195	10	0
240	11	1
210	13	0
250	14	1
220	13	0
265	15	1
231	15	0
250	16	0
280	16	1
295	17	1
260	17	0
275	18	0
315	19	1
290	19	0

Tìm đường hồi qui thể hiện mối quan hệ giữa chỉ số Y với X và Z.

PHỤ LỤC I: KIẾN THỨC BỔ TRỢ

1. Tính chất của các phép toán $\bigcup ; \bigcap$

a) Tính chất giao hoán

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

b) Tính chất kết hợp

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

c) Tính chất phân phối

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

d) Tính đối ngẫu (De-Morgan)

$$\overline{A \cap B} = \overline{A} \cup \overline{B}, \ \ \overline{A \cup B} = \overline{A} \cap \overline{B}\,.$$

2. 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 1, có n_2 cách thực hiện giai đoạn thứ 2,..., có n_k cách thực hiện giai đoạn thứ k. Khi đó ta có $n = n_1.n_2...n_k$ cách thực hiện toàn bộ công việc.

3. Quy tắc cộng

Giả sử một công việc có thể thực hiện được k trường hợp loại trừ lẫn nhau: trường hợp thứ nhất có m_1 kết quả, trường hợp thứ hai có m_2 kết quả,..., trường hợp thứ k có m_k kết quả. Khi đó việc thực hiện công việc trên cho $m = m_1+m_2+...+m_k$ kết quả.

4. Mẫu

a) Mẫu không lặp: các phần tử của mẫu chỉ có mặt một lần.

b) Mẫu có lặp: các phần tử của mẫu có thể lặp lại nhiều lần trong mẫu.

c) Mẫu không thứ tự: khi thay đổi vị trí các phần tử khác nhau của mẫu ta không nhận được mẫu mới.

d) Mẫu có thứ tự: khi thay đổi vị trí các phần tử khác nhau của mẫu ta nhận được mẫu mới.

5. Các công thức thường dùng

a) Hoán vị

Định nghĩa: hoán vị của n phần tử là một nhóm có thứ tự gồm đủ mặt n phần tử đã cho. Số hoán vị của n phần tử được ký hiệu là P_n.

$$P_n = n!$$

Lưu ý:

- Ta quy ước $0! = 1$.
- Số cách sắp xếp n phần tử khác nhau vào n vị trí đã cho trên đường thẳng là $P_n = n!$.
- Số cách sắp xếp n phần tử khác nhau vào n vị trí đã cho trên đường tròn là $P_{n-1} = (n-1)!$.

b) Chỉnh hợp lặp (mẫu có thứ tự)

Định nghĩa: chỉnh hợp lặp chập k của n phần tử là một nhóm có thứ tự gồm k phần tử không nhất thiết khác nhau chọn từ n phần tử đã cho. Số chỉnh hợp lặp chập k của n phần tử kí hiệu là $\widetilde{A_n^k}$.

$$\widetilde{A_n^k} = n^k.$$

Lưu ý: số chỉnh hợp lặp chập k của n phần tử có thể được tính bằng cách áp dụng quy tắc nhân, trong đó có k giai đoạn, mỗi giai đoạn có n cách.

c) Chỉnh hợp (mẫu không lặp, có thứ tự)

Định nghĩa: chỉnh hợp lặp chập k của n phần tử ($k \leq n$) là một nhóm có thứ tự gồm k phần tử khác nhau chọn từ n phần tử đã cho. Số chỉnh hợp chập k của n phần tử kí hiệu là $A_n^k = n^k$.

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

Lưu ý: $A_n^n = P_n$.

d) Tổ hợp (mẫu không lặp, không có thứ tự)

Định nghĩa: Tổ hợp chập k của n phần tử ($k \leq n$) là một nhóm không phân biệt thứ tự gồm k phần tử khác nhau chọn từ n phần tử đã cho. Số tổ hợp chập k của n phần tử kí hiệu là C_n^k.

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

Lưu ý: tổ hợp khác chỉnh hợp ở việc không lưu ý đến thứ tự sắp xếp của các phần tử.

e) Nhị thức Newton

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

6. Một số công thức giải tích quan trọng

a. Một số công thức đạo hàm và nguyên hàm

$$(\arcsin u)' = \frac{u'}{\sqrt{1-u^2}} \qquad \int \frac{dx}{x^2+a^2} = \frac{1}{a}\arctan\frac{x}{a} + C$$

$$(\arccos u)' = \frac{-u'}{\sqrt{1-u^2}} \qquad \int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C$$

$$(\arctan u)' = \frac{u'}{1+u^2} \qquad \int \frac{dx}{\sqrt{x^2 \pm k}} = \ln\left|x + \sqrt{x^2 \pm k}\right| + C$$

$$(arc\cot u)' = \frac{-u'}{1+u^2} \qquad \int \frac{dx}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C$$

b. Công thức Taylor

$$\begin{aligned} f(x) &= f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + ... + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{f^{(n+1)}(x_0)}{(n+1)!}(x-x_0)^{n+1} \\ &= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1} \\ &= P_n(x) + R_n(x) \quad (*) \end{aligned}$$

c. Công thức Euler

$$e^{ix} = \cos x + i\sin x$$

Các ví dụ

Ví dụ 1 Có 3 sinh viên trường A và 6 sinh viên trường B.

a) Chọn ngẫu nhiên 2 sinh viên trường A và 1 sinh viên trường B, hỏi có bao nhiêu cách thực hiện?

b) Chọn ngẫu nhiên 1 sinh viên trường A và 1 sinh viên trường B, cử đi học 2 nước khác nhau, hỏi có bao nhiêu cách thực hiện?

ĐS: a) 18; b) 36.

Ví dụ 2 Một lọ đựng 2 viên thuốc tốt và 3 viên thuốc xấu. chọn ngẫu nhiên 2 viên thuốc từ lọ này. Tính số cách để chọn được ít nhất 1 viên thuốc xấu.
ĐS: 9.

Ví dụ 3 Từ các số 1,2,3,4,5,6,7, có thể thành lập được bao nhiêu số trong các trường hợp sau:

a) Số có 3 chữ số đôi một khác nhau.

b) Số có 4 chữ số đôi một khác nhau, trong đó có chữ số1.

ĐS: a) 210; b) 480.

Ví dụ 4 Xếp 10 sinh viên, trong đó có 5 nam và 5 nữ vào 2 dãy bàn dài. Hỏi có bao nhiêu cách xếp sao cho nam ngồi 1 dãy, nữ ngồi 1 dãy.
ĐS: 28800.

Ví dụ 5 Trong một đợt xổ số, người ta phát hành trên mỗi tờ vé số có 5 chữ số. trong một đợt xổ số khác người ta phát hành trên mỗi tờ vé số có 6 chữ số, khi đó số vé số tăng cho việc phát hành tăng thêm một chữ số là bao nhiêu?

Ví dụ 6 Có 8 người, mỗi người bắt tay một lần với những người khác. Hỏi:

a) Có bao nhiêu cái bắt tay?

b) Nếu trong họ có 3 người không bắt tay lẫn nhau thì còn lại bao nhiêu cái bắt tay?

ĐS: a) 28; b) 25.

Ví dụ 7 Một hộp có 5 bi đỏ, 6 bi vàng. Có bao nhiêu cách lấy từ hộp ra

a) 3 bi.

b) 3 bi trong đó có 2 bi đỏ.

c) 3 bi trong đó có bi đỏ ?

ĐS: a) 165; b) 60; c) 145.

Ví dụ 8 Cho 4 chữ số ; 1, 2, 3, 4, có bao nhiêu cách lập một số:

a) Có 3 chữ số khác nhau từ các chữ số trên.

b) Có 3 chữ số từ các chữ số trên.

ĐS: a) 24; b) 64.

7. Bài tập có hướng dẫn

Bài 1.1. Có bao nhiêu cách xếp 5 quyển sách lên một giá sách?

Giải

Dễ thấy mỗi cách xếp là một hoán vị của 5 phần tử. Từ đó suy ra số cách xếp là

$$P_5 = 5! = 120.$$

Bài 1.2. Một học sinh phải thi 4 môn trong 10 ngày (mỗi ngày thi một môn). Có mấy cách lập chương trình thi?

Giải

Có A_{10}^4 cách lập chương trình thi.

Bài 1.3. Có mấy cách rút 3 quân bài từ bộ bài 52 quân?

Giải

Số cách rút bằng số tổ hợp chập 3 từ 52 phần tử:

$$C_{52}^3 = 4960.$$

Bài 1.4. Hỏi cần phải tổ chức bao nhiêu trận đấu để 12 đội bóng thi đấu theo thể thức vòng tròn?

Giải

Phải tổ chức $C_{12}^2 = 66$ trận đấu.

Bài 1.5. Có 4 quả bóng màu trắng và 6 quả bóng màu xanh.

a) Có bao nhiêu cách chọn 5 quả bóng?

b) Có bao nhiêu cách chọn 3 quả bóng màu trắng và 2 quả bóng màu xanh?

Giải

a) Có C_{10}^5 cách chọn 5 quả bóng.

b) Có $C_4^3.C_6^2$ cách chọn 3 quả bóng màu trắng và 2 quả bóng màu xanh.

Bài 1.6. Nam đi dự đám cưới của hai bạn Hải và Lan. Trước khi ra về, Nam và 4 người bạn nữa cùng chụp hình lưu niệm với cô dâu chú rể. Hãy tính cách xếp các bạn thành một hàng để chụp hình sao cho:

a) Cô dâu đứng cạnh chú rể.

b) Cô dâu không đứng cạnh chú rể.

c) Nam đứng bên phải chú rể.

Giải

a) Ta xem cô dâu chú rể là một "bó". Khi đó, ta có $P_6 = 6! = 720$ cách xếp 5 người bạn cùng cô dâu chú rể thành một hàng. Mặc khác ứng với mỗi cách xếp này ta lại có $P_2 = 2! = 2$ cách hoán vị trong nội bộ giữa cô dâu và chú rể. Vậy có tất cả $6!2! = 1440$ cách xếp theo yêu cầu.

b) Xếp ngẫu nhiên 7 người bạn thành một hàng thì có $P_7 = 7! = 5040$ cách. Vậy theo câu a) ta có $5040 - 1440 = 3600$ cách xếp sao cho cô dâu không đứng cạch chú rể.

c) Tương tự câu a), ta có 720 cách xếp sao cho Nam đứng bên cạch chú rể.

Bài 1.7. Lớp học có 45 sinh viên, trong đó có 43 bạn là đoàn viên. Có bao nhiêu cách chọn ngẫu nhiên 5 đoàn viên để bầu vào ban cán sự lớp gồm 1 lớp trưởng, 1 lớp phó học tập, 1 lớp phó đời sống, 1 bí thư và 1 phó bí thư.

Giải

Vì việc chọn 5 thành viên có phân biệt vị trí nên số cách chọn 5 đoàn viên để bầu vào ban cán sự lớp là số chỉnh hợp không lặp chập 5 của 43 phần tử:

$$A_{43}^5 = \frac{43!}{38!} = 115511760\,(\text{cách})$$

Bài 1.8. Xếp ngẫu nhiên 12 hành khách lên 5 toa tàu. Có bao nhiêu cách:

a) Xếp ngẫu nhiên 12 hành khách lên 5 toa tàu một cách tùy ý.

b) Xếp ngẫu nhiên 12 hành khách lên 5 toa tàu sao cho toa thứ nhất có 3 hành khách.

Giải

a) Số cách xếp ngẫu nhiên 12 hành khách lên 5 toa tàu một cách tùy ý là số chỉnh hợp lặp chập 5 của 12 phần tử:

$$\overline{A}_5^{12} = 5^{12} = 244140625\,\text{cách}$$

b) Ta chia bài toán gồm 2 giai đoạn:

+ Giai đoạn 1: Xếp ngẫu nhiên 3 hành khách (từ 12 hành khách) vào toa thứ nhất, có $C_{12}^{3} = 220$ (cách).

+ Giai đoạn 2: Xếp ngẫu nhiên 9 hành khách vào 4 toa tàu còn lại một cách tùy ý, có $\overline{A}_4^9 = 4^9 = 262144$ (cách). Vậy có $C_{12}^{3}.\overline{A}_4^9 = 57671680$ cách xếp ngẫu nhiên 12 khách lên 5 toa tàu sao cho toa thứ nhất có 3 hành khách.

Bài 1.9. Một thùng có 50 quyển sách, trong đó có 20 quyển sách Tiếng Việt và 30 quyển sách Toán. Có bao nhiêu cách:

a) Lấy ngẫu nhiên ra 10 quyển sách.

b) Lấy ngẫu nhiên ra 9 quyển sách trong đó có 5 quyển sách Toán.

c) Lấy ra 8 quyển sách Toán để trao cho 8 em học sinh.

Giải

a) Số cách lấy ngẫu nhiên ra 10 quyển sách từ 50 quyển sách là số tổ hợp chập 10 của 50 phần tử: $C_50^10 = 10272278170$ (cách).
b)
+ Có C_{20}^{4} cách lấy ra 4 quyển sách Tiếng Việt.

+ Có C_{30}^{5} cách lấy ra 5 quyển sách Toán.

Vậy có $C_{20}^{4}.C_{30}^{5} = 690441570$ cách lấy ngẫu nhiên ra 9 quyển sách trong đó có 5 quyển sách Toán.

c) Có C_{30}^{8} cách lấy ra 8 quyển sách Toán từ 30 quyển sách Toán. ứng với mỗi cách này, ta có 8! cách trao 8 quyển sách cho 8 em học sinh. Vậy có $C_{30}^{8}.8! = 235989936000$ cách thoả yêu cầu.

Bài 1.10. Một lô hàng có 10 sản phẩm, trong đó có 8 sản phẩm tốt và 2 phế phẩm. Có bao nhiêu cách:

a) Lấy ngẫu nhiên ra 4 sản phẩm.

b) Lấy ra ngẫu nhiên 4 sản phẩm, trong đó có 3 sản phẩm tốt.

c) Lấy ra ngẫu nhiên 4 sản phẩm, trong đó có ít nhất 1 phế phẩm

Giải

a) Số cách lấy ngẫu nhiên 4 sản phẩm từ 10 sản phẩm là tổ hợp chập 4 của 10 phần tử: $C_{10}^{4} = 210$ (cách).

b) Ta có thể xem bài toán gồm 2 giai đoạn:

+ Giai đoạn 1: Lấy 3 sản phẩm tốt từ 8 sản phẩm tốt, có $C_8^3 = 56$ (cách).

+ Giai đoạn 2: Lấy 1 phế phẩm từ 2 phế phẩm, có $C_2^1 = 2$ (cách).

Vậy có $C_8^3.C_2^1 = 112$ cách lấy ra ngẫu nhiên 4 sản phẩm, trong đó có 3 sản phẩm tốt.

c) Vì lô hàng chỉ có 2 phế phẩm nên bài toán gồm 2 trường hợp:

+ Trường hợp 1: Lấy được 1 phế phẩm và 3 sản phẩm tốt, có $C_2^1.C_8^3 = 112$ (cách).

+ Trường hợp 2: Lấy được 2 phế phẩm và 2 sản phẩm tốt, có $C_2^2.C_8^2 = 28$ (cách).

Vậy có $C_2^1.C_8^3 + C_2^2.C_8^2 = 140$ cách lấy ra ngẫu nhiên 4 sản phẩm, trong đó có ít nhất 1 phế phẩm.

Bài 1.11. Một ngày học 3 môn học trong số 7 môn học. Hỏi có bao nhiêu cách xếp thời khoá biểu trong một ngày?

Giải

Giả sử có thể chọn tùy ý các môn học trong ngày đó. Việc xếp thời khoá biểu trong ngày chính là việc chọn ra 3 môn trong số 7 môn có để ý đến thứ tự và không có lặp. Do đó, số cách xếp là: $A_7^3 = 210$.

Bài 1.12. Có mấy cách phân phối 15 sản phẩm cho 3 người sao cho người thứ nhất có 2 sản phẩm, người thứ hai có 3 sản phẩm và người thứ ba có 10 sản phẩm.

Giải

Ta có

$$C_{15}^2.C_{13}^3.C_{10}^{10} = \frac{15!}{2!3!10!}$$

Bài 1.13. Trên một vòng tròn có 12 điểm. Có mấy cách vẽ dây cung có các mút là các điểm đã cho? Có mấy tam giác nhận các điểm đó là đỉnh?

Giải

Có thể vẽ được C_{12}^2 dây cung và C_{12}^3 tam giác.

Bài 1.14. Tìm số đường chéo của một đa giác lồi n đỉnh?

Giải

Số đường chéo của một đa giác lồi n đỉnh là $C_n^2 - n$.

Bài 1.15. Có 10 người định cư vào 3 nước. Nước Mỹ: 4 người, Hàn Quốc: 3 người, Việt Nam: 3 người. Hỏi có bao nhiêu cách sắp xếp?

Giải

Công việc sắp xếp 10 người gồm 3 giai đoạn:

+ Giai đoạn 1: Chọn tùy ý 4 người vào nước Mỹ: C_{10}^4 cách

+ Giai đoạn 2: Chọn tùy ý 3 người vào nước Hàn Quốc:C_6^3 cách

+ Giai đoạn 3: Chọn tùy ý 3 người vào nước Việt Nam:C_3^3 cách

Theo nguyên lý nhân, số cách sắp xếp là: $C_{10}^4.C_6^3.C_3^3 = 4200$

Bài 1.16.

1. Có bao nhiêu số chẵn gồm 5 chữ số?
2. Có bao nhiêu số nguyên dương mỗi số gồm 3 chữ số khác nhau?

Giải

1. Số các số chẵn gồm 5 chữ số $9.10^3.5 = 45000$.

2. Số các số nguyên dương gồm 3 chữ số khác nhau $A_{10}^3 - A_9^2 = 648$.

8. Bài tập đề nghị

Bài 1.17. Một lớp học có 30 sinh viên trong đó có 20 nam. Có bao nhiêu cách chọn ra một ban cán sự gồm 4 sinh viên nếu:

a) Có đúng 2 nam.

b) Không có nam.

c) Nhiều nhất 2 Nam.

d) Có ít nhất 1 Nam.

ĐS:

a) $C_{20}^2.C_{10}^2$;

b) C_{10}^4;

c) $C_{20}^0.C_{10}^4 + C_{20}^1.C_{10}^3 + C_{20}^2.C_{10}^2$;

d) $C_{30}^4 - C_{10}^4$; $C_{20}^1.C_{10}^3 + C_{20}^2.C_{10}^2 + C_{20}^3.C_{10}^1 + C_{20}^4$.

Bài 1.18. Trong một buổi dạ vũ có 22 nam và 18 nữ. Hỏi có bao nhiêu cách chọn:

a) Hai người ra khiêu vũ?

b) Một đôi nam nữ ra khiêu vũ?

c) Ba đôi nam nữ ra khiêu vũ?

ĐS:

a) C_{40}^2;

b) $C_{22}^1.C_{18}^1$;

c) $C_{22}^3.C_{18}^3.3!$.

Bài 1.19. Người ta dùng 5 cột cờ để báo hiệu trên biển. Biết rằng có tất cả 7 màu cờ khác nhau. Hỏi có bao nhiêu tín hiệu khác nhau nếu:

a) Dùng màu tùy ý.

b) Dùng 5 màu khác nhau.

c) Hai cột cờ kế tiếp không được cùng màu.

ĐS:

a) 7^5;

b) A_7^5;

c) 7.6^4.

Bài 1.20. Một lớp học 30 sinh viên trong đó có 20 nam. Có bao nhiêu cách chọn ra một ban cán sự lớp gồm: 1 lớp trưởng, 1 lớp phó, 1 bí thư và 1 thủ quỷ. Nếu:

a) Chọn bất kỳ.

b) Toàn là nữ.

c) Có ít nhất một nam.

d) Lớp trưởng là nữ.

e) Có đúng một nam.

ĐS:

a) A_{30}^4;

b) $C_{10}^1.A_{29}^3$;

c) $C_{20}^1.C_{10}^3.4!$;

d) A_{10}^4;

e) $A_{30}^4 - A_{10}^4$.

Bài 1.21. Có 5 người. Hỏi có bao nhiêu cách sắp xếp 5 người này:

a) Ngồi thành hàng dài.

b) Ngồi thành vòng tròn.

c) Ngồi thành vòng tròn có đánh số.

ĐS: 5!; 4!; 5!.

Bài 1.22. Năm người cùng lên một đoàn tàu hỏa có 8 toa xe. Có bao nhiêu cách để:

a) Lên tùy ý.

b) Lên cùng một toa.

c) Lên 5 toa đầu.

d) Lên 5 toa khác nhau.

e) A và B lên cùng toa đầu.

f) A và B lên cùng một toa.

g) A và B lên cùng toa và 3 người còn lại không lên toa này.

ĐS:

a) $\widetilde{A}_8^5$;

b) C_8^1;

c) $\widetilde{A}_5^5$;

d) A_8^5;

e) $\widetilde{A}_3^8$;

f) $C_8^1.\widetilde{A}_3^8$;

f) $C_8^1.\widetilde{A}_3^7$.

Bài 1.23. Một hộp có 5 bi trắng, 3 bi xanh. Lấy từ hộp ra 2 bi. Có 3 cách lấy:

1. Lấy ngẫu nhiên 2 bi.
    a) Có bao nhiêu cách lấy được 2 bi?
    b Có bao nhiêu cách lấy được 2 bi trắng?
    c Có bao nhiêu cách lấy được 1 bi trắng, 1 bi xanh?
2. Lấy lần lượt 2 bi. Hỏi như câu 1.
3. Lấy có hoàn lại 2 bi (chọn lặp). Hỏi như câu 1.

ĐS:

1) 28; 10; 15.

2) 56; 20; 30.

3) 64; 25; 30.

Bài 1.24. Có bao nhiêu số điện thoại của một tổng đài gồm các số có 10 chữ số.

Bài 1.25. Có bao nhiêu đường thẳng có thể vẽ qua 20 điểm (không có 3 điểm nào thẳng hàng) trên mặt phẳng?

ĐS: C_20^2.

Bài 1.26. Có bao nhiêu cách sắp xếp 5 người ngồi quanh một chiếc bàn tròn sao cho 2 người định trước được ngồi cạnh nhau.

ĐS: 60.

Bài 1.27. Một lớp học có 30 sinh viên trong đó có 20 nam. Có bao nhiêu cách chọn ra một ban cán sự lớp gồm: 1 lớp trưởng, 2 lớp phó và 1 thủ quỷ.

Bài 1.28. Có bao nhiêu số $n : 100 \leq n \leq 999$ sao cho n gồm:

a) 3 chữ số khác nhau.

b) 3 chữ số lẽ.

c) n lẽ và thỏa a).

ĐS: 648; 125; 320.

Bài 1.29. Có 5 chữ số 1, 2, 3, 4 và 5. Hỏi có bao nhiêu cách sắp xếp 5 chữ số này thành 2 nhóm chữ số chẵn và chữ số lẻ riêng biệt nhau?

ĐS: 24.

Bài 1.30 Sử dụng công thức nhị thức Newton, hãy chứng minh rằng một tập hợp có n phần tử sẽ có tất cả 2^n tập hợp con.

PHỤ LỤC II: BẢNG TRA THỐNG KÊ

Bảng 1: Bảng giá trị hàm Gauss (hàm mật độ Gauss)

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3986	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	9653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449

Bảng 1: Bảng giá trị hàm Gauss (tiếp theo)

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

x	0	1	2	3	4	5	6	7	8	9
2,1	0440	0431	0422	0413	0404	0396	0388	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0031	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0006	0006	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Bảng 2: Bảng giá trị tích phân Laplace (hàm phân phối xác suất Gauss)

$$\Phi_0\left(x\right) = \frac{1}{\sqrt{2\pi}}\int\limits_0^x \exp\left(-\frac{t^2}{2}\right)dt$$

X	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0	0	0,004	0,008	0,012	0,016	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0389	0438	0478	0517	0557	0396	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2703	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	1492	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2	4772	4778	4783	4788	4793	4793	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4838	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4875	4881	4884	4887	4890

Bảng 2: Bảng giá trị tích phân Laplace (tiếp theo)

X	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
2,3	4893	4896	4898	4901	4904	4904	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4927	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4945	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4959	4961	4962	4963	4964
2,7	4962	4966	4967	4968	4969	4969	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4977	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3	49865	49869	49874	49878	49882	49882	49889	49893	49897	49900
3,1	49903	49906	49909	49912	49915	49915	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49940	49924	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49958	49961	49962	49964	49965
3,4	49966	49967	49969	49970	49971	49971	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49980	49982	49982	49983	49984
3,6	49984	49985	49985	49986	49986	49986	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49993
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

Bảng 3: Bảng phân vị chuẩn tắc U_α

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{U_\alpha} \exp(-\frac{t^2}{2})dt = g$$

g	U_g	g	U_g	g	U_g	g	U_g
0.50	0.000	0.71	0.553	0.92	1.405	0.980	2.054
0.51	0.025	0.72	0.583	0.93	1.476	0.981	2.075
0,52	0.030	0.73	0.613	0.94	1.555	0.982	2.097
0.53	0.075	0.74	0.643	0.95	1.645	0.983	2.120
0.54	0.100	0.75	0.674	0.955	1.695	0.984	2.144
0.55	0.126	0.76	0.706	0.960	1.751	0.985	2.170
0.56	0.151	0.77	0.739	0.965	1.812	0.986	2.197
0.57	0.176	0.78	0.772	0.966	1.825	0.987	2.226
0.58	0.202	0.79	0.806	0.967	1.837	0.988	2.257
0.59	0.228	0.80	0.842	0.968	1.852	0.989	2.290
0.60	0.253	0.81	0.878	0.969	1.866	0.990	2.326
0.61	0.279	0.82	0.915	0.970	1.881	0.991	2.366
0.62	0.305	0.83	0.954	0.971	1.896	0.992	2.409
0.63	0.332	0.84	0.994	0.972	1.911	0.993	2.457
0.64	0.358	0.85	1.036	0.973	1.927	0.994	2.512
0.65	0.385	0.86	1.080	0.974	1.943	0.995	2.576
0.66	0.412	0.87	1.126	0.975	1.960	0.996	2.652
0.67	0.440	0.88	1.175	0.976	1.977	0.997	2.748
0.68	0.468	0.89	1.227	0.977	1.995	0.998	2.878
0.69	0.496	0.90	1.282	0.978	2.014	0.999	2.090
0.70	0.524	0.91	1.341	0.979	2.034		
g	U_g	g	U_g	g	U_g	g	U_g

Bảng 4: Bảng phân vị Student $t_\alpha(n)$

bậc tự do $n-1$, mức xác suất α

$P\left(T > t_\alpha\left(n-1\right)\right) = \alpha$ với $T \sim St(n)$.

$n-1$; α	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.675	66.619
2	1.886	2.920	4.303	6.965	9.925	22.326
3	1.638	2.353	3.182	4.541	5.841	10.213
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.861	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.719	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505

Bảng 4: Bảng phân vị Student $t_\alpha(n)$ (tiếp theo)

$n-1$; α	0.10	0.05	0.025	0.01	0.005	0.001
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
$+\infty$	1.282	1.645	1.960	2.326	2.576	3.090

Bảng 4: Bảng phân vị Khi bình phương bậc tự do n mức xác suất α

$n;\alpha$	0,995	0,99	0,975	0,95	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,0151	0,103	5,911	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	10,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,314	1,646	2,180	2,733	15,507	17,535	20,090	21,995
9	1,735	2,088	2,700	3,322	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,982	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	4,601	5,229	5,262	7,261	24,996	27,488	30,758	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582
20	7,343	8,260	9,591	10,851	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401
22	8,543	9,542	10,982	12,388	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,558

Bảng 4: Bảng phân vị Khi bình phương bậc tự do n mức xác suất α (tiếp theo)

$n;\alpha$	0,995	0,99	0,975	0,95	0,05	0,025	0,01	0,005
25	10,520	11,524	13,120	14,611	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	40,113	43,194	46,963	49,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336
30	13,787	14,930	16,791	18,493	43,773	46,979	50,892	63,672
40	20,707	22,164	24,433	26,509	55,578	5,342	63,691	66,766
50	27,991	29,707	32,307	24,754	67,505	71,420	76,154	79,490
100	67,328	70,065	74,222	77,929	124,34	129,56	135,80	140,16

PHỤ LỤC III: CÁC THUẬT NGỮ CƠ BẢN

Tiếng Việt	**Tiếng Anh**
Phép thử	Experiment
Biến cố	Event
Biến cố chắc chắn	Certain event
Biến cố không thể có	Imposible event
Biến cố ngẫu nhiên	Random event
Xác suất	Probability
Kết cục đồng khả năng	Equaly likely possible outcomes
Kết cục thuận lợi	Favorable outcomes
Định nghĩa cổ điển về xác suất	Clasical definition of probability
Sơ đồ Venn	Venn diagram
Tần suất của biến cố	Relative-frequency of probability
Định nghĩa thống kê về xác suất	Statistical definition of probability
Bảng số ngẫu nhiên	Random number table
Định nghĩa hình học về xác suất	Geometrical definition of probability
Định nghĩa tiên đề về xác suất	Axiomical definition of probability
Định nghĩa xác suất chủ quan	Subjective definition of probabbility
Nguyên lý xác suất lớn	Rule of the grand probability
Nguyên lý xác suất nhỏ	Rule of the small probability
Tổng của các biến cố	Union of events
Biến cố xung khắc	Mutually exclusive events
Biến cố không xung khắc	Non-exclusive events
Biến cố xung khắc từng đôi	Paired mutually exclusive events
Nhóm đầy đủ các biến cố	Universal set of events
Biến cố đối lập	Complement of event
Tích của các biến cố	Intersection of events

Tiếng Việt	**Tiếng Anh**
Biến cố độc lập	Independent events
Biến cố phụ thuộc	Dependent events
Biến cố độc lập từng đôi	Paired independent events
Biến cố độc lập toàn phần	Universal independent events
Xác suất có điều kiện	Conditional probability
Định lý cộng xác suất	Theorem of Addition
Định lý nhân xác suất	Theorem of Multiplication
Công thức Bernoully	Bernoully theorem
Công thức xác suất đầy đủ	Law of total probability
Công thức Bayes	Bayes theorem
Xác suất tiên nghiệm	A priori probability
Xác suất hậu nghiệm	A posteri probability
Biến ngẫu nhiên	Random variable
Biến ngẫu nhiên rời rạc	Discrete random variable
Biến ngẫu nhiên liên tục	Continuos random variable
Quy luật phân phối xác suất	Probability distribution
Bảng phân phối xác suất	Probability distribution table
Hàm phân bố xác suất	Cumulative distribution function
Hàm mật độ xác suất	Density probability function
Kỳ vọng toán	Expected value, Expectation
Trung vị	Median
Mốt	Mode
Phương sai	Variance
Độ lệch chuẩn	Standard deviation
Hệ số biến thiên	Coefficient of variation
Giá trị tới hạn	Critical value

Tiếng Việt	**Tiếng Anh**
Hệ số bất đối xứng	Skewness
Hệ số nhọn	Kurtosis
Quy luật không - một	Alternative (Bernoully) distribution
Dấu hiệu nghiên cứu định tính	Qualitative variable
Quy luật nhị thức	Binomial distribution
Quy luật nhị thức theo tỷ lệ	Binomial distribution of proportion
Quy luật poisson	Poisson distribution
Chuỗi Poisson	Poisson series
Lý thuyết phục vụ công cộng	The queue theory
Lý thuyết quản lý dữ trữ	Inventory managment theory
Quy luật siêu bội	Hypergeometric distribution
Hệ số hiệu chỉnh	Adjustment coefficient
Quy luật phân phối đều	Uniform distribution
Quy luật phân phối lũy thừa	Exponential distribution
Quy luật phân phối chuẩn	Normal distribution
Quy luật phân phối chuẩn hóa	Stadard normal distribution
Phép chuẩn hóa biến ngẫu nhiên	Stadardisation of random variable
Tích phân Poisson	Poisson integral
Hàm Laplace	Laplace function
Quy tắc hai xích ma	Two sigma rule
Quy tắc ba xích ma	Three sigma rule
Định lý địa phương Laplace	Local laplace theorem
Định lý tích phân Laplace	Integral laplace theorem
Quy luật khi bình phương	Chi squares distribution
Số bậc tự do	Degree of freedom
Quy luật Student	T-Student distribution

Tiếng Việt	**Tiếng Anh**
Quy luật Fisher - Snedecor	Fisher-Snedecor distribution
Biến ngẫu nhiên hai chiều	Two random variable
Bảng phân phối xác suất đồng thời	Joint distribution table
Bảng phân phối xác suất biên	Marginal distribution table
Bảng phân phối xác suất có điều kiện	Conditional distribution table
Hàm phân bố xác suất đồng thời	Joint cumulative distribution function
Hàm phân bố xác suất biên	Marginal cumulative function
Hàm phân bố xác suất có điều kiện	Conditional cumulative function
Hàm mật độ xác suất đồng thời	Joint density probability function
Hàm mật độ xác suất biên	Marginal density function
Hàm mật độ xác suất có điều kiện	Conditional density function
Hiệp phương sai	Covariance
Hệ số tương quan	Correlation coefficient
Kỳ vọng toán có điều kiện	Conditional expected value
Hàm hồi quy	Regession function
Hàm các biến ngẫu nhiên	Fuction of random variables
Luật số lớn	Law of grand numbers
Bất đẳng thức Trêbưsep	inegality
Định lý Trêbưsep	theorem
Định lý Bernoully	theorem
Định lý giới hạn trung tâm	Central limit theorem
Định lý Lindenberg - Lewi	Lindenberg- theorem
Hàm đặc trưng	Caracteristic function
Thống kê toán	Mathematical statistic
Phương pháp nghiên cứu toàn bộ	Total study
Phương pháp mẫu	Sampling

Tiếng Việt	**Tiếng Anh**
Thống kê mô tả	Descriptive statistic
Thống kê suy diễn	Inferential statistic
Tổng thể	Population
Bảng phân phối tần số	Frequence distribution table
Bảng phân phối tần suất	Relative-frequence distribution table
Tần số tích lũy	Cumulative frequency
Tần suất tích lũy	Cumulative relative relative-frequency
Trung bình tổng thể	Population mean
Trung bình điều hòa	Harmonical mean
Trung bình nhân	Geometrical mean
Phương sai tổng thể	Population variance
Độ lệch chuẩn tổng thể	Population standard deviation
Tần suất tổng thể	Population Relation-frequency
Mẫu ngẫu nhiên	Random sample
Mẫu đại diện	Representative sample
Mẫu cụ thể	Concret sample
Mẫu hoàn lại	Sampling with replacement
Mẫu không hoàn lại	Sampling without replacement
Mẫu ngẫu nhiên giản đơn	Simple random sample
Mẫu ngẫu nhiên hệ thống	Systematic radom sample
Mẫu chùm	Quota sample
Mẫu phân tổ	Cluster sample
Mẫu phân tầng	Stratiffed sample
Thang định danh	Nominal scale
Thang thứ bậc	Ordinal scale
Thang đo khoảng	Interval scale

Tiếng Việt	**Tiếng Anh**
Thang tỷ lệ	Ratio scale
Đường đa giác	Diagram
Biểu đồ hình cột	Histogram
Biểu đồ hình bánh xe	Pie chart
Đồ thị hình hộp	Box - plot
Thống kê	Statistic
Trung bình mẫu	Sample mean
Trung bình số học	Arithmetic mean
Trung bình có trọng số	Weight arithmetic mean
Kỳ vọng toán của trung bình mẫu	Expected value of sample mean
Phương sai của trung bình mẫu	Variance of sample mean
Sai số chuẩn của trung bình mẫu	Standard error of sample mean
Trung vị mẫu	Sample median
Mốt của mẫu	Sample mode
Khoảng biến thiên	Range
Khoảng tứ phân vị	Inter - quartile range
Tứ phân vị 1	Lower quartile
Tứ phân vị 3	Upper quartile
Tổng bình phương các sai lệch	Sum squares of variation
Độ lệch bình phương trung bình	Mean squares of variation
Phương sai mẫu	Sample variance
Độ lệch chuẩn mẫu	Sample standard deviation
Hệ số biến thiên của mẫu	Sample coefficient of variation
Hệ số bất đối xứng của mẫu	Sample Skewness
Hệ số nhọn của mẫu	Sample kurtosis
Tần suất mẫu	Sample relative frequency

Tiếng Việt	**Tiếng Anh**
Kỳ vọng toán của tần suất mẫu	Expected value of relative frequency
Phương sai của tần suất mẫu	Variance of relative frequency
Sai so chuẩn của tần suất mẫu	Standard error of relative frequency
Mẫu ngẫu nhiên hai chiều	Two dimention random sample
Phân phối thực nghiệm của X	Empiric distribution of X
Phân phối thực nghiệm của Y	Empiric distribution of Y
Phân phối thực nghiệm có điều kiện	Conditional empiric distribution
Trung bình có điều kiện	Conditional mean
Hệ số tương quan mẫu	Sample correlation coefficient
Suy diễn thống kê	Deduction statistatic
Quy nạp thống kê	Induction statistatic
Phương pháp ước lượng	Estimation
Phương pháp ước lượng điểm	Point estimation
Phương pháp hàm ước lượng	Method moment
Ước Lượng không chệch	Unbiased estimator
Ước lượng chệch	Biased estimator
Độ chệch	Bias
Ước lượng tiệm cận không chệch	Assimptotically unbiased estimator
Ước lượng hiệu quả	Efficient estimator
Ước lượng hiệu quả nhất	The most efficient estimator
Bất đẳng thức Cramer Ọ Rao	Cramer- Rao inegality
Độ hiệu quả	Efficacity
Ước lượng vững	Consistent estimator
Ước lượng đủ	Sufficient estimator
Phương pháp ước lượng hợp lý tối đa	Maximun log likelihood estitimation

Tiếng Việt	**Tiếng Anh**
Hàm hợp lý	Likelihood function
Phương pháp ƯL khoảng tin cậy	Confidence interval estimation
Khoảng tin cậy	Confidence interval
Độ tin cậy	Confidence coefficient (level)
Độ chính xác (sai số) của ước lượng	Precision (Error) of estimation
Khoảng tin cậy hai phía	Two side confidence interval
Khoảng tin cậy đối xứng	Symmetric confidence interval
Khoảng tin cậy bên phảI	Right confident interval
Khoảng tin cậy bên tráI	Left confidence interval
kích thước mẫu	Sample size
Phương pháp mẫu kép	Double samples
Mẫu lớn	Large sample
Mẫu nhỏ	Small sample
Mẫu theo cặp	Mached samples
Kiểm định giả thuyết thống kê	Hypothesis tests
Giả thuyết thống kê	Statistical hypothesis
Giả thuyết gốc	Null hypothesis
Giả thuyết đối	Alternative hypothesis
Giả thuyết đơn	Simple hypothesis
Giả thuyết hợp	Compose hypothesis
Tiêu chuẩn kiểm định giả thuyết	Test statistic
Giá trị quan sát của TCKĐ	Observed value of test statistic
Miền bác bỏ giả thuyết	Rejection region
Miền không bác bỏ giả thuyết	Non-rejection region
Miền bác bỏ bên phải	Right regection region
Miền bác bỏ bên trái	Left rejection region

Tiếng Việt	**Tiếng Anh**
Miền bác bỏ hai phía	Two - taied rejection region
Điểm (giá trị tới hạn)	Critical(Value) point
Mức ý nghĩa của kiểm định	Significance level
Sai lầm loại một	Type I error
Sai lầm loại hai	Type II error
Xác suất mắc sai lầm loại 1	Probability of a type I error
Xác suất mắc sai lầm loại 2	Probability of a type II error
Lực kiểm định	Testing power
Thủ tục kiểm định giả thuyết TK	Decision rule
Giá trị xác suất của kiểm định	P - value
Định lý Neyman - Pearson	Neyman- Pearson theorem
Kiểm định tham số	Parametric hypothesis test
Kiểm định phi tham số	Nonparametric hypothsis test
Kiểm định khi bình phương	Chi - squares test
Bảng tiếp liên	Contingent table
Quy luật đa thức	Multinomial distribution
Kiểm định Kolmogorov-Smirnov test	
Kiểm định Lilliefors	Lilliefors test
Kiểm định Jarque - Bera	Jarque - Bera test
Kiểm định theo dấu	Sign test
Kiểm định tổng hạng Wilcoxon	Wilcoxon rank - sum test
Kiểm định tổng hạng theo dấu Wilcoxon	Wilcoxon signed - rank test
Kiểm định Kruskal - Wallis	Kruskal - Wallis test
Đoạn mạch	Runs
Kiểm định đoạn mạch	Runs test
Kiểm định Wald - Wolfowitz	Wald - Wolfowitz test

Tiếng Việt	**Tiếng Anh**
Phân tích phương sai - ANOVA	analysis of variance
Bảng phân tích phương sai	ANOVA table
Nhân tố	Factor
Mô hình hiệu quả cố định	fixed effects models
Mô hình hiệu quả ngẫu nhiên	random effects models
Mô hình một nhân tố	one factor models
Mô hình hai nhân tố	multifactor models
Phân tích phương sai bảng ngẫu nhiên hai chiều	Twoo-Way Analysis of Variance
Bảng ngẫu nhiên hai chiều	Twoo- Way Table
Hàm hồi quy	Regression function
Hàm hồi quy tuyến tính	Linear Regression function
Hàm hồi quy đơn	Simple Regression function
Mô hình hồi quy đơn	Simple Regression Model
Hàm hồi quy bội	Multiple Regression function
Mô hình hồi quy bội	Multiple Regression Model
Hàm hồi quy tổng thể	Population Regression Function
Hàm hồi quy mẫu	Sample Regression function
Biến độc lập	independent variable
Biến phụ thuộc	dependent variable
Phương pháp bình phương bé nhất	method of least squares
Ước lượng (tham số) hồi quy	regression estimators
Sai số chuẩn	standard errors
Tham số hồi quy	regression parameters
Hệ số tương quan	correlation coeficient
Dự báo	prediction

Tiếng Việt	**Tiếng Anh**
Biến dự báo	predictor
Hệ số tương quan riêng phần	partial correlation coeficient
Hệ số xác định	coeficient of determination
Hiệu chỉnh (Điều chỉnh)	adjusted
Chuỗi thời gian	time series
Phân tích xu thế	trend analysis
Phân tích mùa vụ	seasonality analysis
Phân tích chu kỳ	cyclycal analysis
Phương pháp bình quân di động	moving average method
Phương pháp san mũ	exponential smoothing methods
Mô hình nhân	multiplicative model
Mô hình cộng	additive model
Phương pháp san mũ Holt - Winters	Holt - Winters exponential smoothing methods
ảnh của các thể	Image of the case
Biểu đồ thành phần chính	Component Plot
Biểu đồ thành phần chính khi đÓ quay	Component Plot in Rotated Space
Giá trị riêng	Eigenvalue
Ma trận chiếu	Matrix of projection
Ma trận hệ số tương quan	Component matrix
Ma trận trung tâm hóa (qui tâm)	Matrix centre
Nhân tố chính	Principal factor
Phân tích nhân tố	Factor analysis
Phân tích thành phần chính	Principal component analysis
Phương pháp quay	Rotation method
Phép quay trực giao	orthogonal rotation
Phương pháp triết xuất (dữ liệu)	Extration mehod

Tiếng Việt	**Tiếng Anh**
Quán tính	Inertia
Thành phần chính	Principal component
Véc tơ riêng	Eigenvector
Véc tơ trung tâm	Centre vector
Vết ma trận	Trace of matrix
Biến chính tắc	Canonical variable
Biến định tính	Qualitative variable
Độ đo khi bình phương	Chi-Square metrice
Độ đo khoảng cách	Distance Measure
Điểm cột	Point of column
Phân tích chính tắc (tương quan chính tắc)	Canonical correlation analysis
Điểm dòng	Point of row
Mô tả đồng thời	Multi-presentation
Nhân tố chính tắc	Canonical factor
Phân tích phân biệt	Discriminant analysis
Phân tích tương ứng	Correspondence analysis
Phương pháp chuẩn hóa	Normalization Method
Biểu đồ thang bậc	Agglomeration schedule
Các phương pháp phân lớp	Classifications methods
Phân lớp theo thứ bậc	Classification hierarchical
Phương pháp tâm di động	Method to centre dinamic
Phương pháp đám mây di động	Method the nuages dinamic
Biểu diễn Lé vy-Khinchin	Lé vy-Khinchin representation
Biến phân hữu hạn	Finite variation
Biến ngẫu nhiên đa hợp	Compound random variable
Cường độ	Rate

Tiếng Việt	**Tiếng Anh**
Đến trước phục vụ trước	First-come, first-served (FCFS)
Hàm Dirac	DiracÒs delta function
Hàm đáp ứng	The response function
Hàm đặc trưng	The characteristic function
Hàm sinh momment	The moment-generating function
Hàm sinh xác suất	The probability-generating function
Hàm tự hiệp phương sai	The autocovariance function
Hàm tự tương quan	The autocorrelation function
Hồi quy dương	Positive recurrent
Mật độ kích thước nhảy	A jump size density
Phân phối dừng	Stationary distribution
Phân phối đều	Uniform distribution
Phân phối Gamma	Gamma distribution
Phân phối hữu hạn chiều	Finite-dimensional distributions
Phân phối mũ	Exponential distribution
Phân phối nhị thức	Binomial distribution
Phân phối Poisson	Poisson distribution
Phương trình cân bằng	The balance equations
Phương trình Kolmogorov lùi	Kolmogorov backward equantions
Phương trình Kolmogorov tiến	Kolmogorov forward equantions
Quá trình cấp 2	Second order stochastic processes
Quá trình cấp có số gia dừng	Processes with stationary in crements
Quá trình cấp có số gia độc lập	Processes with independent increments
Quá trình dừng theo nghĩa hẹp	Strict-sensestationary processes (SSS)
Quá trình dừng theo nghĩa rộng	Wide-sense stationary processes (WSS)

Tiếng Việt	**Tiếng Anh**
Quá trình đếm	Counting processes
Quá trình Lé vy	Lé vy processes
Quá trình Markov	Markov processes
Quá trình ngẫu nhiên	Stochastic processes
Quá trính Poisson bù	Compensated Poisson processes
Quá trình Poisson đa hợp	Compound Poisson processes
Quá trình đa hợp bù	The compensated compound Poisson processes
Quá trình Poisson được lọc	Filtered Poisson processes
Quá trình Poisson không thuần nhất	The nonhomogeneous Poisson processes
Quá trình Poisson thuần nhất	The homogeneous Poisson processes
Quá trình sinh tử	Birth and death processes
Số gia độc lập	Independent increments
Số gia dừng	Stationary increments
Thời gian đến	The arrival times (Waiting times)
Thời gian đến trung gian	Interarrival times
Thống kê thứ tự	Order statistics
Tỉ lệ chuyển đổi tức thời	Instantaneous transition rates
Trượt	Drift
Xích Markov	Markov Chains

Tài liệu tham khảo

Tiếng Việt

[1] Lê Sĩ Đồng (2013), *Giáo trình xác suất - Thống kê*, Nhà xuất bản Giáo dục Việt Nam.

[2] Trần Lộc Hùng (2005), *Giáo trình xác suất thống kê*, Nhà xuất bản Giáo dục.

[3] Trần Lộc Hùng (2005), *Hướng dẫn giải bài tập xác suất và thống kê*, Nhà xuất bản Giáo dục.

[4] Nguyễn Văn Hữu, Đào Hữu Hồ, Hoàng Hữu Như (2004), *Thống kê toán học*, Nhà xuất bản Đại học Quốc Gia Hà Nội.

[5] Lê Khánh Luận, Nguyễn Thanh Sơn (2011), *Lý thuyết xác suất và thống kê*, Nhà xuất bản Đại học Quốc Gia Tp. Hồ Chí Minh.

[6] Lê Khánh Luận, Nguyễn Thanh Sơn (2013), *Bài tập xác suất và thống kê*, Nhà xuất bản Đại học Quốc Gia Tp. Hồ Chí Minh.

[7] Nguyễn Viết Phú, Nguyễn Duy Tiến (2004), *Cơ sở lý thuyết xác suất*, Nhà xuất bản Đại học Quốc gia Hà Nội.

[8] Nguyễn Duy Tiến, Vũ Viết Yên (2006), *Lý thuyết xác suất*, Nhà xuất bản Giáo dục.

[9] Nguyễn Cao Văn, Trần Thái Ninh, Ngô Văn Thứ (2012), *Giáo trình Lý thuyết xác suất và Thống kê toán*, Nhà xuất bản Đại học Kinh tế Quốc dân.

[10] Nguyễn Cao Văn, Trần Thái Ninh, Nguyễn Thế Hệ (2013), *Bài tập Xác suất và Thống kê toán*, Nhà xuất bản Đại học Kinh tế Quốc dân.

Tiếng Anh

[11] Feller W.(1971),*An introduction to probability theory and its applications*, volume I, third edition, John Wiley and Sons, New York.

[12] Feller W.(1971),*An introduction to probability theory and its applications*, volume II, 2nd edition, John Wiley and Sons, New York.