TRƯỜNG ĐẠI HỌC TÀI CHÍNH – MARKETING
BỘ MÔN TOÁN THỐNG KÊ

Giáo Trình

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ ỨNG DỤNG

(***Dành cho chương trình chất lượng cao***)

Mã số : GT – 15 – 21

Nhóm biên soạn:
Nguyễn Huy Hoàng (Chủ biên)
Nguyễn Trung Đông
Nguyễn Văn Phong
Dương Thị Phương Liên
Nguyễn Tuấn Duy
Võ Thị Bích Khuê

THÀNH PHỐ HỒ CHÍ MINH - 2021

MỤC LỤC

LỜI MỞ ĐẦU

Các bạn đang có trong tay cuốn **"Lý thuyết xác suất và thống kê ứng dụng"** dành cho sinh viên hệ chất lượng cao, trường đại học Tài chính – Maketing. Đây là giáo trình dành cho sinh viên khối ngành kinh tế và quản trị kinh doanh với thời lượng 3 tín chỉ (45 tiết giảng); chính vì vậy chúng tôi cố gắng lựa chọn các nội dung căn bản, trọng yếu và có nhiều ứng dụng trong kinh tế và quản trị kinh doanh; chú trọng ý nghĩa và khả năng áp dụng của kiến thức; giáo trình được biên tập trên cơ sở tham khảo nhiều giáo trình quốc tế cũng như trong nước (xem phần tài liệu tham khảo), cũng như kinh nghiệm giảng dạy nhiều năm của các tác giả; giáo trình dành cho hệ đào tạo chất lượng cao nên chúng tôi cũng rất quan tâm việc giới thiệu thuật ngữ Anh – Việt, giúp sinh viên có thể tự đọc, tự nghiên cứu các tài liệu viết bằng tiếng Anh.

Nội dung giáo trình đã được thiết kế phù hợp với chương trình đào tạo và trình độ của sinh viên khối ngành kinh tế và quản trị kinh doanh. Giáo trình bao gồm 6 chương và một số phụ lục;

Chương 1. Trình bày về biến cố ngẫu nhiên và xác suất.

Chương 2. Trình bày về đại lượng ngẫu nhiên và phân phối xác suất.

Chương 3. Trình bày về mẫu ngẫu nhiên và bài toán ước lượng khoảng tin cậy.

Chương 4. Trình bày về bài toán kiểm định giả thuyết thống kê.

Chương 5. Trình bày về nội dung phân tích phương sai.

Chương 6. Trình bày về phân tích dãy số thời gian.

Cuối mỗi chương, chúng tôi có giới thiệu một số thuật ngữ Anh – Việt và tài liệu tham khảo.

Phần cuối, chúng tôi biên soạn một số đề tham khảo để sinh viên có cơ hội thử sức, tự rèn luyện và một số phụ lục để tiện cho sinh viên có thể tự tra cứu.

Do đối tượng người đọc là sinh viên chuyên ngành kinh tế và quản trị kinh doanh nên chúng tôi chọn cách tiếp cận đơn giản không quá đi sâu về lý thuyết mà chủ yếu quan tâm vào ý nghĩa và áp dụng trong kinh tế quản trị kinh doanh của khái niệm và kết quả lý thuyết xác suất và thống kê toán, chúng tôi cũng sử dụng nhiều ví dụ để người học dễ hiểu, dễ áp dụng; Giáo trình do Giảng viên cao cấp TS. Nguyễn Huy Hoàng và ThS. Nguyễn Trung Đông biên tập phần lý thuyết, TS. Nguyễn Tuấn Duy, TS. Võ Thị Bích Khuê, ThS. Nguyễn Văn Phong và ThS. Dương Thị Phương Liên biên tập phần bài tập các chương, đề tham khảo và phần phụ lục; đây là các giảng viên của Bộ môn Toán – Thống kê, trường đại học Tài chính – Marketing, đã có nhiều năm kinh nghiệm nghiên cứu và giảng dạy Lý thuyết xác suất và Thống kê ứng dụng cho sinh viên khối ngành kinh tế và quản trị kinh doanh.

Lần đầu biên soạn, nên giáo trình này không tránh khỏi sai sót. Rất mong nhận được sự góp ý của các độc giả để lần sau giáo trình được hoàn thiện hơn.

Mọi ý kiến đóng góp xin gởi về địa chỉ email:

hoangtoancb@ufm.edu.vn và nguyendong@ufm.edu.vn.

Xin trân trọng cảm ơn Trường đại học Tài chính – Marketing đã hỗ trợ kinh phí và tạo điều kiện cho giáo trình sớm đến với bạn đọc!

Các tác giả

MỘT SỐ KÝ HIỆU

1. Ω: Không gian mẫu.
2. w: Biến cố sơ cấp.
3. $P(A)$: Xác suất biến cố A.
4. $\mu_X = E(X)$: Kỳ vọng (trung bình) của biến cố X.
5. $\overline{X}$: Trung bình mẫu của X.
6. $\sigma_X^2 = var(X) = D(X)$: Phương sai của biến cố X.
7. S_X^2: Phương sai ngẫu cóa hiệu chỉnh của X.
8. X: Biến ngẫu nhiên X.
9. $X \sim B(n;p)$: Biến ngẫu nhiên X có phân phối nhị thức.
10. $X \sim H(N,K,n)$: Biến ngẫu nhiên X có phân phối siêu bội.
11. $X \sim P(\mu)$: Biến ngẫu nhiên X có phân phối Poisson.
12. $X \sim U[a,b]$: Biến ngẫu nhiên X có phân phối đều.
13. $X \sim Exp(\lambda)$: Biến ngẫu nhiên X có phân phối mũ.
14. $X \sim N(0,1)$: Biến ngẫu nhiên X có phân phối chuẩn tắc.
15. $X \sim N(\mu,\sigma^2)$: Biến ngẫu nhiên X có phân phối chuẩn.
16. $X \sim \Gamma(\alpha,\beta)$: Biến ngẫu nhiên X có phân phối Gamma.
17. $X \sim \chi^2(r)$: Biến ngẫu nhiên X có phân phối Chi bình phương.
18. $X \sim St(n)$: Biến ngẫu nhiên X có phân phối Student.
19. $X \sim F(n,m)$: Biến ngẫu nhiên X có phân phối Fisher.
20. δ: Lượng tăng giảm tuyệt đối liên hoàn.
21. Δ: Lượng tăng giảm tuyệt đối định gốc.
22. Σ: Ký hiệu tổng.
23. Π: Ký hiệu tích.
24. H_0: Giả thuyết H_0.
25. H_1: Đối thuyết (nghịch thuyết) H_1.

Chương 1 | BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT

Mục tiêu chương 1

Chương này giúp sinh viên:

- Phân biệt được sự kiện ngẫu nhiên (đối tượng môn xác suất) và sự kiện tất định (đối tượng của vật lý và hóa học). Nắm được các khái niệm về phép thử, không gian mẫu, biến cố và biến cố sơ cấp cũng như các biến cố đặc biệt.
- Hiểu được thế nào là xác suất và biết một số định nghĩa về xác suất.
- Biết và áp dụng được công thức xác suất đầy đủ và công thức xác suất Bayes.
- Biết áp dụng công thức Bernoulli để tính xác suất.

1.1. Phép thử và các loại biến cố

1.1.1. Sự kiện ngẫu nhiên và phép thử

Sự kiện ngẫu nhiên là những sự kiện dù được thực hiện trong cùng một điều kiện như nhau vẫn có thể cho nhiều kết quả khác nhau. Chẳng hạn, tung một con xúc xắc, ta không thể chắc chắn rằng mặt nào sẽ xuất hiện; lấy ra một sản phẩm từ một lô hàng gồm cả hàng chính phẩm lẫn phế phẩm, ta không chắc chắn sẽ nhận được hàng chính phẩm hay phế phẩm. Sự kiện ngẫu nhiên là đối tượng khảo sát của lý thuyết xác suất.

Mỗi lần cho xảy ra một sự kiện ngẫu nhiên được gọi là thực hiện một *phép thử*, còn sự kiện có thể xảy ra trong kết quả của phép thử đó gọi là *biến cố*. Khi đó, dù ta không thể dự đoán được kết quả nào sẽ xảy ra nhưng thường ta có thể liệt kê tất cả các kết quả có thể xảy ra. Tập hợp tất cả các kết quả có thể xảy ra của một phép thử được gọi là *không gian mẫu*. Ký hiệu Ω.

Ví dụ 1.1. Xét phép thử "tung một con xúc xắc". Ta có thể nhận được mặt 1 chấm, mặt 2 chấm, …, mặt 6 chấm. Vậy không gian mẫu có thể liệt kê và ký hiệu như sau:

$$\Omega = \{1,2,3,4,5,6\}.$$

1.1.2. Các loại biến cố

- Biến cố chắc chắn là biến cố luôn luôn sẽ xảy ra khi thực hiện một phép thử, biến cố chắc chắn thường ký hiệu là U.

- Biến cố không thể có là biến cố không bao giờ xảy ra khi thực hiện một phép thử, biến cố không thể có thường ký hiệu là V.

- Biến cố ngẫu nhiên là biến cố có thể xảy ra hoặc không xảy ra khi thực hiện phép thử. Ta có thể xem biến cố ngẫu nhiên là một tập con của không gian mẫu, các biến cố ngẫu nhiên thường ký hiệu là $A, B, C, ... \subset \Omega$.

- Biến cố sơ cấp là một kết quả (kết cục) của không gian mẫu, ký hiệu $w \in \Omega$. Do đó, không gian mẫu là tập hợp tất cả các biến cố sơ cấp.

Ví dụ 1.2. Thực hiện phép thử tung một con xúc xắc. Ta có:

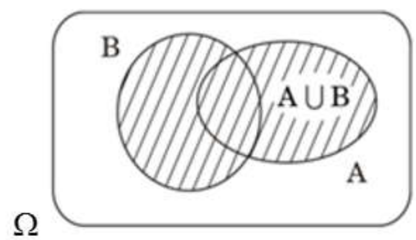
- Không gian mẫu: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Biến cố: "nhận được mặt có số chấm ≤ 6" là biến cố chắc chắn.
- Biến cố: "nhận được mặt có 7 chấm" là biến cố không thể có.
- Biến cố: "nhận được mặt có số chấm là chẵn" là biến cố ngẫu nhiên.
- Biến cố: "nhận được mặt 1 chấm" là biến cố sơ cấp.

1.1.3. Các phép toán giữa các biến cố

1.1.3.1. Tổng các biến cố

Cho hai biến cố bất kỳ $A, B \subset \Omega$, ta có thể thành lập biến cố:

$A \cup B \equiv A + B$ là chỉ biến cố "A xảy ra hay B xảy ra khi thực hiện phép thử".



Hình 1.1. Hình vẽ minh họa tổng hai biến cố.

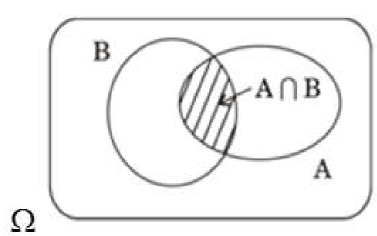
Tổng quát, cho $B_1, B_2, ..., B_n \subset \Omega$, ta có thể thành lập biến cố:

$\bigcup_{i=1}^{n} B_i \equiv \sum_{i=1}^{n} B_i$ là chỉ biến cố "có ít nhất một trong n biến cố đó xảy ra khi thực hiện phép thử".

1.1.3.2. Tích các biến cố

Cho hai biến cố bất kỳ $A, B \subset \Omega$, ta có thể thành lập biến cố:

$A \cap B \equiv A \cdot B$ là chỉ biến cố "A và B cùng xảy ra khi thực hiện phép thử".



Hình 1.2. Hình vẽ minh họa tích hai biến cố.

Tổng quát, cho $B_1, B_2, ..., B_n \subset \Omega$, ta có thể thành lập biến cố:

$\bigcap_{i=1}^{n} B_i \equiv \prod_{i=1}^{n} B_i$ là chỉ biến cố "cả n biến cố đó cùng xảy ra khi thực hiện phép thử".

Ví dụ 1.3. Khảo sát một lớp học về sự yêu thích môn xác suất thống kê và môn kinh tế học. Chọn ngẫu nhiên một sinh viên trong lớp này.

Gọi A là biến cố "nhận được sinh viên thích môn xác suất thống kê" và B là biến cố "nhận được sinh viên thích môn kinh tế học". Suy ra

Biến cố "sinh viên thích ít nhất một môn" là biến cố: $A + B$.

Biến cố "sinh viên thích cả hai môn" là biến cố: AB.

1.1.4. Quan hệ giữa các biến cố

1.1.4.1. Hai biến cố độc lập

Hai biến cố A và B được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra của biến cố A không phụ thuộc vào việc biến cố B xảy ra hay không xảy ra và ngược lại.

Nếu hai biến cố A và B không độc lập với nhau thì ta gọi là hai biến cố phụ thuộc.

Tổng quát,

- $B_1, B_2, ..., B_n$ là họ các biến cố độc lập với nhau từng đôi nếu hai biến cố bất kỳ trong n biến cố đó độc lập với nhau.
- $B_1, B_2, ..., B_n$ là họ các biến cố độc lập toàn phần nếu mỗi biến cố đó độc lập với một tổ hợp bất kỳ của các biến cố còn lại.

1.1.4.2. Hai biến cố xung khắc

Hai biến cố A và B gọi là xung khắc với nhau nếu chúng không thể đồng thời xảy ra trong cùng một phép thử.

A và B xung khắc khi và chỉ khi $A \cap B = \varnothing$.

Tổng quát, cho $B_1, B_2, ..., B_n$ là họ các biến cố xung khắc từng đôi một nếu bất kỳ 2 biến cố nào trong nhóm này cũng xung khắc với nhau, nghĩa là

$$B_i \cap B_j = \varnothing \text{ với } i \neq j \text{ và } i, j = \overline{1, n}.$$

Ví dụ 1.4. Trong một giỏ hàng có hai loại sản phẩm: Sản phẩm loại 1 và sản phẩm loại 2. Lấy ngẫu nhiên từ giỏ hàng đó ra một sản phẩm.

Gọi A là biến cố "nhận được sản phẩm loại 1".

Gọi B là biến cố "nhận được sản phẩm loại 2".

$\Rightarrow$ A và B là 2 biến cố xung khắc.

Ví dụ 1.5. Gieo đồng thời hai con xúc xắc.

Gọi C là biến cố "Con xúc xắc thứ nhất xuất hiện mặt 6 chấm".

Gọi D là biến cố "Con xúc xắc thứ hai xuất hiện mặt 6 chấm".

$\Rightarrow$ C và D không xung khắc.

1.1.4.3. Họ đầy đủ các biến cố

$B_1, B_2, ..., B_n$ là họ đầy đủ các biến cố nếu thỏa mãn hai điều kiện sau:

i) $B_1 \cup B_2 \cup ... \cup B_n = \Omega$ và

ii) $B_i \cap B_j = \varnothing, \; i \neq j.$

Ví dụ 1.6. Gieo một con xúc xắc.

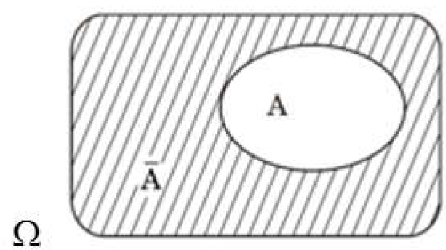
Gọi B_i là biến cố "nhận được mặt có i chấm", $i = \overline{1,6}$.

Các biến cố $B_1, B_2, ..., B_6$ tạo nên một họ đầy đủ các biến cố.

1.1.4.4. Hai biến cố đối lập

Hai biến cố A và $\overline{A}$ gọi là hai biến cố đối lập với nhau nếu chúng tạo nên một họ đầy đủ các biến cố.

A và $\overline{A}$ là hai biến cố đối lập $\Leftrightarrow A \cup \overline{A} = \Omega$ và $A \cap \overline{A} = \varnothing$.



Hình 1.3. Hình vẽ minh họa hai biến cố đối lập.

1.2. Xác suất của biến cố

1.2.1. Khái niệm chung về xác suất

Quan sát các biến cố đối với một phép thử, mặc dù không thể khẳng định một biến cố có xảy ra hay không nhưng người ta có thể phỏng chừng cơ may xảy ra của các biến cố này là ít hay nhiều. Chẳng hạn, với phép thử *"tung xúc xắc"*, biến cố *"nhận được mặt 1"* ít xảy ra hơn biến cố *"nhận được mặt chẵn"*. Do đó, người ta tìm cách định lượng khả năng xuất hiện khách quan của một biến cố mà ta sẽ gọi là *xác suất* của biến cố đó.

Xác suất của một biến cố là một con số đặc trưng cho khả năng xảy ra khách quan của biến cố đó.

Xác suất của biến cố A, ký hiệu là $P(A)$, có thể được định nghĩa bằng nhiều cách.

1.2.2. Định nghĩa cổ điển

Xét một phép thử τ với n kết quả có thể xảy ra, nghĩa là không gian mẫu Ω có n biến cố sơ cấp, và biến cố $A \subset \Omega$ có k phần tử. Nếu *các biến cố sơ cấp có cùng khả năng xảy ra* thì xác suất của A được định nghĩa là

$$P(A) = \frac{|A|}{|\Omega|} = \frac{k}{n}. \tag{1.1}$$

Trong đó $|A|, |\Omega|$ là số khả năng của biến cố A và số khả năng của Ω.

Ví dụ 1.7.

a) Xét phép thử "tung một con xúc xắc" với các biến cố

$A \equiv$ "nhận được mặt 6",

$B \equiv$ "nhận được mặt chẵn".

Theo công thức (1.1), ta có

$$P(A) = \frac{1}{6} \text{ và } P(B) = \frac{3}{6} = 0,5.$$

b) Xét phép thử "lấy ngẫu nhiên một sản phẩm trong một giỏ hàng đựng 4 sản phẩm loại 1 và 6 sản phẩm loại 2" với các biến cố

$C \equiv$ "nhận được sản phẩm loại 1",

$D \equiv$ "nhận được sản phẩm loại 2".

Theo công thức (1.1), ta có

$$P(C) = \frac{4}{10} = 0,4 \text{ và } P(D) = \frac{6}{10} = 0,6.$$

Lưu ý rằng, đối với định nghĩa cổ điển, ta cần hai điều kiện:

Số kết quả của phép thử là hữu hạn,

Các kết quả đồng khả năng xảy ra.

Khi một trong hai điều kiện trên không xảy ra, ta không thể dùng định nghĩa cổ điển để xác định xác suất của một biến cố. Ta có thể định nghĩa xác suất bằng phương pháp thống kê như sau.

1.2.3. Định nghĩa xác suất bằng tần suất

Giả sử phép thử τ có thể lập lại nhiều lần trong điều kiện giống nhau. Nếu trong n lần thực hiện phép thử mà biến cố A xảy ra k lần thì tỷ số $\frac{k}{n}$ được gọi là *tần suất* xảy ra của A trong n phép thử.

Người ta chứng minh được rằng, khi n đủ lớn, tần suất của biến cố A sẽ dao động xung quanh một giá trị cố định nào đó mà ta gọi là xác suất của A, ký hiệu $P(A)$. Ta có

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

Trong thực tế, với n đủ lớn, người ta lấy tần suất của A làm giá trị gần đúng cho xác suất của biến cố A, nghĩa là

$$P(A) \approx \frac{k}{n}. \tag{1.2}$$

Ví dụ 1.8.

a) Thống kê trên 10000 người dân thành phố cho thấy có 51 người bị bệnh cao huyết áp, theo công thức (1.2), ta nói xác suất của biến cố "bị bệnh cao huyết áp" là

$$\frac{51}{10000} \approx 0,005.$$

b) Một nhà máy gồm ba phân xưởng A, B, C. Kiểm tra một lô hàng của nhà máy gồm 1000 sản phẩm, người ta thấy có 252 sản phẩm của phân xưởng A, 349 của phân xưởng B và 399 của phân xưởng C. Theo công thức (1.2), ta nói xác suất

nhận được sản phẩm từ phân xưởng A là $P(A) = \frac{252}{10000} \approx 0,25,$

nhận được sản phẩm từ phân xưởng B là $P(B) = \frac{349}{10000} \approx 0,35,$ và

nhận được sản phẩm từ phân xưởng C là $P(C) = \frac{399}{10000} \approx 0,4.$

Ta còn nói, các phân xưởng A, B, C tương ứng làm ra 25%, 35% và 40% tổng sản lượng nhà máy.

Tương tự, để tìm xác suất làm ra sản phẩm hỏng của phân xưởng A, người ta thống kê trên một số sản phẩm của phân xưởng A và quan sát số sản phẩm hỏng. Chẳng hạn, nếu trong 400 sản phẩm của phân xưởng A nêu trên có 4 sản phẩm hỏng, theo công thức (1.2), ta nói xác suất làm ra một sản phẩm hỏng của phân xưởng A là $\frac{4}{400} = 0,01.$

Ví dụ 1.9. Xét phép thử τ : "tung đồng xu", một cách trực giác, ta cho rằng các biến cố sơ cấp w_1: "nhận được mặt sấp" và w_2: "nhận được mặt ngửa" là đồng khả năng xảy ra, nên do định nghĩa cổ điển, $P(w_1) = P(w_2) = 0,5.$ Khi đó, người ta nói đồng xu này là "công bằng", "đồng chất đẳng hướng", ... Bằng thực nghiệm, một số nhà khoa học đã tung một đồng xu nhiều lần và nhận được kết quả sau:

Người thực hiện	Số lần thảy	Số lần mặt ngửa	Tần suất
Buffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

và khi đó, ta nói xác suất nhận được mặt ngửa $\approx 0,5$.

1.2.4. Định nghĩa hình học về xác suất

Định nghĩa hình học về xác suất có thể sử dụng khi xác suất để một điểm ngẫu nhiên rơi vào một phần nào đó của một miền cho trước tỷ lệ với độ đo của miền đó (độ dài, diện tích, thể tích…) và không phụ thuộc vào dạng thức của miền đó.

Nếu độ đo hình học của toàn bộ miền cho trước là S, còn độ đo hình học của một phần A nào đó của nó là S_A thì xác suất để điểm ngẫu nhiên rơi vào A sẽ bằng:

$$P = \frac{S_A}{S} \tag{1.3}$$

trong đó: $0 < S, S_A < +\infty$.

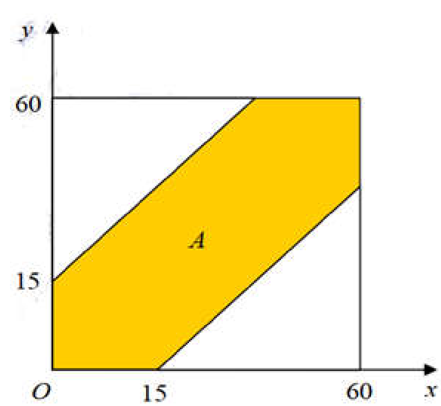
Ví dụ 1.10. Giả sử hai người X và Y hẹn gặp nhau trong khoảng thời gian là 60 phút, với điều kiện người tới trước sẽ đợi người tới sau tối đa 15 phút, sau đó đi khỏi. Tính xác suất để X và Y gặp nhau.

Giải. Gọi x là thời gian đến của X, y là thời gian đến của Y. Khi đó không gian các biến cố sơ cấp sinh ra khi X và Y tới gặp nhau có dạng:

$$\Omega = \left\{(x, y) \in \mathbb{R}^2 : 0 \le x \le 60, 0 \le y \le 60\right\}$$

Gọi A là biến cố hai người gặp nhau, khi đó

$$A = \left\{(x, y) \in \mathbb{R}^2 : |x - y| \le 15\right\}$$



Hình 1.4. Hình vẽ minh họa biến cố hai người gặp nhau.

Theo công thức xác suất hình học (1.3), ta có

$$P(C) = \frac{60 \cdot 60 - 45 \cdot 45}{60 \cdot 60} = \frac{7}{16}.$$

1.2.5. Định nghĩa tiên đề về xác suất

Vào những năm 30 của thế kỷ 20, nhà Toán học người Nga là Kolmogorov đã xây dựng hệ tiên đề làm cơ sở cho việc định nghĩa một cách hoàn chỉnh khái niệm xác suất về mặt lý thuyết. Hệ tiên đề được xây dựng trên cơ sở khái niệm về không gian biến cố sơ cấp $w_1, w_2, ..., w_n$, thực tế là tập hợp tất cả các trường hợp có thể xảy ra của một phép thử. Lúc đó mỗi biến cố A có thể quan niệm như một tập hợp của không gian đó.

Tiên đề 1. Với mọi biến cố A đều có $0 \le P(A) \le 1$.

Tiên đề 2. Nếu $w_1, w_2, ..., w_n$ tạo nên không gian các biến cố sơ cấp thì:

$$P(w_1) + P(w_2) + \cdots + P(w_n) = 1.$$

Tiên đề 3. Nếu các biến cố $A_1, A_2, ..., A_n, ...$ là các tập hợp con không giao nhau của các biến cố sơ cấp thì:

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

Tiên đề 4. Với A, B là hai biến cố bất kỳ, ta có

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

Tiên đề 5. Với hai biến cố A và $\overline{A}$, ta có

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

1.2.6. Nguyên lý xác suất nhỏ và xác suất lớn

Trong nhiều bài toán thực tế, ta thường gặp các biến cố có xác suất rất nhỏ, gần bằng 0. Qua nhiều lần quan sát, người ta thấy rằng: các biến cố có xác suất nhỏ gần như không xảy ra khi thực hiện phép thử. Trên cơ sở đó có thể đưa ra "*Nguyên lý không thực tế không thể có của các biến cố có xác suất nhỏ*" sau đây: *Nếu một biến cố có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử, biến cố đó sẽ không xảy ra.*

Việc quy định một mức xác suất được coi là "rất nhỏ" tuỳ thuộc vào từng bài toán cụ thể. Chẳng hạn: Nếu xác suất để một loại dù không mở khi nhảy dù là 0,01 thì mức xác suất này chưa thể coi là nhỏ và ta không nên sử dụng loại dù đó. Song nếu xác suất để một chuyến tàu đến ga chậm 10 phút là 0,01 thì ta có thể coi mức xác suất đó là nhỏ, tức là có thể cho rằng xe lửa đến ga đúng giờ.

Một mức xác suất nhỏ mà với nó ta có thể cho rằng: biến cố đang xét không xảy ra trong một phép thử được gọi là *mức ý nghĩa*. Tuỳ theo từng bài toán cụ thể, mức ý nghĩa thường được lấy trong khoảng từ 0,01 đến 0,05.

Tương tự như vậy ta có thể nêu ra "Nguyên lý thực tế chắc chắn xảy ra của các biến cố có xác suất lớn" như sau: Nếu một biến cố có xác suất gần bằng 1 thì thực tế có thể cho rằng biến cố đó sẽ xảy ra trong một phép thử.

Cũng như trên, việc quy định mức xác suất được coi là lớn hay nhỏ tuỳ thuộc vào bài toán cụ thể. Thông thường người ta lấy trong khoảng từ 0,95 đến 0,99.

1.3. Xác suất có điều kiện

Xét ví dụ sau: "Tung hai con xúc xắc" với không gian mẫu là

$$\Omega = \{(1,1),(1,2),...,(1,6),(2,1),(2,2),...,(5,6),(6,6)\}$$

(tổng cộng có 36 khả năng (phần tử)) và xét các biến cố

A: "tổng số nút xuất hiện cộng lại bằng 6",

B: "số nút của xúc xắc thứ nhất là số lẻ".

Ta có:

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

$$B = \{(1,1),...,(1,6),(3,1),...,(3,6),(5,1),...,(5,6)\},$$

nên từ định nghĩa cổ điển,

$$P(A) = \frac{|A|}{|\Omega|} = \frac{5}{36} \text{ và } P(B) = \frac{|B|}{|\Omega|} = \frac{18}{36} = 0,5.$$

Bây giờ, ta tung hai con xúc xắc và giả sử ta nhận được thông tin thêm là số nút của xúc xắc thứ nhất đã là số lẻ (nghĩa là biến cố B đã xảy ra). Khi đó, phép thử trên trở thành phép thử: "tung hai con xúc xắc khác nhau với số nút của xúc xắc thứ nhất là số lẻ". Do đó, không gian mẫu Ω bị thu hẹp lại là

$$\Omega' = \{(1,1),...,(1,6),(3,1),...,(3,6),(5,1),...,(5,6)\}$$

và hiện tượng biến cố A xảy ra khi biết biến cố B đã xảy ra trở thành hiện tượng biến cố

$$A' = \{(1,5),(3,3),(5,1)\} = AB$$

xảy ra đối với phép thử τ' và do đó có xác suất là

$$P(A') = \frac{|A'|}{|\Omega'|} = \frac{3}{18} = \frac{1}{6}.$$

Ta ký hiệu $A' = A|B$ và $P(A') = P(A|B)$ được gọi là xác suất để biến cố A xảy ra khi biết biến cố B xảy ra. Từ nhận xét

$$P(A|B) = \frac{1}{6} = \frac{P(AB)}{P(B)}.$$

1.3.1. Định nghĩa

Xét biến cố B với $P(B) > 0$. Xác suất của biến cố A, khi biết biến cố B xảy ra là

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{1.4}$$

Ví dụ 1.11. Trong một bình có 5 quả cầu trắng và 3 quả cầu đen. Lấy ngẫu nhiên lần lượt 2 quả cầu theo phương thức không hoàn lại.

Giải

Gọi A_i là biến cố "nhận được quả cầu trắng lần thứ i", i = 1, 2.

Theo định nghĩa xác suất cổ điển, xác suất để lần thứ nhất lấy được cầu trắng là:

$$P(A_1) = \frac{5}{8}$$

Nếu lần thứ nhất lấy được quả cầu trắng (tức là biến cố A_1 đã xảy ra) thì trong bình còn lại 7 quả cầu, trong đó có 4 quả cầu trắng. Vậy xác suất để lần thứ hai lấy được cầu trắng với điều kiện lần thứ nhất đã lấy được cầu trắng là:

$$P(A_2|A_1) = \frac{4}{7}.$$

Nếu lần thứ nhất lấy được quả cầu đen (tức là biến cố $\overline{A}_1$ đã xảy ra) thì trong bình còn lại 7 quả cầu, trong đó có 5 quả cầu trắng. Vậy xác suất để lần thứ hai lấy được cầu trắng với điều kiện lần thứ nhất đã lấy được cầu đen là:

$$P(A_2|\overline{A}_1) = \frac{5}{7}.$$

1.3.2. Công thức nhân xác suất

Với hai biến cố A và B bất kỳ, ta có

$$P(AB) = P(A)P(B|A). \tag{1.5}$$

Tổng quát, với n biến cố bất kỳ $A_1, A_2, ..., A_n$, ta có

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2...A_{n-1}). \tag{1.6}$$

Ví dụ 1.12. Một thủ quỹ có một chùm chìa khóa gồm 9 chiếc chìa giống hệt nhau trong đó chỉ có 2 chìa có thể mở được tủ sắt. Anh ta thử ngẫu nhiên từng chìa (chìa không trúng được bỏ ra trong lần thử kế tiếp). Tìm xác suất để anh ta mở được tủ vào đúng lần thứ ba.

Giải

Đặt A_i là biến cố "lần thứ i, mở được tủ". Với quy ước rằng khi biến cố A_i xảy ra thì các biến cố $A_1, A_2, ..., A_{i-1}$ vẫn có thể đã xảy ra, biến cố "mở được tủ vào đúng lần thứ ba" là $\overline{A}_1\overline{A}_2A_3$ và do quy tắc nhân xác suất, ta có

$$P\left(\overline{A}_1\overline{A}_2A_3\right) = P\left(\overline{A}_1\right)P\left(\overline{A}_2\middle|\overline{A}_1\right)P\left(A_3\middle|\overline{A}_1\overline{A}_2\right).$$

Do

$$P\left(\overline{A}_1\right) = 1 - P\left(A_1\right) = 1 - \frac{2}{9} = \frac{7}{9},$$

$$P\left(\overline{A}_2\middle|\overline{A}_1\right) = 1 - P\left(A_2\middle|\overline{A}_1\right) = 1 - \frac{2}{8} = \frac{6}{8},$$

$$P\left(A_3\middle|\overline{A}_1\overline{A}_2\right) = \frac{2}{7},$$

ta suy ra:

$$P\left(\overline{A}_1\overline{A}_2A_3\right) = \frac{7}{9}\cdot\frac{6}{8}\cdot\frac{2}{7} = \frac{1}{6}.$$

1.3.3. Công thức xác suất đầy đủ (công thức xác suất toàn phần)

Với hai biến cố A, B bất kỳ, ta có

$$P(A) = P(B)P(A|B) + P(\overline{B})P(A|\overline{B}). \tag{1.7}$$

Tổng quát, cho $B_1, B_2, ..., B_n$ là họ đầy đủ các biến cố và với mọi biến cố A, ta có

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n). \tag{1.8}$$

hay

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i). \tag{1.9}$$

Chứng minh

Do BA và $\overline{B}A$ là hai biến cố xung khắc và $A = BA + \overline{B}A$ nên

$$\begin{aligned} P(A) &= P\left(BA + \overline{B}A\right) = P(BA) + P\left(\overline{B}A\right) \\ &= P(B)P(A|B) + P(\overline{B})P(A|\overline{B}). \end{aligned}$$

Tổng quát, do các biến cố $B_1A, B_2A, ..., B_nA$ xung khắc từng đôi một và $A = B_1A + B_2A + \cdots B_nA$ nên do công thức cộng xác suất:

$$
\begin{aligned}
P(A) &= P(B_1A + B_2A + \cdots B_nA) = P(B_1A) + P(B_2A) + \cdots P(B_nA) \\
&= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n) \\
&= \sum_{i=1}^{n} P(B_i)P(A|B_i).
\end{aligned}
$$

và do công thức nhân xác suất,

$$P(B_iA) = P(B_i)P(A|B_i)$$

với mọi i, ta suy ra

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

Ví dụ 1.13. Có 2 xạ thủ loại I và 8 xạ thủ loại II, xác suất bắn trúng đích của các loại xạ thủ loại I là 0,9 và loại II là 0,7.

a) Chọn ngẫu nhiên ra một xạ thủ và xạ thủ đó bắn một viên đạn. Tìm xác suất để viên đạn đó trúng đích.

b) Chọn ngẫu nhiên ra hai xạ thủ và mỗi người bắn một viên đạn. Tìm xác suất để cả hai viên đạn đó trúng đích.

Giải

a) Gọi A là biến cố "Viên đạn trúng đích".

B_1 là biến cố "Chọn xạ thủ loại I bắn".

B_2 là biến cố "Chọn xạ thủ loại II bắn".

$$P(B_1) = \frac{2}{10} = 0,2, \quad P(A|B_1) = 0,9$$

$$P(B_2) = \frac{8}{10} = 0,8, \quad P(A|B_2) = 0,7$$

Ta có B_1, B_2 tạo thành họ đầy đủ các biến cố. Áp dụng công thức (1.8), ta có:

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) = 0,2 \cdot 0,9 + 0,8 \cdot 0,7 = 0,74.$$

b) Gọi B là biến cố "Cả 2 viên đạn trúng đích".

$B_i, (i = 1,2)$ là biến cố "Chọn được i xạ thủ loại I ".

$$P(B_0) = \frac{C_8^2}{C_{10}^2} = \frac{28}{45}; \; P(B|B_0) = 0,7.0,7 = 0,49$$

$$P(B_1)=\frac{C_2^1.C_8^1}{C_{10}^2}=\frac{16}{45};\ P(B|B_1)=0,9.0,7=0,63$$

$$P(B_2)=\frac{C_2^2}{C_{10}^2}=\frac{1}{45};\ P(B|B_2)=0,9.0,9=0,81.$$

Ta có B_1, B_2, B_3 tạo thành họ đầy đủ các biến cố. Áp dụng công thức(1.9), ta có

$$\begin{aligned}P(B)&=P(B_0)\cdot P(B|B_0)+P(B_1)\cdot P(B|B_1)+P(B_2)\cdot P(B|B_2)\\&=\frac{28}{45}\cdot 0,49+\frac{16}{45}\cdot 0,63+\frac{1}{45}0,81=0,5469.\end{aligned}$$

1.3.4. Công thức Bayes

Cho $B_1,B_2,...,B_n$ là họ đầy đủ các biến cố và xét biến cố A với $P(A)>0.$ Với mỗi $k=1,2,...,n,$ ta có

$$P(B_k|A)=\frac{P(B_k)\cdot P(A|B_k)}{\sum_{i=1}^{n}P(B_i)P(A|B_i)}. \tag{1.10}$$

Chứng minh

Áp dụng công thức nhân xác suất

$$P(A)P(B_k|A)=P(AB_k)=P(B_kA)=P(B_k)P(A|B_k)$$

và công thức xác suất toàn phần

$$P(A)=\sum_{i=1}^{n}P(B_i)P(A|B_i),$$

ta suy ra

$$P(B_k|A)=\frac{P(B_k)\cdot P(A|B_k)}{P(A)}=\frac{P(B_k)\cdot P(A|B_k)}{\sum_{i=1}^{n}P(B_i)P(A|B_i)}.$$

Ví dụ 1.14. Tỷ lệ chính phẩm của máy thứ nhất là 99%, của máy thứ hai là 98%. Một lô sản phẩm gồm 40% sản phẩm của máy thứ nhất và 60% sản phẩm của máy thứ hai. Người ta lấy ngẫu nhiên ra một sản phẩm để kiểm tra thấy là sản phẩm tốt. Tìm xác suất để sản phẩm đó do máy thứ nhất sản xuất.

Giải

Gọi A là biến cố "Sản phẩm kiểm tra là sản phẩm tốt"

B_1 là biến cố "Sản phẩm do máy thứ nhất sản xuất".

B_2 là biến cố "Sản phẩm do máy thứ hai sản xuất".

$$P(B_1) = 40\% = 0,4; \ P(B_2) = 60\% = 0,6$$

$$P(A|B_1) = 99\% = 0,99; \ P(A|B_2) = 98\% = 0,98$$

Do B_1, B_2 là họ đầy đầy đủ các biến cố. Áp dụng công thức (1.10), ta có

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1)+P(B_2)P(A|B_2)}$$
$$= \frac{0,4 \cdot 0,99}{0,4 \cdot 0,99 + 0,6 \cdot 0,98} = 0,4.$$

1.3.5. Sự độc lập của các biến cố

Hai biến cố A, B được gọi là *độc lập* nếu xác suất để biến cố này xảy ra không phụ thuộc vào việc biến cố kia xảy ra, nghĩa là

$$P(A|B) = P(A)$$

và do đó

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

Tổng quát, n biến cố $A_1, A_2, ..., A_n$ được gọi là *độc lập* nếu mỗi biến cố A_i, với $i = 1,2,...,n$, độc lập với tích bất kỳ các biến cố còn lại.

Do định nghĩa, nếu ba biến cố A, B, C là độc lập thì A độc lập với B, C và BC nên

$$P(AB) = P(A)P(B),$$

$$P(AC) = P(A)P(C),$$

$$P[A(BC)] = P(A)P(BC),$$

và vì B, C cũng độc lập với nhau, nên

$$P(BC) = P(B)P(C),$$

và do đó

$$P(ABC) = P(A)P(B)P(C). \tag{1.12}$$

Chú ý. Nếu A và B là biến cố độc lập thì $\overline{A}$ và B; A và $\overline{B}$; $\overline{A}$ và $\overline{B}$ cũng độc lập.

Ví dụ 1.15. Trong một bình có 5 quả cầu trắng và 3 quả cầu đen. Lấy ngẫu nhiên lần lượt 2 quả cầu. Tính xác suất để lấy được 2 quả cầu trắng trong hai trường hợp sau:

a) Lấy hoàn lại.

b) Lấy không hoàn lại.

Giải.

Gọi A là biến cố "Lấy được 2 quả cầu trắng".

A_i là biến cố "Lần thứ i lấy được cầu trắng", i = 1, 2.

Suy ra biến cố lấy được hai của cầu trắng là: $A_1.A_2$

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

a) Nếu lấy 2 quả cầu theo phương thức lần lượt có hoàn lại thì hai biến cố A_1 và A_2 là độc lập với nhau. Theo công thức (1.5), ta có

$$P(A) = P(A_1.A_2) = P(A_1)\cdot P(A_2) = \frac{5}{8}\cdot\frac{5}{8} = \frac{25}{64}.$$

b) Nếu lấy 2 quả cầu theo phương thức lần lượt không hoàn lại thì hai biến cố A_1 và A_2 là phụ thuộc với nhau. Theo công thức (1.5), ta có

$$P(A) = P(A_1.A_2) = P(A_1)\cdot P(A_2|A_1) = \frac{5}{8}\cdot\frac{4}{7} = \frac{5}{14}.$$

Ví dụ 1.16. Tung một đồng xu 3 lần. Tìm xác suất để 3 lần đều được mặt sấp.

Gọi A_i, $(i = 1,2,3)$ là biến cố "nhận được mặt sấp lần tung thứ i",

Ta có

$$P(A_i) = \frac{1}{2}$$

A là biến cố "Tung 3 lần đều được mặt sấp".

$$A = A_1A_2A_3$$

Các biến cố A_1, A_2, A_3 là độc lập toàn phần. Theo công thức (1.12), ta có

$$P(A) = P(A_1A_2A_3) = P(A_1)\cdot P(A_2)\cdot P(A_3) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}.$$

1.4. Công thức Bernoulli

Trong nhiều bài toán thực tế ta thường gặp trường hợp cùng một phép thử được lặp lại nhiều lần. Trong kết quả của mỗi phép thử có thể xảy ra hoặc không xảy ra một biến cố A nào đó và ta không quan tâm đến kết quả của từng phép thử mà quan tâm đến tổng số lần xảy ra của biến cố A trong cả dãy phép thử đó. Chẳng hạn, nếu tiến hành sản xuất hàng loạt một loại chi tiết nào đó thì ta thường quan tâm đến tổng số chi tiết đạt chuẩn của cả quá trình sản xuất. Trong những bài toán như vậy cần phải biết cách xác định xác suất để biến cố A xảy ra một số lần nhất định trong kết quả của cả một dãy phép thử. Bài toán này sẽ được giải quyết khá dễ dàng nếu các phép thử là độc lập với nhau.

Các phép thử được gọi là độc lập với nhau nếu xác suất xảy ra một biến cố nào đó trong từng phép thử sẽ không phụ thuộc vào biến cố đó có xảy ra ở các phép thử khác hay không. Chẳng hạn, tung nhiều lần một đồng xu sẽ tạo nên các phép thử độc lập, lấy nhiều lần sản phẩm từ một lô sản phẩm theo phương thức hoàn lại cũng tạo nên các phép thử độc lập v.v…

Giả sử ta thực hiện n phép thử độc lập. Trong mỗi phép thử chỉ có hai trường hợp: hoặc biến cố A xảy ra, hoặc biến cố A không xảy ra. Xác suất xảy ra của biến cố A trong mỗi phép thử đều bằng p và xác suất không xảy ra của biến cố A trong mỗi phép thử đều bằng $q = 1 - p$. Những bài toán thỏa mãn cả ba giả thiết được gọi là tuân theo lược đồ Bernoulli. Khi đó xác suất để trong n phép thử độc lập biến cố A xuất hiện đúng k lần, ký hiệu $P_n(k)$, được tính bằng công thức Bernoulli:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, 2, ..., n. \tag{1.13}$$

Đặt H_k : "biến cố A xảy ra đúng k lần", với $0 \le k \le n.$ Ta có

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

Chứng minh

Dùng quy nạp trên n. Hiển nhiên công thức đúng với $n = 1$ vì khi đó $H_0 = \overline{A}$ và $H_1 = A.$ Do đó

$$P(H_0) = C_1^0 p^0 (1-p)^{1-0} = 1 - p$$

và

$$P(H_1) = C_1^1 p^1 (1-p)^{1-1} = p.$$

Giả sử công thức đúng với $n \ge 1,$ nghĩa là khi thực hiện n lần phép thử τ một cách độc lập thì xác suất để biến cố A xảy ra đúng k lần là

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

Bây giờ, thực hiện phép thử τ thêm một lần nữa một cách độc lập và gọi X là biến cố: *"A xảy ra trong lần thử thứ* $n+1$" thì biến cố: *"A xảy ra đúng k lần trong* $n+1$ *phép thử*" là

$$H_k\overline{A} + H_{k-1}A.$$

Do các biến cố $H_k\overline{A}$ và $H_{k-1}A$ là xung khắc, H_k và $\overline{A}$ cũng như H_{k-1} và A là các biến cố độc lập nên

$$
\begin{aligned}
P\left(H_k\overline{A}+H_{k-1}A\right) &= P\left(H_k\overline{A}\right)+P\left(H_{k-1}A\right)\\
&= P\left(H_k\right)P\left(\overline{A}\right)+P\left(H_{k-1}\right)P\left(A\right)\\
&= C_n^k p^k\left(1-p\right)^{n-k}\left(1-p\right)+ C_n^{k-1}p^{k-1}\left(1-p\right)^{n-k+1}p\\
&= C_n^k p^k\left(1-p\right)^{n-k+1}+ C_n^{k-1}p^k\left(1-p\right)^{n-k+1}\\
&= \left(C_n^k+ C_n^{k-1}\right)p^k\left(1-p\right)^{n-k+1}\\
&= C_{n+1}^k p^k\left(1-p\right)^{(n+1)-k}.
\end{aligned}
$$

Ví dụ 1.17. Xác suất chữa khỏi bệnh A của một phương pháp điều trị là 95%. Với 10 người bị bệnh A được điều trị bằng phương pháp này, tính xác suất để

a) có 8 người khỏi bệnh.

b) có nhiều nhất 9 người khỏi bệnh.

Giải

Do việc khỏi bệnh của người này và người khác là độc lập nhau nên số người khỏi bệnh trong 10 người điều trị thỏa lược đồ Bernoulli với $n=10$ và $p=0,95$. Theo công thức (1.13). Ta có

$$P\left(H_k\right)= C_{10}^k 0,05^k\left(0,95\right)^{10-k}$$

a) Xác suất để có 8 người khỏi bệnh là $P\left(H_8\right)= C_{10}^8\left(0,05\right)^8\left(0,95\right)^{10-8}=0,0746$.

b) Biến cố: *"có nhiều nhất 9 người khỏi bệnh"* là biến cố đối của biến cố : *"có 10 người khỏi bệnh"* nên có xác suất là

$$P\left(H_{k\le 9}\right)= 1-C_{10}^{10}\left(0,05\right)^{10}\left(0,95\right)^{10-10}=0,4013.$$

1.5. Tóm tắt chương 1

1. Xác suất của biết cố A:

$$P\left(A\right)=\frac{\left|A\right|}{\left|\Omega\right|}$$

($\left|A\right|$ và $\left|\Omega\right|$ lần lượt là số khả năng thuận lợi cho A và Ω).

2. Tính chất:

i) $0\le P\left(A\right)\le 1$.

ii) $P\left(\overline{A}\right)=1-P\left(A\right)$.

3. Công thức cộng:

i) Nếu $A_1, A_2, ..., A_n$ xung khắc với nhau từng đôi một thì

$$P(A_1 + A_2 + \cdots + A_2) = P(A_1) + P(A_2) + \cdots + P(A_2).$$

ii) Với A và B là hai biết cố bất kỳ

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

4. Công thức xác suất có điều kiện:

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

5. Công thức nhân:

i) Nếu $A_1, A_2, ..., A_n$ bất kỳ thì

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1A_2....A_{n-1}).$$

ii) Nếu $A_1, A_2, ..., A_n$ độc lập với nhau từng đôi một thì

$$P(A_1A_2 \cdots A_2) = P(A_1)P(A_2)\cdots P(A_n).$$

6. Công thức đầy đủ (toàn phần) và công thức Bayes:

Với $B_1, B_2, ..., B_n$ là họ đầy đủ các biến cố và với mọi biến cố A, ta có

i) Công thức đầy đủ

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

ii) Công thức Bayes

$$P(B_k|A) = \frac{P(B_k)\cdot P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}, \quad k = 1, 2, ..., n.$$

7. Công thức Bernoulli:

Đặt H_k : "biến cố A xảy ra đúng k lần", với $0 \le k \le n, 0 < p < 1.$ Ta có

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

1.6. Bài tập

Biểu diễn các biến cố

Bài số 1. Kiểm tra 3 sản phẩm. Gọi A_k là biến cố sản phẩm thứ k tốt. Hãy trình bày các cách biểu diễn qua A_k và qua giản đồ Venn các biến cố sau đây:

A: tất cả đều xấu,

B: có ít nhất một sản phẩm xấu,

C: có ít nhất một sản phẩm tốt,

D: không phải tất cả sản phẩm đều tốt,

E: có đúng một sản phẩm xấu,

F: có ít nhất 2 sản phẩm tốt.

Bài số 2. Ba người, mỗi người bắn một phát. Gọi A_i là biến cố thứ i bắn trúng. Hãy biểu diễn qua A_i các biến cố sau :

A: chỉ có người thứ nhất bắn trúng,

B: người thứ nhất bắn trúng và người thứ hai bắn trật,

C: cả 3 người đều bắn trúng,

D: có ít nhất 2 người bắn trúng,

E: chỉ có 2 người bắn trúng,

F: không ai bắn trúng,

G: không có hơn 2 người bắn trúng,

H: người thứ nhất bắn trúng, hoặc người thứ hai và người thứ ba cùng bắn trúng,

I: người thứ nhất bắn trúng hay người thứ hai bắn trúng,

K: có ít nhất 1 người bắn trúng.

Bài số 3. Ba sinh viên A, B, C cùng thi môn xác suất thống kê. Xét các biến cố:

A: sinh viên A đậu,

B: sinh viên B đậu,

C: sinh viên C đậu.

Hãy biểu diễn qua A, B, C các biến cố sau:

a) chỉ có A đậu,

b) A đậu và B rớt,

c) có ít nhất một người đậu,

d) cả 3 cùng đậu,

e) có ít nhất 2 người đậu,

f) chỉ có 2 người đậu,

g) không ai đậu,

h) không có quá 2 người đậu.

Bài số 4. Quan sát 4 sinh viên làm bài thi. Kí hiệu $B_j (j = 1,2,3,4)$ là biến cố sinh viên j làm bài thi đạt yêu cầu. Hãy viết các biến cố sau đây

a) Có đúng một sinh viên đạt yêu cầu,

b) có đúng 3 sinh viên đạt yêu cầu,

c) có ít nhất 1 sinh viên đạt yêu cầu,

d) không có sinh viên nào đạt yêu cầu.

Định nghĩa xác suất, xác suất có điều kiện, công thức cộng, công thức nhân

Bài số 5. Thống kê 2000 sinh viên một khóa của trường đại học theo giới tính và ngành học thu được các số liệu sau:

	Nam	Nữ
Học tài chính ngân hàng	400	500
Học quản trị kinh doanh	800	300

Lấy ngẫu nhiên một sinh viên khóa đó. Tìm xác suất để nhận được:

a) Sinh viên là Nam.

b) Sinh viên học tài chính ngân hàng.

c) Sinh viên nam và tài chính ngân hàng.

d) Hoặc sinh viên nam, hoặc học tài chính ngân hàng.

e) Nếu đã chọn được một sinh viên nam thì xác suất để người đó học tài chính ngân hàng bằng bao nhiêu?

Đáp số: a) 0,6; b) 0,45; c) 0,2; d) 0,85; e) 1/3.

Bài số 6. Một công ty liên doanh cần tuyển một kế toán trưởng, một trưởng phòng tiếp thị, có 40 người dự tuyển trong đó có 15 nữ. Tính xác suất trong 2 người được tuyển có:

a) kế toán trưởng là nữ,

b) ít nhất 1 nữ.

Đáp số: a) 0,375;b) 0,6154.

Bài số 7. Một lô hàng có 10 sản phẩm, trong đó có 7 sản phẩm tốt, 3 sản phẩm xấu. Lấy ngẫu nhiên từ lô hàng ra 4 sản phẩm. Tính xác suất để 4 sản phẩm lấy ra có 3 sản phẩm tốt.

Đáp số: 0,5.

Bài số 8. Một lớp học có 50 học sinh trong kỳ thi giỏi Toán và Văn, trong đó có 20 người giỏi Toán, 25 người giỏi Văn, 10 người giỏi cả Toán lẫn Văn. Chọn ngẫu nhiên 1 học sinh của lớp này. Tính xác suất để học sinh được chọn giỏi Toán hoặc Văn.

Đáp số: 0,7.

Bài số 9. Trong 1 khu phố, tỷ lệ người mắc bệnh tim là 6%; mắc bệnh phổi là 8% và mắc cả hai bệnh là 5%. Chọn ngẫu nhiên 1 người trong khu phố đó. Tính xác suất để người đó không mắc cả 2 bệnh tim và bệnh phổi.

Đáp số: 0,91.

Bài số 10. Trong 100 người phỏng vấn có 40 người thích dùng nước hoa A, 28 người thích dùng nước hoa B, 10 người thích dùng cả 2 loại A, B. Chọn ngẫu nhiên 1 người trong số 100 người trên. Tính xác suất người này :

a) thích dùng ít nhất 1 loại nước hoa trên,

b) không dùng loại nào cả.

Đáp số: a) 0,58; b) 0,42.

Bài số 11. Một cơ quan có 210 người, trong đó có 100 người ở gần cơ quan, 60 người trong 100 người gần cơ quan là nữ, biết rằng số nữ chiếm gấp đôi số nam trong cơ quan.

Chọn ngẫu nhiên 1 người trong cơ quan. Tính xác suất :

a) người này là nam,

b) người này ở gần cơ quan,

c) người này phải trực đêm (người trực đêm phải ở gần cơ quan hoặc là nam).

Đáp số: a) 1/3; b) 0,4762; c) 0,619.

Bài số 12. Cho A và B là 2 biến cố sao cho $P(A)=\frac{1}{2}, P(B)=\frac{1}{3}, P(AB)=\frac{1}{6}$. Hãy tính:

1) $P(A+B)$ 2) $P(\overline{A}+\overline{B})$ 3) $P(\overline{A+B})$

4) $P(\overline{AB})$ 5) $P(A\overline{B})$ 6) $P(\overline{A}B)$

7) $P(\overline{A}+B)$ 8) $P(A|B)$ 9) $P(\overline{A}|B)$

10) $P(AB|B)$ 11) $P(A\overline{B}|B)$ 12) $P(A\overline{B}|\overline{B})$

Bài số 13. Đội tuyển cầu lông của Trường Đại học Tài chính - Marketing có 3 vận động viên, mỗi vận động viên thi đấu một trận. Xác suất thắng trận của các vận viên A, B, C lần lượt là: 0,9; 0,7; 0,8. Tính xác suất :

a) Đội tuyển thắng ít nhất 1 trận,

b) Đội tuyển thắng 2 trận,

c) C thua, biết rằng đội tuyển thắng 2 trận.

Đáp số: a) 0,994; b) 0,398; c) 0,3166.

Bài số 14. Cho 3 biến cố A, B, C sao cho

P(A) = 0,5; P(B) = 0,7; P(C) = 0,6; P(AB) = 0,3; P(BC) = 0,4; P(AC) = 0,2 và P(ABC) = 0,1.

a) Tìm xác suất để cả 3 biến cố A, B, C đều không xảy ra.

b) Tìm xác suất để có đúng 2 trong 3 biến cố đó xảy ra.

c) Tìm xác suất để chỉ có đúng 1 biến cố trong 3 biến cố đó xảy ra.

Đáp số: a) 0; b) 0,6; c) 0,3.

Bài số 15. Một người có 5 con gà mái, 2 con gà trống nhốt chung trong một cái lồng. Một người đến mua, người bán gà bắt ngẫu nhiên 1 con. Người mua chấp nhận con đó.

a) Tính xác suất để người đó mua được con gà mái.

Người thứ hai lại đến mua, người bán gà lại bắt ngẫu nhiên ra 1 con.

b) Tìm xác suất để người thứ hai mua được con gà trống.

c) Xác suất này sẽ bằng bao nhiêu nếu người bán gà quên mất rằng con gà bán cho người thứ nhất là gà trống hay gà mái.

Đáp số: a) $\frac{5}{7}$; *b)* $\frac{1}{3}$; *c)* $\frac{2}{7}$.

Bài số 16. Hai công ty A, B cùng kinh doanh một mặt hàng. Xác suất để công ty A thua lỗ là 0,2; xác suất để công ty B thua lỗ là 0,4. Tuy nhiên trên thực tế, khả năng cả 2 công ty cùng thua lỗ là 0,1. Tìm xác suất để

a) chỉ có một công ty thua lỗ,

b) có ít nhất một công ty làm ăn không thua lỗ.

Đáp số: a) 0,4; b) 0,9.

Bài số 17. Một thủ quỹ có một chùm chìa khóa gồm 12 chiếc bề ngoài giống hệt nhau, trong đó có 4 chìa mở được cửa chính của thư viện. Cô ta thử từng chìa một một cách ngẫu nhiên, chìa nào không trúng thì bỏ ra. Tìm xác suất để cô ta mở được cửa chính của thư viện ở lần mở thứ 5.

Đáp số : 0,071.

Bài số 18. Một lô hàng có 6 sản phẩm tốt, 4 sản phẩm xấu. Lấy ngẫu nhiên không hoàn lại từng sản phẩm cho đến khi lấy được 2 sản phẩm tốt thì ngừng,

a) Tính xác suất để ngừng lại ở lần lấy sản phẩm thứ 2,

b) Biết đã ngừng lại ở lần lấy sản phẩm thứ 4. Tính xác suất để lần lấy thứ nhất lấy được sản phẩm tốt.

Đáp số a) $\frac{1}{3}$; *b)* $\frac{1}{3}$.

Bài số 19. Một chàng trai viết 4 lá thư cho 4 cô gái; nhưng vì đãng trí nên anh ta bỏ 4 lá thư vào 4 phong bì một cách ngẫu nhiên, dán kín rồi mới ghi địa chỉ gửi,

a) Tính xác suất để không có cô nào nhận đúng thư viết cho mình,

b) Tính xác suất để có ít nhất 1 cô nhận đúng thư của mình,

c) Tổng quát hóa với n cô gái. Tính xác suất có ít nhất 1 cô nhận đúng thư. Xấp xỉ giá trị xác suất này khi cho $n \to \infty$.

Đáp số a) 0,625; b) 0,375; c) e^{-1}.

Bài số 20. Trong 1 lô hàng 10 sản phẩm có 2 sản phẩm xấu, chọn không hoàn lại để phát hiện ra 2 sản phẩm xấu, khi nào chọn được sản phẩm xấu thứ 2 thì dừng lại.

a) Tính xác suất dừng lại ở lần chọn thứ 4.

b) Biết rằng đã chọn được sản phẩm xấu ở lần chọn thứ nhất, tính xác suất dừng lại ở lần chọn thứ 4.

c) Nếu việc kiểm tra dừng lại ở lần chọn thứ 3, tính xác suất lần chọn đầu được sản phẩm xấu.

Đáp số : a) $\frac{1}{15}$; *b)* $\frac{1}{9}$; *c)* $\frac{1}{2}$.

Bài số 21. Đội tuyển bóng bàn Thành phố có 4 vận động viên A, B, C, D . Mỗi vận động viên thi đấu 1 trận, với xác suất thắng trận lần lượt la : 0,6, 0,7, 0,8, 0,9. Tính

a) xác suất đội tuyển thắng ít nhất 1 trận,

b) xác suất đội tuyển thắng 2 trận,

c) xác suất đội tuyển thắng 3 trận,

d) xác suất D thua, trong trường hợp đội tuyển thắng 3 trận.

Đáp số: a) 0,9976; b) 0,2144; c) 0,4404; d) 0,763.

Bài số 22. Ở một cơ quan nọ có 3 chiếc ôtô. Khả năng có sự cố của mỗi xe ôtô lần lượt là 0,15 ; 0,20 ; 0,10.

a) Tìm khả năng 3 ôtô cùng bị hỏng.

b) Tìm khả năng có ít nhất 1 ôtô hoạt động tốt.

c) Tìm khả năng cả 3 ôtô cùng hoạt động được.

d) Tìm xác suất có không quá 2 ôtô bị hỏng.

Công thức xác suất đầy đủ và công thức Bayes

Đáp số: a) 0,003; b) 0,997; c) 0,612; d) 0,997.

Bài số 23. Một nhà máy sản xuất bóng đèn, máy A sản xuất 25%, máy B: 35%, máy C: 40% số bóng đèn. Tỉ lệ sản phẩm hỏng của mỗi máy trên số sản phẩm do máy đó sản xuất lần lượt là 3%, 2%, 1%. Một người mua 1 bóng đèn do nhà máy sản xuất.

a) Tính xác suất để sản phẩm này tốt.

b) Biết rằng sản phẩm này là xấu. Tính xác suất để sản phẩm do máy C sản xuất.

Đáp số: a) 0,9815;b) 0,2162.

Bài số 24. Trong một trạm cấp cứu bỏng : 80% bệnh nhân bỏng do nóng, 20% bỏng do hóa chất. Loại bỏng do nóng có 30% bi biến chứng, loại bỏng do hóa chất có 50% bị biến chứng.

a) Chọn ngẫu nhiên một bệnh án. Tính xác suất để gặp một bệnh án của bệnh nhân bị biến chứng.

b) Rút ngẫu nhiên được một bệnh án của một bệnh nhân bị biến chứng. Tính xác suất để bệnh án đó là của bệnh nhân bị biến chứng do nóng gây ra? do hóa chất gây ra?

Đáp số: a) 0,34; b) 0,7059; 0,2941.

Bài số 25. Một lô hạt giống được phân thành ba loại. Loại 1 chiếm 2/3 số hạt cả lô, loại 2 chiếm 1/4, còn lại là loại 3. Loại 1 có tỉ lệ nẩy mầm 80%, loại 2 có tỉ lệ nẩy mầm 60% và loại 3 có tỉ lệ nẩy mầm 40%. Hỏi tỉ lệ nẩy mầm chung của lô hạt giống là bao nhiêu?

Đáp số: 0,72.

Bài số 26. Hai nhà máy cùng sản xuất 1 loại linh kiện điện tử. Năng suất nhà máy hai gấp 3 lần năng suất nhà máy một. Tỷ lệ hỏng của nhà máy một và hai lần lượt là 0,1% và 0,2%. Giả sử linh kiện bán ở Trung tâm chỉ do hai nhà máy này sản xuất. Mua 1 linh kiện ở Trung tâm.

a) Tính xác suất để linh kiện ấy hỏng.

b) Giả sử mua linh kiện và thấy linh kiện bị hỏng. Theo ý bạn thì linh kiện đó do nhà máy nào sản xuất.

Đáp số: a) 0,175%; b) nhà máy 2.

Bài số 27. Có 8 bình đựng bi, trong đó có :

2 bình loại 1: mỗi bình đựng 6 bi trắng 3 bi đỏ,

3 bình loại 2: mỗi bình đựng 5 bi trắng 4 bi đỏ,

3 bình loại 3: mỗi bình đựng 2 bi trắng 7 bi đỏ.

Lấy ngẫu nhiên một bình và từ bình đó lấy ngẫu nhiên 1 bi.

a) Tính xác suất để bi lấy ra là bi trắng.

b) Biết rằng bi lấy ra là bi trắng. Tính xác suất để bình lấy ra là bình loại 3.

Đáp số: a) 0,4583; b) 0,182.

Bài số 28. Một chuồng gà có 9 con gà mái và 1 con gà trống. Chuồng gà kia có 1 con mái và 5 con trống. Từ mỗi chuồng lấy ngẫu nhiên 1 con đem bán. Các con gà còn lại được dồn vào chuồng thứ ba. Nếu ta lại bắt ngẫu nhiên 1 con gà nữa từ chuồng này ra thì xác suất để bắt được con gà trống là bao nhiêu?

Đáp số: 0,3619.

Bài số 29. Có 2 hộp áo; hộp một có 10 áo trong đó có 1 phế phẩm; hộp hai có 8 áo trong đó có 2 phế phẩm. Lấy ngẫu nhiên 1 áo từ hộp một bỏ sang hộp hai; sau đó từ hộp này chọn ngẫu nhiên ra 2 áo. Tìm xác suất để cả 2 áo này đều là phế phẩm.

Đáp số: 1/30.

Bài số 30. Có 3 xạ thủ cùng bắn vào một con thú, mỗi người bắn 1 viên đạn, với xác suất bắn trúng lần lượt là 0,6; 0,7; 0,8. Biết rằng nếu trúng 1 phát đạn thì xác suất để con thú bị tiêu diệt là 0,5; trúng 2 phát thì xác suất để con thú bị tiêu diệt là 0,8; còn nếu trúng 3 phát đạn thì chắc chắn con thú bị tiêu diệt.

a) Tính xác suất con thú bị tiêu diệt.

b) Giả sử con thú bị tiêu diệt. Tính xác suất nó bị trúng 2 phát đạn.

Đáp số: a) 0,7916; b) 0,4568.

Bài số 31. Có 3 hộp bi; hộp một có 10 bi trong đó có 3 bi đỏ; hộp hai có 15 bi trong đó có 4 bi đỏ; hộp ba có 12 bi trong đó có 5 bi đỏ. Gieo một con xúc xắc. Nếu xuất hiện mặt 1 thì chọn hộp một, xuất hiện mặt hai thì chọn hộp 2, xuất hiện các mặt còn lại thì chọn hộp ba. Từ hộp được chọn, lấy ngẫu nhiên 1 bi

a) Tính xác suất để được bi đỏ,

b) Giả sử lấy được bi đỏ. Tính xác suất để bi đỏ này thuộc hộp hai.

Đáp số: a) 0,3722; b) 0,1194.

Bài số 32. Một hộp có 15 quả bóng bàn, trong đó có 9 mới 6 cũ, lần đầu chọn ra 3 quả để sử dụng, sau đó bỏ vào lại, lần hai chọn ra 3 quả.

a) Tính xác suất 3 quả bóng chọn lần hai là 3 bóng mới.

b) Biết rằng lần hai chọn được 3 bóng mới, tính xác suất lần đầu chọn được 2 bóng mới.

Đáp số: a) 0,0893; b) 0,4089.

Bài số 33. Có 3 cái thùng. Thùng 1 có 6 bi trắng, 4 bi đỏ; thùng 2 có 5 bi trắng, 5 bi đỏ và thùng 3 có 10 bi trắng. Giả sử người ta lấy ngẫu nhiên 2 bi từ thùng 1 bỏ vào thùng 2. Sau đó, lại lấy ngẫu nhiên 1 bi từ thùng 2 bỏ vào thùng 3 rồi từ thùng 3 lấy ngẫu nhiên ra 1 bi. Tìm xác suất để bi lấy ra là đỏ.

Đáp số: 0,0439.

Công thức Bernoulli

Bài số 34. Một bác sĩ chữa khỏi bệnh A cho một người với xác suất là 95%. Giả sử có 10 người bị bệnh A đến chữa một cách độc lập nhau. Tính xác suất để

a) Có 8 người khỏi bệnh,

b) Có nhiều nhất 9 người khỏi bệnh.

Đáp số: a) 0,0746; b) 0,4013.

Bài số 35. Một thiết bị có 10 chi tiết với độ tin cậy của mỗi chi tiết là 0,9. (Xác suất làm việc tốt trong khoảng thời gian nào đó).

Tính xác suất để trong khoảng thời gian ấy :

a) Có đúng một chi tiết làm việc tốt,

b) Có ít nhất 2 chi tiết làm việc tốt.

Đáp số: a) $9 \cdot 10^{-9} \approx 0$; *b)* 1.

Bài số 36. Một cầu thủ đá thành công quả phạt 11m với xác suất 80%.

- Đá 4 thành công 2.

- Đá 6 thành công 3.

Công việc nào dễ thực hiện ?

Đáp số: Đá 4 thành công 2 dễ hơn.

Bài số 37. Trong một thành phố có 70% dân cư thích xem bóng đá. Chọn ngẫu nhiên 10 người, tính xác suất có :

a) 5 người thích xem bóng đá,

b) ít nhất 2 người thích xem bóng đá.

Đáp số: a) 0,1029; b) 0,9999.

Bài số 38. Một nhà toán học có xác suất giải được một bài toán khó là 0,9. Cho nhà toán học này 5 bài toán khó được chọn một cách ngẫu nhiên.

a) Tính xác suất để nhà toán học này giải được 3 bài.

b) Tính xác suất để nhà toán học này giải được ít nhất 1 bài.

c) Tính số bài có khả năng nhất mà nhà toán học này giải được.

Đáp số: a) 0,0729; b) 0,99999; c) 5 bài.

Bài số 39. Tỷ lệ mắc bệnh Basedow ở một vùng rừng núi nào đó là 70%. Trong đợt khám tuyển sức khoẻ để xuất cảnh, người ta khám cho 100 người. Tìm xác suất để

a) Trong 100 người có 60 người bị Basedow,

b) Trong 100 người có 75 người bị Basedow,

c) Trong 100 người có ít nhất một người bị Basedow.

Đáp số: a) 0,0085; b) 0,0496; c) 1.

Bài số 40. Một lô hàng với tỷ lệ phế phẩm là 5%. Cần phải lấy mẫu cỡ bao nhiêu sao cho xác suất để bị ít nhất một phế phẩm không bé hơn 0,95.

Đáp số: $n \geq 59$.

Bài số 41. Hai đấu thủ A, B thi đấu cờ. Xác suất thắng của người A trong một ván là 0,6 (không có hòa). Trận đấu bao gồm 5 ván, người nào thắng một số ván lớn hơn là người thắng cuộc. Tính xác suất để người B thắng cuộc.

Đáp số: 0,31744.

Bài số 42. Một máy sản xuất lần lượt từng sản phẩm. Xác suất sản xuất ra một phế phẩm của máy là 0,01.

a) Cho máy sản xuất 10 sản phẩm. Tính xác suất để có 2 phế phẩm.

b) Máy cần sản xuất ít nhất bao nhiêu sản phẩm để xác suất có ít nhất một chính phẩm trên 0,99.

Đáp số: a) 0,0042; b) 2.

Bài số 43. Một xí nghiệp có hai phân xưởng A và B cùng sản xuất một loại sản phẩm với tỉ lệ phế phẩm tương ứng là 2% và 3%. Cho mỗi phân xưởng sản xuất ra 5 sản phẩm. Tính xác suất để số phế phẩm do hai phân xưởng sản xuất là bằng nhau.

Đáp số:0,7885.

1.7. Tài liệu tham khảo

[1] Nguyễn Cao Văn, Trần Thái Ninh, Ngô Văn Thứ, Lý thuyết xác suất và thống kê toán, NXB Đại học Kinh tế Quốc dân, 2012.

[2] Nguyễn Cao Văn, Trần Thái Ninh, Nguyễn Thế Hệ, Bài tập xác suất và thống kê toán, NXB Đại học Kinh tế Quốc dân, 2012.

[3] Phạm Hoàng Uyên, Lê Thị Thiên Hương, Huỳnh Văn Sáu, Nguyễn Phúc Sơn, Huỳnh Tố Uyên, Lý thuyết xác suất, NXB Đại học Quốc gia Thành Phố Hồ Chí Minh, 2015.

[4] Anderson, Sweeney, and William [2010], Statistics for Business and Economics, South-Western Cengage Learning (11^{th} Edition).

[5] Michael Barrow, Statistics for Economics, Accounting and Business Studies-Prentice Hall, 2006.

[6] Newbold Paul - Statistics for Bussiness and Economics, 5th edition - Prentice Hall, 2005.

Thuật ngữ chính chương 1	
Tiếng Anh	**Tiếng Việt**
Axiom	Tiên đề
Addition rule	Công thức cộng
Bayes’ formula	Công thức Bayes
Countable additivity	Công thức cộng
Classical probability	Xác suất cổ điển
Conditional probability	Xác suất có điều kiện
Define	Định nghĩa
Event	Biến cố
Element	Phần tử
Experiment	Phép thử
Experimental probability	Xác suất thực nghiệm
Event A occurs	Biến cố A xảy ra
Independent	Độc lập
Infinite sequence of outcomes	Dãy vô hạn kết cục
Monotonicity	Tính đơn điệu
Mutually exclusive events	Họ biến cố xung khắc
Multiplication rule	Công thức nhân
Outcome	Kết cục
Odds in favor	Tỷ lệ thuận lợi
Relative frequency	Tần số tương đối
Probability	Xác suất
Posterior probability	Xác suất hậu nghiệm
Prior probability	Xác suất tiên nghiệm
Sample space	Không gian mẫu
Subset	Tập con
Sequence of mutually exclusive events	Một dãy các biến cố xung khắc từng đôi một
The finite additivity of the probability	Tính cộng hữu hạn của xác suất
The law of large numbers	Luật số lớn
The number of elements	Số lượng các phần tử

Chương 2 **ĐẠI LƯỢNG NGẪU NHIÊN VÀ PHÂN PHỐI XÁC SUẤT**

Mục tiêu chương 2

Chương này giúp sinh viên:

- Phân biệt được thế nào là đại lượng ngẫu nhiên liên tục và đại lượng ngẫu nhiên rời rạc.
- Thành lập được bảng phân phối xác suất và tính được các hàm xác suất, hàm phân phối.
- Tính và nêu được ý nghĩa các tham số đặc trưng như trung bình, phương sai, mốt, trung vị,...
- Biết áp dụng các luật phân phối như: Nhị thức, Siêu bội, Poisson, Chuẩn, ….

2.1. Đại lượng ngẫu nhiên

2.1.1. Khái niệm

Xét phép thử τ với không gian mẫu Ω. Giả sử, ứng với mỗi biến số sơ cấp $w \in \Omega$, ta liên kết với một số thực $X(w) \in \mathbb{R}$, thì X được gọi là một *biến số ngẫu nhiên*.

Ví dụ 2.1. Với trò chơi sấp ngửa bằng cách tung đồng xu, giả sử nếu xuất hiện mặt sấp, ta được 1 đồng; nếu xuất hiện mặt ngửa, ta mất 1 đồng. Khi đó, ta có

Phép thử τ : "*tung đồng xu*",

Không gian mẫu $\Omega = \{S, N\}$,

Biến số ngẫu nhiên X với $X(S) = 1$ và $X(N) = -1$.

Tổng quát, biến số ngẫu nhiên X của một phép thử τ với không gian mẫu Ω là một ánh xạ

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

2.1.2. Phân loại

- Khi $X(\Omega)$ là một tập hợp hữu hạn $\{x_1, x_2, ..., x_k\}$ hay là một dãy $\{x_1, x_2, ..., x_n, ...\}$, ta nói X là một *biến số ngẫu nhiên rời rạc.*

Ví dụ 2.2. Số chấm xuất hiện ở mặt trên của xúc xắc; số sinh viên vắng mặt trong một buổi học; số máy hỏng trong từng ca sản xuất,… là các biến ngẫu nhiên rời rạc.

- Khi $X(\Omega)$ là một khoảng của $\mathbb{R}$ (hay cả $\mathbb{R}$), ta nói X là một *biến số ngẫu nhiên liên tục*.

Ví dụ 2.3. Gọi X là kích thước của chi tiết do một máy sản xuất ra, X là biến ngẫu nhiên liên tục.

Do các biến số ngẫu nhiên X là các ánh xạ có giá trị trong $\mathbb{R}$ nên với một hàm số $u:\mathbb{R}\to\mathbb{R}$, ta có thể thành lập biến số ngẫu nhiên $u(X)$, với

$$\begin{aligned} u(X):\Omega &\to \mathbb{R} \\ w &\mapsto u\left[X(w)\right] \end{aligned}$$

Chẳng hạn, với $u(x)=x-\mu$, $\mu\in\mathbb{R}$ ta có biến số ngẫu nhiên $u(X)=X-\mu$, với $(X-\mu)(w)=X(w)-\mu.$ với mọi $w\in\Omega$, và với $u(x)=(x-\mu)^2$, $\mu\in\mathbb{R}$ ta có biến số ngẫu nhiên $u(X)=(X-\mu)^2$, với $(X-\mu)^2(w)=(X(w)-\mu)^2$ với mọi $w\in\Omega$.

2.2. Phân phối xác suất của đại lượng ngẫu nhiên

2.2.1. Đại lượng ngẫu nhiên rời rạc

Để xác định một biến số ngẫu nhiên rời rạc, người ta cần xác định các giá trị x_i, $i=1,2,...$có thể nhận được bởi biến ngẫu nhiên này và đồng thời cũng cần xác định xác suất để X nhận giá trị này là bao nhiêu, nghĩa là, cần xác định

$$P\{w\in\Omega: X(w)=x_i\}\equiv P(X=x_i), \text{ với } i=1,2,...$$

2.2.1.1. Bảng phân phối xác suất

Xét biến số ngẫu nhiên rời rạc $X:\Omega\to\mathbb{R}$, với $X(\Omega)=\{x_1,x_2,...,x_n,...\}$. Giả sử $x_1<x_2<\cdots<x_n<\cdots$. Ta lập bảng các giá trị tương ứng

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

với $p_i=P(X=x_i)$, được gọi là *bảng phân phối xác suất* của X.

Ví dụ 2.4. Trong hộp có 10 sản phẩm trong đó có 6 chính phẩm. Lấy ngẫu nhiên 2 sản phẩm. Lập bảng phân phối xác suất cho số chính phẩm được lấy ra.

Giải

Gọi X là số chính phẩm được lấy ra trong 3 sản phẩm, $X\in\{0,1,2\}$.

$$P(X=0)=\frac{C_4^2}{C_{10}^2}=\frac{6}{45}=\frac{2}{15},$$

$$P(X=1)=\frac{C_6^1.C_4^1}{C_{10}^2}=\frac{24}{45}=\frac{8}{15},$$

$$P(X=2)=\frac{C_6^2}{C_{10}^2}=\frac{15}{45}=\frac{1}{3}.$$

Vậy bảng phân phối xác suất của X là:

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

Ví dụ 2.5. Một cơ quan có 3 xe ôtô : 1 xe 4 chỗ; 1 xe 50 chỗ và 1 xe tải. Xác suất để trong một ngày làm việc, các xe được xử dụng là 0,8; 0,4 và 0,9. Hãy lập bảng phân phối xác suất cho số xe được xử dụng trong một ngày của cơ quan.

Giải

Gọi X là số xe được xử dụng trong một ngày của cơ quan. Ta có $X \in \{0,1,2,3\}$.

Gọi A_1, A_2, A_3 lần lượt là biến cố *"xe 4 chỗ"; "xe 50 chỗ"; "xe tải"* được xử dụng trong ngày của cơ quan. Khi đó, A_1, A_2, A_3 là các biến cố độc lập, $P(A_1)=0,8;\ P(A_2)=0,4;\ P(A_3)=0,9$ và

$$\begin{aligned}P(X=0)&=P(\overline{A}_1\overline{A}_2\overline{A}_3)=P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)\\&=0,2\cdot 0,6\cdot 0,1=0,012\end{aligned}$$

$$\begin{aligned}P(X=1)&=P(A_1\overline{A}_2\overline{A}_3+\overline{A}_1A_2\overline{A}_3+\overline{A}_1\overline{A}_2A_3)\\&=P(A_1\overline{A}_2\overline{A}_3)+P(\overline{A}_1A_2\overline{A}_3)+P(\overline{A}_1\overline{A}_2A_3)\\&=P(A_1)P(\overline{A}_2)P(\overline{A}_3)+P(\overline{A}_1)P(A_2)P(\overline{A}_3)+P(\overline{A}_1)P(\overline{A}_2)P(A_3)\\&=0,8\cdot 0,6\cdot 0,1+0,2\cdot 0,4\cdot 0,1+0,2\cdot 0,6\cdot 0,9=0,164\end{aligned}$$

$$\begin{aligned}P(X=2)&=P(A_1A_2\overline{A}_3+A_1\overline{A}_2A_3+\overline{A}_1A_2A_3)\\&=P(A_1A_2\overline{A}_3)+P(A_1\overline{A}_2A_3)+P(\overline{A}_1A_2A_3)\\&=P(A_1)P(A_2)P(\overline{A}_3)+P(A_1)P(\overline{A}_2)P(A_3)+P(\overline{A}_1)P(A_2)P(A_3)\\&=0,8\cdot 0,4\cdot 0,1+0,8\cdot 0,6\cdot 0,9+0,2\cdot 0,4\cdot 0,9=0,536\end{aligned}$$

$$P(X=3)=P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3)=0,8\cdot 0,4\cdot 0,9=0,288$$

Do đó, bảng phân phối xác suất của cho X là

X	0	1	2	3
P	0,012	0,164	0,536	0,288

2.2.1.2. Hàm xác suất

Hàm số $f:\mathbb{R}\to\mathbb{R}$ xác định bởi

$$f(x)=\begin{cases} p_i & \text{khi} \quad x=x_i, \\ 0 & \text{khi} \quad x\neq x_i, \forall i. \end{cases}$$

được gọi là *hàm xác suất* của X. Từ tính chất của bảng phân phối xác suất, ta có

(i) $\forall x\in\mathbb{R}, f(x)\geq 0,$ và

(ii) $\sum\limits_x f(x)=1.$

Ví dụ 2.6. Từ bảng phân phối xác suất của ví dụ 2.5. Ta có hàm xác suất của X như sau:

$$f(x)=\begin{cases} 0,012 & \text{khi} \quad x=0, \\ 0,164 & \text{khi} \quad x=1, \\ 0,536 & \text{khi} \quad x=2, \\ 0,288 & \text{khi} \quad x=3, \\ 0 & \text{khi} \quad x\neq 0,1,2,3. \end{cases}$$

2.2.1.3. Hàm phân phối (tích lũy)

Cho $f:\mathbb{R}\to\mathbb{R}$ là hàm xác suất của biến ngẫu nhiên rời rạc X, hàm số $F:\mathbb{R}\to\mathbb{R}$, được xác định bởi

$$F(x)=P(X\leq x)=\sum_{x\geq x_i} f(x_i),$$

được gọi là *hàm phân phối tích lũy*, hay vắn tắt là *hàm phân phối*, của X.

Bằng cách liệt kê các giá trị của $X(\Omega)$ theo thứ tự tăng dần, khi X lấy giá trị tạo thành một dãy $x_1<x_2<\cdots<x_n<\cdots$ ta có hàm phân phối

$$F(x)=\begin{cases} 0 & \text{khi} \quad x<x_1, \\ f(x_1)+f(x_2)+\cdots+f(x_{n-1}) & \text{khi} \quad x_{n-1}\leq x<x_n, \\ 1 & \text{khi} \quad x\geq x_n. \end{cases}$$

Từ tính chất của hàm xác suất và định nghĩa của hàm phân phối, ta có

(i) $0\leq F(x)\leq 1, \forall x\in\mathbb{R},$

(ii) $\lim\limits_{x\to-\infty} F(x)=0$ và $\lim\limits_{x\to+\infty} F(x)=1,$

(iii) F là hàm tăng, và F liên tục bên phải tại mọi $x\in\mathbb{R}$.

Ví dụ 2.7. Với biến số ngẫu nhiên X cho bởi ví dụ 2.6, ta có hàm phân phối

$$F(x)=\begin{cases}0,012 & \text{khi} \quad x<0,\\ 0,164 & \text{khi} \quad 0\le x<1,\\ 0,536 & \text{khi} \quad 1\le x<2,\\ 0,288 & \text{khi} \quad 2\le x<3,\\ 0 & \text{khi} \quad 3\le x.\end{cases}$$

2.2.2. Đại lượng ngẫu nhiên liên tục

2.2.2.1. Hàm mật độ (xác suất)

Hàm số $f:\mathbb{R}\to\mathbb{R}$ được gọi là *hàm mật độ xác suất*, hay vắn tắt là *hàm mật độ*, của biến số ngẫu nhiên liên tục X nếu

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

với mọi $a,b\in\mathbb{R},\ a\le b.$

Từ định nghĩa, dễ dàng suy ra

(i) $\forall x\in\mathbb{R},\ f(x)\ge 0,$ và

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

Ví dụ 2.8. Cho biến số ngẫu nhiên liên tục X với hàm mật độ xác suất

$$f(x)=\begin{cases}Ax(3-x) & \text{khi} \quad x\in[0,3],\\ 0 & \text{khi} \quad x\notin[0,3].\end{cases}$$

Xác định hằng số A.

Giải

+) Do $f(x)\ge 0,\ \forall x\in\mathbb{R}$ nên $A>0$

+) Ta có

$$\int_{-\infty}^{+\infty} f(x)dx=1 \Leftrightarrow \int_{-\infty}^{0} f(x)dx+\int_0^3 f(x)dx+\int_3^{+\infty} f(x)dx=1$$

$$\Leftrightarrow A\int_0^3\left(3x-x^2\right)dx=1 \Leftrightarrow \frac{9}{2}A=1 \Leftrightarrow A=\frac{2}{9}.$$

2.2.2.2. Hàm phân phối (tích lũy)

Cho $f:\mathbb{R}\to\mathbb{R}$ là hàm mật độ của biến ngẫu nhiên liên tục X, hàm số $F:\mathbb{R}\to\mathbb{R}$, được gọi là *hàm phân phối tích lũy*, hay vắn tắt là *hàm phân phối*, của biến số ngẫu nhiên liên tục X nếu

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

với mọi $x \in \mathbb{R}$.

Trực tiếp từ định nghĩa, ta được

(i) $0 \le F(x) \le 1, \forall x \in \mathbb{R}$,

(ii) $\lim\limits_{x \to -\infty} F(x) = 0$ và $\lim\limits_{x \to +\infty} F(x) = 1$,

(iii) F là hàm tăng, và F liên tục bên phải tại mọi $x \in \mathbb{R}$.

Ví dụ 2.9. Cho biến số ngẫu nhiên liên tục X với hàm mật độ xác suất như ví dụ 2.8. Tìm hàm phân phối của X.

Giải

+) Trường hợp 1. Nếu $x < 0$ thì

$$F(x) = \int_{-\infty}^{x} f(t)dt = 0$$

+) Trường hợp 2. Nếu $0 \le x < 3$ thì

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_{0}^{x} f(t)dt = \int_{0}^{x} (3t - t^2)dt = \frac{3}{2}x^2 - \frac{1}{3}x^3.$$

+) Trường hợp 3. Nếu $3 \le x$ thì

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_{0}^{3} f(t)dt + \int_{3}^{+\infty} f(t)dt = \int_{0}^{3} (3t - t^2)dt = 1.$$

Vậy hàm phân phối của X là

$$F(x) = \begin{cases} 0 & \text{khi } x < 0, \\ \frac{3}{2}x^2 - \frac{1}{3}x^3 & \text{khi } 0 \le x < 3, \\ 1 & \text{khi } x \ge 3. \end{cases}$$

Ví dụ 2.10. Cho X là biến số ngẫu nhiên liên tục có hàm phân phối xác suất

$$F(x) = \begin{cases} 0 & \text{khi } x \le -1, \\ a\left(b + x - \frac{1}{3}x^3\right) & \text{khi } -1 < x \le 1, \\ 1 & \text{khi } x > 1. \end{cases}$$

Tìm các hằng số a, b.

Do X là biến số ngẫu nhiên liên tục nên hàm phân phối xác suất liên tục bên phải tại mọi $x \in \mathbb{R}$. Đặc biệt, tại $x = \pm 1$, $\lim\limits_{x \to -1^+} F(x) = F(-1)$ cho

$$a\left(b - \frac{2}{3}\right) = 0 \qquad (*)$$

và $\lim\limits_{x \to 1^+} F(x) = F(1)$ cho

$$a\left(b + \frac{2}{3}\right) = 1 \qquad (**)$$

Từ (*) và (**), ta suy ra $a = \frac{3}{4}$ và $b = \frac{2}{3}$.

2.3. Các số đặc trưng của đại lượng ngẫu nhiên

2.3.1. Kỳ vọng

Cho X là đại lượng ngẫu nhiên với hàm xác suất (hàm mật độ xác suất) $f(x)$ và $u(X)$ là một hàm theo biến số ngẫu nhiên X. *Kỳ vọng* của $u(X)$ được xác định là

$$E[u(X)] = \sum_i u(x_i) f(x_i) \text{ khi } X \text{ là biến ngẫu nhiên rời rạc.}$$

$$E[u(X)] = \int_{-\infty}^{+\infty} u(x) f(x) dx \text{ khi } X \text{ là biến ngẫu nhiên liên tục.}$$

2.3.2. Trung bình

Cho X là biến ngẫu nhiên với hàm xác suất (hàm mật độ xác suất) $f(x)$, khi $u(X) = X$, thì $E(X)$ được gọi là *trung bình* của X, ký hiệu μ_X, nghĩa là

$$E(X) = \sum_i x_i f(x_i) \text{ khi } X \text{ là biến số ngẫu nhiên rời rạc, và}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \text{ khi } X \text{ là biến ngẫu nhiên liên tục.}$$

Tính chất:

(i) $E(C) = C$ với C là hằng số.

(ii) $E(aX + bY) = aE(X) + bE(Y)$ (với $a, b \in \mathbb{R}$ và X, Y là hai đại lượng ngẫu nhiên).

2.3.3. Phương sai

Cho X là biến ngẫu nhiên với hàm xác suất (hàm mật độ xác suất) $f(x)$, khi đó với $u(X) = (X - \mu_X)^2$, thì $E(X - \mu_X)^2$, được gọi là *phương sai* của X, ký hiệu σ_X^2 hay $\operatorname{var}(X)$, nghĩa là

$$\sigma_X^2 = \sum_i (x_i - \mu_X)^2 f(x_i) \text{ khi } X \text{ là biến số ngẫu nhiên rời rạc, và}$$

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x) dx \text{ khi } X \text{ là biến ngẫu nhiên liên tục.}$$

Tính chất:

(i) $\operatorname{var}(C) = 0$ với C là hằng số.

(ii) $\operatorname{var}(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y)$ (với $a, b \in \mathbb{R}$ và X, Y là hai đại lượng ngẫu nhiên độc lập).

2.3.4. Mệnh đề. Cho X là biến số ngẫu nhiên với trung bình $E(X)$. Ta có

$$\operatorname{var}(X) = E(X^2) - [E(X)]^2$$

Chứng minh. Do E tuyến tính, nghĩa là

$$E(aX + bY) = aE(X) + bE(Y)$$

ta suy ra

$$\begin{aligned}\operatorname{var}(X) &= E\left[(X - \mu_X)^2\right] = E\left[X^2 - 2XE(X) + [E(X)]^2\right] \\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2 = E(X^2) - [E(X)]^2.\end{aligned}$$

2.3.5. Độ lệch chuẩn

Độ lệch chuẩn của biến ngẫu nhiên X, ký hiệu σ_X là căn bậc hai của phương sai.

$$\sigma_X = \sqrt{\sigma_X^2}$$

Chú ý: X, $E(X)$, σ_X có cùng đơn vị đo.

Ví dụ 2.11. Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất:

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

Tính trung bình, phương sai và độ lệch chuẩn của X.

Giải

Trung bình của X:

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

Phương sai của X:

$$\operatorname{var}(X) = (1-3,2)^2 \cdot 0,1 + (3-3,2)^2 \cdot 0,5 + (4-3,2)^2 \cdot 0,4 = 0,76.$$

Độ lệch chuẩn của X:

$$\sigma_X = \sqrt{0,76} \approx 0,872.$$

Ví dụ 2.12. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x) = \begin{cases} 2x & \text{khi} \quad x \in [0;1], \\ 0 & \text{khi} \quad x \notin [0;1]. \end{cases}$$

Tính trung bình, phương sai và độ lệch chuẩn của X.

Giải

Trung bình của X:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}x^3 \Big|_0^1 = \frac{2}{3}.$$

Phương sai của X:

$$\operatorname{var}(X) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x) dx = \int_0^1 \left(x - \frac{2}{3}\right)^2 2x dx$$

$$= \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x\right) dx = \frac{1}{2}x^4 - \frac{8}{9}x^3 + \frac{4}{9}x^2 \Big|_0^1 = \frac{1}{18}.$$

Độ lệch chuẩn của X:

$$\sigma_X = \sqrt{\frac{1}{18}} \approx 0,2357.$$

2.3.6. Ý nghĩa của kỳ vọng và phương sai

Ý nghĩa kỳ vọng :

- Kỳ vọng toán phản ánh giá trị trung tâm của phân phối xác suất của biến ngẫu nhiên.

- Trong kinh tế, kỳ vọng toán đồng thời mang 2 ý nghĩa:

+ Nếu xét trong 1 số lớn phép thử tương tự thì nó phản ánh giá trị trung bình

+ Nếu xét trong 1 phép thử đơn lẻ thì nó phản ánh giá trị mong đợi.

Ý nghĩa phương sai :

- Phương sai phản ánh mức độ phân tán của các giá trị của biến ngẫu nhiên so với giá trị trung bình. Phương sai càng lớn: phân tán càng nhiều quanh giá trị trung bình còn phương sai càng nhỏ: giá trị càng tập trung quanh giá trị trung bình.

- Trong kinh tế, phương sai phản ánh mức độ rủi ro hay độ biến động (kém ổn định).

Ví dụ 2.13. Nhu cầu hàng ngày về rau sạch ở một khu dân cư có bảng phân phối xác suất.

X	20	21	22	23	24	25	26
P	0,05	0,1	0,2	0,3	0,15	0,12	0,08

Mỗi kg rau mua vào giá 2 ngàn đồng, bán ra 2 ngàn rưỡi. Song nếu bị ế phải bán 1 ngàn rưỡi mới hết. Hàng ngày nên đặt mua 22 kg hay 24 kg rau để bán thì tốt hơn.

Giải

Trường hợp 1. Nếu mua 22 kg thì gọi X_1 là số tiền lãi. Ta có

$$P(X_1 = 11) = P(X \geq 22) = 0,85$$

$$P(X_1 = 10) = P(X = 21) = 0,1$$

$$P(X_1 = 9) = P(X = 20) = 0,05$$

Ta có bảng phân phối xác suất

X_1	9	10	11
P	0,05	0,1	0,85

$$E(X_1) = 9 \cdot 0,05 + 10 \cdot 0,1 + 11 \cdot 0,85 = 10,8 \text{ (ngàn đồng)}$$

$$\text{var}(X_1) = (9 - 10,8)^2 \cdot 0,05 + (10 - 10,8)^2 \cdot 0,1 + (11 - 10,8)^2 \cdot 0,85 = 0,26$$

Trường hợp 2. Nếu mua 24 kg thì gọi X_2 là số tiền lãi. Ta có

$$P(X_2 = 12) = P(X \geq 24) = 0,35$$

$$P(X_2 = 11) = P(X = 23) = 0,3$$

$$P(X_2 = 10) = P(X = 22) = 0,2$$

$$P(X_2 = 9) = P(X = 21) = 0,1$$

$$P(X_2 = 8) = P(X = 20) = 0,05$$

Ta có bảng phân phối xác suất

X_2	8	9	10	11	12
P	0,05	0,1	0,2	0,3	0,35

$$E(X_2) = 8 \cdot 0,05 + 9 \cdot 0,1 + 10 \cdot 0,2 + 11 \cdot 0,3 + 12 \cdot 0,35 = 10,8 \text{ (ngàn đồng)}$$

$$\text{var}(X_2) = (8 - 10,8)^2 0,05 + (9 - 10,8)^2 \cdot 0,1 + (10 - 10,8)^2 \cdot 0,2 +$$
$$+ (11 - 10,8)^2 \cdot 0,3 + (12 - 10,8)^2 \cdot 0,35 = 1,36$$

Vậy đặt mua 22 kg hay 24 kg đều có tiền lãi trung bình 10,8 ngàn.

Vì $\text{var}(X_1)=0,26<\text{var}(X_2)=1,36$ nên đặt mua 22 kg thì độ rủi ro thấp hơn đặt mua 24 kg.

Ví dụ 2.14. Khi đầu tư vào 2 thị trường A và B, lãi suất thu được là biến ngẫu nhiên có bảng phân phối xác suất tương ứng:

X_A	−1	5	8
P	0,2	0,5	0,3

X_B	−2	6	9
P	0,2	0,4	0,4

a) Muốn có lãi trung bình cao nên đầu tư vào đâu?

b) Muốn kinh doanh ổn định thì đầu tư vào đâu?

c) Người ta muốn giảm thiểu độ rủi ro bằng cách đầu tư vào cả hai, nên chia tỉ lệ đầu tư như thế nào biết rằng 2 thị trường A và B là độc lập.

Giải

a) Trung bình lãi suất của hai thị trường

$$E(X_A)=-1\cdot 0,2+5\cdot 0,5+8\cdot 0,3=4,7$$

$$E(X_B)=-2\cdot 0,2+6\cdot 0,5+9\cdot 0,3=5,6$$

Vậy muốn có lãi trung bình cao nên đầu tư vào thị trường B.

b) Phương sai lãi suất của hai thị trường

$$\text{var}(X_A)=(-1)^2\cdot 0,2+5^2\cdot 0,5+8^2\cdot 0,3-(4,7)^2=9,81$$

$$\text{var}(X_B)=(-2)^2\cdot 0,2+6^2\cdot 0,4+9^2\cdot 0,4-(5,6)^2=16,24.$$

Vậy muốn kinh doanh ổn định nên đầu tư vào thị trường A.

c) Gọi a là tỷ lệ tiền lãi đầu tư vào thị trường A và $(1-a)$ là tỷ lệ tiền lãi đầu tư vào thị trường B. Khi đó, tiền lãi: $Z=aX_A+(1-a)X_B$.

Ta có

$$\text{var}(Z)=a^2\,\text{var}(X_A)+(1-a)^2\,\text{var}(X_B)=9,81a^2+16,24(1-a)^2$$

Bài toán tìm a sao cho $\text{var}(Z)\to\min$

Đặt $f(a)=9,81a^2+16,24(1-a)^2$

Đạo hàm cấp 1: $f^{/}(a)=52,1a-32,48$

Cho $f^{/}(a)=0\Leftrightarrow a=0,6234$

Đạo hàm cấp 2: $f^{//}(a)=52,1>0$

Với $a = 0,6234$ thì $f(a)$ đạt giá trị nhỏ nhất

Vậy đầu tư 62,34% vốn vào thị trường A và 37,66% vốn vào thị trường B sẽ giảm thiểu được rủi ro.

2.3.7. Mốt và trung vị

Mốt của đại lượng ngẫu nhiên rời rạc X, ký hiệu $M_0(X)$, là giá trị x_0 của X sao cho $P(X = x_0)$ là lớn nhất. Người ta còn nói rằng $M_0(X)$, là *giá trị tin chắc nhất* của X. Trong trường hợp X là biến số ngẫu nhiên liên tục với hàm mật độ $f_X(x)$ thì $M_0(X)$, là giá trị x_0 của X sao cho $f_X(x_0)$ là lớn nhất.

Trung vị của đại lượng ngẫu nhiên (rời rạc hay liên tục), ký hiệu $Me(X)$là giá trị x_0 của X sao cho $P(X \le x_0) = P(X \ge x_0) = 0,5.$

Chú ý rằng Mốt cũng như trung vị của một đại lượng ngẫu nhiên thì không duy nhất.

Ví dụ 2.15. Xét biến số ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

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

Ta có $M_0(X) = 1$ hay $M_0(X) = 2$ và $1 < Me(X) < 2$.

Ví dụ 2.16. Cho biến số ngẫu nhiên liên tục X với hàm mật độ xác suất

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

Tính mốt và trung vị của X.

Giải

+) Tìm mốt của X

$$f(x) = \frac{2}{9}(3x - x^2)$$

Tập xác định: $D = [0,3]$

Ta có: $f'(x) = \frac{2}{9}(3 - 2x)$

Cho $f'(x) = 0 \Leftrightarrow x = \frac{3}{2} \in D$

Ta lại có: $f(0) = f(3) = 0; f\left(\frac{3}{2}\right) = \frac{1}{2}$

Hàm mật độ đạt giá trị lớn nhất tại $x = \frac{3}{2}$. Vậy $M_0(X) = \frac{3}{2}$.

+) Tìm trung vị của X

$$f(x) = \frac{2}{9}\left(3x - x^2\right)$$

Tập xác định: $D = [0,3]$

Gọi $Me(X) = x_0 \in [0,3]$

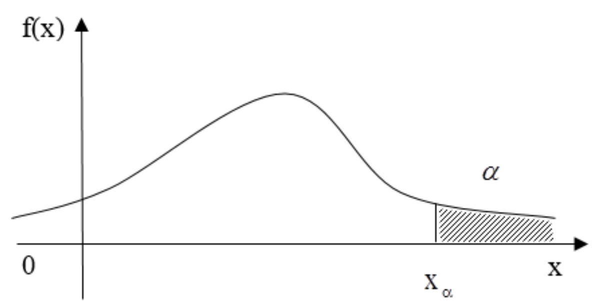
$$P(X \le x_0) = 0,5 \Leftrightarrow \int_{-\infty}^{x_0} f(x)dx = 0,5 \Leftrightarrow \frac{2}{9}\int_{-\infty}^{x_0}\left(3x - x^2\right)dx = 0,5$$

$$\Leftrightarrow -\frac{2}{27}x_0^3 + \frac{1}{3}x_0^2 = \frac{1}{2} \Leftrightarrow \left[\begin{array}{l} x_0 = \frac{3+3\sqrt{3}}{2} \notin [0,3] \\ x_0 = \frac{3}{2} \in [0,3] \\ x_0 = \frac{3-3\sqrt{3}}{2} \notin [0,3] \end{array}\right.$$

Vậy $Me(X) = \frac{3}{2}$.

2.3.8. Giá trị tới hạn

Giá trị tới hạn mức α của một biến ngẫu nhiên liên tục X, ký hiệu x_α là giá trị của X thỏa mãn: $P(X > x_\alpha) = \alpha$.



2.3.9. Hệ số đối xứng và hệ số nhọn

Người ta còn có một số tham số liên quan đến hình dáng của hàm mật độ như sau: Với X là biến số ngẫu nhiên với trung bình μ_X và phương sai σ_X^2, giá trị

$$\gamma_1(X) = \frac{E\left[(x-\mu_X)^3\right]}{\sigma_X^3}$$

được gọi là *hệ số đối xứng* của X và giá trị

$$\gamma_2(X) = \frac{E\left[(x-\mu_X)^4\right]}{\sigma_X^4}$$

được gọi là *hệ số nhọn* của X.

Nếu $\gamma_1(X)=0$ thì phân phối của X là đối xứng; lệch phải khi $\gamma_1(X)>0$ và lệch trái khi $\gamma_1(X)<0$. Ngoài ra giá trị $\gamma_2(X)$ càng lớn thì phân phối của X càng nhọn.

Ví dụ 2.17. Đo đường kính (X) một chi tiết máy (đơn vị mm). Ta có các số liệu : 201; 203; 209; 204; 202; 206; 200; 207; 207. Tính hệ số đối xứng và hệ số nhọn.

$$\mu_X = \frac{1}{9}(201+203+209+\cdots+207+207) = 204,3333,$$

$$\sigma_X^2 = \frac{1}{9}\left((201-204)^2+(203-204)^2+\cdots+(207-204)^2\right) = 8,4444,$$

Hệ số đối xứng

$$\gamma_1(X) = \frac{\frac{1}{9}\left((201-204)^3+(203-204)^3+\cdots+(207-204)^3\right)}{(9,5)^{3/2}} = 0,3939.$$

Hệ số nhọn

$$\gamma_2(X) = \frac{\frac{1}{9}\left((201-204)^4+(203-204)^4+\cdots+(207-204)^4\right)}{(9,5)^2} = 1,8029.$$

2.4. Một số quy luật phân phối xác suất quan trọng

2.4.1. Phân phối nhị thức $B(n;p)$

2.4.1.1. Định nghĩa. Biến ngẫu nhiên rời rạc X được gọi là có phân phối nhị thức, ký hiệu $X \sim B(n;p)$ nếu hàm xác suất của X có dạng sau

$$f(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{khi } x = 0,1,2,...,n \\ 0 & \text{khi } x \neq 0,1,2,...,n \end{cases}$$

Công thức xác suất

$$P(X=k) = C_n^k p^k (1-p)^{n-k}, \text{ với } k = 0,1,2,...,n.$$

Ví dụ 2.18. Trong một vùng dân cư có 70% gia đình có máy giặt, chọn ngẫu nhiên 12 gia đình. Tính xác suất

a) có đúng 5 gia đình có máy giặt.

b) có ít nhất 2 gia đình có máy giặt.

Giải

Gọi X là số gia đình có máy giặt trong số 12 gia đình này, $X \sim B(12;0,7)$. Ta có

$$P(X=k)=C_{12}^{k}(0,7)^{k}(1-0,7)^{12-k}=C_{12}^{k}(0,7)^{k}(0,3)^{12-k}$$

a) Xác suất để nhận được đúng 5 gia đình có máy giặt là

$$P(X=5)=C_{12}^{5}(0,7)^{5}(0,3)^{12-5}=0,0291.$$

b) Xác suất để có ít nhất 2 gia đình có máy giặt là

$$P(X\geq 2)=1-P(X<2)=1-P(X=0)-P(X=1)$$
$$=1-C_{12}^{0}(0,7)^{0}(0,3)^{12}-C_{12}^{1}(0,7)^{1}(0,3)^{11}\approx 1.$$

2.4.1.2. Mệnh đề. Cho $X \sim B(n;p)$, ta có

i) Trung bình: $\mu_X = np.$

ii) Phương sai: $\sigma_X^2 = npq,$ với $q=1-p.$

iii) Giá trị tin chắc nhất : $M_0(X)=k_0,$ với k_0 là số nguyên thỏa bất phương trình $np-q\leq k_0\leq np-q+1.$

Ví dụ 2.19. Một nhân viên tiếp thị bán hàng ở 5 chỗ khác nhau trong ngày. Xác suất bán được hàng ở mỗi nơi đều là 0,4.

a) Tìm xác suất để nhân viên này bán được hàng trong ngày.

b) Mỗi năm nhân viên này đi bán hàng 330 ngày. Gọi Y là số ngày bán được hàng trong năm. Tìm giá trị tin chắc nhất của Y.

Giải

Gọi X là tổng số nơi bán được hàng trong ngày, $X \sim B(5;0,4)$. Ta có

$$P(X=k)=C_{5}^{k}(0,4)^{k}(1-0,4)^{5-k}=C_{5}^{k}(0,4)^{k}(0,6)^{5-k}$$

a) Xác suất để bán được hàng trong ngày là

$$P(X\geq 1)=1-P(X<1)=1-P(X=0)$$
$$=1-C_{5}^{0}(0,4)^{0}(0,6)^{5}=0,92224.$$

b) Tìm số ngày bán được hàng nhiều khả năng nhất trong một năm.

Ta có $Y \sim B(330;0,92224)$ và do đó, số ngày bán được hàng nhiều khả năng nhất là $k_0 = M_0(Y)$, với $k_0\in\mathbb{N}$ thỏa

$$330\cdot 0,92224-(1-0,92224)\le k_0\le 330\cdot 0,92224-(1-0,92224)+1$$

$$\Leftrightarrow 304,26\le k_0\le 305,26\Leftrightarrow k_0=305.$$

Vậy số ngày bán được hàng có nhiều khả năng trong một năm là 305.

2.4.2. Phân phối siêu bội $H(N,K,n)$

2.4.2.1. Định nghĩa. Biến ngẫu nhiên rời rạc X được gọi là có phân phối siêu bội, ký hiệu $X\sim H(N,K,n)$ nếu hàm xác suất của X có dạng sau

$$f(x)=\begin{cases}\dfrac{C_K^x C_{N-K}^{n-x}}{C_N^n} & \text{khi } x\in\left[\max\{0,n-N+K\},\min\{n,K\}\right]\\ 0 & \text{khi } x\notin\left[\max\{0,n-N+K\},\min\{n,K\}\right]\end{cases}$$

Công thức xác suất

$$P\left(X=k\right)=\frac{C_K^k C_{N-K}^{n-k}}{C_N^n},\text{ với } \max\{0,n-N+K\}\le k\le \min\{n,K\}.$$

2.4.2.2. Mệnh đề. Cho $H(N,K,n)$, ta có

i) Trung bình: $\mu_X=np$,

ii) Phương sai: $\sigma_X^2=np(1-p)\left(\dfrac{N-n}{N-1}\right)$.

2.4.2.3. Lưu ý : Nếu $X\sim H(N,K,n)$, trong đó $n<<N$ thì X được xem như có phân phối nhị thức $X\sim B(n;p)$, với $p=\dfrac{K}{N}$.

Ví dụ 2.20. Từ một hộp đựng 15 quả cam trong đó có 5 quả hư, lấy ra 3 quả. Gọi X là số quả hư trong 3 quả lấy ra.

a) Tính xác suất để cả 3 quả đều hư.

b) Tính trung bình và phương sai của X.

Giải

Ta có $X\sim H(15,5,3)$. Công thức tính xác suất

$$P\left(X=k\right)=\frac{C_5^k C_{15-5}^{3-k}}{C_{15}^3}$$

a) Xác suất để cả 3 quả đều hư

$$P\left(X=3\right)=\frac{C_5^3 C_{10}^3}{C_{15}^3}=0,021978.$$

b) Trung bình (kỳ vọng) và phương sai của X

$$\mu_X = np = 3\frac{5}{15} = 1;\ \sigma_X^2 = 3\frac{5}{15}\left(1-\frac{5}{15}\right)\left(\frac{15-3}{15-1}\right) = \frac{4}{7}.$$

Ví dụ 2.21. Có 8000 sản phẩm trong đó có 2000 sản phẩm không đạt tiêu chuẩn kỹ thuật. Lấy ngẫu nhiên (không hoàn lại) 10 sản phẩm. Tính xác suất để trong 10 sản phẩm lấy ra có 2 sản phẩm không đạt tiêu chuẩn.

Giải

Cách 1. Tính trực tiếp theo phân phối của nó

Ta có $N = 8000,\ K = 2000,\ n = 10$

Gọi X số sản phẩm trong đạt tiêu chuẩn trong 10 sản phẩm lấy ra, $X \sim H(8000, 2000, 10)$

Công thức tính xác suất:

$$P(X = k) = \frac{C_{2000}^{k} C_{6000}^{10-k}}{C_{8000}^{10}}$$

Xác suất lấy hai sản phẩm không đạt tiêu chuẩn

$$P(X = 2) = \frac{C_{2000}^{2} C_{6000}^{8}}{C_{8000}^{10}} = 0,281697.$$

Cách 2. Tính xấp xỉ (tính gần đúng)

Do $n = 10 << N = 8000$; $p = \frac{K}{N} = 0,25$

$$X \sim H(8000, 2000, 10) \equiv B(10; 0,25)$$

Xác suất lấy hai sản phẩm không đạt tiêu chuẩn

$$P(X = 2) = C_{10}^{2}(0,25)^2(0,75)^8 = 0,28157.$$

2.4.3. Phân phối Poisson $P(\mu)$

2.4.3.1. Định nghĩa. Biến ngẫu nhiên rời rạc X được gọi là có phân phối poisson, ký hiệu $X \sim P(\mu)$ nếu hàm xác suất của X có dạng sau

$$f(x) = \begin{cases} e^{-\mu}\dfrac{\mu^x}{x!} & \text{khi } x = 0, 1, 2, ..., n, ... \\ 0 & \text{khi } x \neq 0, 1, 2, ..., n, ... \end{cases}$$

Công thức xác suất

$$P(X = k) = e^{-\mu}\frac{\mu^k}{k!},\ \text{với } k = 0, 1, 2, ..., n.$$

2.4.3.2. Mệnh đề. Cho $X \sim P(\mu)$, ta có

i) Trung bình: $\mu_X = \mu$,

ii) Phương sai: $\sigma_X^2 = \mu$,

iii) Độ lệch chuẩn: $\sigma_X = \sqrt{\mu}$.

2.4.3.3. Chú ý: Nếu $X \sim B(n,p)$, trong đó p đủ nhỏ và n đủ lớn thì X được xem như có phân phối Poisson $X \sim P(\mu)$, với $\mu = np$.

Bằng cách viết

$$P(X=k) = C_n^k p^k (1-p)^{n-k} = \frac{n(n-1)...(n-k+1)}{k!} p^k (1-p)^{n-k}$$
$$= \frac{1}{k!} \frac{n(n-1)...(n-k+1)}{n^k} (np)^k (1-p)^{n-k}$$

và với $np = \mu$ không đổi, khi $n \to \infty$ ta có $p \to 0$ và

$$\lim_{n\to\infty} \frac{n(n-1)...(n-k+1)}{n^k} = 1,$$

$$\lim_{n\to\infty} (1-p)^{n-k} = \lim_{p\to 0} (1-p)^{\frac{\mu}{p}-k} = \lim_{p\to 0} \left[(1-p)^{\frac{1}{-p}} \right]^{-\mu+kp} = e^{-\mu}.$$

Từ đó, suy ra $C_n^k p^k (1-p)^{n-k} \approx \dfrac{e^{-\mu}\mu^k}{k!}$ khi n khá lớn.

Trong ứng dụng, khi $X \sim B(n;p)$, trong đó $n > 50$, $p < 0{,}01$ và $np < 5$ thì ta có thể dùng xấp xỉ $X \sim P(np)$.

Ví dụ 2.22. Một máy dệt có 4000 ống sợi. Xác suất để mỗi ống sợi bị đứt trong 1 phút là 0,0005. Tính xác suất để trong 1 phút

a) có 3 ống sợi bị đứt,

b) có ít nhất 2 ống sợi bị đứt.3

Giải

Cách 1. Tính trực tiếp theo luật phân phối của nó

Ta có $n = 4000;\ p = 0{,}0005$

Gọi X là số ống sợi bị đứt trong 1 phút (trong 4000 ống sợi), $X \sim B(4000; 0{,}0005)$

Công thức tính xác suất:

$$P(X=k) = C_{4000}^k (0{,}0005)^k (0{,}9995)^{4000-k}$$

a) có 3 ống sợi bị đứt,

$$P(X=3)=C_{4000}^{3}(0,0005)^{3}(0,9995)^{3997}=0,1804822$$

b) có ít nhất 2 ống sợi bị đứt.

$$\begin{aligned}P(X\ge 2)&=1-P(X<2)=1-P(X=0)-P(X=1)\\&=1-C_{4000}^{0}(0,0005)^{0}(0,9995)^{4000}-C_{4000}^{1}(0,0005)^{1}(0,9995)^{3999}\\&=0,594062.\end{aligned}$$

Cách 2. Tính xấp xỉ (tính gần đúng)

Ta có trung bình : $\mu=np=4000\cdot 0,0005=2<5$ nên phân phối nhị thức được xấp xỉ bằng phân phối Poisson như sau:

$$X\sim B(4000;0,0005)\equiv P(2)$$

Công thức tính xác suất:

$$P(X=k)=e^{-2}\frac{2^{k}}{k!}$$

a) có 3 ống sợi bị đứt,

$$P(X=3)=e^{-2}\frac{2^{3}}{3!}=0,180447.$$

b) có ít nhất 2 ống sợi bị đứt.

$$\begin{aligned}P(X\ge 2)&=1-P(X<2)=1-P(X=0)-P(X=1)\\&=1-e^{-2}\frac{2^{0}}{0!}-e^{-2}\frac{2^{1}}{1!}=0,59399.\end{aligned}$$

2.4.4. Phân phối đều $U[a,b]$

2.4.4.1. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối đều, ký hiệu $X\sim U[a,b]$ nếu hàm mật độ của X có dạng sau

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

Công thức tính xác suất

$$P(\alpha\le X\le\beta)=\frac{\beta-\alpha}{b-a}$$

2.4.4.2. Mệnh đề. Cho $X\sim U[a,b]$, ta có

i) Trung bình: $\mu_X = \dfrac{a+b}{2}$,

ii) Phương sai: $\sigma_X^2 = \dfrac{(b-a)^2}{12}$.

Ví dụ 2.23. Một loại sản phẩm do một nhà máy sản xuất được đóng thành từng hộp. Trọng lượng của hộp là biến ngẫu nhiên phân phối đều trong khoảng $\left[1,9;2,1\right]$ (kg). Tính trọng lượng trung bình của một hộp và tỷ lệ hộp có trọng lượng từ 1,95 kg trở lên.

Giải

Gọi X là trọng lượng của mỗi hộp sản phẩm, $X \sim U\left[1,9;2,1\right]$

Trọng lượng trung bình của một hộp chính là

$$E(X) = \frac{a+b}{2} = \frac{1,9+2,1}{2} = 2\,(\text{kg})$$

Tỷ lệ hộp có trọng lượng từ 1,95 kg trở lên là:

$$P(X \geq 1,95) = P(1,95 \leq X \leq 2,1) = \frac{2,1-1,95}{2,1-1,9} = 0,75 = 75\%.$$

2.4.5. Phân phối mũ

2.4.5.1. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối mũ, ký hiệu $X \sim \text{Exp}(\lambda)$ nếu hàm mật độ của X có dạng sau

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{khi} \quad x > 0, \\ 0, & \text{khi} \quad x \leq 0. \end{cases}$$

2.4.5.2. Mệnh đề. Cho $X \sim \text{Exp}(\lambda)$, ta có

i) Trung bình: $\mu_X = \dfrac{1}{\lambda}$,

ii) Phương sai: $\sigma_X^2 = \dfrac{1}{\lambda^2}$.

Ví dụ 2.24. Tuổi thọ (tính theo giờ) của một trò chơi điện tử bấm tay là một biến số ngẫu nhiên liên tục có hàm mật độ xác suất như sau:

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

a) Tìm hằng số k.

b) Tính xác suất tuổi thọ của trò chơi này nằm trong khoảng từ 50 đến 150 giờ.

c) Tính xác suất tuổi thọ của trò chơi này ít hơn 100 giờ.

Giải

a) Tìm hằng số k.

Ta có

+) Vì $f(x)$ là hàm mật độ nên $f(x) \geq 0 \Rightarrow k > 0$

+) $\int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow \int_{0}^{+\infty} ke^{-\frac{x}{100}} dx = 1$

$$\Leftrightarrow k\int_{0}^{+\infty} e^{-\frac{x}{100}} dx = 1 \Leftrightarrow \left.\left(-100k\, e^{-\frac{x}{100}}\right)\right|_{0}^{+\infty} = 1 \Leftrightarrow 100k = 1 \Leftrightarrow k = \frac{1}{100}$$

b) Tính xác suất tuổi thọ của trò chơi này nằm trong khoảng từ 50 đến 150 giờ.

Ta có

$$P(50 \leq X \leq 150) = \int_{50}^{150} f(x)dx = \frac{1}{100}\int_{50}^{150} e^{-\frac{x}{100}}dx$$

$$= -e^{-\frac{x}{100}}\Big|_{50}^{150} = e^{-\frac{1}{2}} - e^{-\frac{3}{2}} \approx 0,3834.$$

c) Tính xác suất tuổi thọ của trò chơi này ít hơn 100 giờ.

Ta có

$$P(X < 100) = \int_{-\infty}^{100} f(x)dx = \frac{1}{100}\int_{0}^{100} e^{-\frac{x}{100}}dx$$

$$= -e^{-\frac{x}{100}}\Big|_{0}^{100} = 1 - e^{-1} \approx 0,632.$$

2.4.6. Phân phối chuẩn tắc $N(0,1)$

2.4.6.1. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối chuẩn tắc, ký hiệu $X \sim N(0,1)$ nếu hàm mật độ của X có dạng sau

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty,$$

Với $a, b \in \mathbb{R}$, $a \leq b$, ta có

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}}\int_{a}^{b} e^{-\frac{x^2}{2}} dx = \phi_0(b) - \phi_0(a).$$

Trong đó $\phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$

2.4.6.2. Mệnh đề. Cho $X \sim N(0,1)$, ta có

i) Trung bình: $\mu_X = 0$,

ii) Phương sai: $\sigma_X^2 = 1$,

iii) Độ lệch chuẩn: $\sigma_X = 1$.

2.4.7. Phân phối chuẩn $N(\mu, \sigma^2)$

2.4.7.1. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối chuẩn, ký hiệu $X \sim N(\mu, \sigma^2)$ nếu hàm mật độ của X có dạng sau

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$$

Với $a, b \in \mathbb{R}$, $a \le b$, ta có

$$P(a \le X \le b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

2.4.7.2. Mệnh đề. Cho $X \sim N(\mu, \sigma^2)$, ta có

i) Trung bình: $\mu_X = \mu$,

ii) Phương sai: $\sigma_X^2 = \sigma^2$,

iii) Độ lệch chuẩn: $\sigma_X = \sigma$.

2.4.7.3. Chú ý

i) Nếu $X \sim N(\mu, \sigma^2)$ thì đặt $Y = \frac{X-\mu}{\sigma}$, ta có $Y \sim N(0,1)$. Do đó

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Y \le \frac{b-\mu}{\sigma}\right) = \phi_0\left(\frac{b-\mu}{\sigma}\right) - \phi_0\left(\frac{a-\mu}{\sigma}\right)$$

ii) Phân phối chuẩn dùng để khảo sát các hiện tượng bình thường. Cụ thể, nếu $X \sim B(n;p)$ với tích np lớn thì ta xấp xỉ phân phối nhị thức $B(n;p)$ bằng phân phối chuẩn $N(\mu;\sigma^2)$, với $\mu = np$, $\sigma^2 = npq$. Ta có $X \sim B(n;p) \equiv N(np, npq)$

+) $P\left(X=k\right)\approx\dfrac{1}{\sqrt{npq}\sqrt{2\pi}}e^{-\frac{\left(k-np\right)^2}{2npq}}$,

+) $P\left(a\le X\le b\right)\approx\phi_0\left(\dfrac{b-np}{\sqrt{npq}}\right)-\phi_0\left(\dfrac{a-np}{\sqrt{npq}}\right)$.

Ví dụ 2.25. Trọng lượng X (tính bằng gam) một loại trái cây có phân phối chuẩn $N\left(\mu;\sigma^2\right)$, với $\mu=500(\text{gam})$ và $\sigma^2=16(\text{gam}^2)$. Trái cây thu hoạch được phân loại theo trọng lượng như sau :

a) loại 1 : trên 505 gam,

b) loại 2 : từ 495 đến 505 gam,

c) loại 3 : dưới 495 gam.

Tính tỷ lệ mỗi loại.

Giải

Gọi X là trọng lượng trái cây thì $X\sim N\left(\mu;\sigma^2\right)=N\left(500;4^2\right)$.

a) Tỷ lệ trái cây loại 1 là

$$P\left(X>505\right)=\phi_0\left(+\infty\right)-\phi_0\left(\frac{505-500}{4}\right)=0,5-\phi_0\left(1,25\right)=0,10565.$$

b) Tỷ lệ trái cây loại 2 là

$$P\left(495\le X\le 505\right)=\phi_0\left(\frac{505-500}{4}\right)-\phi_0\left(\frac{495-500}{4}\right)=0,7887.$$

c) Và tỷ lệ của loại 3 là

$$P\left(X<495\right)=P\left(\frac{X-500}{4}<\frac{495-500}{4}\right)=\phi_0\left(\frac{495-500}{4}\right)-\phi_0\left(-\infty\right)$$
$$=\phi_0\left(-1,25\right)+\phi_0\left(+\infty\right)=-\phi_0\left(1,25\right)+0,5=0,10565.$$

Vậy, trái cây thu hoạch được có khoảng 11% loại 1, 78% loại 2 và 11% loại 3.

Ví dụ 2.26. Thời gian chạy 1000m của mỗi sinh viên là một biến ngẫu nhiên phân phối chuẩn có trung bình là 80 giây, độ lệch chuẩn là 10 giây. Một sinh viên phải chạy tối đa là bao nhiêu thời gian để đứng trong 10% số người đứng đầu.

Giải

Gọi X là thời gian chạy hết 1000m của mỗi sinh viên, $X\sim N\left(\mu=80,\sigma^2=10^2\right)$.

t là thời gian chạy tối đa để đứng trong 10% số người đứng đầu, ta có:

$$P(X \le t) = 0,1 \Leftrightarrow P(0 \le X \le t) = 0,1$$

$$\Leftrightarrow \phi_0\left(\frac{t-80}{10}\right) - \phi_0\left(\frac{0-80}{10}\right) = 0,1 \Leftrightarrow \phi_0\left(\frac{t-80}{10}\right) + 0,5 = 0,1$$

$$\Leftrightarrow \phi_0\left(\frac{80-t}{10}\right) = 0,4 = \phi_0(1,28) \Leftrightarrow t = 67,2.$$

Vậy thời gian tối đa của sinh viên hoàn thành 1000m để đứng trong 10% số người đứng đầu.

Ví dụ 2.27. Xác suất để một sản phẩm không được kiểm tra chất lượng sau khi sản xuất là 0,2. Tìm xác suất để trong 400 sản phẩm sản xuất ra có:

a) 80 sản phẩm không được kiểm tra chất lượng.

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

Giải

Gọi X là số sản phẩm không được kiểm tra chất lượng, $X \sim B(400;0,2)$

Ta có: $\mu = np = 400 \cdot 0,2 = 80; \sigma^2 = npq = 400 \cdot 0,2 \cdot 0,8 = 8^2$

nên có thể coi $X \sim N(80;8^2)$

a) $P(X = 80) \approx \dfrac{1}{\sqrt{400.0,2.0,8}} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(80-80)^2}{2 \cdot 64}} = \dfrac{1}{8\sqrt{2\pi}} = 0,049868.$

b) $P(70 \le X \le 100) \approx \phi_0\left(\dfrac{100\ -80}{8}\right) - \phi_0\left(\dfrac{70\ -80}{8}\right)$

$$= \phi_0(2,5) - \phi_0(-1,25) = \phi_0(2,5) + \phi_0(1,25) = 0,8882.$$

2.4.8. Phân phối Gamma và phân phối Chi bình phương

Định nghĩa: Hàm Gamma $\Gamma : (0,\infty) \to \mathbb{R}$

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

2.4.8.1. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối Gamma, ký hiệu $X \sim \Gamma(\alpha,\beta)$, với $\alpha,\beta > 0$, nếu hàm mật độ của X có dạng sau

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{khi } x > 0 \\ 0 & \text{khi } x \le 0 \end{cases},$$

2.4.8.2. Mệnh đề. Cho $X \sim \Gamma(\alpha,\beta)$, ta có

i) Trung bình: $\mu_X = \alpha\beta$,

ii) Phương sai: $\sigma_X^2 = \alpha\beta^2$.

2.4.8.3. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối Chi bình phương, ký hiệu $X \sim \chi^2(r)$, nếu hàm mật độ của X có dạng sau

$$f(x) = \begin{cases} \dfrac{1}{\Gamma\left(\dfrac{r}{2}\right)2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} & \text{khi } x > 0 \\ 0 & \text{khi } x \le 0 \end{cases},$$

nghĩa là $X \sim \Gamma\left(\dfrac{r}{2}, 2\right)$

2.4.8.4. Mệnh đề. Cho $X \sim \chi^2(r)$, ta có

i) Trung bình: $\mu_X = r$,

ii) Phương sai: $\sigma_X^2 = 2r$.

2.4.9. Phân phối Student: $St(n)$

2.4.9.1. Định nghĩa: Biến ngẫu nhiên liên tục T được gọi là có phân phối Student, ký hiệu $T \sim St(n)$, nếu hàm mật độ của T có dạng sau

$$f(t) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\dfrac{n}{2}\right) \cdot \left(1 + \dfrac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

2.4.9.2. Mệnh đề. Cho X là biến ngẫu nhiên liên tục có phân phối Gauss, $X \sim N(0,1)$; Y là biến ngẫu nhiên liên tục có phân phối Chi bình phương với n bậc tự do, $Y \sim \chi^2(n)$ và X, Y là hai biến độc lập.

Đặt $T = \dfrac{X}{\sqrt{Y/n}}$ thì T có phân phối Student với n bậc tự do, $T \sim St(n)$.

2.4.9.3. Mệnh đề. Cho $T \sim St(n)$, ta có

i) Trung bình: $\mu_T = 0$,

ii) Phương sai: $\sigma_T^2 = \dfrac{n}{n-2}$.

2.4.9.4. Chú ý: Nếu $X \sim St(n)$, với $n \ge 30$, thì $X \sim N(0,1)$.

2.4.10. Phân phối Fisher: $F(n,m)$

2.4.10.1. Định nghĩa. Biến ngẫu nhiên liên tục F có phân phối Fisher, ký hiệu $F \sim F(n,m)$ nếu hàm mật độ của F có dạng như sau

$$f(x) = \begin{cases} \dfrac{\sqrt{n^n \cdot m^m} \cdot \Gamma\left(\dfrac{n+m}{2}\right) x^{\frac{n-m}{2}}}{\Gamma\left(\dfrac{n}{2}\right)(m+nx)^{\frac{n+m}{2}}} & \text{khi } x > 0, \\ 0 & \text{khi } x \le 0. \end{cases}$$

với n, m là hai bậc tự do.

2.4.10.2. Mệnh đề. Cho X, Y là hai biến ngẫu nhiên liên tục có phân phối Chi bình phương, $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$ và X, Y là hai biến độc lập.

Đặt $F = \dfrac{X/n}{Y/m}$ thì F có phân phối Fisher với n, m bậc tự do, $F \sim F(n,m)$.

2.4.10.3. Mệnh đề. Cho $F \sim F(n,m)$, ta có

i) Trung bình: $\mu_F = \dfrac{n}{n-2}$, $n > 2$,

ii) Phương sai: $\sigma_F^2 = \dfrac{2m^2(n+m^2-2)}{n(m-2)^2(m-4)}$, $m > 4$.

2.5. Tóm tắt chương 2

A. Đại lượng ngẫu nhiên rời rạc

1. Bảng phân phối xác suất

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

với $p_i = P(X = x_i)$, được gọi là *bảng phân phối xác suất* của X.

Tính chất: $\sum_i p_i = 1$.

2. Hàm xác suất của biến ngẫu nhiên X: Hàm số $f : \mathbb{R} \to \mathbb{R}$ xác định bởi

$$f(x) = \begin{cases} p_i & \text{khi} \quad x = x_i, \\ 0 & \text{khi} \quad x \ne x_i, \forall i. \end{cases}$$

Tính chất : $\forall x \in \mathbb{R}, f(x) \ge 0$, và $\sum_x f(x) = 1$.

3. Hàm phân phối xác suất: Cho $f:\mathbb{R}\to\mathbb{R}$ là hàm xác suất của biến ngẫu nhiên rời rạc X, hàm số $F:\mathbb{R}\to\mathbb{R}$, được xác định bởi

$$F(x)=P(X\le x)=\sum_{x\ge x_i} f(x_i),$$

B. Đại lượng ngẫu nhiên liên tục

1. Hàm mật độ (xác suất): Hàm số $f:\mathbb{R}\to\mathbb{R}$ được gọi là *hàm mật* của biến số ngẫu nhiên liên tục X nếu $P(a\le X\le b)=\int_a^b f(x)dx$

với mọi $a,b\in\mathbb{R}$, $a\le b$. Ta có : $\forall x\in\mathbb{R}$, $f(x)\ge 0$, và $\int_{-\infty}^{+\infty} f(x)dx=1$.

2. Hàm phân phối (tích lũy): Cho $f:\mathbb{R}\to\mathbb{R}$ là hàm mật độ của biến ngẫu nhiên liên tục X, hàm số $F:\mathbb{R}\to\mathbb{R}$, được gọi là *hàm phân phối* của biến số ngẫu nhiên liên tục X nếu $F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)dt$ với mọi $x\in\mathbb{R}$.

C. Trung bình và phương sai: Cho X là biến ngẫu nhiên với hàm xác suất (hàm mật độ xác suất) $f(x)$,

1. Trung bình:

$E(X)=\sum_i x_i f(x_i)$ khi X là biến số ngẫu nhiên rời rạc, và

$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$ khi X là biến ngẫu nhiên liên tục.

2. Phương sai

$\sigma_X^2=\sum_i (x_i-\mu_X)^2 f(x_i)$ khi X là biến số ngẫu nhiên rời rạc, và

$\sigma_X^2=\int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)dx$ khi X là biến ngẫu nhiên liên tục.

3. Độ lệch chuẩn:

$\sigma_X=Se(X)$ gọi là độ lệch chuẩn của X.

4. Mệnh đề. Cho X là biến số ngẫu nhiên với trung bình $E(X)$. Ta có

$$var(X)=E(X^2)-[E(X)]^2$$

D. Các quy luật thường gặp

1. Phân phối nhị thức: $X \sim B(n;p)$

i) Công thức xác suất

$$P(X=k)=C_n^k p^k (1-p)^{n-k}\text{, với } k=0,1,2,\ldots,n.$$

ii) Trung bình: $\mu_X = np$,

iii) Phương sai: $\sigma_X^2 = np(1-p)$,

iv) Giá trị tin chắc nhất : $M_0(X)=k_0$, với k_0 là số nguyên thỏa bất phương trình $np-q \le k_0 \le np-q+1$.

2. Phân phối siêu bội: $X \sim H(N,K,n)$

i) $P(X=k)=\dfrac{C_K^k C_{N-K}^{n-k}}{C_N^n}$, với $\max\{0,n-N+K\} \le k \le \min\{n,K\}$.

ii) Trung bình: $\mu_X = np$,

iii) Phương sai: $\sigma_X^2 = np(1-p)\left(\dfrac{N-n}{N-1}\right)$.

3. Phân phối Poisson : $X \sim P(\mu)$

i) Công thức xác suất

$$P(X=k)=e^{-\mu}\frac{\mu^k}{k!}\text{, với } k=0,1,2,\ldots,n.$$

ii) Trung bình: $\mu_X = \mu$,

iii) Phương sai: $\sigma_X^2 = \mu$.

4. Phân phối chuẩn tắc: $X \sim N(0,1)$

i) $P(a \le X \le b)=\phi_0(b)-\phi_0(a)$, với $\phi_0(x)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^x e^{-\frac{t^2}{2}}dt$.

ii) Trung bình: $\mu_X = 0$,

iii) Phương sai: $\sigma_X^2 = 1$,

5. Phân phối chuẩn: $X \sim N(\mu,\sigma^2)$

i) $P(a \le X \le b)=\phi_0\left(\dfrac{b-\mu}{\sigma}\right)-\phi_0\left(\dfrac{a-\mu}{\sigma}\right)$,

i) Trung bình: $\mu_X = \mu$,

ii) Phương sai: $\sigma_X^2 = \sigma^2$.

2.6. Bài tập

Bài số 1. Cho X là một biến số ngẫu nhiên có phân phối xác suất như sau

X	1	2	3	4	5	6	7
P_X	a	2a	2a	3a	a^2	$2a^2$	$7a^2+a$

a) Xác định a.

b) Tính $P[X \geq 5]$, $P[X < 3]$.

c) Tính k nhỏ nhất sao cho $P[X \leq k] \geq 0,5$.

Đáp số: a) $a = 0,1$*; b)* $P[X \geq 5] = 0,2$; $P[X < 3] = 0,3$*; c)* $k = 3$.

Bài số 2. Xét trò chơi, tung một con xúc xắc ba lần: nếu cả ba lần được 6 nút thì thưởng 6 ngàn đồng, nếu hai lần 6 nút thì thưởng 4 ngàn đồng, một lần 6 nút thì thưởng 2 ngàn đồng, và nếu không có 6 nút thì không thưởng gì hết. Mỗi lần chơi phải đóng A ngàn đồng. Hỏi

a) A bao nhiêu thì người chơi về lâu về dài huề vốn (gọi là trò chơi công bằng),

b) A bao nhiêu thì trung bình mỗi lần người chơi mất 1 ngàn đồng.

Đáp số: a) $A = 1$; *b)* $A = 2$.

Bài số 3. Một nhà đầu tư có 3 dự án. Gọi X_i $(i = 1,2,3)$ là số tiền thu được khi thực hiện dự án thứ i (giá trị âm chỉ số tiền bị thua lỗ). X_i là biến số ngẫu nhiên. Qua nghiên cứu, giả sử có số liệu như sau : (Đơn vị tính : 100 triệu đồng)

X_1	-20	30	60
P	0,3	0,2	0,5

X_2	-20	-10	100
P	0,4	0,2	0,4

X_3	-25	-30	80
P	0,2	0,3	0,5

Theo anh (chị), ta nên chọn dự án nào ?

Đáp số: Nên chọn dự án 1.

Bài số 4. Có 3 xạ thủ cùng bắn vào một mục tiêu, mỗi người bắn 1 viên, trong cùng một số điều kiện nhất định. Xác suất để mỗi xạ thủ bắn trúng mục tiêu lần lượt là 0,6; 0,7; 0,9. Gọi X là số viên đạn trúng mục tiêu. Tính $E(X)$; $Var(X)$ và $Mod(X)$.

Đáp số: $EX = 2,2$; $Var(X) = 0,54$; $Mod(X) = 2$.

Bài số 5. Một phân xưởng có ba máy M_1, M_2, M_3. Trong một giờ, mỗi máy sản xuất được 10 sản phẩm. Số sản phẩm không đạt tiêu chuẩn trong 10 sản phẩm của M_1, M_2, M_3 lần lượt là 1, 2, 1. Lấy ngẫu nhiên 1 sản phẩm từ 10 sản phẩm do mỗi máy sản xuất. Gọi X là số sản phẩm không đạt tiêu chuẩn trong 3 sản phẩm được lấy ra.

a) Lập bảng phân phối xác suất của X.

b) Tìm E(X), Var(X), Mod(X).

c) Tính $P(X \le 1)$.

Đáp số: a)

X	0	1	2	3
P	0,648	0,306	0,044	0,002

b) $EX = 0,4$; $Var(X) = 0,34$; $Mod(X) = 0$; c) 0,954.

Bài số 6. Tỷ lệ khách hàng phản ứng tích cực đối với một chiến dịch quảng cáo là biến số ngẫu nhiên có bảng phân phối xác suất như sau :

X (%)	0	10	20	30	40	50
P	0,1	0,2	0,35	0,2	0,1	0,05

a) Tính tỷ lệ khách hàng bình quân phản ứng tích cực đối với chiến dịch quảng cáo đó.

b) Tìm xác suất để có trên 20% khách hàng phản ứng tích cực đối với chiến dịch quảng cáo.

Đáp số: a) 21,5%; b) 0,35.

Bài số 7. Qua theo dõi trong nhiều năm kết hợp với sự đánh giá của các chuyên gia tài chính thì lãi suất đầu tư vào một công ty là biến số ngẫu nhiên có bảng phân phối xác suất như sau:

X (%)	9	10	11	12	13	14	15
P	0,05	0,15	0,3	0,2	0,15	0,1	0,05

a) Tính xác suất để khi đầu tư vào công ty đó thì sẽ đạt được lãi suất ít nhất là 12%.

b) Tính lãi suất kỳ vọng khi đầu tư vào công ty đó.

c) Mức độ rủi ro khi đầu tư vào công ty đó có thể đánh giá bằng cách nào?

Đáp số a) 0,5; b) $EX = 11,75$ *; c)* $\sigma_X^2 = 2,2875$.

Bài số 8. Cho biến số ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

X	1	4	8
P	0,3	0,1	0,6

Tính $P\left(\left|X - E\left(X\right)\right| < 4\right)$.

Đáp số: 0,7.

Bài số 9. Lợi nhuận X thu được khi đầu tư vào một dự án có bảng phân phối xác suất như sau (đơn vị : tỷ đồng).

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

a) Tìm mức lợi nhuận có khả năng nhiều nhất khi đầu tư vào dự án đó.

b) Việc đầu vào dự án này có hiệu quả hay không? Tại sao?

c) Làm thế nào để đo được mức độ rủi ro của vụ đầu tư này? Hãy tìm mức độ rủi ro đó.

Đáp số: a) $\text{Mod}(X) = 2$ *; b)* $EX = 0,8$ *;* $\sigma_X^2 = 2,16$.

Bài số 10. Tại một cửa hàng bán xe máy Honda người ta thống kê được số xe máy bán ra hàng tuần (X) với bảng phân phối xác suất như sau :

X	0	1	2	3	4	5	6	7	8	9	10	11
P	0,05	0,12	0,17	0,08	0,12	0,2	0,07	0,02	0,07	0,02	0,03	0,05

a) Tìm số xe trung bình bán được mỗi tuần.

b) Tìm phương sai và độ lệch chuẩn của số xe bán được mỗi tuần và giải thích ý nghĩa của kết quả nhận được.

Đáp số: a) 4,33; b) $\sigma_X^2 = 8,3411; \sigma_X = 2,89$.

Bài số 11. Sản phẩm nhà máy được đóng thành từng hộp, mỗi hộp có 10 sản phẩm. Gọi X là số sản phẩm loại một có trong hộp. Cho biết X có bảng phân phối xác suất như sau:

X	6	7
P	0,7	0,3

Khách hàng chọn cách kiểm tra để mua hàng như sau : Từ mỗi hộp lấy ngẫu nhiên ra 3 sản phẩm để kiểm tra, nếu thấy có ít nhất 2 sản phẩm loại một thì mua hộp đó. Lấy ngẫu nhiên 3 hộp để kiểm tra. Tính xác suất để có 2 hộp được mua.

Đáp số: 0,438.

Bài số 12. Thống kê số khách hàng trên một xe buýt tại một tuyến giao thông ta thu được các số liệu sau:

Số khách trên một chuyến	30	40	45	50	60
Tần suất tương ứng	0,15	0,2	0,3	0,25	0,1

a) Tìm kỳ vọng và phương sai của số khách hàng đi mỗi chuyến và giải thích ý nghĩa của kết quả nhận được.

b) Chi phí cho mỗi chuyến xe là 400 ngàn đồng không phụ thuộc vào số khách đi trên xe thì công ty xe buýt có thể thu lãi bình quân cho mỗi chuyến xe là 312 ngàn đồng. Công ty phải quy định giá vé là bao nhiêu?

Đáp số : a) $E(X) = 44,5;\ Var(X) = 67,25$. *b) 16 ngàn đồng.*

Bài số 13. Tuổi thọ của một loại bóng đèn nào đó là một biến số ngẫu nhiên liên tục X (đơn vị năm) với hàm mật độ như sau

$$f(x) = \begin{cases} kx^2(4-x) & \text{khi } \ 0 \le x \le 4 \\ 0 & \text{khi } \ x \notin [0,4] \end{cases}$$

a) Tìm k và vẽ đồ thị $f(x)$.

b) Tìm xác suất để bóng đèn hỏng trước khi nó được 1 năm tuổi.

Đáp số: a) $k = \frac{3}{64}$*; b) 0,0508.*

Bài số 14. Khối lượng của một con vịt 6 tháng tuổi là một biến số ngẫu nhiên X (đơn vị tính là Kg) có hàm mật độ

$$f(x) = \begin{cases} k(x^2-1) & \text{khi } \ 1 \le x \le 3 \\ 0 & \text{khi } \ x \notin [1,3] \end{cases}$$

a) Tìm k.

b) Với k tìm được, tính

(i) khối lượng trung bình của vịt 6 tháng tuổi,

(ii) tỷ lệ vịt chậm lớn, biết vịt 6 tháng tuổi chậm lớn là vịt có khối lượng nhỏ hơn 2Kg,

(iii) hàm phân phối xác suất của X.

Đáp số: a) $k = \frac{3}{20}$*; b)* $EX = 2,4$*;* $P(X < 2) = 0,2$

$$F(x) = \begin{cases} 0 & \text{khi } x \le 1 \\ \dfrac{x^3 - 3x + 2}{20} & \text{khi } 1 < x \le 3 \\ 1 & \text{khi } \ x > 3 \end{cases}$$

Bài số 15. Cho hàm mật độ của biến ngẫu nhiên X có dạng

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

a) Tìm a và xác định hàm phân phối xác suất của X.

b) Tính xác suất để X nhận giá trị trong khoảng $\left(\frac{\pi}{4},\pi\right)$.

Đáp số: a) $a=0,5;\ F(x)=\begin{cases} 0 & \text{khi } x\le -\frac{\pi}{2} \\ 0,5(\sin x+1) & \text{khi } -\frac{\pi}{2}<x\le\frac{\pi}{2} \\ 1 & \text{khi } x>\frac{\pi}{2} \end{cases}$

b) 0,1465.

Bài số 16. Cho X là biến số ngẫu nhiên liên tục có hàm phân phối xác suất sau:

$$F(x)=\frac{1}{4}+\frac{1}{\pi}\arctan x$$

a) Tính $P(0<X<1)$.

b) Tìm hàm mật độ xác suất của X.

Đáp số: a) 0,25; b) $f(x)=\dfrac{1}{\pi\left(1+x^2\right)}$.

Bài số 17. Cho X là biến số ngẫu nhiên liên tục có hàm phân phối xác suất sau:

$$F(x)=\frac{1}{2}+\frac{1}{\pi}\arctan\frac{x}{2}$$

Tìm giá trị x_1 thỏa mãn điều kiện: $P\left(X>x_1\right)=\frac{1}{4}$.

Đáp số: $x_1=2$.

Bài số 18. Thời gian xếp hàng chờ mua hàng của khách là biến số ngẫu nhiên liên tục X với hàm phân phối xác suất như sau:

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

a) Tìm hệ số a.

b) Tìm thời gian trung bình.

c)Tìm xác suất để trong 3 người xếp hàng thì có không quá 2 người phải chờ quá 0,5 phút.

Đáp số: a) $a = 2$ *; b)* $EX = 0,5$ *; c) 0,875.*

Bài số 19. Tỷ lệ mắc một loại bệnh trong một vùng dân cư là biến số ngẫu nhiên liên tục X có hàm mật độ như sau:

$$f(x) = \begin{cases} \dfrac{1}{20} & \text{khi } x \in [5,25] \\ 0 & \text{khi } x \notin [5,25] \end{cases}$$

a) Tính $P(|X-10| > 2,5)$.

b) Tính tỷ lệ mắc bệnh trung bình và phương sai của X.

Đáp số: a) 0,75; b) $EX = 15;\ \sigma_X^2 = 33,3$.

Bài số 20. Cho biến số ngẫu nhiên liên tục X có hàm mật độ xác suất như sau:

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

Tính trung bình và phương sai của X.

Đáp số : $EX = \sigma_X^2 = m + 1$.

Bài số 21. Xác suất để một con gà đẻ trong ngày là 0,6. Nuôi 5 con.

1) Tính xác suất để trong một ngày :

a) không con nào đẻ,

b) cả 5 con đẻ,

c) có ít nhất 1 con đẻ,

d) có ít nhất 2 con đẻ.

2) Nếu muốn mỗi ngày có trung bình 100 trứng thì phải nuôi bao nhiêu con gà.

Đáp số: 1) a) 0,01024; b) 0,07776; c) 0,98976; d) 0,91296; 2) 167 con.

Bài số 22. Một sọt cam có 10 trái trong đó có 4 trái hư. Lấy ngẫu nhiên ra 3 trái.

a) Tính xác suất lấy được 3 trái hư.

b) Tính xác suất lấy được 1 trái hư

c) Tính xác suất lấy được ít nhất 1 trái hư.

d) Tính xác suất lấy được nhiều nhất 2 trái hư.

Đáp số: a) 0,033; b) 0,5; c) 0,83; d) 0,967.

Bài số 23. Một tổng đài bưu điện có các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và có tốc độ trung bình 2 cuộc gọi trong 1 phút. Biết rằng số cuộc gọi trong một khoảng thời gian cố định có phân phối Poisson. Tìm xác suất để

a) có đúng 5 cuộc điện thoại trong 2 phút,

b) không có cuộc điện thoại nào trong khoảng thời gian 30 giây,

c) có ít nhất 1 cuộc điện thoại trong khoảng thời gian 10 giây.

Đáp số: a) 0,1563; b) 0,3679; c) 0,2835.

Bài số 24. Xác suất để một máy sản xuất ra phế phẩm là 0,02.

a) Tính xác suất để trong 10 sản phẩm do máy sản xuất có không quá 1 phế phẩm.

b) Một ngày máy sản xuất được 250 sản phẩm. Tìm số phế phẩm trung bình và số phế phẩm tin chắc nhất của máy đó trong một ngày.

Đáp số: a) 0,9838; b) $\mathrm{E}(\mathrm{X}) = 5$; $\mathrm{Mod}(\mathrm{X}) = 5$.

Bài số 25. Xác suất để một máy sản xuất ra sản phẩm loại A là 0,25. Tính xác suất để trong 80 sản phẩm do máy sản xuất ra có từ 25 đến 30 sản phẩm loại A.

Đáp số: 0,11927.

Bài số 26. Gieo 100 hạt giống của một loại nông sản. Xác suất nảy mầm của mỗi hạt là 0,8. Tính xác suất để có ít nhất 90 hạt nảy mầm.

Đáp số: 0,0057.

Bài số 27. Giả sử xác suất trúng số là 1%. Mỗi tuần mua một vé số. Hỏi phải mua vé số liên tiếp trong tối thiểu bao nhiêu tuần để có không ít hơn 95% hy vọng trúng số ít nhất 1 lần.

Đáp số: 299.

Bài số 28. Bưu điện dùng một máy tự động đọc địa chỉ trên bì thư để phân loại từng khu vực gởi đi, máy có khả năng đọc được 5000 bì thư trong 1 phút. Khả năng đọc sai 1 địa chỉ trên bì thư là 0,04% (xem như việc đọc 5000 bì thư này là 5000 phép thử độc lập).

a) Tính số bì thư trung bình mỗi phút máy đọc sai.

b) Tính số bì thư tin chắc nhất trong mỗi phút máy đọc sai.

c) Tính xác suất để trong một phút máy đọc sai ít nhất 3 bì thư.

Đáp số: a) 2; b) 2; c) 0,3233.

Bài số 29. Giả sử tỷ lệ dân cư mắc bệnh A trong vùng là 10%. Chọn ngẫu nhiên 1 nhóm 400 người.

a) Viết công thức tính xác suất để trong nhóm có nhiều nhất 50 người mắc bệnh A.

b) Tính xấp xỉ xác suất đó bằng phân phối chuẩn.

Đáp số: b) 0,9599.

Bài số 30. Sản phẩm sau khi hoàn tất được đóng thành kiện, mỗi kiện gồm 10 sản phẩm với tỷ lệ thứ phẩm là 20%. Trước khi mua hàng, khách hàng muốn kiểm tra bằng cách từ mỗi kiện chọn ngẫu nhiên 3 sản phẩm.

1) Lập bảng phân phối xác suất của số sản phẩm tốt trong 3 sản phẩm lấy ra.

2) Nếu cả 3 sản phẩm được lấy ra đều là sản phẩm tốt thì khách hàng sẽ đồng ý mua kiện hàng đó. Tính xác suất để khi kiểm tra 100 kiện

a) có ít nhất 80 kiện hàng được mua,

b) có ít nhất 60 kiện được mua.

Đáp số: 1)

X	0	1	2	3
P	0	0,066	0,467	0,467

2a) $8,2\cdot 10^{-12}$; *2b) 0,0052.*

Bài số 31. Một trạm cho thuê xe Taxi có 3 chiếc xe. Hàng ngày trạm phải nộp thuế 8USD cho 1 chiếc xe (bất kể xe đó có được thuê hay không). Mỗi chiếc được cho thuê với giá 20USD. Giả sử số xe được yêu cầu cho thuê của trạm trong 1 ngày là đại lượng ngẫu nhiên có phân phối Poisson với $\mu = 2,8$.

a) Tính số tiền trung bình trạm thu được trong một ngày.

b) Giải bài toán trên trong trường hợp trạm có 4 chiếc xe.

c) Theo bạn, trạm nên có 3 hay 4 chiếc xe ?

Đáp số: a) 32; b) 24; c) Thuê 3 xe .

Bài số 32. Đường kính của một loại chi tiết do một máy sản xuất có phân phối chuẩn, kỳ vọng 20mm, phương sai $(0,2\text{mm})^2$. Lấy ngẫu nhiên 1 chi tiết máy. Tính xác suất để

a) có đường kính trong khoảng 19,9mm đến 20,3mm,

b) có đường kính sai khác với kỳ vọng không quá 0,3mm.

Đáp số: a) 0,6247; b) 0,8664.

Bài số 33. Trong hệ thống tỷ giá hối đoái thả nổi, sự biến động của tỷ giá hối đoái chịu sự tác động của nhiều nhân tố và có thể xem như là biến ngẫu nhiên có phân phối chuẩn. Giả sử ở giai đoạn nào đó tỷ giá của USD với VND có trung bình là 18000đ và độ lệch chuẩn là 800đ. Tìm xác suất để trong một ngày nào đó.

a) Tỷ giá sẽ cao hơn 19000đ,

b) Tỷ giá sẽ thấp hơn 17500đ,

c) Tỷ giá nằm trong khoảng từ 17500đ đến 19500.

Đáp số: a) 0,1057; b) 0,266; c) 0,7036.

Bài số 34. Khối lượng của một gói đường (đóng bằng máy tự động) có phân phối chuẩn. Trong 1000 gói đường có 70 gói có khối lượng lớn hơn 1015. Hãy ước lượng xem có bao nhiêu gói đường có khối lượng ít hơn 1008g. Biết rằng khối lượng trung bình của 1000 gói đường là 1012g.

Đáp số: 24,4 gói.

Bài số 35. Lãi suất (%) đầu tư vào một dự án năm 2000 được coi như 1 đại lượng ngẫu nhiên có phân phối theo quy luật chuẩn. Theo đánh giá của uỷ ban đầu tư thì lãi suất cao hơn 20% có xác suất 0,1587, và lãi suất cao hơn 25% có xác suất là 0,0228. Vậy khả năng đầu tư mà không bị thua lỗ là bao nhiêu?

Đáp số: 0,9987.

Bài số 36. Một công ty kinh doanh mặt hàng A dự định sẽ áp dụng một trong 2 phương án kinh doanh. Ký hiệu X_1 là lợi nhuận thu được khi áp dụng phương án thứ 1, X_2 là lợi nhuận thu được khi áp dụng phương án thứ 2. X_1, X_2 đều được tính theo đơn vị triệu đồng/ tháng) và $X_1 \sim N(140,2500)$, $X_2 \sim N(200,3600)$. Nếu biết rằng, để công ty tồn tại và phát triển thì lợi nhuận thu được từ mặt hàng kinh doanh 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?.

Đáp số: $P(X_1 \geq 80) = 0,8849 < P(X_2 \geq 80) = 0,9772$, *chọn phương án thứ 2.*

Bài số 37. Độ dài của 1 chi tiết máy được tiện ra có phân phối chuẩn $N(\mu cm;(0,2cm)^2)$. Sản phẩm coi là đạt nếu độ dài sai lệch so với độ dài trung bình không quá 0,3cm.

a) Tính xác suất chọn ngẫu nhiên 1 sản phẩm thì được sản phẩm đạt yêu cầu.

b) Chọn ngẫu nhiên 3 sản phẩm. Tính xác suất có ít nhất 2 sản phẩm đạt yêu cầu .

c) Nếu sản phẩm tốt mà bị loại trong kiểm tra thì mắc phải sai lầm loại 1, nếu sản phẩm không đạt mà được nhận thì mắc phải sai lầm loại 2. giả sử khả năng mắc phải sai lầm loại 1, loại 2 lần lượt là 0,1 và 0,2. Tính xác suất để trong 3 lần kiểm tra hoàn toàn không nhầm lẫn.

Đáp số a) 0,8664; b) 0,9512; c) 0,697.

Bài số 38. Khối lượng của 1 loại trái cây có quy luật phân phối chuẩn với khối lượng trung bình là 250g, độ lệch chuẩn về khối lượng là 5g.

a) Một người lấy 1 trái từ trong sọt trái cây ra. Tính xác suất người này lấy được trái loại 1 (trái loại 1 là trái có khối lượng > 260g).

b) Nếu lấy được trái loại 1 thì người này sẽ mua sọt đó. Người này kiểm tra 100 sọt, tính xác suất mua được 6 sọt.

Đáp số: a) 0,0228; b) 0,019.

Bài số 39. Có hai thị trường A và B, lãi suất của cổ phiếu trên hai thị trường này là các đại lượng ngẫu nhiên có phân phối chuẩn, độc lập với nhau, có kỳ vọng và phương sai được cho trong bảng dưới đây:

	Trung bình	Phương sai
Thị trường A	19%	36
Thị trường B	22%	100

a) Nếu mục đích là đạt lãi suất tối thiểu bằng 10% thì nên đầu tư vào loại cổ phiếu nào?

b) Để giảm rủi ro đến mức thấp nhất thì nên đầu tư vào cổ phiếu trên cả hai thị trường theo tỷ lệ như thế nào?

Đáp số: a) nên đầu tư vào cổ phiếu trên thị trường loại A.

b) 74% vào thị trường A còn lại là thị trường B.

Bài số 40. Nghiên cứu chiều cao của những người trưởng thành, người ta nhận thấy rằng chiều cao đó tuân theo quy luật phân bố chuẩn với trung bình là 175cm và độ lệch tiêu chuẩn 4cm. Hãy xác định :

a) Tỷ lệ người trưởng thành có chiều cao trên 180cm.

b) Tỷ lệ người trưởng thành có chiều cao từ 166cm đến 177cm.

c) Giá trị h_0, nếu biết rằng 33% người trưởng thành có chiều cao dưới mức h_0.

d) Giới hạn biến động chiều cao của 90% người trưởng thành xung quanh giá trị trung bình của nó.

Đáp số: a) 0,1056; b) 0,6793; c) 173,24; d) 6,56.

Bài số 41. Chiều dài của chi tiết được gia công trên máy tự động là biến ngẫu nhiên tuân theo quy luật phân phối chuẩn với độ lệch tiêu chuẩn là 0,01mm. Chi tiết được coi là đạt tiêu chuẩn nếu kích thước thực tế của nó sai lệch so với kích thước trung bình không vượt quá 0,02mm.

a) Tìm tỷ lệ chi tiết không đạt tiêu chuẩn.

b) Xác định độ đồng đều (phương sai) cần thiết của sản phẩm để tỷ lệ chi tiết không đạt tiêu chuẩn chỉ còn 1%.

Đáp số: a) 0,9545; b) $\left(7,752.10^{-3}\right)^2$.

Bài số 42. Khối lượng X của một loại trái cây ở nông trường được biết có kỳ vọng 250gr và phương sai $81\left(\text{gr}\right)^2$. Trái cây được đóng thành sọt, mỗi sọt 100 trái. Mỗi sọt được gọi là loại A nếu khối lượng không dưới 25kg. Kiểm tra ngẫu nhiên 100 sọt. Tính xác suất :

a) có nhiều nhất 60 sọt loại A,

b) ít nhất 45 sọt loại A.

Đáp số: a) 0,9824; b) 0,8644.

Bài số 43. Việc kiểm tra các viên bi được tiến hành như sau: nếu viên bi không lọt qua lỗ có đường kính d_1 song lọt qua lỗ có đường kính d_2 thì viên bi được coi là đạt tiêu chuẩn, nếu không thì viên bi bị loại. Biết đường kính các viên bi sản xuất ra là biến ngẫu nhiên có phân phối chuẩn với trung bình là $\frac{d_1 + d_2}{2}$ và độ lệch chuẩn là $\frac{d_2 - d_1}{4}$. Tìm xác suất để viên bi bị loại.

Đáp số: 0,0456.

Bài số 44. Một đề thi trắc nghiệm có 4 câu hỏi lý thuyết và 3 bài tập độc lập nhau. Khả năng để một sinh viên trả lời đúng một câu hỏi lý thuyết là 0,7 và đúng một bài tập là 0,5. Trả lời đúng một câu hỏi lý thuyết được 1 điểm, sai được 0 điểm. Trả lời đúng mỗi bài tập được 2 điểm, sai được 0 điểm. Tìm số điểm trung bình mà sinh viên đó đạt được.

Đáp số: 5,8.

Bài số 45. Trọng lượng sản phẩm do một máy sản xuất là ĐLNN tuân theo luật phân phối chuẩn với độ lệch chuẩn là 1,6gam. Sản phẩm được coi là đạt tiêu chuẩn nếu trọng lượng của nó sai lệch so với trung bình không quá 2gam.

a) Tính tỉ lệ sản phẩm đạt tiêu chuẩn do máy đó sản xuất.

b) Cần sản xuất tối thiểu bao nhiêu sản phẩm để xác suất "có ít nhất 100 sản phẩm đạt tiêu chuẩn" không bé hơn 90%.

Đáp số : a)0,7888; b) 134.

2.7. Tài liệu tham khảo

[1] Nguyễn Cao Văn, Trần Thái Ninh, Ngô Văn Thứ, Lý thuyết xác suất và thống kê toán, NXB Đại học Kinh tế Quốc dân, 2012.

[2] Nguyễn Cao Văn, Trần Thái Ninh, Nguyễn Thế Hệ, Bài tập xác suất và thống kê toán, NXB Đại học Kinh tế Quốc dân, 2012.

[3] Phạm Hoàng Uyên, Lê Thị Thiên Hương, Huỳnh Văn Sáu, Nguyễn Phúc Sơn, Huỳnh Tố Uyên, Lý thuyết xác suất, NXB Đại học Quốc gia Thành Phố Hồ Chí Minh, 2015.

[4] Anderson, Sweeney, and William [2010], Statistics for Business and Economics, South-Western Cengage Learning (11^{th} Edition).

[5] Michael Barrow, Statistics for Economics, Accounting and Business Studies- Prentice Hall, 2006.

[6] Newbold Paul - Statistics for Bussiness and Economics, 5th edition - Prentice Hall, 2005.

Thuật ngữ chính chương 2	
Tiếng Anh	**Tiếng Việt**
Binomial random variables	Biến ngẫu nhiên có phân phối nhị thức
continuous random variable	Biến ngẫu nhiên liên tục
Corollary	Hệ quả
Compute	Tính toán
Chi-squared distribution	Phân phối Chi bình phương
Cumulative distribution function	Hàm phân phối
continuous random variable	Biến ngẫu nhiên liên tục
Discrete random variable	Biến ngẫu nhiên rời rạc
Density function	Hàm mật độ
Gamma and Beta distributions	Phân phối Gamma và Beta
Hypergeometric Random Variable	Biến ngẫu nhiên có phân phối siêu bội
Poisson random variable	Biến ngẫu nhiên có phân phối Poisson
Random variable	Biến ngẫu nhiên
Integration by parts	Tích phân từng phần
Mean	Trung bình
Median	Trung vị
Measurable	Đo được
Normal distribution	Phân phối chuẩn
Normal random variables	Biến ngẫu nhiên có phân phối chuẩn
Expected value	Giá trị kỳ vọng
Exponential random variable	Biến ngẫu nhiên có phân phối mũ
Standard deviation	Độ lệch chuẩn
Parameter	Tham số
Probability density function	Hàm mật độ xác suất
Probability function	Hàm xác suất
Pareto distribution	Phân phối Pareto
Theorem	Định lý
Uniform distribution function	Hàm phân phối đều
Variance	Phương sai

Chương 3

MẪU NGẪU NHIÊN VÀ BÀI TOÁN ƯỚC LƯỢNG

Mục tiêu chương 3

Chương này giúp sinh viên:

- Phân biệt được thế nào là tổng thể và mẫu ngẫu nhiên.
- Tính được các tham số đặc trưng mẫu như trung bình, phương sai, tỷ lệ...
- Nắm được bài toán ước lượng điểm và ước lượng khoảng.
- Hiểu được bài toán ước lượng khoảng và tìm được lượng khoảng như ước lượng trung bình, phương sai, tỷ lệ.

3.1. Mẫu ngẫu nhiên

3.1.1. Tổng thể nghiên cứu

Trong thực tế thường phải nghiên cứu một tập hợp các phần tử đồng nhất theo một hay nhiều dấu hiệu định tính hoặc định lượng đặc trưng cho các phần tử đó. Chẳng hạn một doanh nghiệp phải nghiên cứu tập hợp các khách hàng thì dấu hiệu định tính có thể là mức độ hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, còn dấu hiệu định lượng là nhu cầu của khách hàng về số lượng sản phẩm của doanh nghiệp.

3.1.1.1. Định nghĩa

Toàn bộ tập hợp các phần tử đồng nhất theo một dấu hiệu nghiên cứu định tính hoặc định lượng nào đó được gọi là tổng thể nghiên cứu hay tổng thể.

Ví dụ 3.1.

a) Nghiên cứu tập hợp các khách hàng của một doanh nghiệp theo dấu hiệu định tính - mức độ hài lòng về sản phẩm, hay định lượng - nhu cầu về số lượng sản phẩm.

b) Nghiên cứu tập hợp học sinh của một lớp: định tính: học lực; định lượng: chiều cao/ cân nặng.

3.1.1.2. Các phương pháp mô tả tổng thể

Giả sử trong tổng thể, dấu hiệu nghiên cứu X nhận các giá trị $X_1, X_2, ..., X_k$ với các tần số tương ứng $N_1, N_2, ..., N_k$; $\sum_{i=1}^{k} N_i = N$ (N còn gọi là kích thước tổng thể).

a. Bảng phân phối tần số của tổng thể

X	X_1	X_2	…	X_k
Tần số	N_1	N_2	…	N_k

b. Bảng phân phối tần suất

Đặt $p_i = \frac{N_i}{N}$ $(i = 1, 2, ..., k)$, gọi là tần suất xuất hiện giá trị X_i

X	X_1	X_2	…	X_k
Tần suất	p_1	p_2	…	p_k

Trong đó, $\sum_{i=1}^{k} p_i = 1; \quad 0 \le p_i \le 1$

Nhận xét : Việc mô tả dấu hiệu X trên một tổng thể bằng các phương pháp trên cho phép chúng ta có thể coi dấu hiệu X như một biến ngẫu nhiên.

Ví dụ 3.2. Điểm thi của học sinh một trường:

X	0	1	…
Tần suất	p_0	p_1	…

Mặc dù kết quả thi đã có (tất nhiên) nhưng dựa trên số liệu thống kê, điểm của một thí sinh A nào đó được coi như 1 biến ngẫu nhiên.

3.1.1.3. Các tham số đặc trưng của tổng thể:

a. Trung bình tổng thể: Là trung bình số học của các giá trị của dấu hiệu trong tổng thể, ký hiệu là μ và được tính bởi công thức:

$$\mu = \frac{1}{N}\sum_{i=1}^{k} N_i X_i = \sum_{i=1}^{k} p_i X_i \tag{3.1}$$

Nếu xem dấu hiệu nghiên cứu như biến ngẫu nhiên X thì trung bình tổng thể chính là kỳ vọng toán của biến ngẫu nhiên đó.

b. Phương sai tổng thể: Là trung bình số học của bình phương các sai lệch giữa các giá trị của các dấu hiệu trong tổng thể và trung bình tổng thể, ký hiệu σ^2 được tính bởi công thức:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} N_i (X_i - \mu)^2 = \sum_{i=1}^{k} p_i (X_i - \mu)^2 \tag{3.2}$$

c. Độ lệch chuẩn tổng thể: ký hiệu là σ và được tính bởi công thức:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{k} p_i (X_i - \mu)^2} \tag{3.3}$$

d. Tỷ lệ tổng thể: là tỷ số giữa số phần tử mang dấu hiệu nghiên cứu, M và kích thước của tổng thể N, ký hiệu p và được xác định

$$p = \frac{M}{N}. \tag{3.4}$$

3.1.2. Mẫu ngẫu nhiên

Các tham số đặc trưng của tổng thể có thể xác định được một cách trực tiếp nếu áp dụng phương pháp nghiên cứu toàn bộ tổng thể. Tuy nhiên trong thực tế việc áp dụng phương pháp này gặp phải những khó khăn chủ yếu sau:

- Phải chịu chi phí rất lớn về thời gian, nhân lực, tiền bạc và phương tiện.
- Nếu quy mô của tập hợp quá lớn có thể xảy ra trường hợp trùng hoặc bỏ sót các phần tử, sai sót trong quá trình thu thập thông tin ban đầu, hạn chế độ chính xác của kết quả phân tích.
- Nếu các phần tử của tập hợp bị phá huỷ trong quá trình nghiên cứu thì phương pháp nghiên cứu toàn bộ trở thành vô nghĩa.
- Chưa thể xác định được toàn bộ N phần tử của tổng thể.

Vì thế trong thực tế phương pháp nghiên cứu toàn bộ thường chỉ được áp dụng đối với các tập hợp có quy mô nhỏ, còn chủ yếu người ta áp dụng phương pháp nghiên cứu không toàn bộ, đặc biệt là phương pháp mẫu bằng cách chọn ra từ tổng thể n phần tử và chỉ tập trung nghiên cứu các phần tử đó. Tập hợp n phần tử này được gọi là mẫu kích thước n.

Phương pháp chọn mẫu:

- Mỗi lần lấy vào mẫu chỉ một phần tử.
- Lấy phần tử nào đưa vào mẫu là hoàn toàn ngẫu nhiên.
- Các phần tử được lấy vào mẫu theo phương thức hoàn lại.

Định nghĩa: Mẫu ngẫu nhiên kích thước n là tập hợp của n biến ngẫu nhiên độc lập $X_1, X_2, ..., X_n$ được thành lập từ biến ngẫu nhiên X trong tổng thể và có cùng quy luật phân phối xác suất với X, ký hiệu: $T = T(X_1, X_2, ..., X_n)$. Ta có

$$E(X_i) = \mu,\ i = 1,2,...,n \text{ và } var(X_i) = \sigma^2,\ i = 1,2,...,n.$$

Ví dụ 3.3. Gọi X là số chấm thu được khi gieo một xúc xắc, X là biến ngẫu nhiên với bảng phân phối xác suất :

X	1	2	3	4	5	6

P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Nếu gieo con xúc xắc 3 lần và gọi X_i là số chấm xuất hiện ở lần gieo thứ i $(i=\overline{1,3})$ thì ta có 3 biến ngẫu nhiên X_1, X_2, X_3 độc lập, cùng phân phối xác suất với X tạo nên một mẫu ngẫu nhiên kích thước $n=3$.

$$T = T(X_1, X_2, X_3).$$

Hơn nữa, mỗi biến ngẫu nhiên X_i trong mẫu đều có bảng phân phối xác suất giống như bảng phân phối xác suất của biến ngẫu nhiên X, do đó:

$$E(X_i) = \frac{1}{6}(1+2+3+4+5+)6 = \frac{21}{6} = E(X),\ i=1,2,3.$$

$$\mathrm{var}(X_i) = \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) - \left(\frac{21}{6}\right)^2 = \frac{35}{12} = \mathrm{var}(X),\ i=1,2,3.$$

3.1.3. Các đặc trưng quan trọng của mẫu

3.1.3.1. Thống kê

Thống kê là một biểu thức theo mẫu $X_1, X_2, ..., X_n$ và không phụ thuộc vào các tham số chưa biết. Ký hiệu $T = T(X_1, X_2, ..., X_n)$.

Khi đã quan sát được mẫu, ta có thể tính ra giá trị của một thống kê. Tùy theo từng vấn đề nghiên cứu, ta có thể đặt ra một hay nhiều thống kê khác nhau. Các thống kê thường dùng là

a. Trung bình mẫu :

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.5}$$

b. Phương sai mẫu có hiệu chỉnh:

$$S_X^2 = \frac{\left(X_1-\overline{X}\right)^2 + \left(X_2-\overline{X}\right)^2 + \cdots + \left(X_n-\overline{X}\right)^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 \tag{3.6}$$

c. Giá trị nhỏ nhất của mẫu : $Y_1 = \min\{X_1, X_2, ..., X_n\}$,

d. Giá trị lớn nhất của mẫu : $Y_n = \max\{X_1, X_2, ..., X_n\}$,

e. Khoảng biến thiên của mẫu : $Y_n - Y_1$.

Ví dụ 3.4. Quan sát chiều cao X (cm) của 10 người, ta ghi được 158, 163, 157, 162, 154, 152, 160, 159, 165, 156.

Với mẫu trên, ta tính được

- Chiều cao trung bình mẫu : $\overline{X} = 158,6\text{cm}$,
- Phương sai của mẫu : $S_X^2 = 16,49\text{cm}^2$,
- Người thấp nhất : $Y_1 = 152\text{cm}$,
- Người cao nhất : $Y_2 = 165\text{cm}$,
- Khoảng biến thiên của mẫu : $R = 13\text{cm}$.

Tương tự như các tham số đặc trưng cho một biến ngẫu nhiên, trung bình mẫu là giá trị mà ta hy vọng nhận được khi xem xét một phần tử của mẫu, phương sai mẫu cho ta biết mức độ phân tán của số liệu mẫu. Phương sai càng nhỏ, số liệu càng ít phân tán.

Để có thể tính toán phương sai một cách nhanh chóng, ta có thể dùng kết quả sau

3.1.3.2. Mệnh đề. *Xét mẫu* $X_1, X_2, ..., X_n$ *với trung bình* $\overline{X}$ *và phương sai* S_X^2, *ta có*

$$S_X^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right). \tag{3.7}$$

Lưu ý rằng trong trường hợp số liệu mẫu có lặp lại và ta ghi nhận các tần số xuất hiện số liệu như sau

X	X_1	X_2	...	X_k
Tần số	n_1	n_2	...	n_k

thì cỡ mẫu là $n = n_1 + n_2 + \cdots + n_k$ và các công thức tính trung bình cũng như phương sai trở thành

$$\overline{X} = \frac{n_1X_1 + n_2X_2 + \cdots + n_kX_k}{n} = \frac{1}{n}\sum_{i=1}^{k} n_iX_i,$$

$$S_X^2 = \frac{n_1\left(X_1 - \overline{X}\right)^2 + n_2\left(X_2 - \overline{X}\right)^2 + \cdots + n_k\left(X_k - \overline{X}\right)^2}{n-1}$$

$$= \frac{1}{n-1}\sum_{i=1}^{k} n_i\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} n_iX_i^2 - n\overline{X}^2\right).$$

Ví dụ 3.5. Một máy tự động đóng bột vào bao. Cân ngẫu nhiên 15 bao được các trọng lượng sau:

39,75	40,25	39,50	40,25	40,50
40,00	39,75	40,00	40,00	39,25
39,25	39,50	40,00	39,50	39,50

a) Lập bảng phân phối tần số thực nghiệm của trọng lượng các bao bột.

b) Tính giá trị trung bình và phương sai mẫu hiệu chỉnh.

Giải

a) Bảng phân phối tần số thực nghiệm:

Trọng lượng (kg)	39,25	39,50	39,75	40,00	40,25	40,50
Số bao	2	4	2	4	2	1

b) Gọi X là trọng lượng các bao bột.

Ta có:

Cỡ mẫu: $n = 2 + 4 + 2 + 4 + 2 + 1 = 15$

Trung bình mẫu:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i X_i = \frac{2\cdot 39,25 + 4\cdot 39,5 + 2\cdot 39,75 + 4\cdot 40 + 2\cdot 40,25 + 1\cdot 40,50}{15} = 39,8$$

Phương sai mẫu có hiệu chỉnh

$$\sum_{i=1}^{k} n_i X_i^2 = 2\cdot 39,25^2 + 4\cdot 39,5^2 + 2\cdot 39,75^2 + 4\cdot 40^2 + 2\cdot 40,25^2 + 1\cdot 40,50^2$$

$$= 23762,625$$

Áp dụng (3.7), ta có

$$S_X^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} n_i X_i^2 - n\overline{X}^2\right) = \frac{1}{14}\left(23762,625 - 15\cdot 39,8^2\right) = 0,1446.$$

Ví dụ 3.6. Gặt ngẫu nhiên 100 điểm trồng lúa của một vùng nông thôn ta thu được bảng số liệu như sau:

Năng suất	30	33	34	36	40
Số điểm	15	20	41	18	6

Xác định các thống kê đặc trưng mẫu.

Giải

Gọi X là năng suất lúa (tạ/ha).Ta có mẫu cụ thể kích thước $n = 100$.

Ta có thể tính toán dựa vào bảng sau:

x_i	n_i	$n_i X_i$	$n_i X_i^2$
30	15	450	13500
33	20	660	21780
34	41	1394	47396
36	18	648	23328
40	6	240	9600

	$\sum n_i = n = 100$	$\sum n_i x_i = 3392$	$\sum n_i x_i^2 = 115604$

Từ đó ta có:

Năng suất lúa trung bình: $\overline{X} = \dfrac{3392}{100} = 33,92$

Phương sai mẫu có hiệu chỉnh của năng suất lúa:

$$S_X^2 = \frac{1}{99}\left(115604 - 100 \cdot 33,92^2\right) = 5,5289.$$

3.1.3.3. Định lý Lindeberg-Lévy. Nếu mẫu $X_1, X_2, ..., X_n$ lấy từ tổng thể có phân phối chuẩn $N\left(\mu, \sigma^2\right)$ nghĩa là $X_i \sim N\left(\mu, \sigma^2\right)$, với mọi i, thì

i) $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

ii) $\dfrac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1)$.

3.2. Trình bày kết quả điều tra

Đối với dữ liệu định tính hoặc dữ liệu định lượng ít biểu hiện, ta sử dụng:

- Bảng tần số, tần suất, tần số tích lũy, tần suất tích lũy;
- Biểu đồ hình tròn, biểu đồ hình cột, biểu đồ hình thanh.

3.2.1. Trình bày kết quả điều tra dưới dạng bảng

Bảng tần số, tần suất

Trị số của biến (X_i)	**Tần số** (f_i)	**Tần suất (%)**
X_1	f_1	$\dfrac{f_1}{n}$
X_2	f_2	$\dfrac{f_2}{n}$
…	…	…
X_k	f_k	$\dfrac{f_k}{n}$
Tổng số	$\sum_{i=1}^{k} f_i = n$	100%

Bảng tần số gồm hai phần

- Trị số của biến nghiên cứu, ký hiệu X_i.

- Số lần xuất hiện của trị số gọi là tần số, ký hiệu f_i.

Cũng có thể thể hiện tần số bằng hình thức phần trăm (%)

Ví dụ 3.7. Kết quả khảo sát 200 người về lựa chọn màu của xe máy mà họ yêu thích như sau:

Màu xe	Tần số	Tần suất (%)
Đỏ	30	15
Đen	60	30
Xanh	50	25
Trắng	20	10
Khác	40	20
Tổng số	200	100

Việc cộng dồn các giá trị tần số và tần suất cho ta các giá trị tần số tích lũy và tần suất tích lũy cho dữ liệu định tính dạng thứ bậc nhằm cung cấp thông tin thêm cho người đọc.

Trị số của biến (X_i)	Tần số tích lũy (f_i)	Tần suất tích lũy (%)
X_1	f_1	$\frac{f_1}{n}$
X_2	$f_1 + f_2$	$\frac{f_1 + f_2}{n}$
…	…	…
X_k	$f_1 + f_2 + \cdots + f_k$	$\frac{f_1 + f_2 + \cdots + f_k}{n}$

Ví dụ 3.8. Khảo sát 50 người về mức lương hằng tháng mà họ nhận được trong vòng 5 năm trở lại đây như sau:

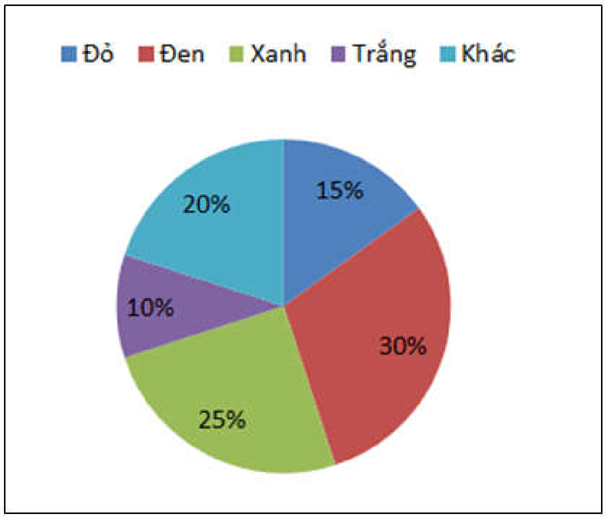
Mức lương (triệu đồng/tháng)	Tần số	Tần số tích lũy	Tần suất (%)	Tần suất tích lũy (%)
<3	5	5	10	10
[3;5)	15	20	30	40
[5;10)	20	40	40	80
>10	10	50	20	100
Tổng số	50		100	

Ý nghĩa: Giá trị 40 của tần số tích lũy cho biết có 40 người trong tổng số những người được khảo sát có thu nhập nhiều nhất là 10 triệu đồng/tháng.

3.2.2. Trình bày kết quả điều tra bằng biểu đồ

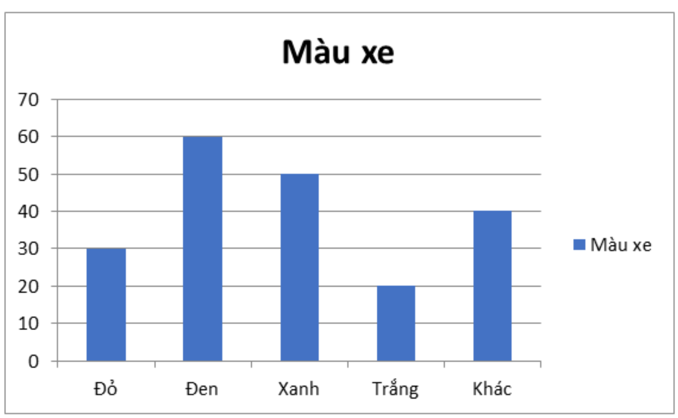
3.3.2.1. Biểu đồ hình tròn: là biểu đồ mà dữ liệu thể hiện các nhóm giá trị khác nhau được phân biệt dựa vào mà sắc, nhóm giá trị nào có tần số càng lớn thì phần màu sắc tượng trưng cho nhóm giá trị đó sẽ càng to hơn các nhóm giá trị khác. Mỗi một màu sắc được thể hiện là một hình quạt

Ví dụ 3.9. Với dữ liệu ở ví dụ 3.7, ta có dạng biểu diễn số lượng màu xe như hình vẽ sau: trong đó mỗi màu xe là hình quạt.



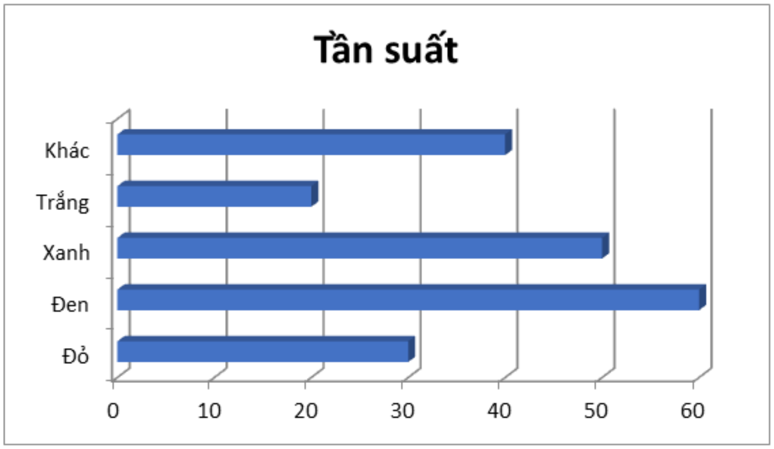
3.3.2.2. Biểu đồ hình cột: là dạng biểu đồ được sử dụng cho các dữ liệu định tính hoặc định lượng nhưng có ít biểu hiện hoặc định lượng đã được phân khoảng. Trục hoành thể hiện giá trị của vấn đề nghiên cứu; trục tung thể hiện tần số. Mỗi cột thể hiện một giá trị; độ cao của cột là tần số.

Ví dụ 3.10. Với dữ liệu ở ví dụ 3.7, ta có dạng biểu diễn số lượng màu xe được thống kê khi khảo sát như hình vẽ sau:



3.3.2.3. Biểu đồ hình thanh: là dạng biểu đồ hình cột mà khi quay ngang, hai trục đổi vị trí cho nhau, nó hay sử dụng khi giá trị của vấn đề nghiên cứu dài.

Ví dụ 3.11. Với dữ liệu ở ví dụ 3.7, ta có dạng biểu diễn số lượng màu xe như hình vẽ sau:



3.2.3. Tính giá trị của các đặc trưng mẫu qua số liệu điều tra

3.2.3.1. Các đặc trưng đo lường khuynh hướng tập trung

a. Số trung bình: là mức độ đại diện điển hình cho 1 tiêu thức nào đó của tổng thể mà các đơn vị của tổng thể biểu hiện nhiều mức độ khác nhau.

Số trung bình cộng đơn giản: với mẫu khảo sát : $X_1, X_2, ..., X_n$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.8}$$

Số trung bình cộng có trọng số : với số liệu mẫu có lặp lại và ta ghi nhận các tần số xuất hiện số liệu như sau

X	X_1	X_2	...	X_k
Tần số	f_1	f_2	...	f_k

thì cỡ mẫu là $n = f_1 + f_2 + \cdots + f_k$ và công thức tính trung bình trở thành

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} f_i X_i. \tag{3.9}$$

Lưu ý: Nếu số liệu được cho dưới dạng có khoảng cách tổ thì ta lấy điểm giữa đại diện.

Ví dụ 3.12. Năng suất lao động của công nhân trong một phân xưởng sau

Mức năng suất lao động (tạ/người) (X_i)	21	23	25	27	29
Số công nhân (f_i)	5	10	30	15	5

Tính Năng suất lao động trung bình. Áp dụng công thức (3.9), ta có:

$$\overline{X} = \frac{5\times21+10\times23+30\times25+15\times27+5\times29}{5+10+30+15+5} = 25{,}154.$$

b. Mốt (Mode): là biểu hiện của một tiêu thức được gặp nhiều nhất trong tổng thể. Đối với một dãy số lượng biến, mốt là lượng biến có tần số lớn nhất. Ký hiệu M_0

Cách xác định mốt

+) Đối với đại lượng biến rời rạc: Mốt là lượng biến có tần số lớn nhất.

Ví dụ 3.13. Với dữ liệu ví dụ 3.12. Ta có tần số lớn nhất là 30 nên $M_0 = 25$.

+) Đối với đại lượng biến có khoảng cách tổ

Bước 1: Xác định tổ chứa mốt, là tổ có mật độ phân phối lớn nhất.

Mật độ phân phối là tỷ số giữa các tần số và trị số khoảng cách tổ tương ứng

$$\text{Mật độ phân phối} = \frac{\text{Tần số}}{\text{Trị số khoảng cách tổ}} = \frac{f_i}{d_i} (i = \overline{1,n})$$

Bước 2. Tính trị số Mốt gần đúng theo CT

$$M_0 = x_{M_{0(\min)}} + d_{M_0} \frac{F_{M_0} - F_{M_{0-1}}}{\left(F_{M_0} - F_{M_{0-1}}\right) + \left(F_{M_0} - F_{M_{0+1}}\right)} \qquad (3.10)$$

- $x_{M_0(\min)}$: là giới hạn dưới của khoảng cách tổ có Mốt
- d_{M_0} : là trị số khoảng cách tổ có Mốt
- F_{M_0} : là mật độ phân phối của tổ có Mốt
- $F_{M_{0-1}}$: là mật độ phân phối của tổ đứng trước tổ có Mốt
- $F_{M_{0+1}}$: là mật độ phân phối của tổ đứng sau tổ có Mốt

Lưu ý: Nếu khoảng cách tổ bằng nhau thì công thức xác định mốt là

$$M_0 = x_{M_{0(\min)}} + d_{M_0} \frac{f_{M_0} - f_{M_{0-1}}}{\left(f_{M_0} - f_{M_{0-1}}\right) + \left(f_{M_0} - f_{M_{0+1}}\right)} \qquad (3.11)$$

Ví dụ 3.14. Năng suất lao động của công nhân trong một phân xưởng sau

Phân tổ công nhân theo mức năng suất lao động (tạ/người)	Số công nhân (f_i)
20-22	5
22-24	10
24-26	20
26-28	15

28-30	5

Do khoảng cách tổ bằng nhau nên áp dụng công thức (3.11), ta có

$$M_0 = 24 + (26-24)\frac{10-5}{(10-5)+(10-7,5)} = 25,33.$$

c. Số trung vị : là lượng biến của đơn vị tổng thể ở vị trí giữa trong dãy số lượng biến. Số trung vị chia dãy số lượng biến thành hai phần, mỗi phần có số đơn vị tổng thể bằng nhau, ký hiệu Me.

Đối với dãy số lượng biến rời rạc:

+) Trường hợp số đơn vị tổng thể lẻ ($n = 2m+1$). Số trung vị sẽ là đại lượng ở vị trí thứ ($m+1$): $Me = X_{m+1}$.

Ví dụ 3.15. Ta có mức năng suất lao động của năm công nhân trong một tổ là 20, 22, 25, 27, 29 tạ/người. Vậy số trung vị: Me = 25 tạ/ người.

+) Trường hợp số đơn vị tổng thể chẵn ($n = 2m$). Số trung vị sẽ là đại lượng giữa vị trí thứ (m) và ($m+1$): $Me = \dfrac{X_m + X_{m+1}}{2}$.

Ví dụ 3.16. Ta có mức năng suất lao động của sáu công nhân trong một tổ là: 20, 22, 25, 27, 29, 30 tạ/người. vậy số trung vị: Me = 26 tạ/ người.

Đối với dãy số có lượng biến là khoảng cách tổ

Bước 1. Xác định số tổ chứa trung vị là số tổ có tần số tích lũy bằng hoặc lớn hơn hay bằng một nửa tổng các tần số cộng 1.

Bước 2. Xác định giá trị gần đúng của trung vị

$$M_e = X_{M_{e(\min)}} + d_{M_e}\frac{\frac{1}{2}\sum f_i - S_{M_{e-1}}}{f_{M_e}} \tag{3.12}$$

Trong đó

- $X_{Me(\min)}$: là giá trị giới hạn dưới của tổ chứa trung vị
- d_{Me} : là trị số khoảng cách tổ chứa trung vị
- $\sum f_i$:là tổng tần số
- S_{Me-1} : là tần số tích lũy của tổ đứng trước tổ chứa trung vị.
- f_{Me} : Tần số của tổ chứa trung vị.

Ví dụ 3.17. Về mức lương của công nhân trong một phân xưởng của xí nghiệp X như sau:

Mức lương 1 công nhân (triệu đồng / người)	Số công nhân (người)	Tần số Tích lũy
2,7 – 2,9	5	5
2,9 – 3,1	10	15
3,1 – 3,3	20	35
3,3 – 3,5	15	50
3,5 – 3,7	5	55

Ta có $35 \geq 55:2$ nên tổ chứa trung vị là tổ thứ 3. Áp dụng công thức (3.12), ta có

$$M_e = 3,1 + 0,2\frac{55:2-15}{20} = 3,225.$$

3.2.3.2. Các đặc trưng đo lường khuynh hướng phân tán

a. Khoảng biến thiên: là hiệu số giữa lượng biến lớn nhất và lượng biến nhỏ nhất của tiêu thức nghiên cứu.

$$R = X_{max} - X_{min} \tag{3.13}$$

b. Phương sai hiệu chỉnh: là số bình quân số học của bình phương các độ chênh lệch giữa các lượng biến với số bình quân số học của các biến đó

Phương sai có hiệu chỉnh đơn giản: với mẫu khảo sát : $X_1, X_2, ..., X_n$

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \tag{3.14}$$

Phương sai có hiệu chỉnh có trọng số : với số liệu mẫu có lặp lại và ta ghi nhận các tần số xuất hiện số liệu như sau

X	X_1	X_2	...	X_k
Tần số	f_1	f_2	...	f_k

thì cỡ mẫu là $n = f_1 + f_2 + \cdots + f_k$ và công thức tính phương sai có hiệu chỉnh trở thành

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i\left(X_i - \overline{X}\right)^2. \tag{3.15}$$

Lưu ý: Nếu số liệu được cho dưới dạng có khoảng cách tổ thì ta lấy điểm giữa đại diện.

c. Độ lệch chuẩn có hiệu chỉnh: là căn bậc 2 của phương sai, hay nói cách khác là số bình quân toàn phương của các độ lệch giữa các lượng biến với số bình quân số học của chúng.

$$S_X = \sqrt{S_X^2}. \tag{3.16}$$

Ví dụ 3.18. Với dữ liệu ví dụ 3.12:

Phương sai có hiệu chỉnh: $S_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{k} f_i\left(X_i - \overline{X}\right)^2 = 4,0385.$

Độ lệch chuẩn có hiệu chỉnh: $S_X = 2,01.$

d. Tứ phân vị (Quartiles): là chia dãy số thành 4 phần, mỗi phần có số đơn vị bằng nhau.

Cách xác định tứ phân vị

+) Tài liệu phân tổ không có khoảng cách tổ

Q_1 : Tứ phân vị thứ 1 là giá trị đứng ở vị trí (n+1)/4, là phân vị thứ 25.

Q_2 : Tứ phân vị thứ 2 chính là trung vị Me đứng ở vị trí (n+1)/2, là phân vị thứ 50.

Q_3 : Tứ phân vị thứ 3 là giá trị đứng ở vị trí 3(n+1)/4, là phân vị thứ 75.

Ví dụ 3.19. Giả sử ta có dãy số liệu như sau

5	5	6	7	8	8	9

Ta có $Q_1 = 5$; $Q_2 = Me = 7$; $Q_3 = 8.$

Nếu (n+1) không chia hết cho 4 thì tứ phân vị được xác định bằng cách cộng thêm vào như ví dụ sau:

Ví dụ 3.20. Xét tiền lương của 8 công nhân như sau

3600	3800	4000	4200	4400	5000	5400	5600

Ta có $\dfrac{n+1}{4} = \dfrac{9}{4} = 2,25$ nên các xác định tứ phân vị như sau:

Phân vị thứ 1: là giá trị nằm giữa quan sát thứ 2 và thứ 3 theo tọa độ lệch 0,25 gần phía quan sát thứ 2 nên cách xác định Q_1 như sau:

$$Q_1 = 3800 + 0,25(4000 - 3800) = 3850$$

Phân vị thứ 2: là giá trị nằm giữa quan sát thứ 4 và thứ 5 nên cách xác định Q_2 như sau

$$Q_2 = Me = 0,5(4200 + 4400) = 4300$$

Phân vị thứ 3: là giá trị nằm giữa quan sát thứ 6 và thứ 7 theo tọa độ lệch 0,75 gần phía quan sát thứ 6 nên cách xác định Q_3 như sau

$$Q_3 = 5000 + 0,75(5400 - 5000) = 5300$$

+) Tài liệu phân tổ có khoảng cách tổ

Tổ chứa phân vị thứ i có tần số tích lũy $\geq \frac{n+1}{4}i$

Tứ phân vị thứ 1:

$$Q_1 = X_{Q_{1(\min)}} + d_{Q_1}\frac{\frac{1}{4}\sum f_i - S_{Q_{1-1}}}{f_{Q_1}} \quad (3.17)$$

Tứ phân vị thứ 3 :

$$Q_3 = X_{Q_{3(\min)}} + d_{Q_3}\frac{\frac{3}{4}\sum f_i - S_{Q_{3-1}}}{f_{Q_3}} \quad (3.18)$$

Ví dụ 3.21. Khảo sát doanh thu của các cửa hàng ta có bảng số liệu sau:

Doanh thu (tr.đ)	Cửa hàng (f_i)	Tần số tích lũy (S_i)
200-400	8	8
400-500	12	20
500-600	25	45
600-800	25	70
800-1000	9	79
Tổng	79	

Tính phân vị thứ 1 và phân vị thứ 3 của doanh thu.

Giải

Xác định phân vị thứ 1: Vì $\frac{n+1}{4} = \frac{80}{4} = 20$. Ta có $S_2 = 20 \geq 20$. Vậy tổ chứa phân vị thứ 1 là tổ 2. Áp dụng công thức (3.17), ta có

$$Q_1 = 400 + 100\frac{79:4-8}{12} = 497,92.$$

Xác định phân vị thứ 3: Vì $\frac{n+1}{4}3 = \frac{80}{4}3 = 60$. Ta có $S_4 = 70 \geq 60$. Vậy tổ chứa phân vị thứ 3 là tổ 4. Áp dụng công thức (3.18), ta có

$$Q_3 = 600 + 200\frac{(3:4)79-45}{25} = 714.$$

3.3. Ước lượng tham số

Giả sử biến ngẫu nhiên gốc X có quy luật phân phối xác suất đã biết nhưng chưa biết tham số θ nào đó của nó. Phải ước lượng (xác định 1 cách gần đúng) giá trị của θ.

Có hai phương pháp là phương pháp ước lượng điểm và phương pháp ước lượng bằng khoảng tin cậy.

3.3.1. Phương pháp ước lượng điểm

Phương pháp ước lượng điểm dùng một thống kê $\hat{\theta}$ nào đó của mẫu ngẫu nhiên để thay thế cho tham số θ chưa biết của tổng thể. Có nhiều cách chọn thống kê $\hat{\theta}$ khác nhau tạo nên những phương pháp ước lượng điểm khác nhau.

3.3.1.1. Phương pháp hàm ước lượng (phương pháp mômen)

a. Khái niệm: Giả sử cần ước lượng tham số θ của biến ngẫu nhiên gốc X. Từ tổng thể lập mẫu ngẫu nhiên kích thước $n: W=\left(X_1,X_2,...X_n\right)$. Chọn lập thống kê $\hat{\theta}=\left(X_1,X_2,...X_n\right)$ mà thực chất là một thống kê đặc trưng mẫu tương ứng với tham số θ cần ước lượng. Chẳng hạn, để ước lượng kỳ vọng toán μ của biến ngẫu nhiên gốc thì chọn thống kê trung bình mẫu $\overline{X}$, để ước lượng phương sai σ^2 của biến ngẫu nhiên gốc thì chọn thống kê S_X^2,…Nếu lập một mẫu cụ thể và tính được giá trị $\hat{\theta}=f\left(X_1,X_2,...,X_n\right)$ của thống kê $\hat{\theta}$ trên mẫu cụ thể đó thì nó là ước lượng của θ.

Thống kê $\hat{\theta}$ được gọi là *hàm ước lượng* của θ.

b. Các tiêu chuẩn lựa chọn hàm ước lượng

Ước lượng không chệch: Thống kê $\hat{\theta}$ của mẫu được gọi là ước lượng không chệch của tham số θ của biến ngẫu nhiên gốc X nếu $E\left(\hat{\theta}\right)=\theta$. Ngược lại nếu $E\left(\hat{\theta}\right)\neq\theta$. thì $\hat{\theta}$ được gọi là ước lượng chệch của θ.

Nhận xét:

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch của trung bình tổng thể μ của biến ngẫu nhiên gốc, nghĩa là $E\left(\overline{X}\right)=\mu$.

- Tỷ lệ mẫu f là ước lượng không chệch của tỷ lệ tổng thể p của biến ngẫu nhiên gốc, nghĩa là $E\left(f\right)=p$.

- Phương sai mẫu S_X^2 là ước lượng không chệch của phương sai tổng thể σ^2.

Ước lượng hiệu quả: Thống kê $\hat{\theta}$ của mẫu được gọi là ước lượng hiệu quả nhất của tham số θ của biến ngẫu nhiên gốc X nếu nó là ước lượng không chệch và có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng mẫu đó.

Khi hai ước lượng $\hat{\theta}_1$ và $\hat{\theta}_2$ nào đó đều là các ước lượng không chệch của θ song không phải là ước lượng hiệu quả nhất thì có thể so sánh phương sai của hai ước lượng đó để tìm ra ước lượng hiệu quả hơn. Giả sử $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$, lúc đó độ hiệu quả của $\hat{\theta}_1$ so với $\hat{\theta}_2$ được xác định bằng biểu thức:

$$EF = \frac{\text{var}(\hat{\theta}_2)}{\text{var}(\hat{\theta}_1)} \tag{3.19}$$

Ước lượng vững: Thống kê $\hat{\theta}$ của mẫu được gọi là ước lượng vững của tham số θ của biến ngẫu nhiên gốc X nếu $\hat{\theta}$ tiến về θ khi mẫu lớn.

Ước lượng đủ: Một ước lượng $\hat{\theta}$ được gọi là ước lượng đủ nếu nó chứa đựng toàn bộ các thông tin trong mẫu về tham số θ của ước lượng.

c. Một vài kết luận của phương pháp hàm ước lượng

Dùng những tiêu chuẩn trên để đánh giá các thống kê đặc trưng mẫu khác nhau cho phép ta lựa chọn được những thống kê tốt nhất, tức là ước lượng một cách chính xác nhất các tham số đặc trưng của tổng thể. Có thể đưa ra một số kết luận chung sau đây:

- Vì trung bình mẫu $\overline{X}$ là ước lượng không chệch, hiệu quả nhất và vững của trung bình tổng thể μ và đồng thời là ước lượng tuyến tính không chệch tốt nhất, do đó nếu chưa biết μ có thể dùng $\overline{X}$ để ước lượng nó.

- Vì tần suất mẫu f là ước lượng không chệch, hiệu quả nhất và vững của tần suất tổng thể p và đồng thời là ước lượng tuyến tính không chệch tốt nhất, do đó nếu chưa biết p có thể dùng f để ước lượng nó.

- Vì phương sai S_X^2 là các ước lượng không chệch của phương sai tổng thể σ^2, do đó nếu chưa biết σ^2 có thể dùng S_X^2.

3.3.1.2. Phương pháp ước lượng hợp lý tối đa

Giả sử đã biết quy luật phân phối xác suất của biến ngẫu nhiên gốc X dưới dạng hàm mật độ $f(x,\theta)$ hoặc biểu thức xác suất nếu X là biến ngẫu nhiên rời rạc. Cần phải ước lượng tham số θ nào đó của X.

Lập mẫu ngẫu nhiên kích thước n: $W=(X_1,X_2,...X_n)$ và xây dựng hàm hợp lý tại một giá trị cụ thể của mẫu: $L(X_1,X_2,...,X_n,\theta)=f(X_1,\theta)\cdot f(X_2,\theta)\cdots f(X_n,\theta)$. Giá trị của thống kê θ tại điểm đó: $\hat{\theta}=f(X_1,X_2,...,X_n)$ được gọi là ước lượng hợp lý tối đa của θ nếu ứng với giá trị này hàm hợp lý đạt cực đại.

Cách tìm giá trị của θ để hàm hợp lý đạt cực đại:

- Tìm đạo hàm bậc nhất của $\ln\left[L(X,\theta)\right]$ theo θ.

- Giải phương trình: $\dfrac{d\ln\left[L(X,\theta)\right]}{d\theta}=0,$ giả sử có nghiệm $\theta=\hat{\theta}=f(X_1,X_2,...,X_n)$.

- Tìm đạo hàm bậc hai $\dfrac{d^2\ln\left[L(X,\theta)\right]}{d\theta^2}$, nếu tại điểm $\theta=\hat{\theta}$ đạo hàm bậc hai âm thì tại điểm này hàm lnL đạt cực đại, do đó $\hat{\theta}=f(X_1,X_2,...,X_n)$ là ước lượng điểm hợp lý tối đa cần tìm của θ.

3.3.2. Phương pháp ước lượng bằng khoảng tin cậy

Trong phần này, ta chỉ đề cập đến ước lượng trung bình và phương sai trong phân phối chuẩn $N(\mu,\sigma^2)$ và ước lượng tỷ lệ trong phân phối Bernoulli $B(1;p)$.

Cụ thể, để ước lượng tham số θ của một tổng thể có phân phối chuẩn hay có phân phối Bernoulli, ta lập mẫu ngẫu nhiên $X_1,X_2,...,X_n$ và xác định một thống kê $G=f(X_1,X_2,...,X_n,\theta)$ sao cho mặc dù θ chưa được xác định nhưng quy luật phân phối của G vẫn được xác định. Từ phân phối của G và với mức xác suất α cho trước, ta tìm được khoảng tin cậy $[a,b]$ sao cho

$$P(a\le G\le b)=1-\alpha \tag{3.20}$$

và từ các bất đẳng thức $a\le f(X_1,X_2,...,X_n,\theta)\le b,$ ta suy ra ước lượng khoảng cho θ.

3.3.3. Bài toán ước lượng khoảng tin cậy cho giá trị trung bình

3.3.3.1. Ước lượng μ trong phân phối $N(\mu,\sigma_0^2)$, (σ_0^2 biết)

Để ước lượng trung bình μ trong phân phối chuẩn có phương sai đã biết σ_0^2, ta dùng $X_1,X_2,...,X_n$ là mẫu độc lập, $X_i\sim N(\mu,\sigma_0^2)$ để ước lượng μ. Ta có, do định lý Lindeberg-Lévy,

$$\overline{X} \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$$

nên

$$Z = \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{\sigma_0} \sim N(0,1).$$

Với độ tin cậy bằng $(1-\alpha)$ cho trước tìm được cặp giá trị α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$, từ đó tìm được hai giá trị tới hạn tương ứng của phân phối chuẩn hóa là $a = z_{1-\alpha_1}$ và $b = z_{\alpha_2}$ thoả mãn: $P(Z < a) = 1 - \alpha_1$ và $P(Z > b) = \alpha_2$.

$$\Leftrightarrow P(a \le Z \le b) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha.$$

$$\Leftrightarrow P(-a \le Z \le b) = 1 - \alpha.$$

Thay biểu thức của Z vào và biến đổi ta được:

$$P\left(\overline{X} - \frac{b\sigma_0}{\sqrt{n}} \le \mu \le \overline{X} + \frac{a\sigma_0}{\sqrt{n}}\right) = 1 - \alpha.$$

Với độ tin cậy bằng $(1-\alpha)$ tham số μ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng

$$\left[\overline{X} - \frac{b\sigma_0}{\sqrt{n}} \le \mu \le \overline{X} + \frac{a\sigma_0}{\sqrt{n}}\right]. \tag{3.21}$$

Biểu thức (3.21) mới chỉ cho ta khoảng tin cậy tổng quát. Với độ tin cậy $(1-\alpha)$ cho trước, có thể tìm được vô số cặp giá trị α_1 và α_2 thỏa mãn điều kiện: $\alpha_1 + \alpha_2 = \alpha$, từ đó sẽ có vô số khoảng tin cậy tương ứng. Trong thực tế người ta thường dùng các trường hợp đặc biệt sau:

Khoảng tin cậy đối xứng: Chọn $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$, ta có giá trị tới hạn $C = z_{\alpha/2}$, khoảng tin cậy của μ là:

$$\left[\overline{X} - \frac{C\sigma_0}{\sqrt{n}}; \overline{X} + \frac{C\sigma_0}{\sqrt{n}}\right] \tag{3.22}$$

Với sai số ước lượng : $\varepsilon = \dfrac{C\sigma_0}{\sqrt{n}}.$

ε được gọi là *sai số* hay *độ chính xác* của ước lượng. Nó phản ánh mức độ sai lệch của trung bình mẫu so với trung bình tổng thể với xác suất $(1-\alpha)$ cho trước.

Khoảng tin cậy phía phải (ước lượng tối thiểu): Chọn $\alpha_1 = 0$ và $\alpha_2 = \alpha$ thì $z_{\alpha_1} = z_0 = +\infty$, ta có giá trị tới hạn $C = z_\alpha$, khoảng tin cậy của μ là:

$$\left[\overline{X} - \frac{C\sigma_0}{\sqrt{n}}; +\infty\right) \tag{3.23}$$

Khoảng tin cậy phía trái (ước lượng tối đa): Chọn $\alpha_1 = \alpha$ và $\alpha_2 = 0$ thì $z_{\alpha_2} = z_0 = +\infty$, ta có giá trị tới hạn $C = z_\alpha$, khoảng tin cậy của μ là:

$$\left(-\infty; \overline{X} + \frac{C\sigma_0}{\sqrt{n}}\right] \tag{3.24}$$

Với cùng độ tin cậy $(1-\alpha)$, khoảng tin cậy nào ngắn hơn sẽ tốt hơn. Trong trường hợp này độ dài khoảng tin cậy I sẽ là ngắn nhất khi khoảng tin cậy là đối xứng.

Khi đó:

$$I = 2\varepsilon = 2\frac{C\sigma_0}{\sqrt{n}} \tag{3.25}$$

Ví dụ 3.22. Trọng lượng một loại sản phẩm là biến ngẫu nhiên phân phối chuẩn với độ lệch chuẩn là 1 gam. Cân thử 25 sản phẩm loại này ta thu được kết quả sau:

Trọng lượng (gam)	18	19	20	21
Số sản phẩm	3	5	15	2

Với độ tin cậy 95%, hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình của loại sản phẩm nói trên.

Giải

Gọi X là trọng lượng sản phẩm, $X \sim N(\mu, 1)$. Đây là bài toán ước lượng tham số μ của biến ngẫu nhiên phân phối chuẩn bằng khoảng tin cậy đối xứng khi đã biết σ_0^2

Ta có: $n = 25$, $\overline{X} = 19,64$; $\sigma_0 = 1$

Với độ tin cậy: $\gamma = 0,95$ suy ra mức ý nghĩa $\alpha = 1 - \gamma = 1 - 0,95 = 0,05$

Ta có $\phi_0(C) = \frac{1-\alpha}{2} = \frac{1-0,05}{2} = 0,475 \to C = 1,96$

Vậy khoảng tin cậy đối xứng của trọng lượng trung bình là:

$$\mu \in \left[\overline{X} - \frac{C\sigma_0}{\sqrt{n}}; \overline{X} + \frac{C\sigma_0}{\sqrt{n}}\right] = [19,248; 20,032].$$

Kết luận: Với độ tin cậy 95%, từ mẫu cụ thể đã cho, trọng lượng trung bình của sản phẩm nói trên nằm trong khoảng [19,248; 20,032].

3.3.3.2. Ước lượng μ trong phân phối $N(\mu, \sigma^2)$, (σ^2 chưa biết)

Để ước lượng trung bình μ trong phân phối chuẩn có phương sai σ^2 chưa biết, ta cũng dùng $X_1, X_2, ..., X_n$ là mẫu độc lập, $X_i \sim N(\mu, \sigma^2)$ để ước lượng μ. Ta có, do định lý Lindeberg - Lévy,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

nên

$$Z = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1).$$

Ta không thể dùng biến số này để ước lượng μ vì trong biểu thức còn chứa tham số σ chưa biết. Do đó, ta dùng phương sai mẫu S_X^2 làm ước lượng điểm thay thế σ^2, với lưu ý rằng, cũng do định lý Lindeberg-Lévy,

$$Y = \frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1).$$

Do Y, Z độc lập nên bằng cách đặt

$$T = \frac{Z}{\sqrt{\frac{Y}{n-1}}} = \frac{(\overline{X} - \mu)\sqrt{n}}{S_X}$$

thì T có phân phối $St(n-1)$.

Với độ tin cậy bằng $(1-\alpha)$ cho trước tìm được cặp giá trị α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$, từ đó tìm được hai giá trị tới hạn Student tương ứng là $a = t_{1-\alpha_1}(n-1)$ và $b = t_{\alpha_2}(n-1)$ thoả mãn: $P(T > a) = 1 - \alpha_1$ và $P(T > b) = \alpha_2$.

$$\Leftrightarrow P(a \le T \le b) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha.$$

$$\Leftrightarrow P(-a \le T \le b) = 1 - \alpha.$$

Thay biểu thức của T vào và biến đổi ta được:

$$P\left(\overline{X} - \frac{bS_X}{\sqrt{n}} \le T \le \overline{X} + \frac{aS_X}{\sqrt{n}}\right) = 1 - \alpha. \tag{3.26}$$

Với độ tin cậy bằng $(1-\alpha)$ tham số μ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng

$$\left[\overline{X}-\frac{bS_X}{\sqrt{n}};\overline{X}+\frac{aS_X}{\sqrt{n}}\right] \tag{3.27}$$

Biểu thức (3.27) mới chỉ cho ta khoảng tin cậy tổng quát. Với độ tin cậy $(1-\alpha)$ cho trước, có thể tìm được vô số cặp giá trị α_1 và α_2 thỏa mãn điều kiện: $\alpha_1+\alpha_2=\alpha$, từ đó sẽ có vô số khoảng tin cậy tương ứng. Trong thực tế người ta thường dùng các trường hợp đặc biệt sau:

Khoảng tin cậy đối xứng: Chọn $\alpha_1=\alpha_2=\frac{\alpha}{2}$, ta có giá trị tới hạn $C=t_{\alpha/2}(n-1)$, khoảng tin cậy của μ là:

$$\left[\overline{X}-\frac{CS_X}{\sqrt{n}};\overline{X}+\frac{CS_X}{\sqrt{n}}\right]. \tag{3.28}$$

Với sai số ước lượng: $\varepsilon=\frac{CS_X}{\sqrt{n}}$.

ε được gọi là *sai số* hay *độ chính xác* của ước lượng. Nó phản ánh mức độ sai lệch của trung bình mẫu so với trung bình tổng thể với xác suất $(1-\alpha)$ cho trước.

Khoảng tin cậy phía phải (ước lượng tối thiểu): Chọn $\alpha_1=0$ và $\alpha_2=\alpha$, ta có giá trị tới hạn $C=t_\alpha(n-1)$, khoảng tin cậy của μ là:

$$\left[\overline{X}-\frac{CS_X}{\sqrt{n}};+\infty\right) \tag{3.29}$$

Khoảng tin cậy phía trái (ước lượng tối đa): Chọn $\alpha_1=\alpha$ và $\alpha_2=0$, ta có giá trị tới hạn $C=t_\alpha(n-1)$, khoảng tin cậy của μ là:

$$\left(-\infty;\overline{X}+\frac{CS_X}{\sqrt{n}}\right] \tag{3.30}$$

Trong trường hợp này độ dài khoảng tin cậy I cũng là ngắn nhất khi khoảng tin cậy là đối xứng, khi đó:

$$I=2\varepsilon=2\frac{CS_X}{\sqrt{n}} \tag{3.31}$$

Ví dụ 3.23. Năng suất một loại cây trồng là biến ngẫu nhiên có phân phối chuẩn. Thu hoạch tại một số điểm được kết quả sau:

Năng suất (tấn/ha)	30	31	32	33	34	35
Số điểm	4	5	6	3	4	3

Với độ tin cậy 95%, hãy ước lượng năng suất trung bình bằng khoảng tin cậy đối xứng.

Giải

Gọi X là năng suất loại cây trồng

Từ mẫu đã cho, ta tính được: $n = 25$; $\overline{X} = 32,28$; $S_X = 1,646$

Với độ tin cậy: $\gamma = 0,95$ suy ra mức ý nghĩa $\alpha = 1 - \gamma = 1 - 0,95 = 0,05$. Suy ra giá trị tới hạn $C = t_{0,025}(24) = 2,064$

Vậy khoảng tin cậy đối xứng của năng suất trung bình có dạng:

$$\mu \in \left[\overline{X} - \frac{CS_X}{\sqrt{n}}; \overline{X} + \frac{CS_X}{\sqrt{n}}\right] = [31,6; 32,96].$$

Kết luận: Với độ tin cậy 95%, từ mẫu cụ thể đã cho, năng suất trung bình của loại cây trồng nói trên nằm trong khoảng [31,6; 32,96].

Ví dụ 3.24. Để xác định kích thước trung bình của chi tiết do một máy sản xuất người ta lấy ngẫu nhiên 200 chi tiết để đo kích thước và thu được bảng số liệu sau:

Kích thước chi tiết (cm)	Số chi tiết tương ứng
54,795 – 54,805	6
54,805 – 54,815	14
54,815 – 54,825	33
54,825 – 54,835	47
54,835 – 54,845	45
54,845 – 54,855	33
54,855 – 54,865	15
54,865 – 54,875	7

Giả thiết kích thước chi tiết là biến ngẫu nhiên phân phối chuẩn.

a) Với độ tin cậy 95% hãy ước lượng bằng khoảng tin cậy đối xứng kích thước trung bình của chi tiết do máy đó sản xuất.

b) Với độ tin cậy 95% hãy ước lượng bằng khoảng tin cậy tối thiểu và tối đa kích thước trung bình của chi tiết do máy đó sản xuất.

Giải

Gọi X là kích thước chi tiết do máy đó sản xuất.

Ta có $n = 200;\ \overline{X} = 54,8353;\ S_X = 0,0164$

a) Đây là bài toán ước lượng tham số μ của biến ngẫu nhiên phân phối chuẩn bằng khoảng tin cậy đối xứng trong trường hợp chưa biết σ^2.

Với độ tin cậy: $\gamma = 0,95$ suy ra mức ý nghĩa $\alpha = 1-\gamma = 1-0,95 = 0,05.$ Suy ra giá trị tới hạn $C = t_{0,025}(199) = 1,96$

Khoảng tin cậy đối xứng kích thước giá trị trung bình là:

$$\mu \in \left[\overline{X} - \frac{CS_X}{\sqrt{n}};\ \overline{X} + \frac{CS_X}{\sqrt{n}}\right] = [54,83294; 54,83752].$$

b) Khoảng ước lượng bằng khoảng tin cậy tối thiểu và tối đa kích thước trung bình của chi tiết do máy đó sản xuất

Với độ tin cậy: $\gamma = 0,95$ suy ra mức ý nghĩa $\alpha = 1-\gamma = 1-0,95 = 0,05.$ Suy ra giá trị tới hạn $C = t_{0,05}(199) = 1,64$

Khoảng ước lượng bằng khoảng tin cậy tối thiểu kích thước giá trị trung bình là:

$$\mu \in \left[\overline{X} - \frac{CS_X}{\sqrt{n}};\ +\infty\right) = [54,8334;\ +\infty).$$

Khoảng ước lượng bằng khoảng tin cậy tối đa kích thước giá trị trung bình là:

$$\mu \in \left(-\infty;\ \overline{X} + \frac{CS_X}{\sqrt{n}}\right] = (-\infty;\ 54,8372].$$

3.3.4. Bài toán ước lượng khoảng tin cậy cho phương sai

3.3.4.1. Ước lượng σ^2 trong phân phối $N(\mu_0, \sigma^2)$, (μ_0 biết).

Để ước lượng phương sai σ^2 trong phân phối chuẩn có trung bình μ_0 biết, ta cũng dùng mẫu độc lập $X_1, X_2, ...X_n \sim N(\mu_0, \sigma^2)$ để ước lượng σ^2. Trước hết, lưu ý rằng

$$X_i \sim N(\mu_0, \sigma^2) \Rightarrow \frac{X_i - \mu_0}{\sigma} \sim N(0,1)$$

và

$$\left(\frac{X_i - \mu_0}{\sigma}\right)^2 \sim \chi^2(1).$$

Chọn thống kê:

$$Y = \sum_{i=1}^{n} \frac{(X_i - \mu_0)^2}{\sigma^2} \sim \chi^2(n)$$

Với độ tin cậy $1-\alpha$ cho trước, ta tìm được α_1 và α_2 sao cho $\alpha_1+\alpha_2=\alpha$. Từ đó tìm được hai giá trị tới hạn chi bình phương $a=\chi^2_{1-\alpha_1}(n)$ và $b=\chi^2_{\alpha_2}(n)$ sao cho:

$$P(a \le Y \le b)=1-\alpha$$

$$\Leftrightarrow P\left[a \le \sum_{i=1}^{n}\frac{(X_i-\mu_0)^2}{\sigma^2} \le b\right]=1-\alpha$$

$$\Leftrightarrow P\left[\frac{1}{b}\sum_{i=1}^{n}(X_i-\mu_0)^2 \le \sigma^2 \le \frac{1}{a}\sum_{i=1}^{n}(X_i-\mu_0)^2\right]=1-\alpha.$$

Như vậy, với độ tin cậy bằng $(1-\alpha)$ tham số σ^2 của biến ngẫu nhiên gốc X sẽ nằm trong khoảng

$$\left[\frac{1}{b}\sum_{i=1}^{n}(X_i-\mu_0)^2;\frac{1}{a}\sum_{i=1}^{n}(X_i-\mu_0)^2\right] \tag{3.32}$$

Biểu thức (3.32) mới chỉ cho ta khoảng tin cậy tổng quát. Với độ tin cậy $(1-\alpha)$ cho trước, có thể tìm được vô số cặp giá trị α_1 và α_2 thỏa mãn điều kiện: $\alpha_1+\alpha_2=\alpha$, từ đó sẽ có vô số khoảng tin cậy tương ứng. Trong thực tế người ta thường dùng các trường hợp đặc biệt sau:

Khoảng tin cậy hai phía: Chọn $\alpha_1=\alpha_2=\frac{\alpha}{2}$ suy ra giá trị tới hạn $a=\chi^2_{1-\frac{\alpha}{2}}(n)$ và $b=\chi^2_{\frac{\alpha}{2}}(n)$, khoảng tin cậy của σ^2 là:

$$\left[\frac{1}{b}\sum_{i=1}^{n}(X_i-\mu_0)^2;\frac{1}{a}\sum_{i=1}^{n}(X_i-\mu_0)^2\right] \tag{3.33}$$

Khoảng tin cậy phía phải (ước lượng tối thiểu): Chọn $\alpha_1=0$ và $\alpha_2=\alpha$, khoảng tin cậy của σ^2 là:

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{\alpha}(n)};+\infty\right) \tag{3.34}$$

Khoảng tin cậy phía trái (ước lượng tối đa): Chọn $\alpha_1=\alpha$ và $\alpha_2=0$, khoảng tin cậy của σ^2 là:

$$\left(-\infty\ ;\ \frac{\sum_{i=1}^{n}\left(X_i-\mu_0\right)^2}{\chi^2_{1-\alpha}(n)}\right). \tag{3.35}$$

3.3.4.2. Ước lượng σ^2 trong phân phối $N\left(\mu,\sigma^2\right)$, ($\mu$ chưa biết).

Để ước lượng phương sai σ^2 trong phân phối chuẩn có trung bình μ chưa biết, ta cũng dùng mẫu độc lập $X_1, X_2, ...X_n \sim N\left(\mu,\sigma^2\right)$ để ước lượng σ^2.

Ta có, do định lý Lindeberg-Lévy

$$Y=\frac{(n-1)S_X^2}{\sigma^2}\sim\chi^2(n-1).$$

Với độ tin cậy $1-\alpha$ cho trước, ta tìm được α_1 và α_2 sao cho $\alpha_1+\alpha_2=\alpha$. Từ đó tìm được hai giá trị tới hạn chi bình phương $a=\chi^2_{1-\alpha_1}(n-1)$ và $b=\chi^2_{\alpha_2}(n-1)$ sao cho:

$$\begin{aligned}
&P(a\le Y\le b)=1-\alpha\\
&\Leftrightarrow P\left[a\le\frac{(n-1)S_X^2}{\sigma^2}\le b\right]=1-\alpha\\
&\Leftrightarrow P\left[\frac{(n-1)S_X^2}{b}\le\sigma^2\le\frac{(n-1)S_X^2}{a}\right]=1-\alpha.
\end{aligned}$$

Như vậy, với độ tin cậy bằng $(1-\alpha)$ tham số σ^2 của biến ngẫu nhiên gốc X sẽ nằm trong khoảng

$$\left[\frac{(n-1)S_X^2}{b};\frac{(n-1)S_X^2}{a}\right] \tag{3.36}$$

Biểu thức (3.36) mới chỉ cho ta khoảng tin cậy tổng quát. Với độ tin cậy $(1-\alpha)$ cho trước, có thể tìm được vô số cặp giá trị α_1 và α_2 thỏa mãn điều kiện: $\alpha_1+\alpha_2=\alpha$, từ đó sẽ có vô số khoảng tin cậy tương ứng. Trong thực tế người ta thường dùng các trường hợp đặc biệt sau:

Khoảng tin cậy hai phía: Chọn $\alpha_1=\alpha_2=\frac{\alpha}{2}$, suy ra hai giá trị tới hạn $a=\chi^2_{1-\frac{\alpha}{2}}(n-1)$ và $b=\chi^2_{\frac{\alpha}{2}}(n-1)$, khoảng tin cậy của σ^2 là:

$$\left[\frac{(n-1)S_X^2}{b};\frac{(n-1)S_X^2}{a}\right] \tag{3.37}$$

Khoảng tin cậy phía phải (ước lượng tối thiểu): Chọn $\alpha_1=0$ và $\alpha_2=\alpha,$ khoảng tin cậy của σ^2 là:

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{\alpha}(n-1)}\ ;+\infty\right) \tag{3.38}$$

Khoảng tin cậy phía trái (ước lượng tối đa): Chọn $\alpha_1=\alpha$ và $\alpha_2=0,$ khoảng tin cậy của σ^2 là:

$$\left(-\infty\ ;\ \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{1-\alpha}(n-1)}\right) \tag{3.39}$$

Ví dụ 3.25. Khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau

X(cm)	11 - 15	15 - 19	19 – 23	23 - 27	27 – 31	31 - 35	35 - 39
Số sản phẩm	8	9	20	16	16	13	18

Giả sử X có phân phối chuẩn. Hãy ước lượng phương sai của X với độ tin cậy 95% trong các trương hợp sau:

a) Biết trung bình tổng thể của X là 25 cm.

b) Chưa biết giá trị trung bình của X.

Giải

a) Để ước lượng phương sai tổng thể khi biết trung bình tổng thể $\mu_0=25,$ ta dùng thống kê

$$Y=\sum_{i=1}^{n}\frac{(X_i-\mu_0)^2}{\sigma^2}\sim\chi^2(n).$$

Với độ tin cậy $\gamma=95\%,$ ta chọn khoảng tin cậy cho Y : $[a,b]$

Với $a=\chi^2_{0,975}(100)=77,929;\ a=\chi^2_{0,025}(100)=129,561$

Ta lập bảng

$X_i-\mu_0$	-12	-8	-4	0	4	8	12

n	8	9	20	16	16	13	18

Từ đó ta tìm được cỡ mẫu $n = 100; \sum_{i=1}^{k} n_i (X_i - \mu_0)^2 = 5728$

Khoảng ước lượng của phương sai tổng thể σ^2

$$\sigma^2 \in \left[\frac{1}{b}\sum_{i=1}^{n}(X_i - \mu_0)^2; \frac{1}{a}\sum_{i=1}^{n}(X_i - \mu_0)^2\right] = [44,211; 73,503].$$

b) Để ước lượng phương sai tổng thể khi chưa biết trung bình của tổng thể, ta dùng thống kê

$$Y = \frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1),$$

Với độ tin cậy $\gamma = 95\%,$ ta chọn khoảng tin cậy cho Y : $[a, b]$

Với $a = \chi^2_{0,975}(99) = 77,929; a = \chi^2_{0,025}(99) = 129,561$

Khoảng ước lượng của phương sai tổng thể σ^2

$$\sigma^2 \in \left[\frac{(n-1)S_X^2}{b}; \frac{(n-1)S_X^2}{a}\right] = [42,784; 71,130].$$

3.3.5. Bài toán ước lượng khoảng tin cậy cho tỷ lệ

Để ước lượng tỷ lệ p trong phân phối Bernoulli, ta cũng dùng mẫu độc lập $X_1, X_2, ..., X_n \sim B(1; p)$ để ước lượng p. Đặt

$$f = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

thì

$$Z = \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}} \sim N(0,1)$$

Với độ tin cậy $\gamma = 1 - \alpha$ cho trước, từ bảng phân phối Gauss, ta tìm được C sao cho

$$P(-C \le Z \le C) = 1 - \alpha$$

$$\Leftrightarrow P\left(-C \le \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}} \le C\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(f - C\sqrt{\frac{p(1-p)}{n}} \le p \le f + C\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

Biểu thức này chưa cho ta khoảng ước lượng của p vì các đầu mút còn lệ thuộc vào p chưa biết. Do đó, ta thay p bằng ước lượng điểm, $\overline{p} = f$, của nó và ta có

Khoảng ước lượng đối xứng : Với độ tin cậy $\gamma = 1-\alpha$ cho trước, ta tìm được giá trị tới hạn C thỏa mãn $\phi_0(C) = \dfrac{1-\alpha}{2}$, khoảng tin cậy đối xứng của p là:

$$p \in \left[f - C\sqrt{\frac{f(1-f)}{n}} ; f + C\sqrt{\frac{f(1-f)}{n}} \right]. \tag{3.40}$$

Khoảng tin cậy phía phải (ước lượng tối thiểu): Với độ tin cậy $\gamma = 1-\alpha$ cho trước, ta tìm được giá trị tới hạn C thỏa mãn $\phi_0(C) = \dfrac{1-2\alpha}{2}$, khoảng tin cậy tối tiểu của p là:

$$p \in \left[f - C\sqrt{\frac{f(1-f)}{n}} \; ; \; 1 \right] \tag{3.41}$$

Khoảng tin cậy phía trái (ước lượng tối đa): Với độ tin cậy $\gamma = 1-\alpha$ cho trước, ta tìm được giá trị tới hạn C thỏa mãn $\phi_0(C) = \dfrac{1-2\alpha}{2}$, khoảng tin cậy tối đa của p là:

$$p \in \left[0 \; ; \; f + C\sqrt{\frac{f(1-f)}{n}} \right] \tag{3.42}$$

3.4. Bài toán xác định cỡ mẫu

Ta cũng có thể giải quyết bài toán tìm cỡ mẫu n khi muốn ước lượng tỷ lệ p với độ tin cậy γ và sai số ε_0 cho trước như sau

$$C\sqrt{\frac{f(1-f)}{n}} \le \varepsilon_0 \Leftrightarrow n \ge \frac{C^2 f(1-f)}{\varepsilon_0^2} \tag{3.43}$$

Xét rằng hàm số $y = x(1-x)$, hàm số này đạt giá trị lớn nhất là $\dfrac{1}{4}$ tại điểm $x = \dfrac{1}{2}$, cho nên từ sai số ước lượng ta có thể viết

$$C\sqrt{\frac{f(1-f)}{n}} \le C\sqrt{\frac{1}{4n}} \le \varepsilon_0 \Leftrightarrow n \ge \frac{C^2}{4\varepsilon_0^2}$$

Do đó, ta có thể chọn cỡ mẫu

$$n \ge \frac{C^2}{4\varepsilon_0^2}. \tag{3.44}$$

Vậy

- Nếu biết tỷ lệ mẫu f, ta tìm n thỏa mãn $n \geq \frac{C^2 f\left(1-f\right)}{\varepsilon_0^2}$.

- Nếu chưa biết tỷ lệ mẫu f, ta tìm n thỏa mãn $n \geq \frac{C^2}{4\varepsilon_0^2}$.

Ví dụ 3.26. Để ước lượng tỷ lệ bệnh sốt rét ở Đồng Bằng Sông Cửu Long.

1) Nếu muốn sai số ước lượng không quá 2% ở độ tin cậy 95% thì cần quan sát ít nhất bao nhiêu người ?

2) Ta quan sát ngẫu nhiên 200 người, thấy có 24 người mắc bệnh sốt rét.

a) Tìm khoảng ước lượng đối xứng của p với độ tin cậy 95%,

b) Tìm khoảng ước lượng tối đa, tối thiểu của p với độ tin cậy 95%.

c) Nếu muốn sai số ước lượng không quá 2% ở độ tin cậy 95% thì cần quan sát ít nhất mấy người ?

Giải

1) Với độ tin cậy $\gamma = 1 - \alpha = 0,95 \to \alpha = 0,05$.

Ta có $\phi_0\left(C\right) = \frac{1-\alpha}{2} = 0,475 \to C = 1,96$ và với $\varepsilon_0 = 0,02$ nên áp dụng công thức (3.44), ta có

$$n \geq \frac{C^2}{4\varepsilon_0^2} = \frac{(1,96)^2}{4(0,02)^2} = 2401.$$

Vậy cần quan sát ít nhất 2041 người.

2) Ta có $n = 200,\ k = 24,\ \ f = \frac{24}{200} = 0,12$,

a) Khoảng ước lượng đối xứng của p.

Với độ tin cậy $\gamma = 95\% = 0,95$ suy ra mức ý nghĩa $\alpha = 5\% = 0,05$.

Ta có $\phi_0\left(C\right) = \frac{1-\alpha}{2} = 0,475 \to C = 1,96$

Ta có ước lượng của tỷ lệ p là

$$p \in \left[f - C\sqrt{\frac{f\left(1-f\right)}{n}}; f + C\sqrt{\frac{f\left(1-f\right)}{n}} \right] = \left[0,075;\ 0,165\right].$$

b) Khoảng ước lượng tối đa, tối thiểu của p

Với độ tin cậy $\gamma = 95\% = 0,95$ suy ra mức ý nghĩa $\alpha = 5\% = 0,05$.

Ta có $\phi_0(C) = \frac{1-2\alpha}{2} = 0,45 \to C = 1,64$

Khoảng ước lượng tối đa của p

$$p \in \left[0\ ;\ f + C\sqrt{\frac{f(1-f)}{n}}\right] = [0;\ 0,1577]$$

Khoảng ước lượng tối thiểu của p

$$p \in \left[f - C\sqrt{\frac{f(1-f)}{n}};\ 1\right] = [0,0823;\ 1].$$

c) Ta có $\phi_0(C) = \frac{1-\alpha}{2} = 0,475 \to C = 1,96$ và với $\varepsilon_0 = 0,02$ nên áp dụng công thức (3.43), ta có

$$n \geq \frac{C^2 f(1-f)}{\varepsilon_0^2} = 1014,18.$$

Vậy cần quan sát ít nhất 1015 người.

3.5. Tóm tắt chương 3

1. Các đặc trưng đo lường khuynh hướng tập trung

i) Số trung bình mẫu: $\overline{X} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i,\ n = f_1 + f_2 + \cdots + f_k.$

ii) Mốt (Mode): Mốt là lượng biến có tần số lớn nhất.

+) Đối với đại lượng biến có khoảng cách tổ

$$M_0 = x_{M_{0(\min)}} + d_{M_0}\frac{F_{M_0} - F_{M_{0-1}}}{\left(F_{M_0} - F_{M_{0-1}}\right) + \left(F_{M_0} - F_{M_{0+1}}\right)}$$

ii) Số trung vị:

+) Trường hợp số đơn vị tổng thể lẻ ($n = 2m+1$). Số trung vị sẽ là đại lượng ở vị trí thứ ($m+1$): $Me = X_{m+1}$.

+) Trường hợp số đơn vị tổng thể chẵn ($n = 2m$). Số trung vị sẽ là đại lượng giữa vị trí thứ (m) và ($m+1$): $Me = \frac{X_m + X_{m+1}}{2}$.

Đối với dãy số có lượng biến là khoảng cách tổ

$$M_e = X_{M_{e(\min)}} + d_{M_e}\frac{\frac{1}{2}\sum f_i - S_{M_{e-1}}}{f_{M_e}}$$

2. Các đặc trưng đo lường khuynh hướng phân tán

i) Phương sai hiệu chỉnh: $S_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{k} f_i\left(x_i - \overline{x}\right)^2.$

ii) Độ lệch chuẩn có hiệu chỉnh: $S_X = \sqrt{S_X^2}.$

ii) Tứ phân vị (Quartiles): là chia dãy số thành 4 phần, mỗi phần có số đơn vị bằng nhau.

Cách xác định tứ phân vị cho tài liệu phân tổ có khoảng cách tổ

Tổ chứa phân vị thứ i có tần số tích lũy $\geq \dfrac{n+1}{4}i.$

Tứ phân vị thứ 1: $Q_1 = X_{Q_{1(min)}} + d_{Q_1}\dfrac{\frac{1}{4}\sum f_i - S_{Q_{1-1}}}{f_{Q_1}}.$

Tứ phân vị thứ 3 : $Q_3 = X_{Q_{3(min)}} + d_{Q_3}\dfrac{\frac{3}{4}\sum f_i - S_{Q_{3-1}}}{f_{Q_3}}.$

3. Bài toán ước lượng khoảng tin cậy cho giá trị trung bình

i) Ước lượng μ trong phân phối $N\left(\mu,\sigma_0^2\right)$, ($\sigma_0^2$ biết)

+) Với mức ý nghĩa α. Ta có : $\phi_0\left(C\right) = \dfrac{1-\alpha}{2}$

Khoảng tin cậy đối xứng: $\left[\overline{X} - \dfrac{C\sigma_0}{\sqrt{n}}; \overline{X} + \dfrac{C\sigma_0}{\sqrt{n}}\right].$

+) Với mức ý nghĩa α. Ta có : $\phi_0\left(C\right) = \dfrac{1-2\alpha}{2}$

Khoảng tin cậy phía phải: $\left[\overline{X} - \dfrac{C\sigma_0}{\sqrt{n}}; +\infty\right).$

Khoảng tin cậy phía trái: $\left(-\infty; \overline{X} + \dfrac{C\sigma_0}{\sqrt{n}}\right].$

ii) Ước lượng μ trong phân phối $N\left(\mu,\sigma^2\right)$, ($\sigma^2$ chưa biết)

+) Với mức ý nghĩa α. Ta có: $C = t_{\alpha/2}\left(n-1\right)$

Khoảng tin cậy đối xứng: $\left[\overline{X} - \dfrac{CS_X}{\sqrt{n}}; \overline{X} + \dfrac{CS_X}{\sqrt{n}}\right].$

+) Với mức ý nghĩa α. Ta có: $C = t_{\alpha}\left(n-1\right)$

Khoảng tin cậy phía phải: $\left[\overline{X} - \frac{CS_X}{\sqrt{n}}; +\infty\right)$.

Khoảng tin cậy phía trái: $\left(-\infty; \overline{X} + \frac{CS_X}{\sqrt{n}}\right]$.

4. Bài toán ước lượng khoảng tin cậy cho phương sai

i) Ước lượng σ^2 trong phân phối $N\left(\mu_0, \sigma^2\right)$, ($\mu_0$ biết).

+) Với mức ý nghĩa α, ta có $a = \chi^2_{1-\frac{\alpha}{2}}\left(n\right)$ và $b = \chi^2_{\frac{\alpha}{2}}\left(n\right)$

Khoảng tin cậy hai phía: $\left[\frac{1}{b}\sum_{i=1}^{n}\left(X_i - \mu_0\right)^2; \frac{1}{a}\sum_{i=1}^{n}\left(X_i - \mu_0\right)^2\right]$.

+) Với mức ý nghĩa α, ta có $\chi^2_\alpha(n)$

Khoảng tin cậy phía phải: $\left(\frac{\sum_{i=1}^{n}\left(X_i - \mu_0\right)^2}{\chi^2_\alpha(n)}; +\infty\right)$.

Khoảng tin cậy phía trái: $\left(-\infty; \frac{\sum_{i=1}^{n}\left(X_i - \mu_0\right)^2}{\chi^2_{1-\alpha}(n)}\right)$.

ii) Ước lượng σ^2 trong phân phối $N\left(\mu, \sigma^2\right)$, ($\mu$ chưa biết).

+) Với mức ý nghĩa α, ta có $a = \chi^2_{1-\frac{\alpha}{2}}\left(n-1\right)$ và $b = \chi^2_{\frac{\alpha}{2}}\left(n-1\right)$

Khoảng tin cậy hai phía: $\left[\frac{(n-1)S_X^2}{b}; \frac{(n-1)S_X^2}{a}\right]$.

+) Với mức ý nghĩa α, ta có $\chi^2_\alpha(n-1)$

Khoảng tin cậy phía phải: $\left(\frac{\sum_{i=1}^{n}\left(X_i - \mu_0\right)^2}{\chi^2_\alpha(n-1)}; +\infty\right)$.

Khoảng tin cậy phía : $\left(-\infty\ ;\ \dfrac{\sum_{i=1}^{n}\left(X_i-\mu_0\right)^2}{\chi^2_{1-\alpha}(n-1)}\right).$

5. Bài toán ước lượng khoảng tin cậy cho tỷ lệ

i) Với mức ý nghĩa α. Ta có : $\phi_0\left(C\right)=\dfrac{1-\alpha}{2}$

Khoảng ước lượng đối xứng : $p\in\left[f-C\sqrt{\dfrac{f\left(1-f\right)}{n}};f+C\sqrt{\dfrac{f\left(1-f\right)}{n}}\right].$

ii) Với mức ý nghĩa α. Ta có : $\phi_0\left(C\right)=\dfrac{1-2\alpha}{2}$

Khoảng tin cậy phía phải: $p\in\left[f-C\sqrt{\dfrac{f\left(1-f\right)}{n}}\ ;1\right].$

Khoảng tin cậy phía trái: $p\in\left[0\ ;\ f+C\sqrt{\dfrac{f\left(1-f\right)}{n}}\right].$

3.6. Bài tập

Bài số 1. Quan sát thời gian cần thiết để sản xuất một chi tiết máy, ta thu được số liệu cho bảng sau

Khoảng thời gian (phút)	Số lần quan sát
20 – 25	2
25 – 30	14
30 – 35	26
35 – 40	32
40 – 45	14
45 – 50	8
50 – 55	4

Tính trung bình mẫu $\overline{X}$, phương sai mẫu có hiệu chỉnh S_X^2.

Đáp số: $\overline{X}=36{,}6;\ S_X^2=45{,}1414.$

Bài số 2. Có tài liệu phân tổ về năng suất lao động của công nhân một doanh nghiệp trong kỳ nghiên cứu như sau:

Năng suất lao động (Sp/ca)	Số công nhân

20 – 22	10
22 – 24	40
24 – 26	80
26 – 28	50
28 – 30	20

Hãy tính trung bình, phương sai có hiệu chỉnh, mốt, tứ phân vị của năng suất lao động một công nhân.

Đáp số: 35,3; 4,131; 25,143; 24; 25,25; 26,8.

Bài số 3. Có tài liệu dưới đây của một doanh nghiệp

Năng suất lao động (kg)	Số công nhân
110 – 120	10
120 – 130	30
130 – 140	50
140 – 150	60
150 – 160	145
160 – 170	110
170 – 180	80
180 – 190	15

Hãy tính trung bình, phương sai có hiệu chỉnh, mốt, tứ phân vị của năng suất lao động một công nhân.

Đáp số: 155,5; 251,253; 157,083; 145,83; 156,9; 167,3.

Bài số 4. Chiều dài của một loại sản phẩm được xuất khẩu hàng loạt là biến ngẫu nhiên phân phối chuẩn với $\mu = 100\text{mm}; \sigma^2 = 4^2\text{mm}^2$. Kiểm tra ngẫu nhiên 25 sản phẩm. Khả năng chiều dài trung bình của số sản phẩm kiểm tra nằm trong khoảng từ 98mm đến 101mm là bao nhiêu.

Đáp số: 0,8828.

Bài số 5. Đo độ dài của một loại trục xe, ta có kết quả

Nhóm	18,4-18,6	18,6-18,8	18,8 -19	19 -19,2	19,2-19,4	19,4-19,6	19,6-19,8
n_i	1	4	20	41	19	8	4

Hãy ước lượng điểm độ dài trung bình và phương sai của trục xe.

Đáp số: $\overline{X} = 19{,}133;\ S_X^2 = 0{,}539$.

Bài số 5. Đo sức bền chịu lực của 1 loại ống thí nghiệm, người ta thu được bộ số liệu sau

4500	6500	5200	4800	4900	5125	6200	5375

Từ kinh nghiệm nghề nghiệp, người ta cũng biết rằng sức bền đó có phân phối chuẩn với độ lệch chuẩn $\sigma = 300$. Hãy ước lượng sức bền trung bình của loại ống trên, với độ tin cậy 90%.

Đáp số: $\mu \in [5151,0517; 5498,9483]$.

Bài số 7. Trên tập mẫu gồm 100 số liệu, người ta tính được $\overline{X} = 0,1; S_X = 0,014$. Hãy ước lượng giá trị trung bình tổng thể, với độ tin cậy 95%.

Đáp số: $\mu \in [0,0973; 0,1027]$.

Bài số 8. Chọn ngẫu nhiên 36 công nhân của xí nghiệp thì thấy lương trung bình là 380 ngàn đ/tháng. Giả sử lương công nhân tuân theo luật chuẩn với $\sigma = 14$ ngàn đồng. Với độ tin cậy 95%, hãy ước lượng mức lương trung bình của công nhân trong toàn xí nghiệp.

Đáp số: $\mu \in [375,4267; 384,5733]$.

Bài số 9. Điểm trung bình môn toán của 100 thí sinh dự thi vào ĐHKT là 5 với độ lệch chuẩn mẫu đã điều chỉnh $S_X = 2,5$.

a) Ước lượng điểm trung bình môn toán của toàn thể thí sinh với độ tin cậy là 95%.

b) Với sai số 0,25 điểm. Hãy xác định độ tin cậy.

Đáp số: a) $\mu \in [4,51; 5,49]$*; b) 68,26%.*

Bài số 10. 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 đèn để thử nghiệm, thấy mỗi bóng tuổi thọ trung bình là 1000 giờ. Hãy ước lượng tuổi thọ trung bình của bóng đèn xí nghiệp A sản xuất với độ tin cậy là 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 nghiêu bóng.

Đáp số: a) $\mu \in [980,4; 1019,6]$*; b) 86,64%; c) 62.*

Bài số 11. Khối lượng các bao bột mì tại một cửa hàng lương thực theo quy luật chuẩn. Kiểm tra 20 bao, thấy khối lượng trung bình của mỗi bao bột mì là 48kg, và phương sai mẫu có điều chỉnh là $S_X^2 = (0,5\text{kg})^2$.

a) Với độ tin cậy 95%, hãy ước lượng khối lượng trung bình của một bao bột mì thuộc cửa hàng.

b) Với độ chính xác là 0,26kg. Hãy xác định độ tin cậy.

c) Với độ chính xác là 160g và độ tin cậy là 95%, tính cỡ mẫu.

Đáp số: a) $\mu \in [47,766;48,234]$*; b) 97%; c) 43.*

Bài số 12. 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 kho đồ hộp với độ tin cậy 94%.

b) Với sai số cho phép $\varepsilon = 3\%$, hãy xác định độ tin cậy.

Đáp số: a) $p \in [0,051;0,169]$*; b) 66,3%.*

Bài số 13. a) Muốn ước lượng tỷ lệ bệnh sốt xuất huyết ở Tp. Hồ Chí Minh với sai số không quá 3% ở độ tin cậy 95% thì phải quan sát ít nhất bao nhiêu người ?

b) Giả sử quan sát 100 người thấy có 20 người bị bệnh sốt xuất huyết. Hãy ước lượng tỷ lệ bệnh sốt xuất huyết ở Tp. Hồ Chí Minh ở độ tin cậy 97%. Nếu muốn sai số ước lượng không quá 3% ở độ tin cậy 95% thì phải quan sát ít nhất bao nhiêu người ?

Đáp số: a) 1068 người; b) $p \in [0,1132;0,2868]$*; 683 người.*

Bài số 14. Để ước lượng xác suất mắc bệnh gan với độ tin cậy 90% và sai số không vượt quá 2% thì cần phải khám ít nhất bao nhiêu người, biết rằng tỷ lệ mắc bệnh gan thực nghiệm đã cho bằng 0,9.

Đáp số: 606 người.

Bài số 15. Trước bầu cử, người ta phỏng vấn ngẫu nhiên 2000 cử tri thì thấy có 1380 người ủng hộ một ứng cử viên K. Với độ tin cậy 95%, hỏi ứng cử viên đó thu được tối thiểu bao nhiêu phần trăm phiếu bầu ?

Đáp số: 67,3%.

Bài số 16. Gieo thử 400 hạt giống thì thấy có 20 hạt không nảy mầm. Tỷ lệ hạt giống không nảy mầm tối đa là bao nhiêu, với độ tin cậy 95%.

Đáp số: 6,78715%.

Bài số 17. Muốn biết trong hồ có bao nhiêu cá, người ta bắt lên 2000 con, đánh dấu xong lại thả xuống hồ. Sau một thời gian, người ta bắt lên 500 con và thấy có 20 con cá có đánh dấu của lần bắt trước. Dựa vào kết quả đó, hãy ước lượng số cá có trong hồ với độ tin cậy 95%.

Đáp số: $[34978;87642]$.

Bài số 18. Để có thể dự đoán được số lượng chim thường nghỉ tại vườn nhà mình, người chủ bắt 89 con, đem đeo khoen cho chúng rồi thả đi. Sau một thời gian, ông bắt ngẫu nhiên được 120 con và thấy có 7 con có đeo khoen. Hãy dự đoán số chim giúp ông chủ vườn ở độ tin cậy 99%.

Đáp số: $[785;28525]$.

Bài số 19. Sản lượng mỗi ngày của một phân xưởng là biến ngẫu nhiên tuân theo luật chuẩn. Kết quả thống kê của 9 ngày cho ta :

27	26	21	28	25	30	26	23	26

Hãy ước lượng sản lượng trung bình và phương sai mỗi ngày, với độ tin cậy 95%.

Đáp số: $\mu \in [23,7522;27,8034]$; $\sigma^2 \in [3,1682;25,4840]$.

Bài số 20. Cân thử 100 quả cam, ta có bộ số liệu sau :

Khối lượng (g)	32	33	34	35	36	37	38	39	40
Số quả	2	3	15	26	28	6	8	8	4

a) Hãy ước lượng khối lượng trung bình các quả cam ở độ tin cậy 95%.

b) Cam có khối lượng dưới 34g được coi là cam loại 2. Tìm ước lượng tỷ lệ cam loại 2 với độ tin cậy 90% .

Đáp số: a) $\mu \in [35,539;36,241]$*; b)* $p \in [0,0143;0,0857]$.

Bài số 21. Lô trái cây của một chủ cửa hàng được đóng thành sọt mỗi sọt 100 trái. Kiểm tra 50 sọt thấy có 450 trái không đạt tiêu chuẩn.

a) Ước lượng tỷ lệ trái cây không đạt tiêu chuẩn của lô hàng với độ tin cậy 95%.

b) Muốn ước lượng tỷ lệ trái cây không đạt tiêu chuẩn với độ chính xác 0,5%, độ tin cậy đạt được là bao nhiêu.

c) Muốn ước lượng tỷ lệ trái cây không đạt tiêu chuẩn với độ tin cậy 99% và độ chính xác 1% thì cần kiểm tra bao nhiêu sọt.

d) Muốn ước lượng tỷ lệ trái cây không đạt tiêu chuẩn với độ tin cậy 99% thì độ chính xác đạt được là bao nhiêu?

Đáp số: a) $p \in [0,082;0,098]$*; b) 78,5%; c) 0,012; d) 55 sọt.*

Bài số 22. Điều tra năng suất lúa trên diện tích 100 hec ta trồng lúa của một vùng, ta thu được bảng số liệu sau :

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích (ha)	10	20	30	15	10	10	5

a) Hãy ước lượng năng suất lúa trung bình của vùng đó với độ tin cậy 95%?

b) Những thửa ruộng có năng suất từ 48tạ/ha trở lên là những thửa có năng suất cao. Hãy ước lượng tỉ lệ diện tích có năng suất cao trong vùng với độ tin cậy 97%.

Đáp số: a) $\mu \in [45,353;46,647]$*; b)* $p \in [0,156;0,344]$.

Bài số 23. Đo đường kính của 100 chi tiết do một máy sản xuất kết quả cho ở bảng sau :

Đường kính (mm)	Số chi tiết
19,80 - 19,85	3
19,85 - 19,90	5
19,90 - 19,95	16
19,95 - 20,00	28
20,00 - 20,05	23
20,05 - 20,10	14
20,10 - 20,15	7
20,15 - 20,20	4

Quy định những chi tiết có đường kính 19,9mm đến 20,1mm là những chi tiết đạt tiêu chuẩn.

a) Ước lượng đường kính trung bình của những chi tiết đạt tiêu chuẩn với độ tin cậy 95%.

b) Ước lượng tỷ lệ chi tiết đạt tiêu chuẩn với độ tin cậy 95%.

c) Muốn ước lượng đường kính trung bình của chi tiết đạt tiêu chuẩn muốn độ chính xác đạt 0,02mm và khi ước lượng tỷ lệ chi tiết đạt tiêu chuẩn muốn độ chính xác là 5%, với cùng độ tin cậy là 99% thì cần đo thêm bao nhiêu chi tiết nữa.

Đáp số: a) $\mu \in [19,986;20,008]$*; b)* $p \in [0,733;0,887]$*; c) 310.*

Bài số 24. Kích thước của một chi tiết máy là một đại lượng ngẫu nhiên có phân phối chuẩn. Trong một mẫu gồm 30 chi tiết máy được kiểm tra, ta tính được $\overline{X} = 0,47\text{cm}$ và $S_X = 0,032\text{ cm}$. Tìm khoảng tin cậy cho phương sai và trung bình chuẩn của kích thước của toàn bộ các chi tiết máy với độ tin cậy 95%.

Đáp số: $\mu \in [0,482;0,458]$; $\sigma^2 \in [0,00065;0,00185]$.

Bài số 25. Lấy 28 mẫu xi măng của một nhà máy sản xuất xi măng để kiểm tra. Kết quả kiểm tra về sức chịu lực R (kg/cm^2) như sau:

10,0	13,0	13,7	11,5	11,0	13,5	12,2

13,0	10,0	11,0	13,5	11,5	13,0	12,2
13,5	10,0	10,0	11,5	13,0	13,7	14,0
13,0	13,7	13,0	11,5	10,0	11,0	13,0

Với độ tin cậy 95% hãy ước lượng:

a) Sức chịu lực trung bình của xi măng do nhà máy sản xuất.

b) Phương sai của sức chịu lực.

Đáp số: a) $\mu \in [11,64;12,64]$; $\sigma^2 \in [1,156;3,427]$.

Bài số 26. Để định mức thời gian gia công một chi tiết máy người ta theo dõi ngẫu nhiên quá trình gia công 25 chi tiết và thu được bảng số liệu sau:

Thời gian (phút)	Số chi tiết
15-17	1
17-19	3
19-21	4
21-23	12
23-25	3
25-27	2

Với độ tin cậy 95%, hãy ước lượng thời gian gia công trung bình một chi tiết máy. Giả thiết thời gian gia công chi tiết là biến ngẫu nhiên phân phối chuẩn.

Đáp số: $\mu \in [20,5293;21,5107]$.

Bài số 27. Kiểm tra 300 gói hàng do máy tự động đóng gói thì thấy trọng lượng trung bình là 1404g, với độ lệch tiêu chuẩn là 83,4g.

a) Hãy tính độ tin cậy với sai số cho phép là 5g.

b) Với độ tin cậy là 99 % hãy ước lượng trọng lượng trung bình của toàn bộ gói hàng do máy đó đóng gói.

Đáp số: a) 70,2%; b) $\mu \in [1391,577;1416,423]$.

Bài số 28. Với độ tin cậy 95%, hãy ước lượng lượng xăng hao phí trung bình cho một ô tô chạy từ A đến B nếu chạy thử 30 lần trên đoạn đường này người ta ghi nhận được lượng xăng hao phí như sau:

Lượng xăng hao phí (lít)	Số lần tương ứng
9,6-9,8	3
9,8-10,0	5

10,0-10,2	10
10,2-10,4	8
10,4-10,6	4

Biết rằng lượng xăng hao phí là biến ngẫu nhiên phân phối chuẩn.

Đáp số: $\mu \in [10,0454;10,2212]$.

Bài số 29. Điều tra doanh số hàng tháng của 100 hộ kinh doanh một loại hàng, ta có bảng số liệu sau:

Doanh số (triệu đồng)	Số hộ tương ứng
11,5	10
11,6	15
11,7	20
11,8	30
11,9	15
12,0	10

Hãy ước lượng doanh số trung bình hàng tháng của các hộ kinh doanh mặt hàng này bằng khoảng tin cậy đối xứng với hệ số tin cậy 95%, với giả thiết doanh số là biến ngẫu nhiên phân phối chuẩn.

Đáp số: $\mu \in [11,7268;11,7832]$.

Bài số 30. Để xác định giá trung bình đối với một loại hàng hoá trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng thu được bảng số liệu sau:

Giá (ngàn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng	6	7	12	15	30	10	8	6	4	2

Với độ tin cậy 95%, hãy ước lượng giá trung bình của loại hàng đó tại thời điểm đang xét.

Đáp số: $\mu \in [89,903;91,537]$.

Bài số 31. Để nghiên cứu độ ổn định của một máy gia công, người ta lấy ngẫu nhiên 25 chi tiết do máy đó gia công, đem đo và thu được các kích thước sau:

24,1	27,2	26,7	23,6	26,4
25,8	27,3	23,2	26,9	27,1
22,7	26,9	24,8	24,0	23,4
24,5	26,1	25,9	25,4	22,9

26,4	25,4	23,3	23,0	24,3

Với độ tin cậy 95% hãy ước lượng độ phân tán của kích thước các chi tiết do máy đó gia công. Biết kích thước các chi tiết là biến ngẫu nhiên phân phối chuẩn.

Đáp số: $\sigma^2 \in [1,46683;\ 4,7781]$.

Bài số 32. Lãi suất cổ phiếu của một công ty trong vòng 5 năm qua là 15%, 10%, 20%, 7%, 14%. Với độ tin cậy 90% hãy ước lượng độ phân tán của lãi suất cổ phiếu của công ty đó. Biết lãi suất cổ phiếu là biến ngẫu nhiên phân phối chuẩn.

Đáp số: $\sigma^2 \in [10,4132;\ 138,959]$.

Bài số 33. Để nghiên cứu một loại bệnh ở một địa phương, người ta đã khám 2500 người và thấy 500 người mắc bệnh.

a) Hãy ước lượng tỷ lệ mắc bệnh này ở toàn bộ địa phương đó với độ tin cậy 95%.
b) Cần khám cho bao nhiêu người để có thể nhận định về tỷ lệ mắc bệnh của toàn bộ địa phương với sai số không quá 0,01 và với độ tin cậy 99 %.

Đáp số: a) $p \in [0,18432; 0,21568]$*; b) 10651.*

Bài số 34. Một doanh nghiệp có dự định đưa một sản phẩm mới vào một thị trường có 1500000 người tiêu dùng. Nghiên cứu thị trường đối với 2500 khách hàng thấy 800 người sẵn sàng mua sản phẩm đó.

a) Với độ tin cậy 95% hãy ước lượng thị phần tiềm năng của doanh nghiệp.
b) Số lượng khách hàng tiềm năng mà doanh nghiệp hy vọng sẽ có được ở thị trường mới là bao nhiêu?

Đáp số: a) $p \in [0,301714; 0,338286]$*; b)* $[452572; 507429]$.

Bài số 35. Khảo sát ngẫu nhiên 400 sản phẩm của một nhà máy sản xuất, thấy có 92 sản phẩm đạt chất lượng loại A.

a) Với độ tin cậy 95% hãy ước lượng tỷ lệ sản phẩm loại A của nhà máy.
b) Nếu nhà máy sản xuất tổng cộng 100000 sản phẩm thì có tối đa bao nhiêu sản phẩm loại A?

Đáp số: a) $p \in [0,1888; 0,2712]$*; b)* 26460.

3.7. Tài liệu tham khảo

[1] Nguyễn Cao Văn, Trần Thái Ninh, Ngô Văn Thứ, Lý thuyết xác suất và thống kê toán, NXB Đại học Kinh tế Quốc dân, 2012.

[2] Nguyễn Cao Văn, Trần Thái Ninh, Nguyễn Thế Hệ, Bài tập xác suất và thống kê toán, NXB Đại học Kinh tế Quốc dân, 2012.

[3] Phạm Văn Chững, Lê Thanh Hoa, Nguyễn Đình Uông, Thống kê ứng dụng, NXB Đại học Quốc gia Thành Phố Hồ Chí Minh, 2016.

[4] Hà văn Sơn, Giáo trình Lý thuyết Thống kê, ứng dụng trong Quản trị và kinh tế.

[5] Anderson, Sweeney, and William [2010], Statistics for Business and Economics, South-Western Cengage Learning (11^{th} Edition).

[6] Michael Barrow, Statistics for Economics, Accounting and Business Studies-Prentice Hall, 2006.

[7] Newbold Paul - Statistics for Bussiness and Economics, 5th edition - Prentice Hall, 2005.

Thuật ngữ chính chương 3	
Tiếng Anh	**Tiếng Việt**
Approximately	Xấp xỉ
Biased Estimator	Ước lượng chệch
Confidence interval	Khoảng tin cậy
Central limit theorem	Định lý giới hạn trung tâm
Convenience Sampling	Lấy mẫu thuận tiện
Infinite population	Tổng thể vô hạn
Expected value	Giá trị kỳ vọng
Finite population	Tổng thể hữu hạn
Random sample	Mẫu ngẫu nhiên
Relative efficiency	Hiệu quả tương đối
Judgment sampling	Lấy mẫu phán đoán
Point Estimation	Ước lượng điểm
Population Parameter	Tham số tổng thể
Sample mean	Trung bình mẫu
Population proportion	Tỷ lệ tổng thể
Properties of point estimator	Tính chất của ước lượng điểm
Target population	Tổng thể mục tiêu
Simple random sample	Mẫu ngẫu nhiên đơn giản
Sample proportion	Tỷ lệ mẫu
Sampling distribution	Phân phối mẫu
Sample size	Cỡ mẫu
Sampling method	Phương pháp lấy mẫu
Standard deviation of the population	Độ lệch chuẩn tổng thể
Stratified random sampling	Lấy mẫu ngẫu nhiên phân tầng
Standard error	Độ lệch chuẩn
Statistic	Thống kê
Sample statistic	Thống kê mẫu
Sample variance	Phương sai mẫu
Unbiased estimator	Ước lượng không chệch