BÀI GIẢNG

XÁC SUẤT THỐNG KÊ

Mục lục

Chương 1

GIẢI TÍCH TỔ HỢP

1.1 HOÁN VỊ

Cho tập hợp M gồm $n(n \geq 1)$ phần tử.

+ Mỗi cách sắp xếp của n phần tử của M theo một thứ tự nhất định gọi là một hoán vị của n phần tử đã cho.

+ Kí hiệu số các hoán vị khác nhau của n phần tử là P_n

+ $P_n = n!$

Ví dụ 1.1.1. Có bao nhiêu cách khác nhau để sắp xếp 4 em học sinh vào 4 ghế ngồi. Giải: Có $P_4 = 4! = 24$ cách.

1.2 Chỉnh hợp

1.2.1 Chỉnh hợp không lặp

Cho tập M gồm $n(n \geq 1)$ phần tử và k là một số nguyên dương

+ Mỗi cách sắp xếp k phần tử của tập hợp M theo một tứ tự nhất định được gọi là một chỉnh hợp không lặp chập k của n phần tử đã cho.

+ Ký hiệu số các chỉnh hợp không lặp chập k khác nhau của n phần tử đã cho là A_n^k.

+ Hai chỉnh hợp không lặp chập k khác nhau khi chúng có một phần tử khác nhau hoặc chúng có thứ tự khác nhau.

$$A_n^k = \frac{n!}{(n-k)!} = n(n-1)...(n-k+1)$$

Ví dụ 1.2.1. Có bao nhiêu số khác nhau gồm 3 chữ số được thiết lập từ $\{1, 2, 3, 4, 5\}$.

Giải: Một số có 3 chữ số được lập từ $\{1, 2, 3, 4, 5\}$ tương ứng với một chỉnh hợp không lặp chập 3 của 5 số $1, 2, 3, 4, 5$.

Vậy các số khác nhau gồm 3 chữ sỗ được thiết lập từ $1, 2, 3, 4, 5$ bằng:

$$A_5^3 = \frac{5!}{(5-3)!} = \frac{5.4.3.2.1}{2.1} = 60$$

1.2.2 Chỉnh hợp lặp

+ Gọi chỉnh hợp lặp chập k của n phần tử của tập M là một tập hợp có thứ tự gồm k phần tử lấy từ tập M, mà mỗi phần tử của nó có thể có mặt tới k lần.

+ Số các chỉnh hợp lặp chập k của n phần tử là n^k.

Ví dụ 1.2.2. Lập tất cả các chỉnh hợp lặp chập 3 của 2 phần tử $\{1, 2\}$.

Giải: $\{1, 1, 1\}; \{1, 1, 2\}; \{1, 2, 1\}; \{1, 2, 2\}; \{2, 1, 1\}; \{2, 1, 2\}; \{2, 2, 1\}; \{2, 2, 2\}$. Nghĩa là có $2^3 = 8$ chỉnh hợp lặp chập 3 khác nhau của 2 phần tử.

Ví dụ 1.2.3. Có bao nhiêu cách trao 15 phần thưởng cho 5 người dự thi.

Giải: Mỗi cách phân 15 sản phẩm cho 5 người là một chỉnh hợp chập 15 của 5. Vậy số cách để phân ngẫu nhiên 15 phần thưởng cho 5 người là: 5^{15}.

1.3 Tổ hợp

Cho tập M gồm n phần tử.

+ Một tổ hợp (không kể thứ tự) gồm k phần tử $(k \leq n)$ của tập M được gọi là một tổ hợp chập k của n phần tử đã cho.

+ Hai tổ hợp chập k của n phần tử đã cho được gọi là khác nhau nếu chúng có một phần tử khác nhau.

+ Số các tổ hợp chập k khác nhau của n phần tử đã cho được kí hiệu là: C_n^k.

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

+ Chú ý: $0! = 1$.

Ví dụ 1.3.1. Có bao nhiêu cách để chọn 3 cuốn sách trong số 10 cuốn.

Giải: mỗi cách chọn 3 cuốn sách trong số 10 cuốn là một tổ hợp chập 3 của 10. Vậy số cách chọn là:

$$C_{10}^3 = \frac{10!}{3!(10-3)!} = \frac{10.9.8}{3.2.1} = 120$$

Ví dụ 1.3.2. Có bao nhiêu cách để chọn ra 5 người trong lớp có 45 người để đi lao động. Giải: $C_{45}^5 = \dfrac{45!}{5!(45-5)!}$

1.4 Công thức nhị thức Newton

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

1.5 Tích đề các

+ Tích đề các của các tập hợp $\mathrm{A}_1, ..., A_k$ được định nghĩa và kí hiệu bởi

$$A_1 \times A_2 \times .... \times A_k = \{(\mathrm{a}_1, ..., a_k) | a_i \in A_i\}$$

+ Nếu A_i có n_i phần tử thì tích đề các $A_1, ..., A_k$ có $n_1...n_k$ phần tử.

Ví dụ 1.5.1. Có 5 con đường để đi từ A đến B và 6 con đường để đi từ B đến C. Hỏi có bao nhiêu cách để đi từ A đến C mà qua B.

Giải: Mỗi cách lựa chọn là một phần tử của tích đề các của các tập hợp tương ứng có 5 và 6 phần tử. Vậy số cách lựa chọn là: $5.6 = 30$.

BÀI TẬP CHƯƠNG 1

1.1. Một lớp có 50 học sinh, cần bầu ra 5 chức vụ là: lớp trưởng, lớp phó học tập, lớp phó văn thể, bí thư chi đoàn, phó bí thư chi đoàn. Hỏi có bao nhiêu cách lựa chọn nếu

a) Mỗi người có thê kiêm nhiệm tối đa 5 nhiệm vụ.

b) Mỗi người có thể kiêm nhiệm tối đa 3 nhiệm vụ.

c) Mỗi người có thể kiêm nhiệm tối đa 2 nhiệm vụ.

d) Mỗi người có thể kiêm nhiệm tối đa 1 nhiệm vụ.

1.2. Có 14 cuốn sách sắp đặt trên một giá sách. Hỏi có bao nhiêu cách sắp xếp nếu:

a) Giá chỉ có một ngăn đủ cho 14 cuốn.

b) Giá có hai ngăn, mỗi ngăn đủ chổ cho 14 cuốn.

c) Giá có hai ngăn, mỗi ngăn đủ chổ cho 7 cuốn.

d) Giá chỉ có một ngăn đủ chổ cho 10 cuốn (4 cuốn không được để trên giá).

e) Giá chỉ có hai ngăn, mỗi ngăn đủ chỏ cho 5 cuốn sách (có 4 cuốn sẽ không được bỏ lên giá).

1.3. Có k chậu hoa và m cái đôn để đặt chậu hoa. Hỏi có bao nhiêu cách để đặt chậu hoa lên đôn (mỗi đôn chỉ đặt được một chậu hoa) nếu:

a) $k = 6, m = 3$. b) $k = 3, m = 6$. c) $k = m = 9$.

1.4. Một học sinh phi thi 4 môn trong 10 ngày (mỗi ngày thi một môn). Có bao nhiêu cách để lập chương trình thi?

1.5. Trong lô 100 sản phẩm có 80 sản phẩm tốt và 20 sản phẩm xấu. Hỏi

a) Có bao nhiêu cách lấy ra 10 sản phẩm.

b) Có bao nhiêu khả năng lấy ra 10 sản phẩm trong đó có 7 sản phẩm tốt và 3 sản phẩm xấu.

1.6. Phân ngẫu nhiên 18 hành khách lên 5 toa tàu.

a) Có bao nhiêu cách phân ngẫu nhiên 18 hành khách lên 5 toa tàu.

b) Có bao nhiêu cách phân ngẫu nhiên 18 hành khách lên 5 toa tàu mà toa thứ nhất có đúng 5 người.

c) Có bao nhiêu cách phân ngẫu nhiên 18 hành khách lên 5 toa tàu mà mỗi toa không quá 4 người.

1.7. Trong sàn nhảy có 10 nam và 8 nữ. Hỏi có bao nhiêu cách chọn ra 2 đôi nam nữ để nhảy.

1.8. Tỉnh A có 5 đội bóng nam và 3 đội bóng nữ, tỉnh B có 6 đội bóng nam và 2 đội bóng nữ. Dự định diễn ra một trận đấu giữa hai đội bóng nam và một trận đấu của hai đội bóng nữa giữa hai tỉnh. Hỏi có bao nhiêu phương án khác nhau về lựa chọn các đội thi đấu.

1.9. Có bao nhiêu cách để chia 16 đội bóng đá thành 4 bng, mỗi bng chỉ có 4 đội.

1.10. Giải ngoại hạng Anh có tất cả 20 đội tham dự. Trong một mùa giả tất cả các đội đều gặp nhau 2 trận (trận lượt đi và trận lượt về). Hỏi trong một mùa Giải có bao nhiêu trận đấu diễn ra.

Chương 2

BIẾN CỐ VÀ XÁC SUẤT CỦA BIẾN CỐ

2.1 PHÉP THỬ NGẪU NHIÊN VÀ KHÔNG GIAN MẪU

Trong thực tế ta thường gặp rất nhiều hành động mà các kết quả của nó không thể dự báo trước được, chẳng hạn như làm một thí nghiệm hay quan sát một hiện tượng nào đó. Ta gọi chúng là các phép thử ngẫu nhiên, ký hiệu là T.

Các kết quả của T là ngẫu nhiên, không thể xác định trước. Tuy nhiên ta có thể liệt kê ra tấc cả các kết quả có thể của phép thử T.

+ Không gian mẫu: Tập hợp tất cả các kết quả có thể xảy ra của phép thử T, ký hiệu là Ω.

Biến cố ngẫu nhiên: Mỗi tập con $A \subset \Omega$ được gọi là một biến cố.

+ Biến cố sơ cấp: Mỗi phần tử $\omega \in \Omega$ được gọi là biến cố sơ cấp.

+ Biến cố không thể là biến cố không bao giờ xảy ra khi phép thử được thực hiện, ký hiệu $\emptyset$.

+ Biến cố chắc chắn là biến cố luôn luôn xảy ra khi phép thử được thực hiện, ký hiệu là Ω.

Ví dụ 2.1.1. Quan sát tình hình hoạt động của một dây chuyền máy móc là làm một phép thử. Việc dây chuyền máy móc hoạt động tốt hay hỏng hóc là hai biến cố.

Ví dụ 2.1.2. Gieo một con xúc sắc và quan sát số nốt trên mặt xuất hiện của con xúc sắc là một làm một phép thử. Ta không biết trước được mặt nào của con xúc sắc xuất hiện. Không gian mẫu của T là $\Omega = \{1, 2, 3, 4, 5, 6\}$.

+ $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ là những biến cố sơ cấp.

+ $A = \{1, 3, 5\}$ là biến cố mặt lẻ

+ $B = \{2, 4, 6\}$ là biến cố mặt chẵn.

+ $C = \{1, 2, 3, 4, 5, 6\}$ là biến cố chắc chắn, tức là $C = \Omega$.

+ $D = \{7\}$ là biến cố không thể có, tức là $C = \emptyset$.

Ví dụ 2.1.3. Kiểm tra 3 sản phẩm. Biến cố "không có quá 3 sản phẩm tốt có trong 3 sản phẩm kiểm tra" là biến cố chắc chắn. Biến cố "có 4 phế phẩm có trong 3 sản phẩm kiểm tra" là biến cố không thể. Biến cố "có 2 sản phẩm tốt trong 3 sản phẩm kiểm tra" là biến cố ngẫu nhiên.

2.2 Quan hệ giữa các biến cố và các phép toán

Để giải các bài toán xác suất, ta thường diễn tả biến cố phức tạp theo các biến cố đơn giản hơn.

+ Tương đương: Biến cố A và B được gọi là hai biến cố tương đương, ký hiệu $A = B$ nếu biến cố A xảy ra thì B cũng xảy ra và ngược lại.

Ví dụ 2.2.1. Tung một con xúc xắc, biến cố "xúc xắc ra mặt lẻ" và biến cố "xúc xắc ra một trong ba mặt: $1, 3, 5$" là hai biến cố tương đương.

+ Xung khắc: Hai biến cố A và B được gọi là xung khắc nếu chúng không thể cùng xảy ra trong một phép thử.

Ví dụ 2.2.2. Khi kiểm tra 5 sản phẩm, biến cố "có 1 phế phẩm" và biến cố "có 2 phế phẩm" là hai biến cố xung khắc nhau.

Các biến cố $A_1, A_2, , A_n$ được gọi là xung khắc từng đôi nếu 2 biến cố bất kỳ trong số n biến cố này đều xung khắc với nhau.

+ Biến cố đối lập: Biến cố được gọi là biến cố đối của biến cố A nếu nó xảy ra khi và chỉ khi A không xảy ra, ký hiệu là $\bar{A}$. Ta có $\bar{A} = \Omega \backslash A$.

Ví dụ 2.2.3. Kiểm tra một sản phẩm, biến cố "sản phẩm kiểm tra là sản phẩm tốt" và biến cố "sản phẩm kiểm tra là sản phẩm xấu" là hai biến cố đối lập nhau.

Nhận xét:

+ Đặc biệt, trong trường hợp A là biến cố chắc chắn thì biến cố đối lập với A là biến không thể.

+ So sánh với điều kiện xung khắc ta thấy: Hai biến cố đối lập thì xung khắc nhưng hai biến cố cung khắc chưa chắc đã đối lập.

2.2.1 Các phép toán của biến cố

a. Phép hợp (tổng)

+ Hợp của hai biến cố A và B là biến cố xảy ra khi và chỉ khi có ít nhất một trong hai biến cố A hoặc B xảy ra, ký hiệu là $A \cup B$ hoặc $A + B$.

Tổng quát, biến cố A được gọi là tổng của n biến cố $A_1, A_2, ..., A_n$ nếu A xảy ra khi và chỉ khi có ít nhất một trong n biến cố đó xảy ra. Ký hiệu $A = A_1 \cup A_2 \cup ... \cup A_n$ hoặc $A = \sum\limits_{i=1}^{n} A_i$.

Ví dụ 2.2.4. Xét phép thử quan sát hai xạ thủ cùng bắn vào một bia (mỗi xạ thủ bắn một viên đạn).

Gọi A là biến cố "xạ thủ thứ nhất bắn trúng bia", Gọi B là biến cố "xạ thủ thứ hai bắn trúng bia", và Gọi C là biến cố "bia trúng đạn".

Rõ ràng $C = A \cup B$

b. Phép giao (tích)

+ Tích của hai biến cố A và B là biến cố xảy ra nếu cả A và B đều xảy ra, ký hiệu là AB hoặc $A \cap B$.

Tổng quát, biến cố A được gọi là tích của n biến cố $A_1, A_2, ..., A_n$ nếu A xảy ra nếu tất cả các biến cố $A_1, A_2, ..., A_n$ đều xảy ra. Ký hiệu $A = A_1A_2...A_n$.

Ví dụ 2.2.5. Xét phép thử quan sát hai xạ thủ cùng bắn vào một bia (mỗi xạ thủ bắn một viên đạn). Gọi D là biến cố "xạ thủ thứ hai bắn trật". Gọi E là biến cố "xạ thủ thứ hai bắn trật" và gọi F là biến cố "bia không trúng đạn". Rõ ràng $F = D \cap E$.

c. Nhóm đầy đủ các biến cố

Hệ n các biến cố $A_1, A_2, ..., A_n$ là một hệ đầy đủ và xung khắc từng đôi nếu các biến cố xung khắc từng đôi và tổng của chúng là biến cố chắc chắn.

Ví dụ 2.2.6. Kiểm tra 3 sản phẩm, gọi tương ứng là các biến cố có "$0, 1, 2, 3$ sản phẩm tốt trong 3 sản phẩm kiểm tra". Các biến cố là một hệ đầy đủ và xung khắc từng đôi.

d. Phép hiệu

+ Hiệu của biến cố B với biến cố A là biến cố "B xảy ra nhưng A không xảy ra", ký hiệu là $B \backslash A$. Đặc biệt, ta có $\bar{A} = \Omega \backslash A$.

Nhận xét. Khi giải nhiều bài toán xác suất, ta thường biểu diễn các biến cố phức hợp thành tổng và tích các biến cố đơn giản hơn.

Ví dụ 2.2.7. Một nhà máy sản xuất 3 sản phẩm. Gọi $A_i, i = 1, 2, 3$ là biến cố "sản phẩm thứ i là sản phẩm tốt". Khi đó $\bar{A}_i, i = 1, 2, 3$ là biến cố "sản phẩm thứ i là phế phẩm".

Nếu gọi A là biến cố "có 1 sản phẩm tốt trong 3 sản phẩm do nhà máy sản xuất" thì $A = A_1.\bar{A}_2.\bar{A}_3 \cup \bar{A}_1A_2\bar{A}_3 \cup \bar{A}_1\bar{A}_2A_3$.

Nếu gọi B là biến cố "có 2 sản phẩm tốt trong 3 sản phẩm do nhà máy sản xuất" thì $B = \bar{A}_1.A_2.A_3 \cup A_1\bar{A}_2A_3 \cup A_1A_2\bar{A}_3$.

Nếu gọi C là biến cố "có 3 sản phẩm tốt trong 3 sản phẩm do nhà máy sản xuất" thì $C = A_1.A_2.A_3$.

2.3 Xác suất của các biến cố

2.3.1 Khái niệm về xác suất

Giả sử là biến cố của phép thử nào đó. Mặc dù khi tiến hành phép thử ta không thể nói trước biến cố A xảy ra hay không nhưng ta thừa nhận rằng: có một số đo khả năng xảy ra của biến cố A, ký hiệu $p(A)$. Khi đó $p(A) = 1$ nếu là biến cố chắc chắn và $p(A) = 0$ nếu A là biến cố không thể.

Vậy xác suất của một biến cố là một số biểu thị khả năng xảy ra của biến cố đó khi thực hiện phép thử.

a. Định nghĩa cổ điển về xác suất

Phần này chúng ta xây dựng mô hình xác suất cho những phép thử "đối xứng" như tung đồng xu hay gieo xúc sắc hoặc chọn ngẫu nhiên k phần tử từ tập hợp có hữu hạn phần tử.

Định nghĩa 2.3.1. Xác suất của biến cố A là tỉ số giữa số trường hợp thuận lợi cho A và số trường hợp đồng khả năng có thể xảy ra khi thực hiện phép thử, hay $p(A) = \dfrac{m}{n}$ trong đó, m: là số trường hợp thuận lợi cho n: là số trường hợp đồng khả năng có thể xảy ra khi thực hiện phép thử.

b. Tính chất

Nếu A là biến cố ngẫu nhiên thì $0 < P(A) < 1$:

Nếu A là biến cố chắc chắn thì $P(\Omega) = 1$:

Nếu A là biến cố không thể thì $P(\emptyset) = 0$:

Như vậy nếu A là biến cố bất kỳ thì $0 \leqslant P(A) \leqslant 1$.

c. phương pháp tính xác suất

+ phương pháp tính xác suất bằng định nghĩa cổ điển

+ phương pháp suy luận trực tiếp

+ phương pháp sử dụng sơ đồ

+ phương pháp sử dụng các khái niệm Giải tích tổ hợp

Ví dụ 2.3.2. Gieo đồng thời 3 con xúc sắc được chế tạo cân đối, đồng chất. Tính xác suất để tổng số nốt xuất hiện của 3 con là 9.

Giải. Mỗi kết quả của phép thử là một bộ ba (a, b, c), trong đó a, b, c là các số nguyên dương từ 1 đến 6. Vậy số trường hợp đồng khả năng là $n = 6^3 = 216$.

Các bộ ba có tổng bông 9 là:

$(1, 2, 6)$ và 5 hoán vị của nó; $(1, 3, 5)$ và 5 hoán vị của nó $(1, 4, 4)$ và 2 hoán vị của nó; $(2, 2, 5)$ và 2 hoán vị của nó $(2, 3, 4)$ và 5 hoán vị của nó; $(3, 3, 3)$. Suy ra số trường hợp thuận lợi là $m = 6 + 6 + 3 + 6 + 3 + 1 = 25$. Vậy $p(A) = \dfrac{25}{216}$.

Ví dụ 2.3.3. Trước cổng trường có 3 quán cơm bình dân chất lượng ngang nhau. Ba sinh viên Hồng, Hà, Hoa độc lập với nhau chọn một quán để ăn trưa. Tính xác suất để:

a) 3 sinh viên cùng vào một quán;

b) 2 sinh viên cùng vào một quán, còn người kia thì vào quán khác.

Giải: Ta đánh số ba quán cơm là $1, 2, 3$. Gọi a, b, c là quán cơm mà Hồng, Hà, Hoa chọn. Như vậy số trường hợp đồng khả năng là $n = 3^3 = 27$.

Gọi A là biến cố 3 sinh viên cùng vào một quán.

Số trường hợp thuận lợi của A là $(1, 1, 1)$; $(2, 2, 2)$ và $(3, 3, 3)$.

Vậy $p(A) = \dfrac{3}{27} = \dfrac{1}{9}$.

Gọi B là biến cố 2 sinh viên cùng vào một quán, còn sinh viên kia thì vào quán khác. $(1, 1, 2)$ và 2 hoán vị của nó;

$(1, 1, 3)$ và 2 hoán vị của nó;

$(2, 2, 1)$ và 2 hoán vị của nó;

$(2, 2, 3)$ và 2 hoán vị của nó;

$(3, 3, 1)$ và 2 hoán vị của nó;

$(3, 3, 2)$ và 2 hoán vị của nó;

Số trường hợp thuận lợi của B là 6.3=18. Vậy $p(B) = \dfrac{18}{27} = \dfrac{2}{3}$.

Ví dụ 2.3.4. Một túi đựng 10 quả cầu, trong đó có 6 quả màu xanh và 4 quả màu vàng. Lấy ngẫu nhiên (không hoàn lại) từ túi ra 3 quả cầu.

a) Tính xác suất để có 2 quả cầu xanh trong 3 quả cầu lấy ra từ túi.

b) Tính xác suất để trong 3 quả cầu lấy ra có ít nhất 1 quả cầu xanh.

Giải. a) Gọi A là biến cố có 2 quả cầu màu xanh trong 3 quả cầu lấy ra từ túi. Số trường hợp đồng khả năng có thể xảy ra là một nhóm gồm 3 quả cầu (không phân biệt thứ tự) được chọn từ 10 quả cầu. Vậy $n = C_{10}^3 = 120$.

Số trường hợp thuận lợi cho A là số nhóm gồm 3 quả cầu, trong đó có 2 quả xanh. nên $m = C_6^2.C_4^1 = 60$. Vậy $p\left(A\right) = \dfrac{60}{120} = 0.5$.

b) Gọi B là biến cố có ít nhất một quả cầu xanh trong 3 quả cầu lấy ra từ túi. Gọi B_i là biến cố có i quả cầu xanh trong 3 quả cầu lấy ra từ túi $(0 \leqslant i \leqslant 3)$. Ta có B_i là họ xung khắc nên $B = B_1 \cup B_2 \cup B_3$. Suy ra

$$p\left(B\right) = p\left(B_1\right) + p\left(B_2\right) + p\left(B_3\right) = \frac{C_6^1.C_4^2}{C_{10}^3} + \frac{C_6^2.C_4^1}{C_{10}^3} + \frac{C_6^3}{C_{10}^3} = \frac{36 + 60 + 20}{120}.$$

d. Ưu điểm và hạn chế của định nghĩa cổ điển

Ưu điểm: Tìm xác suất của biến cố ta không cần thực hiện phép thử (phép thử chỉ thực hiện một cách giả định).

Nhược điểm: Chỉ áp dụng để tính xác suất khi phép thử có tính "đối xứng" và đòi hỏi phép thử phải xác định số trường hợp thuận lợi và số trường hợp đồng khả năng và đó là những số hữu hạn nhưng trong thực tế đa số các phép thử không thỏa mãn các yêu cầu đó.

e. Định nghĩa xác suất theo thống kê.

Tần số - Tần suất

Xét A là biến cố của một thí nghiệm ngẫu nhiên. Ta lặp lại thí nghiệm này n lần. Khi đó: số lần xuất hiện biến cố A trong n phép thử được gọi là tần số của A, ký hiệu là $\mu_n(A)$ và tỉ số $f(A) = \dfrac{\mu_n(A)}{n}$ được gọi là tần suất của biến cố A trong n phép thử.

Ví dụ 2.3.5. Khi kiểm tra ngẫu nhiên 60 sản phẩm ở một lô hàng, người ta phát hiện ra 3 phế phẩm. Gọi A là biến cố "sản phẩm kiểm tra là phế phẩm" thì tần suất xuất hiện biến cố A là $f(A) = \dfrac{3}{60} = 5\%$.

Ví dụ 2.3.6. Để nghiên cứu khả năng xuất hiện mặt sấp khi tung đồng xu, người ta tiến hành tung đồng xu nhiều lần và thu được kết quả sau:

Người làm thí nghiệm	Số lần tung	Số lần được mặt sấp	Tần suất
Buyffon	4040	2048	0.5069
Pearson	1200	6019	0.5016
Pearson	24000	12012	0.5005

Từ thí nghiệm trên ta thấy khi phép thử tăng lên, tần suất xuất hiện mặt sấp tiến dần đến 0.5 (xác suất xuất hiện mặt sấp khi tung đồng xu là 0.5). Vậy tần suất dần đến xác suất khi số phép thử tăng lên.

Định nghĩa. Tần suất xuất hiện biến cố sẽ hội tụ về xác suất xuất hiện biến cố khi số phép thử tăng lên vô hạn.

Trong thực tế với số phép thử đủ lớn ta có $p(A) \approx f(A)$.

2.4 Các Công thức tính xác suất

2.4.1 Công thức cộng xác suất

Định lý 2.4.1. *Nếu A và B là hai biến cố xung khắc thì:*

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

Tổng quát, nếu $A_1, A_2, ..., A_n$ là n biến cố xung khắc từng đôi thì

$$p(A_1 \cup A_2 \cup ... \cup A_n) = p(A_1) + p(A_2) + ... + p(A_n)$$

Hệ quả 2.4.2. *Nếu $A_1, A_2, ..., A_n$ là n biến cố đầy đủ và xung khắc từng đôi thì:*

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

Hệ quả 2.4.3. *Nếu A và $\bar{A}$ là hai biến cố đối lập với nhau thì: $p(A) = 1 - p(\bar{A})$.*

Ví dụ 2.4.4. Một hộp có 10 sản phẩm (trong đó có 2 phế phẩm). Lấy ngẫu nhiên (không hoàn lại) từ hộp ra 6 sản phẩm. Tính xác suất để có không quá 1 phế phẩm trong 6 sản phẩm lấy ra.

Giải: Gọi A là biến cố "không có phế phẩm nào trong 6 sản phẩm lấy ra" và gọi B là biến cố "có 1 phế phẩm trong 6 sản phẩm lấy ra". Gọi C là biến cố "có không quá 1 phế phẩm trong 6 sản phẩm lấy ra". Suy ra $C = A \cup B$.

Do A, B là hai biến cố xung khắc (vì nó không thể đồng thời xảy ra trong phép thử lấy ngẫu nhiên 6 sản phẩm từ hộp). Vậy

$$p(C) = p(A \cup B) = p(A) + p(B) = \frac{C_8^6}{C_{10}^6} + \frac{C_2^1 C_8^5}{C_{10}^6} = \frac{14 + 56}{105} = \frac{2}{3} \approx 0.66.$$

Định lý 2.4.5. *Nếu A và B là hai biến cố bất kỳ thì:*

$$p(A \cup B) = p(A) + p(B) - p(AB).$$

Ta có thể mở rộng định lý trên cho trường hợp 3 biến cố:

$$p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(A.B) - p(AC) - p(BC) + p(ABC).$$

Ví dụ 2.4.6. Theo kho sát của tổ chức y tế WHO, trong một vùng dân cư tỉ lệ người mắc bệnh tim là 9%, bệnh huyết áp là 12% và mắc cả hai bệnh là 7%. Chọn ngẫu nhiên một người trong vùng đó. Tính xác suất để người đó không mắc cả bệnh tim và huyết áp.

Giải. Gọi A là biến cố "người đó mắc bệnh tim" và gọi B là biến cố "người đó mắc bệnh huyết áp". Ta có $p(A) = 0.09$, $p(B) = 0.12$, $(AB) = 0.07$.

Gọi N là biến cố "người đó không mắc c bệnh tim và bệnh huyết áp". Suy ra $\bar{N} = A \cup B$

Nên

$$p(\bar{N}) = p(A \cup B) = p(A) + p(B) - p(AB) = 0.09 + 0.12 + 0.07 = 0.14$$

Vậy $p(N) = 1 - p(\bar{N}) = 1 - 0.14 = 0.86$.

2.4.2 Công thức nhân xác suất

2.4.2.1 Xác suất có điều kiện

Định nghĩa. Xác suất của biến cố A được tính với điều kiện biến cố B đã xãy ra được gọi là xác suất có điều kiện của A, ký hiệu $P(A/B)$.

Ví dụ 2.4.7. Một túi đựng 5 quả cầu, (trong đó có 2 quả màu trắng). Lấy ngẫu nhiên (không hoàn lại) lần lượt từ túi ra 2 quả cầu. Tính xác suất để lần thứ hai được quả cầu trắng biết rông lần thứ nhất lấy được quả cầu trắng.

Giải. Gọi A là biến cố "lần thứ hai lấy được quả cầu trắng" và gọi B là biến cố "lần thứ hai nhất lấy được quả cầu trắng". Ta cần tìm (A/B) .

Ta thấy lần thứ nhất đã lấy được quả cầu trắng (B đã xãy ra) nên trong túi còn 4 quả cầu, trong đó có 1 quả trắng. Vậy $p(A/B) = \dfrac{1}{4} = 0.25$

Công thức tính: $P(A/B) = \dfrac{P(AB)}{P(B)}$

Nếu A, B là hai biến cố bất kỳ, ta có:

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

Tổng quát ta có:

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

2.4.2.2 Biến cố độc lập

+ Hai biến cố A và B được gọi là độc lập với nhau nếu $p(A/B) = p(A)$ (tức là biến cố B xảy ra hay không cũng không ảnh hưởng đến khả năng xảy ra của biến cố A).

Dựa vào định nghĩa của xác suất có điều kiện, từ điều trên ta có: $p(AB) = p(A).p(B)$.

Nhận xét: A và B độc lập $\Leftrightarrow$ $\bar{A}$ và B độc lập $\Leftrightarrow$ A và $\bar{B}$ độc lập $\Leftrightarrow$ $\bar{A}$ và $\bar{B}$ độc lập.

Nếu $p(C) = 0$ hoặc $p(C) = 1$ thì C độc lập với mọi biến cố.

Định nghĩa. Nếu các biến cố $A_1, A_2, ..., A_n$ độc lập với nhau thì

$$p(A_1...A_n) = p(A_1)...p(A_n).$$

Ví dụ 2.4.8. Một phân xưởng có 3 máy. Xác suất các máy bị hỏng trong ngày tương ứng là $0.1; 0.2$ và 0.15. Tính các xác suất sau đây:

a). Có một máy bị hỏng trong ngày.

b). Có ít nhất một máy bị hỏng trong ngày.

Giải. Gọi A_1, A_2, A_3 tương ứng là các biến cố máy thứ nhất, thứ hai, thứ ba bị hỏng trong ngày.

a). Gọi A là biến cố có một máy bị hỏng trong ngày. Khi đó ta có

$$A = A_1\bar{A}_2\bar{A}_3 \cup \bar{A}_1A_2\bar{A}_3 \cup \bar{A}_1\bar{A}_2A_3$$

Vì các biến cố tích xung khắc từng đôi và độc lập toàn phần nên

$$\begin{aligned}
p(A) &= p\left(A_1\bar{A}_2\bar{A}_3 \cup \bar{A}_1A_2\bar{A}_3 \cup \bar{A}_1\bar{A}_2A_3\right)\\
&= p(A_1)\,p(\bar{A}_2)\,p(\bar{A}_3) + p(\bar{A}_1)\,p(A_2)\,p(\bar{A}_3) + p(\bar{A}_1)\,p(\bar{A}_2)\,p(A_3)\\
&= 0,1.0,8.0,85 + 0,9.0,2.0,85 + 0,9.0,8.0,15 = 0,329
\end{aligned}$$

b). Gọi B là biến cố "có ít nhất một máy bị hỏng trong ngày" nên $\bar{B}$ là biến cố "cả ba máy đều tốt" hay $p(\bar{B}) = \bar{A}_1\bar{A}_2\bar{A}_3$.

Suy ra

$$P(B) = 1 - P(\bar{B}) = 1 - P(\bar{A}_1)\,P(\bar{A}_2)\,P(\bar{A}_3) = 1 - 0,9.0,8.0,85 = 0,388.$$

2.5 Công thức Bernoulli

Trong nhiều bài toán thực tế, ta thường gặp cùng một phép thử được lặp đi lặp lại nhiều lần. Trong mỗi phép thử có thể xảy ra hay không xảy ra biến cố A nào đó và ta quan tâm đến tổng số lần xảy ra biến cố A trong dãy phép thử đó.

Định nghĩa. Các phép thử được gọi là độc lập với nhau nếu xác suất để xảy ra một biến cố nào đó trong từng phép thử sẽ không phụ thuộc vào việc biến cố đó có xảy ra ở phép thử khác hay không.

Ví dụ 2.5.1. Tung nhiều lần một đồng xu, lấy ngẫu nhiên (có hoàn lại) n sản phẩm từ một lô hàng là các phép thử độc lập.

Bài toán. Giả sử tiến hành n phép thử độc lập. Trong mỗi phép thử chỉ có thể xảy ra hai trường hợp: hoặc biến cố A xảy ra, hoặc biến cố A không xảy ra. Xác suất để biến cố A xảy ra trong mỗi phép thử đều bằng p và xác suất A không xảy ra là $q = 1 - p$. Ta gọi n phép thử này được gọi là dãy phép thử Bernoulli. Khi đó xác suất để trong n phép thử độc lập nói trên biến cố A xảy ra đúng k lần là:

$$B(k, n, p) = C_n^k p^k q^{n-k}; k = 0, 1, .., n$$

Ví dụ 2.5.2. Một cầu thủ bóng rỗ có khả năng ném lọt rỗ với xác suất $p = 0,95$. Tìm xác suất để trong 10 lần ném có đúng 8 lần lọt rỗ.

Giải. Rõ ràng đây là dãy phép thử Bernoulli với xác suất ném lọt rỗ là $p = 0,95$. Do đó xác suất cần tìm là $B(8; 0,95; 10) = C_{10}^8.(0,95)^8.(0,05)^2 \simeq 0,0746$.

2.6 Công thức xác suất đầy đủ

Giả sử B là biến cố bất kỳ có thể xảy ra đồng thời với một trong các biến cố $A_1, A_2, ..., A_n$-hệ biến cố xung khắc từng đôi và đầy đủ. Khi đó, xác suất của B là:

$$p(B) = \sum_{i=1}^{n} p(A_i).p(B/A_i).$$

Các xác suất $p(A_1), p(A_2), ..., p(A_n)$ được gọi là xác suất của giả thiết hay còn gọi là xác suất tiên nghiệm.

Ví dụ 2.6.1. Một cơ sở sản xuất mũ gồm có 3 tổ cùng sản xuất mũ (độc lập nhau) với tỉ lệ sản phẩm trong tổng số sản phẩm lần lượt là 20%, 30% và 50%, trong đó tỉ lệ phế phẩm tương ứng là 5%, 2% và 1%. Tất cả sản phẩm làm ra đều được xếp chung vào một kho. Hỏi tỉ lệ phế phẩm của kho là bao nhiêu?

Giải. Lấy ngẫu nhiên một mũ từ trong kho ra kiểm tra. Gọi B là biến cố gặp mũ phế phẩm và A_i là biến cố mũ lấy ra do tổ i sản xuất $(1 \leq i \leq 3)$, ta có

$$\begin{aligned} p(B) &= p(A_1).p(B/A_1) + p(A_2).P(B/A_2) + p(A_3).P(B/A_3) \\ &= \frac{20}{100} \times \frac{5}{100} + \frac{30}{100} \times \frac{2}{100} + \frac{50}{100} \times \frac{1}{100} = 0,021 \end{aligned}$$

2.7 Công thức Bayes

Giả sử B là biến cố bất kỳ có thể xảy ra đồng thời với một trong các biến cố $A_1, A_2, ..., A_n$ -hệ biến cố xung khắc từng đôi và đầy đủ. Giả thiết rằng B đã xảy ra khi đó, với $i = 1, ..., n$ ta có:

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

Nhận xét: Các xác suất $p(A_i/B)$ được xác định sau khi đã biết kết quả của phép thử B đã xảy ra nên gọi là các xác suất hậu nghiệm. Như vậy công thức Bayes cho phép ta xác định các xác suất tiên nghiệm khi biết thông tin B đã xảy ra khi thực hiện phép thử.

Ví dụ 2.7.1. Hai máy tiện cùng sản xuất ra một loại trục xe đạp như nhau. Các trục xe được đóng chung vào một kiện. Năng suất của máy tiện thứ hai gấp đôi năng suất của máy tiện thứ nhất. Máy tiện thứ nhất sản xuất trung bình được 64% trục loại tốt, còn máy tiện thứ hai được 80% trục loại tốt. Lấy ngẫu nhiên từ kiện một trục ra kiểm tra thì được trục loại tốt. Tìm xác suất để trục đó do máy tiện thứ nhất sản xuất.

Giải. Gọi B là biến cố trục đó là trục loại tốt và A_1, A_2 tương ứng là các biến cố trục đó do máy thứ nhất và máy thứ hai sản xuất ra. Ta có $p(A_1) = \frac{1}{3}; P(A_2) = \frac{2}{3}$; và $p(B/A_1) = 0,64; p(B/A_2) = 0,8$

Theo công thức xác suất đầy đủ thì

$$\begin{aligned} p(B) &= p(A_1).p(B/A_1) + p(A_2).p(B/A_2) \\ &= 0,64 \times \frac{1}{3} + 0,8 \times \frac{2}{3} \simeq 0,7467. \end{aligned}$$

Vậy $p(A_1/B) = \dfrac{p(A_1).p(B/A_1)}{p(B)} \simeq \dfrac{0,64 \times \frac{1}{3}}{0,7467} \simeq 0,2857$.

BÀI TẬP CHƯƠNG 2

2.1. Ba xạ thủ, mỗi người bắn một phát. Gọi A_i là biến cố người thứ i bắn trúng. Hãy biểu diễn qua các biến cố sau:

a. A: chỉ có người thứ nhất bắn trúng.

b. B: người thứ nhất bắn trúng còn người thứ hai bắn trật.

c. C : có ít nhất một người bắn trúng.

d. D : cả 3 người đều bắn trúng.

e. E : có ít nhất hai người bắn trúng.

f. F : chỉ có hai người bắn trúng.

g. G : không có ai bắn trúng

h. H : không có hơn 2 người bắn trúng.

i. I : người thứ nhất bắn trúng, hoặc người thứ hai và người thứ ba cùng bắn trúng.

j. K : người thứ nhất bắn trúng hay người thứ hai bắn trúng.

2.2. Gieo đồng thời hai con xúc xắc được chế tạo cân đối, đồng chất. Tìm xác suất để:

a. Tổng số nốt là 7.

b. Tổng số nốt là 8.

2.3. Một công ty cần tuyển hai nhân viên. Có 6 người nộp đơn (trong đó có 4 nữ và 2 nam). Giả sử khả năng trúng tuyển của 6 người là như nhau. Tính xác suất để:

a. cả 2 người trúng tuyển đều là nam.

b. cả 2 người trúng tuyển đều là nữ.

c. có ít nhất một nữ trúng tuyển.

2.4. Trong 30 đề thi, trong đó có 10 đề khó, 20 đề trung bình. Tìm xác suất để:

a. Một học sinh bốc 1 đề, gặp đề trung bình.

b. Một học sinh bốc 2 đề, gặp ít nhất một đề trung bình.

2.5. Lớp X có 60 sinh viên, trong đó có 20 nam và 40 nữ. Chọn ngẫu nhiên một nhóm gồm 8 sinh viên để đi chiến dịch sinh viên tình nguyện. Tính xác suất để:

a. Có 4 nam trong số 8 sinh viên được chọn?

b. Có nhiều nhất 3 sinh viên nam trong 8 sinh viên được chọn?

c. Có ít nhất 1 sinh viên nam trong 8 sinh viên được chọn?

d. Không có sinh viên nam trong 8 sinh viên được chọn?

2.6. Trên giá sách có 50 cuốn sách, trong đó có 3 cuốn sách của cùng một tác giả. Tìm xác suất để không có hai cuốn nào trong 3 cuốn đứng cạnh nhau.

2.7. Tại thành phố Đồng Hới biết rằng biển số xe gắn máy có 4 chữ số. Chọn ngẫu nhiên một biển số xe. Tính xác suất để chọn được biển số xe:

a. Có 4 chữ số khác nhau.

b. Có 2 chữ số giống nhau.

c. Có 3 chữ số giống nhau.

d. Có 4 chữ số giống nhau.

2.8. Một dãy ghế trong hội trường rạp chiếu phim có 20 chổ ngồi, xếp 20 người vào ngồi một cách ngẫu nhiên, trong đó có Lan và Điệp. Tính xác suất để:

a. Lan được ngồi ở hai đầu dãy ghế.

b. Lan và Điệp được ngồi gần nhau.

2.9. Trường X có số sinh viên học học tốt Toán và Anh văn của lớp A và B (biết mỗi lớp có 45 sinh viên) được cho như sau:

Học tốt \ lớp	A	B
Toán	25	25
Anh văn	30	30
Toán và Anh văn	20	10

Có đoàn kiểm định chất lượng đến thanh tra. Theo bạn, giáo viên nên mời đoàn vào lớp nào để khả năng gặp được một sinh viên học tốt ít nhất một môn là cao nhất?

2.10. Ba xạ thủ A, B, C độc lập với nhau cùng nổ súng vào một mục tiêu. Xác suất bắn trúng của xạ thủ tương ứng là 0,4; 0,5 và 0,6. Tính xác suất để:

a. Chỉ có duy nhất một xạ thủ bắn trúng.

b. ít nhất một xạ thủ bắn trúng.

2.11. Có hai túi đựng các quả cầu. Túi thứ nhất đựng 3 quả trắng, 7 quả đỏ và 15 quả xanh. Túi thứ hai đựng 10 quả trắng, 6 quả đỏ và 9 quả xanh. Từ mỗi túi chọn ngẫu nhiên một quả cầu. Tính xác suất để 2 quả cầu được chọn đều có cùng màu.

2.12. Một công ty có 60 nhân viên, trong đó có 20 nam và 40 nữ. Tỷ lệ nhân viên nữ có thể nói tiếng Anh lưu loát là 15% và tỷ lệ này đối với nam là 20%

a. Gặp ngẫu nhiên một nhân viên của công ty. Tìm xác suất để gặp được nhân viên nói tiếng Anh lưu loát?

b. Gặp ngẫu nhiên hai nhân viên của công ty. Tìm xác suất để có ít nhất một người nói tiếng Anh lưu loát trong số 2 người gặp?

2.13. Ba sinh viên An, Hùng, Oanh cùng làm bài thi môn Giải tích. Xác suất làm được bài của từng người lần lượt là $0,7; 0,6$ và $0,5$. Tìm xác suất để:

a. Có hai sinh viên làm được bài thi?

b. Nếu có hai sinh viên làm được bài thi, tìm xác suất để An không làm được bài?

2.14. Trong hồ có 10 con cá carnh (trong đó có 3 cá có đuôi màu đỏ và 7 cá có đuôi màu xanh). Bắt ngẫu nhiên từ hồ ra một con cá. Nếu bắt ra cá có đuôi màu đỏ thì bỏ vào hồ một con cá có đuôi màu xanh. Nếu bắt ra cá có đuôi màu xanh thì bỏ vào một cá có đuôi màu đỏ. Sau đó từ hồ bắt tiếp ra một con cá.

a. Tính xác suất để cá được bắt lần sau có đuôi màu đỏ?

b. Nếu hai con cá được bắt ra (lần 1 và lần 2) có đuôi cùng màu. Tính xác suất để hai con cá này có đuôi cùng màu xanh?

2.15. Trong một kỳ thi tốt nghiệp, mỗi sinh viên phi thi hai môn cơ sở và chuyên nghành. Giả sử xác suất bạn thi đạt môn c sở là 80%. Nếu thi đạt yêu cầu môn c sở thì xác suất thi đạt môn chuyên nghành là 60%. Nếu không đạt yêu cầu môn c sở thì hy vọng thi đạt môn chuyên nghành là 30%. Tìm xác suất để:

a. Thi đạt cả hai môn cơ sở và chuyên ngành?

b. Thi đạt môn chuyên nghành?

c. Thi đạt ít nhất một môn?

2.16. Chị Lan có một chùm chìa khóa gồm 9 chiếc bề ngoài rất giống nhau nhưng trong đó chỉ có 2 chiếc mở được cửa tủ. Chị Lan thử ngẫu nhiên từng chìa (chìa nào không đúng thì bỏ ra). Tìm xác suất để chị Lan mở được cửa ở lần thử thứ 3.

2.17. Hai anh em Hùng và An chi trò chi như sau: Mỗi người lần lượt rút một viên bi từ một hộp đựng 2 bi đỏ và 4 bi xanh. Bi được rút ra không trả lại vào hộp. Người nào lấy được bi đỏ trước thì thắng cuộc. Tính xác suất thắng cuộc của người rút trước.

2.18. Có 3 sinh viên nhưng chỉ có 2 vé đi xem phim. Họ làm 3 lá thăm, trong đó có 2 thăm có đánh dấu. Mỗi người lần lượt rút một thăm. Nếu ai rút được thăm có đánh dấu thì được vé đi xem phim. Hãy chứng minh sự công bông của cách làm này.

2.19. Trong một nhà máy có ba phân xưởng I, II, III tương ứng làm ra 25%, 35% và 40% tổng sản phẩm của nhà máy. Giả sử xác suất làm ra phế phẩm của phân xưởng I là 0,01, của phân xưởng II là 0,02 và phân xưng III là 0,025. Chọn ngẫu nhiên một sản phẩm của nhà máy. Tính xác suất để đó là phế phẩm?

2.20. Một phân xưởng có 3 máy. Xác suất để mỗi máy sản xuất ra sản phẩm đạt tiêu chuẩn kỹ thuật lần lượt là 0,9 ; 0,8 và 0,7. Trong một giờ mỗi máy sản suất được 5 sản phẩm. Tìm xác suất để trong một giờ c 3 máy sản xuất được ít nhất 14 sản phẩm đạt tiêu chuẩn kỷ thuật?

2.21. Có hai hộp sản phẩm, biết rằng:

Hộp thứ 1: có 7 chính phẩm và 3 phế phẩm.

Hộp thứ 2 : có 5 chính phẩm và 3 phế phẩm.

Lấy ngẫu nhiên một sản phẩm ở hộp thứ 1 bỏ vào hộp thứ 2 rồi sau đó từ hộp thứ 2 lấy ngẫu nhiên ra một sản phẩm thì được chính phẩm.

Tính xác suất để sản phẩm lấy ra từ hộp từ hộp thứ 2 là sản phẩm của hộp thứ nhất bỏ vào?

2.22. Có hai kiện hàng:

Kiện thứ nhất: có 5 sản phẩm loại A, 1 sản phẩm loại B.

Kiện thứ hai : có 2 sản phẩm loại A, 4 sản phẩm loại B.

Từ mỗi kiện hàng chọn ngẫu nhiên ra 1 sản phẩm đem giao cho khách hàng. Sau đó các sản phẩm còn lại được dồn vào kiện thứ ba (trống).

a. Chọn ngẫu nhiên một sản phẩm từ kiện hàng thứ ba. Tính xác suất để lấy được là sản phẩm loại B?

b. Nếu ta chọn ngẫu nhiên 2 sản phẩm từ kiện ba. Tính xác suất để có ít nhất một sản phẩm loại B từ 2 sản phẩm đã chọn.

2.23. Có ba hộp đựng các quả cầu:

Hộp thứ 1 có 10 quả cầu màu đỏ. Hộp thứ 2 có 5 quả cầu màu đỏ và 5 quả cầu màu xanh. Hộp thứ 3 có 10 quả cầu màu xanh.

Chọn ngẫu nhiên một hộp rồi từ hộp đó lấy ngẫu nhiên (không hoàn lại) ra 2 quả cầu thì được quả cầu màu xanh. Sau đó cũng từ hộp này lấy ngẫu nhiên ra một quả cầu. Tính xác suất để lấy được quả cầu màu xanh?

2.24. Một hộp có 10 sản phẩm (hoàn toàn không biết chất lượng của các sản phẩm trong hộp này). Mọi giả thiết về số sản phẩm tốt có trong hộp được xem là đồng khả năng. Lấy ngẫu nhiên(không hoàn lại) từ hộp ra 3 sản phẩm để kiểm tra thì thấy có cả 3 sản phẩm tốt. Theo bạn, có bao nhiêu sản phẩm tốt có trong 7 sản phẩm còn lại trong hộp? vì sao?

2.25. Một nhà máy có 3 phân xưởng cùng sản xuất một loại sản phẩm. Phân xưởng 1 sản xuất 25%; phân xưởng 2 sản xuất 25% và phân xưởng 3 sản xuất 50% sản phẩm của toàn nhà máy. Tỷ lệ phế phẩm của các phân xưởng 1, 2 và 3 lần lượt là: 1%, 5% và 10%. Lấy ngẫu nhiên một sản phẩm từ lô hàng do nhà máy sản xuất.

a. Tìm xác suất để lấy được phế phẩm? nêu ý nghĩa thực tế của xác suất này?

b. Nếu lấy được một chính phẩm, theo bạn sản phẩm đó do phân xưởng nào sản xuất? tại sao?

2.26. Có 3 lô hàng, mỗi lô gồm 10.000 sản phẩm. Tỷ lệ sản phẩm loại I của từng lô tương ứng là: 70%, 80% và 90%. Lấy từ mỗi lô ra 10 sản phẩm để kiểm tra (không hoàn lại). Nếu trong 10 sản phẩm lấy ra kiểm tra có từ 8 sản phẩm loại I trở lên thì mua lô hàng đó.

a. Tìm xác suất để lô hàng có tỷ lệ sản phẩm loại I là 70% được mua?

b. Tìm xác suất để có ít nhất một lô hàng được mua?

c. Nếu chỉ có một lô hàng được mua. Tìm xác suất để đó là lô hàng có tỷ lệ sản phẩm loại I là 70%?

Chương 3

ĐẠI LƯỢNG NGẪU NHIÊN VÀ QUY LUẬT PHÂN PHỐI XÁC SUẤT

3.1 ĐỊNH NGHĨA VÀ PHÂN LOẠI CÁC ĐẠI LƯỢNG NGẪU NHIÊN

3.1.1 Định nghĩa

Một đại lượng mà giá trị của nó là ngẫu nhiên, không dự đoán trước được gọi là một đại lượng ngẫu nhiên (ĐLNN), ký hiệu là X, Y, Z.

Khi thực hiện một phép thử, bằng một quy tắc (hay một hàm) ta có thể gán các giá trị bông số cho những kết quả của một phép thử. Các giá trị này được coi là giá trị của đại lượng ngẫu nhiên, ký hiệu là $x_1, x_2, ..., x_n, ...$

Ví dụ 3.1.1. Kiểm tra 3 sản phẩm. Gọi X là số phế phẩm trong 3 sản phẩm kiểm tra. X là ĐLNN vì khi kiểm tra 3 sản phẩm X sẽ nhận một và chỉ một trong số các giá trị: $0, 1, 2, 3$.

3.1.2 Phân loại đại lượng ngẫu nhiên

Một ĐLNN được gọi là rời rạc nếu tập giá trị mà nó có thể nhận là tập hữu hạn hoặc vô hạn đếm được. Đối với ĐLNN rời rạc ta có thể liệt kê được các giá trị của nó.

Ví dụ 3.1.2. Gieo một con xúc sắc. Gọi X là số nốt xuất hiện trên con xúc sắc, X là ĐLNN rời rạc. Ta có tập các giá trị có thể có của X là $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$.

Một ĐLNN được gọi là liên tục nếu tập giá trị mà nó có thể nhận được có thể lấp kín cả một khoảng trên trục số. Đối với ĐLNN liên tục, không thể liệt kê được các giá trị của nó.

Ví dụ 3.1.3. Chiều cao của học sinh trong một lớp học, khối lượng của một loại hoa quả là những ĐLNN liên tục.

3.2 Qui luật phân phối xác suất của ĐLNN

Để xác định một đại lượng ngẫu nhiên ta phải biết đại lượng ngẫu nhiên ấy có thể nhận giá trị nào và nó nhận giá trị ấy với xác suất tương ứng bao nhiêu. Một hệ thức cho phép biểu diễn mối quan hệ giữa các giá trị có thể nhận của ĐLNN với các xác suất tương ứng gọi là qui luật phân phối xác suất của ĐLNN.

Để thiết lập quy luật phân phối xác suất của ĐLNN, ta có thể dùng: bảng phân phối xác suất hay hàm phân phối xác suất hoặc hàm mật độ xác suất.

3.2.1 Bảng phân phối xác suất

Bảng phân phối xác suất dùng để thiết lập qui luật phân phối xác suất của ĐLNN rời rạc X. Giả sử ĐLNN X có thể nhận một trong các giá trị: $x_1, x_2, ..., x_n$ với xác suất tương ứng là: $p_1, p_2, ...p_n$. Hay là $p_i = p\left(X = x_i\right), i = 1..n$.

Khi đó bảng phân phối xác suất của X có dạng:

X	x_1	x_2	...	x_n
p	p_1	p_1	...	p_n

Ta luôn có $\sum\limits_{i=1}^{n} p_i = 1$.

Ví dụ 3.2.1. Một xạ thủ được phép bắn 3 viên đạn vào mục tiêu với xác suất trúng mục tiêu là 0,6. Anh ta bắn cho tới khi hết đạn hoặc trúng mục tiêu thì thôi. Gọi X là số viên đạn bắn ra. Lập bảng phân phối xác suất của X.

Giải. Tập giá trị có thể của X là $X = 1, 2, 3$.

Khi đó, các biến cố $(X = 1), (X = 2), (X = 3)$ lập thành một nhóm biến cố đầy đủ. Gọi A_i là biến cố bắn trúng mục tiêu ở lượt đạn thứ i $(1 \leqslant i \leqslant 3)$. Ta có

$$p(X = 1) = p(\bar{A}_1 A_2) = p(\bar{A}_1).P(A_2) = 0,4 \times 0,6 = 0,24$$
$$p(X = 2) = p(\bar{A}_1.A_2) = p(\bar{A}_1).P(A_2) = 0,4 \times 0,6 = 0,24$$
$$P(X = 3) = 1 - p(X = 1) - p(X = 2) = 1 - 0,6 - 0,24 = 0,16$$

Vậy bảng phân phối xác suất của X là:

X	1	2	3
p	0,6	0,24	0,16

Ví dụ 3.2.2. Trong hộp có 10 sản phẩm (trong đó có 6) chính phẩm. Lấy ngẫu nhiên không hoàn lại từ hộp ra 2 sản phẩm. Lập bảng phân phối xác suất của số chính phẩm được lấy ra.

Giải. Gọi X là số chính phẩm được lấy ra từ hộp thì X là ĐLNN rời rạc có tập giá trị có thể có là $X\left(\Omega\right) = \left\{0, 1, 2\right\}$.

Ta có $p_1 = p\left(X = 0\right) = \frac{C_4^2}{C_{10}^2} = \frac{2}{15}$; $p_2 = P\left(X = 1\right) = \frac{C_6^1 C_4^1}{C_{10}^2} = \frac{8}{15}$; $p_3 = P\left(X = 2\right) = \frac{C_6^2}{C_{10}^2} = \frac{5}{15}$

Vậy bảng phân phối xác suất của X:

X	0	1	2
p	2/15	8/15	5/15

3.2.2 Hàm phân phối xác suất

Hàm phân phối xác suất dùng để thiết lập qui luật phân phối xác suất của cả ĐLNN rời rạc và liên tục.

3.2.2.1 Định nghĩa.

Hàm phân phối của ĐLNN X được định nghĩa bởi biểu thức: $F(x) = p(X < x)$
Khi X là ĐLNN rời rạc thì: $F(x) = \sum\limits_{x_i<x} p(X < x_i) = \sum\limits_{x_i<x} p_i$

3.2.2.2 Tính chất.

Tính chất 1. Hàm phân phối xác suất luôn luôn nhận giá trị trong $[0, 1]$, tức là $0 \leqslant F(x) \leqslant 1$.

Tính chất 2. $F(x)$ là hàm không giảm.

Hệ quả 1. $p(a \leqslant X < b) = F(b) - F(a)$

Hệ quả 2. Nếu X là ĐLNN liên tục thì: $p(a \leqslant X \leqslant b) = p(a \leqslant X < b) = p(a < X \leqslant b) = p(a < X < b)$.

Tính chất 3. $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0; \quad F(+\infty) = \lim\limits_{x\to+\infty} F(x) = 1.$

3.2.2.3 ý nghĩa của hàm phân phối xác suất

Hàm phân phối xác suất $F(x)$ phản ánh mức độ tập trung xác suất về phía bên trái của điểm x. Giá trị của $F(x)$ cho biết cho biết có bao nhiêu phần của đơn vị xác suất phân phối trong khoảng $(-\infty,\ x)$.

Ví dụ 3.2.3. Tìm hàm phân phối xác suất cho trong ví dụ 3.2.2. Ta có hàm phân phối xác suất là

$$F(x) = \begin{cases} 0 & \text{khi } x \leqslant 0 \\ \frac{2}{15} & \text{khi } 0 < x \leqslant 1 \\ \frac{10}{15} & \text{khi } 1 < x \leqslant 2 \\ 1 & \text{khi } x > 2 \end{cases}$$

Ví dụ 3.2.4. Cho hàm phân phối của biến số ngẫu nhiên X là

$$F(x) = \begin{cases} 0 & \text{khi } x \leqslant a \\ \frac{x-a}{b-a} & \text{khi } a < x \leqslant b \\ 1 & \text{khi } x > b \end{cases}$$

với $0 < a < b$. Tính $P(\frac{a+b}{2} < X < a + b)$.

Giải. Ta có

$$p(\frac{a+b}{2} < X < a + b) = F(a + b) - F(\frac{a+b}{2}) = 1 - \frac{1}{2} = \frac{1}{2}.$$

3.2.3 Hàm mật độ xác suất

3.2.3.1 Định nghĩa

Hàm mật độ xác suất của đại lượng ngẫu nhiên liên tục X được định nghĩa như sau:

$$f(x) := F'(x)$$

3.2.3.2 Tính chất.

Giar sử biến số ngẫu nhiên X có hàm mật độ f và hàm phân phối F, ta có các tính chất sau:

i) $f(x) \geqslant 0$,

ii) $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$,

iii) Nếu F liên tục trên $\mathbb{R}$ thì $P(a \leqslant X \leqslant b) = \int\limits_{a}^{b} f(x)dx$,

3.2.3.3 ý nghĩa của hàm mật độ.

Từ định nghĩa hàm mật độ, ta có hệ thức xấp xĩ: $p(x \leqslant X \leqslant x+\Delta x) \approx f(x).\Delta x$

Tức là xác suất để X nhận giá trị thuộc một lân cận $(x,\ x+\Delta x)$ gần như tỷ lệ với giá trị của hàm mật độ $f(x)$ tại điểm x. Vậy, với cùng độ dài Δx, tại điểm x nào mà giá trị của hàm $f(x)$ lớn hơn thì lân cận của điểm ấy sẽ tập trung một xác suất lớn hơn nên $f(x)$ có tên là hàm mật độ xác suất.

3.3 Các tham số đặc trưng của đại lượng ngẫu nhiên

Khi ta xác định được qui luật phân phối xác suất của đại lượng ngẫu nhiên thì ta nắm được toàn bộ thông tin về đại lượng ngẫu nhiên đó. Tuy nhiên trong thực tế cũng rất khó và không cần thiết để nắm toàn bộ thông tin này mà chúng ta cần quan tâm đến những thông tin quan trọng nhất phn ánh đầy đủ các đặc trưng cơ bản của đại lượng ngẫu nhiên mà ta đang nghiên cứu.

3.3.1 Kỳ vọng toán

3.3.1.1 Định nghĩa.

Cho X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị $x_1, x_2, ..., x_n$ với các xác suất tương ứng là $p_1, p_2, ..., p_n$. Khi đó kỳ vọng toán của ĐLNN X được định nghĩa là: $E(X) = \sum\limits_{i=1}^{n} x_i p_i$

Nếu X là đại lượng ngẫu nhiên liên tục với hàm mật độ xác suất $f(x)$ thì: $E(X) = \int\limits_{-\infty}^{+\infty} xf(x)dx$

3.3.1.2 Tính chất.

1. $E(C) = C$,
2. $E(CX) = C.E(X)$
3. $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$
4. Nếu $X_1, X_2, ..., X_n$ là ĐLNN độc lập thì $E(X_1.X_2...X_n) = E(X_1).E(X_2)...E(X_n)$.

Chú ý. Hai ĐLNN được gọi là độc lập với nhau nếu quy luật phân phối xác suất của ĐLNN này không phụ thuộc vào ĐLNN kia nhận giá trị bông bao nhiêu.

3.3.1.3 ý nghĩa của kỳ vọng.

Để tìm hiểu bản chất của kỳ vọng, ta xét các ví dụ sau:

Ví dụ 3.3.1. Nghiên cứu về thu nhập của 500 công nhân nghành may, ta có số liệu như sau:

Thu nhập(tr.đ/năm)	7	8	9	10	11	12	14
Số công nhân	50	70	150	120	55	30	25

Gọi X là thu nhập của công nhân ngành may, từ số liệu trên ta có bảng phân phối xác suất:

X	7	8	9	10	11	12	14
p	0.1	0.14	0.3	0.24	0.11	0.26	0.05

Vậy ta có

$$\begin{aligned} E\left(X\right) &= 7.0,1 + 8.0,14 + 9.0,3 + 10.0,24 + 11.0,11 + 12.0,06 + 14.0,05 = 9,55 \\ &= (7.50 + 8.70 + 9.150 + 10.120 + 11.55 + 12.30 + 14.25)/500 \end{aligned}$$

Kết luận: Thu nhập trung bình của một công nhân nghành may là 9,55 triệu đ/năm. Vậy kỳ toán của ĐLNN chính là giá trị trung bình của ĐLNN đó.

Ví dụ 3.3.2. Một đợt xổ số người ta phát hành 10000 vé với giá vé là 1000 đồng/1vé. Trong đó chỉ có một vé trúng thưởng với trị giá là 50000 đồng. Một người nào đó mua mỗi lần 2 vé (lấy ngẫu nhiên). Tính số tiền trung bình người đó kiếm được trong mỗi lần xổ số.

Giải. Gọi X là số tiền người đó kiếm được trong mỗi lần xổ số.

Ta có tập giá trị của X là $-2000, 48000$.

Suy ra $p(X = -2000) \simeq C_2^0.(\frac{1}{10000})^0.(\frac{9999}{10000})^2 \simeq 0,9998$ và

$$p(X = 48000) \simeq 1 - 0,9998 = 0,0002$$

Vậy bảng phân phối của X:

X	-2000	48000
p	0,9998	0,0002

Suy ra $E(X) = (-2000) \times 0,9998 + 48000 \times 0,0002 = -1990.$

Vậy số tiền trung bình người mất đi 1990 đồng trong mỗi đợt sổ xố.

3.3.2 phương sai

Trong thực tế, nhiều khi nếu chỉ xác định kỳ vọng toán của ĐLNN thì chưa đủ. Để xác định một ĐLNN ta còn phi xác định mức độ phân tán các giá trị của ĐLNN xung quanh giá trị trung bình của nó.

3.3.2.1 Định nghĩa.

phương sai của ĐLNN ngẫu nhiên X, ký hiệu $D(X)$ (hay $Var(X)$), được xác định bởi:

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

Trong thực tế, người ta thường tính bông công thức:

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

Nếu X là ĐLNN rời rạc thì: $D(X) = \sum\limits_{i=1}^{n} [x_i - E(X)]^2 p_i$

Nếu X là ĐLNN liên tục thì: $D(X) = \int\limits_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx$

3.3.2.2 Tính chất.

i. $D(C) = 0$

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

iii. Nếu X, Y là các ĐLNN độc lập thì $D(X + Y) = D(X) + D(Y)$

iv. Nếu X, Y là các ĐLNN độc lập thì $D(X - Y) = D(X) + D(Y)$

v. $D(C + X) = D(X)$, C là hằng số.

3.3.2.3 ý nghĩa của phương sai.

Từ định nghĩa ta thấy phương sai là kỳ vọng toán của bình phương các sai lệch hay phương sai là sai lệch bình phương trung bình. Nó phản ánh mức độ phân tán các giá trị của ĐLNN xung quanh giá trị trung bình. Đại lượng nào có nhiều giá trị sai lệch lớn so với giá trị trung bình thì phương sai sẽ lớn và ngược lại.

Nhận xét: Ta thấy phương sai không cùng đơn vị đo với kỳ vọng cũng như X. Do đó, người ta đưa ra một đại lượng khác là $\sigma(X) = \sqrt{D(X)}$, được gọi là độ lệch chuẩn của biến số ngẫu nhiên X. Ta thấy cả $\sigma(X)$ và $D(X)$ đều đánh giá được mức độ phân tán hay mức độ tập trung của các giá trị của X xung quanh giá trị trung bình của nó.

3.3.3 Giá trị tin chắc nhất

Định nghĩa. Giá trị tin chắc nhất của ĐLNN rời rạc X, ký hiệu là $Mod(X)$ là giá trị của X ứng với xác suất lớn nhất trong bng phân phối xác suất.

Nếu X là ĐLNN liên tục có hàm mật độ xác suất $f(x)$ thì $Mod(X)$ là giá trị của X mà tại đó hàm mật độ đạt giá trị cực đại. Nhận xét. Từ định nghĩa $Mod(X)$ ta thấy $Mod(X)$ chính là giá trị có khả năng xảy ra nhiều nhất trong các giá trị mà ĐLNN X có thể nhận.

Ví dụ 3.3.3. ĐLNN X có qui luật phân phối xác suất như sau:

X	7	8	9	10	11	12	14
p	0.1	0.14	0.3	0.24	0.11	0.26	0.05

Ta thấy $P(X=9) = 0,3$ là lớn nhất vì vậy $Mod(X) = 9$.

3.3.4 Một số tham số đặc trưng khác

3.3.4.1 Mômen

Định nghĩa. Mômen cấp k được định nghĩa là: $\alpha_k = E\left[(X)^k\right]$.

Định nghĩa. Mô men trung cấp k được định nghĩa là: $\eta_k = E\left\{[X - E(X)]^k\right\}$.

Nhận xét: $\alpha_1 = E[(X)]$; $\eta_2 = E\left\{[X - E(X)]^2\right\} = D(X)$ và

$$\begin{aligned}
\eta_2 &= E\left(X^2\right) - [E(X)]^2 = \alpha_2 - (\alpha_1)^2 \\
\eta_3 &= \alpha_3 - 3\alpha_1.\alpha_2 + 2(\alpha_1)^3 \\
\eta_4 &= \alpha_4 - 4\alpha_1.\alpha_3 + 6(\alpha_1)^2\alpha_2 - 3(\alpha_1)^4
\end{aligned}$$

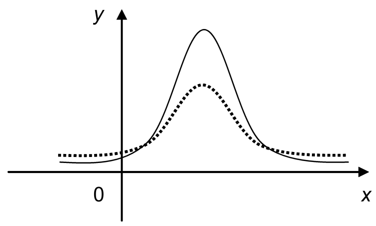
3.3.4.2 Hệ số bất đối xứng:

$$S = \frac{\mu_3}{\sigma^3}$$

3.3.4.3 Hệ số nhọn:

$$K = \frac{\mu_4}{\sigma^4} - 3$$

Nếu K càng lớn thì đồ thị của hàm mật độ của ĐLNN đó càng nhọn.



3.4 Một số qui luật phân phối xác suất thông dụng

3.4.1 Qui luật nhị thức

Giả sử tiến hành n phép thử độc lập, gọi A là biến cố nào đó mà ta cần quan tâm. Khi đó A có thể xảy ra hoặc không xảy ra. Giả sử $P(A) = p$, gọi X là số lần biến cố A xảy ra trong n phép thử. Khi đó X là ĐLNN rời rạc với miền giá trị là $X(\Omega) = \{0, 1, 2, \cdots, n\}$. Theo công thức Bernoulli ta có:

$$P_k = P\{X = k\} = C_n^k p^k (1-p)^{n-k} \tag{3.4.1}$$

Định nghĩa. ĐLNN X có phân bố nhị thức với tham số n, p, ký hiệu $X \sim B(n,p)$ nếu X là ĐLNN rời rạc nhận một trong các giá trị: $0, 1, ..., n$ và các xác suất tương ứng được tính theo công thức (3.4.1).

Nhận xét. Trong nhiều trưng hợp ta cần tính xác suất để ĐLNN X phân phối theo qui luật chuẩn nhận giá trị trong khoảng $(k, k+h)$, ta có công thức sau:

$$P(k \leqslant X \leqslant k+h) := P_k + P_{k+1} + \cdots + P_{k+h}$$

Định lý 3.4.1. *Nếu $X \sim B(n,p)$ thì*

i. $E(X) = np$

ii. $D(X) = npq$

iii. $np - q \leqslant Mod\,(X) \leqslant np + p$

Ví dụ 3.4.2. Trong thành phố Đồng Hới có 65% gia đình có máy giặt. Chọn ngẫu nhiên 12 gia đình và gọi X là số gia đình có máy giặt.

a) Tính xác suất để có đúng 5 gia đình có máy giặt.

b) Tính xác suất để có ít nhất 2 gia đình có máy giặt.

Giải. Nếu coi việc điều tra máy giặt trên là một phép thử thì ta có 12 phép thử độc lập. trong mỗi phép thử chỉ có hai khả năng xảy ra: hoặc có máy giặt hoặc không. Xác suất có máy giặt là $p = 0.65$. Ta có $X \sim B(12, 0.65)$. Vậy:

a) $P\{X = 5\} = C_{12}^{5}\,(0.65)^{5}\,(0.35)^{7} = 0,0591.$

b) $p\{X \geqslant 2\} = 1 - P\{X = 0\} - p\{X = 1\} = 1 - (0.35)^{12} - C_{12}^{1}\,(0.65)\,(0.35)^{11} = 0.999.$

Ví dụ 3.4.3. Một nhà máy sản xuất được 200 sản phẩm trong một ngày. Xác suất để sản xuất ra phế phẩm là 0.05. Tìm số phế phẩm trung bình và số phế phẩm tin chắc nhất của máy đó trong một ngày.

Giải. Gọi X là số phế phẩm của nhà máy trong một ngày thì $X \sim B(200, 0.05)$. Suy ra $E\,(X) = np = 200 \times 0.05 = 10$. Số phế phẩm tin chắc nhất của máy là $np - q \leqslant Mod\,(X) \leqslant np + p$ hay $9.05 \leqslant Mod\,(X) \leqslant 10.05$. Suy ra $Mod(X) = 10$.

3.4.2 Qui luật Poisson

Gi sử tiến hành n phép thử độc lập, gọi A là biến cố nào đó mà ta cần quan tâm. Khi đó A có thể xảy ra hoặc không xảy ra. Giả sử $P\,(A) = p$, gọi X là số lần biến cố A xảy ra trong n phép thử. Trường hợp n lớn p nhỏ ($p < 0.1$) và tích $np = \lambda$ không đổi. Ta có:

$$p_k = p\{X = k\} = C_n^k p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \tag{3.4.2}$$

Định nghĩa. ĐLNN X có phân bố Poisson với tham số λ, ký hiệu $X \sim \wp(\lambda)$ nếu X là ĐLNN rời rạc nhận một trong các giá trị: $0, 1, ., n$ và các xác suất tương ứng tính theo công thức (3.4.2).

Định lý 3.4.4. *Nếu thì*

i. $E\,(X) = \lambda$

ii. $D\,(X) = \lambda$

iii. $\lambda - 1 \leqslant Mod\,(X) \leqslant \lambda$

Ví dụ 3.4.5. Một xe ti vận chuyển 4000 chai Cocacola. Xác suất một chai Cocacola bị vỡ khi vận chuyển là 0,001.

a). Tìm xác suất để có không quá 3 chai bị vỡ khi vận chuyển.

b). Tìm số chai cocacola bị vỡ trung bình và số chai bị vỡ tin chắc nhất khi vận chuyển?

Giải. Gọi X là số chai cocacola bị vỡ khi vận chuyển 4000 chai. Khi đó X là ĐLNN và $X \sim B(4000, 0.001)$. Do $n = 4000$ khá lớn, $p = 0,001$ quá nhỏ và $\lambda = np = 4000 \times 0,001 = 4$ nên ta coi $X \sim \wp(4)$.

a) Xác suất để có không quá 3 chai bị vỡ khi vận chuyển là:

$$P\left(0 \leqslant X \leqslant 3\right) = P_0 + P_1 + P_2 + P_3.$$

Ta có $P_0 = P\left(X = 0\right) = \frac{4^0}{0!}e^{-4}$, $P_1 = \frac{4^1}{1!}e^{-4}$, $P_2 = P\left(X = 2\right) = \frac{4^2}{2!}e^{-4}$, $P_3 = P\left(X = 3\right) = \frac{4^3}{3!}e^{-4}$.

Vậy

$$p\left(0 \leqslant X \leqslant 3\right) = \left(1 + 4 + 8 + \frac{32}{3}\right)e^{-4} \approx ....$$

b). Số cocacola bị vỡ trung bình là hay có trung bình 4 chai cocacola bị vỡ khi vận chuyển 4000 chai. Số chai bị vỡ tin chắc nhất khi vận chuyển 4000 chai là

Suy ra $E\left(X\right) = \lambda = 4$.

Vậy số chai cocacola có khả năng bị vỡ nhiều nhất là 3 hoặc 4 chai khi vận chuyển 4000 chai.

3.4.3 Qui luật siêu bội

Từ một tập hợp gồm N phần tử, (trong đó có M phần tử có tính chất A nào đó). Lấy ngẫu nhiên không hoàn lại ra n phần tử. Gọi X là số phần tử có tính chất A có trong n phần tử lấy ra thì X là ĐLNN rời rạc có thể nhận các giá trị: $0, 1, 2, , n$ với xác suất tương ứng là:

$$P_k = P\left\{X = k\right\} = \frac{C_M^k.C_{N-M}^{n-k}}{C_N^n} \tag{3.4.3}$$

Trong đó,

$$\max\left\{0, M + n - N\right\} \leqslant k \leqslant \min\left\{n, M\right\}$$

Định nghĩa. ĐLNN X có siêu bội với tham số N, M, n, ký hiệu $X \sim H\left(N, M, n\right)$ nếu X là ĐLNN rời rạc nhận một trong các giá trị: $0, 1, ., n$ và các xác suất tương ứng tính theo công thức (3.4.3).

Định lý 3.4.6. *Nếu* $X \sim H\left(N, M, n\right)$ *thì*

1. $E\left(X\right) = np, \quad p = \frac{M}{N}$
2. $D\left(X\right) = npq\frac{N-n}{N-1}$

Nhận xét. Khi n bé so với N thì ta có thể chứng minh công thức xấp xỉ sau:

$$\frac{C_M^k.C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k q^{n-k}$$

Ví dụ 3.4.7. Chọn ngẫu nhiên 7 lá bài từ một bộ bài có 52 lá. Tính xác suất để có 3 lá rô.

Giải. Gọi X là số lá rô trong 7 lá bài được chọn ra, thì X là ĐLNN rời rạc có thể nhận các giá trị $0, 1, 2, 3, 4, 5, 6, 7$. Do trong bộ bài 52 con có 13 lá rô nên $X \sim H(52, 13, 7)$. Vậy

$$P_3 = P\{X = 3\} = \frac{C_{13}^3.C_{39}^4}{C_{52}^7}$$

Ví dụ 3.4.8. Một công ty nhập về một lô hàng có 1000 sản phẩm, trong đó có 800 sản phẩm loại A và 200 sản phẩm loại B. Lấy ngẫu nhiên (không hoàn lại) từ lô hàng 10 sản phẩm để kiểm tra. Tìm xác suất để có ít nhất 8 sản phẩm loại A trong 10 sản phẩm lấy ra kiểm tra?

Giải. Gọi X là số sản phẩm loại A có trong 10 sản phẩm lấy ra để kiểm tra, vì lấy không hoàn lại nên X là ĐLNN có $X \sim H(1000, 800, 10)$.

Vì $n = 10$ khá bé so với $N = 1000$ nên ta có thể xấp xỉ $X \sim B(10\,; 0,8)$. Ta có

$$\begin{aligned} p(X \geqslant 8) &= P(X = 8) + P(X = 9) + P(X = 10) \\ &= C_{10}^8 (0,8)^8 (0,2)^2 + C_{10}^9 (0,8)^9 (0,2) + C_{10}^{10} (0,8)^{10} \\ &= 0,30199 + 0,268435 + 0,107374 = 0,6778. \end{aligned}$$

BÀI TẬP CHƯƠNG 3

3.1. Một hợp tác xã nông nghiệp có hai ô tô vận ti hoạt động. Xác suất bị hỏng trong ngày làm việc của các ô tô tương ứng là 0,1 và 0,2. Gọi X là số ô tô bị hỏng trong thời gian làm việc.

a. Lập bảng phân phối xác suất của X.

b. Thiết lập hàm phân phối xác suất và vẽ đồ thị của nó.

3.2. Một dây chuyền gồm 3 bộ phận hoạt động độc lập với nhau. Xác suất trong thời gian 1 tuần các bộ phận bị hỏng tương ứng là 0,4; 0,2 và 0,3. Gọi X là số bộ phận bị hỏng trong thời gian 1 tuần làm việc.

a. Lập bảng phân phối xác suất của X.

b. Thiết lập hàm phân phối xác suất và vẽ đồ thị của nó.

c. Tính xác suất trong thời gian 1 tuần có không quá 2 bộ phận bị hỏng.

d. Tính giá trị tin chắc nhất của X.

3.3. Có hai lô sản phẩm:

Lô 1: Có 8 chính phẩm và 2 phế phẩm.

Lô 2: Có 7 chính phẩm và 3 phế phẩm.

Từ lô thứ nhất lấy ngẫu nhiên 2 sản phẩm bỏ sang lô thứ hai, sau đó từ lô thứ hai lấy ra 2 sản phẩm. a. Tìm qui luật phân phối xác suất của số chính phẩm được lấy ra.

b. Xây dựng hàm phân phối xác suất của số chính phẩm được lấy ra và vẽ đồ thị.

3.4. Có hai kiện hàng.

Kiện thứ nhất: Có 4 sản phẩm loại A và 8 sản phẩm loại B.

Kiện thứ hai : Có 3 sản phẩm loại A và 5 sản phẩm loại B

Lấy ngẫu nhiên 2 sản phẩm từ kiện thứ nhất bỏ vào kiện thứ hai, sau đó từ kiện thứ hai lấy không hoàn lại ra 3 sản phẩm. Gọi X là số sản phẩm loại A có trong 3 sản phẩm lấy ra từ kiện thứ hai.

a. Lập bảng phân phối xác suất của X.

b. Tính $E(\mathrm{X});v\mathrm{ar}(\mathrm{X})$.

3.5. Có hai kiện hàng, mỗi kiện có 5 sản phẩm.

Kiện thứ nhất: Có 2 sản phẩm loại A.

Kiện thứ hai: Có 3 sản phẩm loại A.

Lấy ngẫu nhiên 1 sản phẩm từ kiện thứ nhất bỏ vào kiện thứ hai, sau đó từ kiện thứ hai lấy ngẫu nhiên 2 sản phẩm bỏ vào kiện 1 thứ nhất. Lập bảng phân phối xác suất của số sản phẩm loại A trong kiện thứ nhất?

3.6. Có hai hộp đựng bi:

Hộp thứ nhất: có 2 bi xanh và 4 bi vàng.

Hộp thứ hai: có 4 bi xanh

Rút ngẫu nhiên 2 bi từ hộp thứ nhất bỏ sang hộp thứ hai, sau đó từ hộp thứ hai lấy ngẫu nhiên 3 bi bỏ vào hộp thứ nhất. Gọi lần lược là số bi xanh có ở hộp thứ nhất, thứ hai sau khi thực hiện phép thử. Tìm qui luật phân phối xác suất của X_1, X_2.

3.7. Có 3 kiện hàng.

Kiện thứ nhất: Có 8 sản phẩm loại A và 2 sản phẩm loại B.

Kiện thứ hai: Có 5 sản phẩm loại A và 5 sản phẩm loại B.

Kiện thứ ba: Có 3 sản phẩm loại A và 7 sản phẩm loại B.

a. Chọn ngẫu nhiên một kiện rồi từ kiện đã chọn lấy ngẫu nhiên không hoàn lại ra 2 sản phẩm thì được 2 sản phẩm loại A. Lấy tiếp từ kiện đã chọn ra 2 sản phẩm.

Tìm qui luật phân phối xác suất của số sản phẩm loại A có trong 2 sản phẩm lấy ra lần sau?

b. Chọn ngẫu nhiên 2 kiện rồi từ 2 kiện đã chọn lấy ngẫu nhiên không hoàn lại mỗi kiện 1 sản phẩm.Tìm qui luật phân phối xác suất của số sản phẩm loại A có trong 2 sản phẩm lấy ra?

3.8. Có 3 kiện hàng, mỗi kiện chứa 10 sản phẩm. Số sản phẩm loại B trong mỗi kiện tương ứng là 1,2,3.

a. Lấy ngẫu nhiên từ mỗi kiện ra một sản phẩm. Tìm qui luật phân phối xác suất của số sản phẩm loại A có trong 3 sản phẩm lấy ra?

b. Chọn ngẫu nhiên một kiện rồi từ kiện đã chọn lấy ngẫu nhiên không hoàn lại 3 sản phẩm. Tìm qui luật phân phối xác suất của số sản phẩm loại B có trong 3 sản phẩm lấy ra?

3.9. Lãi suất thu được trong một năm (tính theo %) khi đầu tư vào công ty A, công ty B tương ứng là các ĐLNN X và Y (X và Y độc lập). Cho biết qui luật phân phối xác suất của X và Y như sau:

X	4	6	8	10	12
p	0.05	0.1	0.3	0.4	0.15

Y	-4	2	8	10	12	16
p	0.1	0.2	0.2	0.25	0.15	0.1

a. Đầu tư vào công ty nào có lãi suất kỳ vọng cao hơn?

b. Đầu tư vào công ty nào có mức độ rủi ro ít hơn? Vì sao?

c. Nếu muốn đầu tư vào cả hai công ty thì nên đầu tư theo tỷ lệ như thế nào để cho:

(1) Thu được lãi suất kỳ vọng cao nhất

(2) Mức độ rủi ro về lãi suất ít nhất.

3.10. Số tiền lời trong năm tới (tính theo đơn vị: triệu đồng) thu được khi đầu tư 100 triệu đồng vào hai ngành A và B tùy thuộc vào tình hình kinh tế trong nước và cho ở bảng sau:

Số tiền lời╲ Tình hình kinh tế	Kém phát triển	ổn định	Phát triển
Ngành A	10	40	80
Ngành B	-30	70	110

a. Số tiền lời kỳ vọng ngành nào là cao hơn?

b. Mức độ rủi ro ngành nào là ít hơn?

Chương 4

ĐẠI CƯƠNG VỀ THỐNG KÊ TOÁN

4.1 TỔNG THỂ VÀ MẪU NGẪU NHIÊN

4.1.1 Tổng thể

Tập hợp các phần tử mà ta cần kho sát về một hay một số dấu hiệu nào đó gọi là một tổng thể.

Giả sử ta nghiên cứu về dấu hiệu X^* nào đó của các phần tử trong tổng thể. Dấu hiệu X^* này thường được cụ thể hóa thành các giá trị thay đổi từ phần tử này qua phần tử khác. Chúng lập nên một đại lượng ngẫu nhiên X nào đó. Khi đó có thể coi việc nghiên cứu dấu hiệu X* trên các phần tử của tổng thể là xét ĐLNN X trên tổng thể.

4.1.2 phương pháp mẫu.

Gi sử cần nghiên cứu ĐLNN X trên một tổng thể gồm N phần tử.

Người ta thường sẽ chọn ngẫu nhiên có hoàn lại n phần tử. Ta gọi đó là một mẫu kích thước n (nói chung n là nhỏ so với N).

Lần lượt đo giá trị X trên các phần tử cụ thể của mẫu, nghiên cứu kết quả và bông các phương pháp khoa học, rút ra kết luận về ĐLNN X trên tổng thể đó.

Ghi chú: Khi N quá lớn so với n, người ta có thể chọn ngẫu nhiên n phần tử không hoàn lại.

Từ đây ta qui ước chỉ xét các tổng thể có số phần tử N rất lớn.

4.1.3 Mẫu ngẫu nhiên kích thước n

Giả sử ĐLNN X (xét trên một tổng thể) có qui luật phân phối cụ thể nào đó. Ta dự định sẽ chọn ngẫu nhiên n phần tử.

Giá trị X (chưa biết trước) trên phần tử thứ 1 là ĐLNN X_1 có cùng qui luật phân phối với X.

Giá trị X (chưa biết trước) trên phần tử thứ n là ĐLNN X_n có cùng qui luật phân phối với X.

Ta gọi bộ thứ tự gồm n ĐLNN $X_1, X_2, ..., X_n$ có cùng phân phối với ĐLNN X là một mẫu ngẫu nhiên kích thước n của ĐLNN X. Ký hiệu $W_X = (X_1, X_2, , X_n)$.

Lấy ngẫu nhiên n phần tử cụ thể trong tổng thể và đo giá trị X trên cụ thể các phần tử thứ 1, thứ 2,... thứ n này, ta được một mẫu cụ thể là $(x_1, x_2, ..., x_n)$.

X ĐLNN trên tổng thể	Mẫu $(X_1, X_2, ..., X_n)$	Mẫu cụ thể $(x_1, x_2, ..., x_n)$
+Trung bình tổng thể $E(X) \equiv M(X) = \mu$	+Trung bình mẫu $\overline{X} = \frac{X_1+...+X_n}{n}$	+Trung bình mẫu cụ thể $\overline{x} = \frac{x_1+...+x_n}{n}$
+phương sai tổng thể $D(X) = \sigma^2$	Phương sai mẫu (hiệu chỉnh) $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\left(\sum\limits_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$	+Phương sai mẫu cụ thể $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$
+Độ lệch tổng thể $\sigma = \sqrt{D(X)}$	+Độ lệch mẫu $S = \sqrt{S^2}$	+Độ lệch mẫu cụ thể $s = \sqrt{s^2}$
+Tỉ lệ tổng thể $P(A) = \dfrac{M_A}{N} = p = E(X)$	+Tỉ lệ mẫu $F_n = \dfrac{X_1 + ... + X_n}{n}$	+Tỉ lệ mẫu cụ thể $f = \dfrac{x_1 + ... + x_n}{n} = \dfrac{m_A}{n}$

M_A là số phần tử có tính chất A trong tổng thể; m_A: số phần tử có tính chất A trong mẫu cụ thể; ở đây X: ĐLNN với X = 1 nếu phần tử có tính chất A, X = 0 nếu phần tử không có tính chất $A(\bullet)$.

+Có thể nói $\overline{x}, s^2, s,\ f$ ở trên lần lượt là giá trị của trung bình mẫu, phương sai mẫu, độ lệch mẫu, tỉ lệ mẫu ứng với mẫu cụ thể (kích thước n).

+Có thể hiểu: tỉ lệ mẫu F_n là tỉ lệ phần tử có tính chất A trong mẫu sẽ lấy ngẫu nhiên n phần tử.

Giá trị tỉ lệ mẫu f là tỉ lệ phần tử có tính chất A trong mẫu cụ thể n phần tử.

Tính chất. Trung bình mẫu, phương sai mẫu, độ lệch mẫu, tỉ lệ mẫu là các ĐLNN và

i) $E\left(\overline{X}\right) = E(X) = \mu$; $D\left(\overline{X}\right) = D(X)/n = 2/n$

ii) $E(S^2) = D(X) = \sigma^2$;

iii) Cho X là ĐLNN xác định bởi $(\bullet)$. Ta có $E(F_n) = E(X) = p$ và $D(F_n) = D(X)/n = p(1-p)/n$.

Chứng minh. i)

$$E\left(\overline{X}\right) = E([X_1 + ... + X_n]/n) = (E(X_1) + ... + E(X_n))/n$$
$$= (E(X) + ... + E(X))/n = E(X)$$

$$D\left(\overline{X}\right) = (D(X_1) + ... + D(X_n))/n^2 = (D(X) + ... + D(X))/n^2 = D(X)/n.$$

ii) Xem [1,2,3]. iii) Dễ chứng minh $E(X) = p$ và $D(X) = p(1-p)$. Tiếp theo, từ i) suy ra iii).

Ví dụ 4.2.1. Điều tra năng suất lúa trong 1 vùng lớn. Kiểm tra 100ha trồng lúa được chọn ngẫu nhiên, người ta thu được số liệu:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích tưng ứng	10	20	30	15	10	10	5

Tính giá trị trung bình mẫu và phương sai mẫu, độ lệch mẫu của mẫu cụ thể trên. Ta lập bảng:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
41	10	410	16810
44	20	880	38720
45	30	1350	60750
46	15	690	31740
48	10	480	23040
52	10	520	27040
54	5	270	14580
Σ	$n = 100$	4600	212680

Giá trị trung bình mẫu $\overline{x} = \dfrac{4600}{100} = 46$ (tạ/ha)

Giá trị phương sai mẫu $s^2 = \frac{1}{99}\left[212680 - 100.46^2\right] = 10,91.$

Giá trị độ lệch mẫu $s = \sqrt{10,91} \approx 3,3$ (tạ/ha)

Ghi chú. Khi tính $\overline{x}$ và s^2, nếu số liệu của mẫu cho theo khoảng thì ta thay khoảng đó bằng giá trị trung tâm của khoảng (tức là trung bình cộng của 2 đầu khoảng).

Ví dụ 4.2.2. Để tìm hiểu số lượng mủ cao su cho mỗi ngày của mỗi cây trên một vùng (trong năm đầu khai thác), người ta kiểm tra ngẫu nhiên 100 cây cao su và có số liệu

Lượng mủ (g)	200-210	210-220	220-230	230-240	240-250	250-260	260-270
Số cây	2	8	14	30	25	12	9

Tìm trung bình, phương sai và độ lệch mẫu cụ thể trên.

Ta có bảng

x_i	n_i	n_ix_i	$n_ix_i^2$
205	2	410	84050
215	8	1720	369800
225	14	3150	708750
235	30	7050	165675
245	25	6125	150062
255	12	3060	780300
265	9	2385	632025
Σ	$n = 100$	23900	5732300

Từ bảng ta suy ra

Giá trị trung bình mẫu $\overline{x} = \dfrac{23900}{100} = 239(g)$

Giá trị phương sai mẫu

$$s^2 = \frac{1}{99}(5732300 - 100 \times 239^2) = 204,0404$$

Giá trị độ lệch mẫu

$$s = \sqrt{204,0404} \approx 14,28(g)$$

4.3 ƯỚC LƯỢNG ĐIỂM

Giả sử ĐLNN X của tổng thể có một đặc trưng số θ chưa biết và cần ước lượng. Ta có thể ước lượng θ bông một con số nào đó xấp xỉ θ. Cách ước lượng như vậy gọi là ước lượng điểm.

4.3.1 phương pháp hàm ước lượng điểm.

Giả sử ĐLNN X của tổng thể có tham số θ cần phải ước lượng điểm.

Từ ĐLNN X ta lập mẫu ngẫu nhiên kích thước n là $(X_1, ..., X_n)$. Lập hàm ước lượng $\widehat{\theta} = f(X_1, ..., X_n)$ là hàm theo các ĐLNN $X_1, ..., X_n$.

Chọn mẫu cụ thể $(x_1, ..., x_n)$. Giá trị hàm ước lượng là $f(x_1, ..., x_n)$ và đây là ước lượng điểm của θ.

Trong thực tế,

+ Để ước lượng trung bình tổng thể $E(X)(M(X))$ ta dùng hàm ước lượng là trung bình mẫu $\overline{X}$.

+ Để ước lượng tỉ lệ tổng thể $p = p(A)$ ta dùng hàm ước lượng là tỉ lệ mẫu F_n.

+ Để ước lượng phương sai tổng thể $D(X)$ ta dùng hàm ước lượng là phương sai mẫu S^2.

Chọn mẫu cụ thể $(x_1, ..., x_n)$. Rồi tính giá trị hàm ước lượng. Đó là ước lượng điểm cần tìm.

4.3.2 Ước lượng không chệch.

Hàm ước lượng $\widehat{\theta} = f(X_1, ..., X_n)$ gọi là ước lượng không chệch của tham số θ nếu giá trị trung bình $E(\widehat{\theta}) = \theta$.

ý nghĩa. Nếu $\widehat{\theta} = f(X_1, ..., X_n)$ là ước lượng không chệch của tham số θ thì sai số trung bình của ước lượng là $E(\widehat{\theta} - \theta) = \theta - \theta = 0$.

Các hàm ước lượng sử dụng trong thực tế đã nói ở trên đều là ước lượng không chệch.

Do $E\left(\overline{X}\right) = E(X)$; $E(S^2) = D(X)$; $E(F_n) = p$ (từ tính chất của $\overline{X}$, S^2 và F_n).

Ghi chú. Người ta chứng minh được với mẫu cụ thể có kích thước n đủ lớn,

+ Trung bình mẫu cụ thể $\overline{x}$ xấp xỉ trung bình tổng thể $E(X)$.

+ Tỉ lệ mẫu cụ thể f xấp xỉ tỉ lệ tổng thể $p = p(A)$.

+ phương sai mẫu cụ thể s^2 xấp xỉ phương sai tổng thể $D(X)$.

Ví dụ 4.3.1. Xét lại ví dụ 4.2.1 ở trên.

a)Hãy ước lượng năng suất trung bình của lúa trong vùng.

b)Hãy ước lượng độ thiếu đồng đều của năng suất lúa trong vùng.

Từ kết quả ví dụ 4.2.1, ta có năng suất lúa trong vùng ước chừng là $\overline{x} = 46$ (tạ/ha).

Độ thiếu đồng đều năng suất lúa trong vùng ước chừng $s \approx 3,3$(tạ/ha) (hoặc $s^2 = 10,91$).

4.4 ƯỚC LƯỢNG KHOẢNG

Giả sử ĐLNN X của tổng thể có đặc trưng số θ chưa biết và cần ước lượng. Ta chọn một khoảng (θ_1, θ_2) sao cho nó chứa θ với xác suất cao (tức là $p(\theta_1 < \theta < \theta_2) = 1 - \alpha$, với $\alpha > 0$ đủ bé).

Cách ước lượng như vậy gọi là ước lượng khoảng.

Ghi chú. Khoảng (θ_1, θ_2) gọi là khoảng tin cậy và $1 - \alpha$ gọi là độ tin cậy của ước lượng trên.

4.4.1 Ước lượng trung bình tổng thể.

Giả sử ĐLNN X có trung bình tổng thể $E(X) = \mu$ chưa biết. Cần ước lượng khoảng tin cậy của μ với độ tin cậy $1-\alpha$, với $\alpha > 0$ đủ bé.

Lập mẫu ngẫu nhiên $(X_1, ..., X_n)$.

$n > 30$ **(hoặc** $n < 30$ **nhưng X có phân phối chuẩn) biết** σ^2	$n \geq 30$ **và chưa biết phưng sai** σ^2**.**	$n < 30$, **chưa biết phưng sai** σ^2 **và X có phân phối chuẩn.**
Xét $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$. Người ta chứng minh được $Z \sim N(0,1)$ Có thể lập luận(*) đưa ra **Qui tắc tìm khoảng tin cậy** Gsử có mẫu kích thước n +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ +tính độ chính xác $\varepsilon = z_\alpha . \frac{\sigma}{\sqrt{n}}$ +Khoảng tin cậy của μ là $(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$với độ tin cậy $1-\alpha$	Xét $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ với S độ lệch mẫu. C/m được $Z \sim N(0,1)$. Tương tự lập luận(*)đưa ra **Qui tắc tìm khoảng tin cậy** Gsử có mẫu kích thước n +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ +tính độ chính xác $\varepsilon = z_\alpha . \frac{s}{\sqrt{n}}$ +Khoảng tin cậy của μ là $(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$với độ tin cậy $1-\alpha$	Xét T = với S độ lệch mẫu. C/m được T có phân phối Student với $n-1$ bậc tự do. **Qui tắc tìm khoảng tin cậy** Gsử có mẫu kích thước n +Tra bảng phân phối Student $(n-1)$ bậc tự do tìm t_α sao cho $P(\|T\| > t_\alpha) = \alpha$ +tính độ chính xác $\varepsilon = t_\alpha . \frac{s}{\sqrt{n}}$ +Khoảng tin cậy của μ là $(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$với độ tin cậy $1-\alpha$

Lập luận(*)

$P(|Z| > z) = \alpha \Leftrightarrow P(|Z| \leq z_\alpha) = 1-\alpha \Leftrightarrow P(\overline{X} - z_\alpha.\sigma/\sqrt{n} \leqslant \mu \leqslant \overline{X} + z_\alpha.\sigma/\sqrt{n}) = 1-\alpha$.

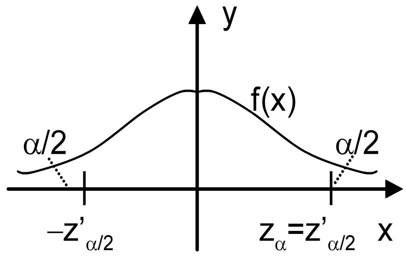
Ví dụ 4.4.1. Điều tra năng suất lúa trên 100 ha lúa được chọn ngẫu nhiên của một vùng lớn, người ta tính được trung bình mẫu cụ thể $\overline{x} = 46$ tạ/ha và độ lệch mẫu cụ thể $s = 3,3$ tạ.

Hãy ước lượng năng suất lúa trung bình cả vùng với độ tin cậy 95%. Gọi X (tạ/ha) là năng suất lúa trên các ha lúa của vùng lớn đó. $E(X) = \mu$ (tạ/ha) là năng suất lúa trung bình toàn vùng. Cần ước lượng μ với độ tin cậy $1-\alpha = 0,95$.

TH này $n = 100 > 30$ và σ^2 chưa biết. Ta làm như sau

+ $\alpha = 0,05$. Tìm z_α sao: $P(|Z| > z_\alpha) = \alpha \Rightarrow z_\alpha = 1,96$.

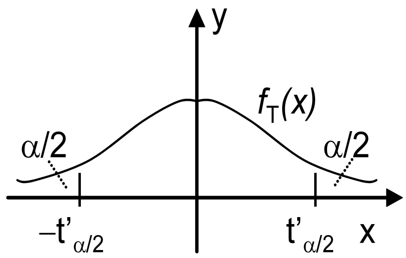
$$\varepsilon = z_\alpha . \frac{s}{\sqrt{n}} = 1,96.\frac{3,3}{\sqrt{100}} = 0,6468 \approx 0,65$$



+Khoảng tin cậy của μ là $(\overline{x} - \varepsilon, \overline{x} + \varepsilon) = (460,65; 46 + 0,65)$.

Năng suất lúa trung bình cả vùng ước chừng 45,35 đến 46,65 (tạ/ha) với độ tin cậy 95%.

Ví dụ 4.4.2. Lấy ngẫu nhiên 15 bao thuốc bột do một công ty dược sản xuất ra. Ta tính được trung bình mẫu cụ thể 39,8 g và phương sai mẫu cụ thể là 0,144. Giả thiết trọng lượng các bao thuốc bột là ĐLNN có phân phối chuẩn. Hãy ước lượng trọng lượng trung bình của các bao thuốc do công ty sản xuất với độ tin cậy 95%. Gọi X (g) là trọng lượng mỗi bao thuốc do công ty sản xuất. $E(X) = \mu(g)$ là trọng



lượng trung bình của mỗi bao. Cần ước lượng μ với độ tin cậy 95%.

TH này $n = 15 < 30$, σ^2 chưa biết và X có phân phối chuẩn.

+Ta có $\alpha = 0,05$.

Tra bảng phân phối Student với bậc tự do $n - 1 = 14$, Tìm t_α sao:$P(|T| > t_\alpha) = 1 - \alpha$, ta có $t_\alpha = 2,145$.

$$\varepsilon = t_\alpha \,.\frac{s}{\sqrt{n}} = 2,145.\frac{\sqrt{0,144}}{\sqrt{15}} = 0,21$$

+Khoảng tin cậy $(\overline{x} - \varepsilon, \overline{x} + \varepsilon) = (39,80,21; 39,8 + 0,21)$.

Vậy trọng lượng trung bình của các bao thuốc khoảng 39,59 đến 40,01g với độ tin cậy 95%.

4.4.2 Ước lượng tỉ lệ tổng thể.

Giả sử tỉ lệ các phần tử có tính chất A trong một tổng thể là $p = P(A)$ mà ta chưa biết.

Cần ước lượng khoảng tin cậy của tỉ lệ p với độ tin cậy $1-\alpha$, với $\alpha > 0$ đủ bé.

Xét ĐLNN X trên tổng thể xác định như sau $X = 1$ nếu phần tử kiểm tra có tính chất $A, X = 0$ nếu phần tử kiểm tra không có tính chất A. Lập mẫu ngẫu nhiên kích thước n đủ lớn $(X1, ..., Xn)$.

Xét ĐLNN $Z = \dfrac{F_n - p}{\sqrt{p(1-p)}/\sqrt{n}}$, với F_n là tỉ lệ mẫu.

Người ta chứng minh được $Z \sim N(0,1)$. Lập luận tương tự trên và để ý với n đủ lớn, ta có $f \approx p$,ta suy ra cách tìm khoảng tin cậy:

+Tìm tần suất f xuất hiện tính chất A trong mẫu cụ thể.

+Tra bảng phân phối chuẩn tắc tìm trị z_α sao cho $P(|Z| > z_\alpha) = \alpha$.

+Tính độ chính xác

$$\varepsilon = z_\alpha . \frac{\sqrt{f(1-f)}}{\sqrt{n}}$$

+Khoảng tin cậy của tỉ lệ p là $(f-\varepsilon, f+\varepsilon)$ với độ tin cậy $1-\alpha$.

Ví dụ 4.4.3. Nghiên cứu nhu cầu tiêu dùng một mặt hàng ở một thành phố, người ta điều tra trên 1000 người được chọn ngẫu nhiên của thành phố và thấy có 600 người có nhu cầu. Hãy ước lượng tỉ lệ người có nhu cầu về mặt hàng đó trong toàn thành phố với độ tin cậy 95%.

Gọi tỉ lệ người có nhu cầu về mặt hàng này là p. Ta cần ước lượng p với độ tin cậy $1-\alpha = 0,95$.

+Tỉ lệ người có nhu cầu này trong mẫu cụ thể là $f = 600/1000 = 0,6$.

+Tìm z_α sao $P(|Z| > z_\alpha) = \alpha = 0,05 \Rightarrow z_\alpha = 1,96$. Và

$$\varepsilon = z_\alpha . \frac{\sqrt{f(1-f)}}{\sqrt{n}} = 1,96 . \frac{\sqrt{0,6.0,4}}{\sqrt{1000}} = 0,03$$

+Khoảng tin cậy của p là $(0,6-0,03; 0,6+0,03) = (0,57; 0,63)$.

Vậy tỉ lệ người trong thành phố có nhu cầu về mặt hàng là 57% đến 63% với độ tin cậy là 95%.

4.4.3 Ước lượng phương sai tổng thể.

Giả sử ĐLNN X có phân phối chuẩn và chưa biết phương sai σ^2 của nó. Hãy ước lượng khong tin cậy của phương sai σ^2 với độ tin cậy $1-\alpha$ với α đủ bé.

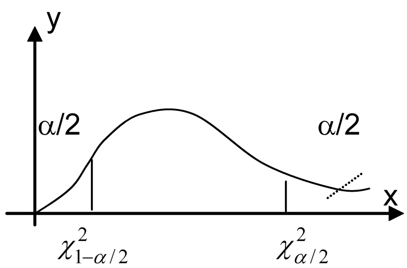
Từ X ta lập mẫu ngẫu nhiên kích thước n là $(X1, ..., Xn)$.

TH1. Biết trung bình tổng thể $E(X)=\mu$.	**TH2. Không biết trung bình tổng thể $E(X)$.**
Xét ĐLNN $\chi^2 = \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$. Người ta c/m nó có phân phối χ^2 với n bậc tự do **Qui tắc tìm khoảng tin cậy sau** +Tìm các số $\chi^2_{\alpha/2}$ và $\chi^2_{1\text{ - }\alpha/2}$ trong bảng phân phối χ^2 với n bậc tự do. +Tính trị số $u=\sum_{i=1}^{n}(x_i-\mu)^2$ +Khoảng tin cậy của phương sai σ^2 là $\left(\frac{u}{\chi^2_{\alpha/2}};\frac{u}{\chi^2_{1-\alpha/2}}\right)$	Xét ĐLNN $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}$. Người ta c/m nó có phân phối "Khi bình phương" với $n-1$ bậc tự do. **Qui tắc tìm khoảng tin cậy sau** +Tìm các số $\chi^2_{\alpha/2}$ và $\chi^2_{1\text{ - }\alpha/2}$ trong bảng phân phối χ^2 với $n-1$ bậc tự do. +Tính trị số $u=(n-1)s^2$ +Khoảng tin cậy của phương sai σ^2 là $\left(\frac{u}{\chi^2_{\alpha/2}};\frac{u}{\chi^2_{1-\alpha/2}}\right)$

Ví dụ 4.4.4. Lượng hao phí nguyên liệu cho một đơn vị sản phẩm là ĐLNN X có phân phối chuẩn với giá trị trung bình là 20 (g). Quan sát ngẫu nhiên 25 sản phẩm, ta có số liệu cho ở bảng

Trọng lượng nguyên liệu hao phí (g)	19.5	20	20.5
Số sản phẩm	5	18	2

Với độ tin cậy 90%, hãy ước lượng phương sai $D(X)=\sigma^2$. Đây là bài toán ước



lượng phương sai biết giá trị trung bình $\mu=20$.

+ $1+\alpha=0,9\Rightarrow\alpha=0,1$, $\alpha/2=0,05$ và $1-\alpha/2=0,95$

Tra bảng χ^2 với bậc tự do $n=25$, ta có $\chi^2_{\alpha/2}=37,65$ và $\chi^2_{1-\alpha/2}=14,61$.

Ta có $u=\Sigma n_i.(x_i-\mu)^2=1,75$.

+ Khoảng tin cậy của phương sai σ^2 là

$$\left(\frac{u}{\chi^2_{\alpha/2}};\frac{u}{\chi^2_{1-\alpha/2}}\right)=\left(\frac{1,75}{37,65};\frac{1,75}{14,61}\right)=(0,0465;0,12).$$

Vậy phương sai của X là khoảng 0,0465 đến 0,12 với độ tin cậy 90%.

Ví dụ 4.4.5. Xét lại ví dụ 4.4.4 với chút thay đổi là không cho biết giá trị trung bình μ.

Đây là bài toán ước lượng phương sai khi không biết giá tri trung bình $E(X)$.

+ $1-\alpha = 0,9 \Rightarrow \alpha = 0,1, \alpha/2 = 0,05$ và $1-\alpha/2 = 0,95$ Tra bảng χ^2 với bậc tự do $n-1 = 24$, ta có $\chi^2_{\alpha/2} = 36,42$ và $\chi^2_{1-\alpha/2} = 13,85$.

Ta có $\overline{x} = 498,5/25 = 19,94$ $s^2 = \frac{1}{24}(9941,75 - 25 \times 19,94^2) = 0,0692$. u $u = (n-1)s^2 = 240,0692 = 1,66$

+Khoảng tin cậy $\left(\frac{u}{\chi^2_{\alpha/2}}; \frac{u}{\chi^2_{1-\alpha/2}}\right) = \left(\frac{1,66}{36,42}; \frac{1,66}{13,85}\right) = (0,046; 0,12)$ Vậy phương sai của X khoảng 0,046 đến 0,12 với độ tin cậy 90%.

4.4.4 Xác định kích thước mẫu.

Giả sử đã biết khoảng tin cậy có độ chính xác ε và độ tin cậy $1-\alpha$. Cần xác định kích thước mẫu n_1 tối thiểu là bao nhiêu.

Xác định kích thước mẫu khi ước lượng trung bình tổng thể.

Với α ta xác định z_α sao cho $P(|Z| > z_\alpha) = \alpha$.

Ta có $\varepsilon = z_\alpha . \frac{\sigma}{\sqrt{n}}$. (TH biết σ^2) hoặc $\varepsilon = z_\alpha . \frac{s}{\sqrt{n}}$ (TH không biết σ^2).

Suy ra $n = \frac{(z_\alpha)^2.\sigma^2}{\varepsilon^2}$ (TH biết σ^2) hoặc $n = \frac{(z_\alpha)^2.s^2}{\varepsilon^2}$ (TH không biết σ^2).

Chọn kích thước mẫu là số tự nhiên $n_1 > n$ này.

Ví dụ 4.4.6. Giả sử cần ước lượng năng suất lúa trung bình cho một vùng lớn. Sau khi điều tra 100ha lúa chọn ngẫu nhiên trong vùng người ta có ước lượng năng suất trung bình của vùng ước chừng 45,35 đến 46,65 (tạ/ha) với độ tin cậy 95% và có độ lệch mẫu cụ thể $s = 3,3$ tạ. Tìm số ha lúa phải điều tra thêm để khong tin cậy (45,35 ; 46,65) như trên có độ tin cậy 99%.

Với $\alpha = 0,01$, tìm z_α sao: $P(|Z| > z_\alpha) = \alpha$. Chọn $z_\alpha = 2,576$.

Ta có $\varepsilon = (46,6545,35)/2 = 0,65$. Biết $\varepsilon = z_\alpha . \frac{s}{\sqrt{n}}$ (TH không biết σ^2)

$n = \frac{(z_\alpha)^2.s^2}{\varepsilon^2} = \frac{2,576^2 \times 3,3^2}{0,65^2} = 171,04 \Rightarrow$ Số ha cần điều tra ít nhất là $n_1 = 172$.

Số ha lúa phải điều tra thêm $n_1 - 100 = 72$ (ha).

Xác định kích thước mẫu khi ước lượng tỉ lệ tổng thể.

Với α ta xác định z_α sao cho $P(|Z| > z_\alpha) = \alpha$. Ta có $\varepsilon = z_\alpha . \frac{\sqrt{f(1-f)}}{\sqrt{n}}$.

Suy ra $n = \dfrac{(z_\alpha)^2.f(1-f)}{\varepsilon^2}$

Chọn kích thước mẫu là số tự nhiên $n_1 > n$ này.

Ví dụ 4.4.7. Nghiên cứu nhu cầu tiêu dùng một mặt hàng ở một thành phố, người ta điều tra trên 1000 người của thành phố và biết tỉ lệ người trong thành phố có nhu cầu về mặt hàng này là 57% đến 63% với độ tin cậy là 95%. Người ta muốn tăng độ tin cậy lên 99% thì số người phải điều tra thêm là bao nhiêu?

Ta có $\varepsilon = z_\alpha.\frac{\sqrt{f(1-f)}}{\sqrt{n}}$ Suy ra $n = \dfrac{(z_\alpha)^2.f(1-f)}{\varepsilon^2}$. Với $\alpha = 0,01$. Tìm z_α sao: $P(|Z| > z_\alpha) = \alpha$, ta có $z_\alpha = 2,576$.

Ta có $f = (0,63 + 0,57)/2 = 0,6$ và $\varepsilon = (0,630,57)/2 = 0,03$.

suy ra $n = (2,57620,60,4)/0,032 = 1769,54 \Rightarrow$ Số người phi điều tra ít nhất $n_1 = 1770$.

Số người phải điều tra thêm $1770 - 1000 = 770$ người.

4.4.5 Xác định độ tin cậy.

Giả sử đã biết khong tin cậy có độ chính xác ε và biết kích thước mẫu n. Cần xác định độ tin cậy $1 - \alpha$ là bao nhiêu.

Xác định độ tin cậy khi ước lượng trung bình tổng thể.

Để đơn giản ta chỉ xét trường hợp kích thước mẫu $n > 30$.

Ta có $\varepsilon = z_\alpha\ .\dfrac{\sigma}{\sqrt{n}}$ (TH biết σ^2) hoặc $\varepsilon = z_\alpha\ .\dfrac{s}{\sqrt{n}}$ (TH không biết σ^2).

Suy ra $z_\alpha = \frac{\varepsilon.\sqrt{n}}{\sigma}$ (TH biết σ^2) hoặc (TH không biết σ^2).

Từ giá trị z_α, ta có $P(|Z| > z) = \alpha \Rightarrow$ độ tin cậy $1 - \alpha$.

Ví dụ 4.4.8. Để ước lượng trọng lượng trung bình sản phẩm của nhà máy, người ta điều tra 121 sản phẩm, được phương sai mẫu cụ thể $s^2 = 15{,}3$ và trung bình mẫu cụ thể là 98 Kg. Ước lượng với độ chính xác là 1Kg thì độ tin cậy là bao nhiêu.

Ta có $\varepsilon = z_\alpha\ .\dfrac{s}{\sqrt{n}}$ (TH không biết σ^2). Suy ra $z_\alpha = \dfrac{\varepsilon.\sqrt{n}}{s}$. $\varepsilon = 1; n = 121$ và $s = \sqrt{15,3} = 3,9115 \Rightarrow; z = (1.11)/3,9115 = 2,8122$.

Từ $P(|Z| > z) = \alpha$ có $\alpha \approx 0,0049 \Rightarrow 1 - \alpha = 0,9951$.

Vậy ước lượng trọng lượng trung bình sản phẩm khong 97 đến 99 Kg thì độ tin cậy là 99,51%

Xác định độ tin cậy khi ước lượng tỉ lệ tổng thể.

Xét trường hợp kích thước n đủ lớn.

Ta có $\varepsilon = z_\alpha.\dfrac{\sqrt{f(1-f)}}{\sqrt{n}}$. Suy ra $z_\alpha = \frac{\varepsilon.\sqrt{n}}{\sqrt{f(1-f)}}$

Từ giá trị z_α , ta có $P(|Z| > z_\alpha) = \alpha \Rightarrow$ độ tin cậy $1 - \alpha$.

Ví dụ 4.4.9. Nghiên cứu nhu cầu tiêu dùng 1 mặt hàng ở một thành phố, người ta điều tra trên 1000 người được chọn ngẫu nhiên và được tỉ lệ mẫu f = 60%. Ước lượng tỉ lệ người (trong thành phố) có nhu cầu về mặt hàng đó khoảng 55% đến 65% thì độ tin cậy bao nhiêu?

Ta có $f = 0,6; \varepsilon = 0,05; n = 1000$.

Biết $\varepsilon = z_\alpha.\frac{\sqrt{f(1-f)}}{\sqrt{n}}$. Suy ra

$$z_\alpha = \frac{\varepsilon.\sqrt{n}}{\sqrt{f(1-f)}} = \frac{0,05 \times \sqrt{1000}}{\sqrt{0,6 \times 0,4}} \approx 3,2275.$$

Ta có $P(|Z| > z_\alpha) = \alpha 0,0013 \Rightarrow 1 - \alpha = 0,9987$.

Vậy ước lượng tỉ lệ người có nhu cầu khoảng 55% đến 65% thì độ tin cậy là 99,87%.

Chú ý. TH: tra bảng $P(Z > z'_\alpha) = \alpha$.

Thay $z_\alpha = z'_\alpha/2$ vào toàn bộ các công thức có sử dung z_α ở trên.

Để ý $P(|Z| > z_\alpha) = P(|Z| > z'_\alpha/2) = 2P(z > z'_\alpha/2) = \alpha$.

4.5 KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

4.5.1 Khái niệm

a. Giả thiết thống kê.

+ Giả thiết thống kê là các giả thiết về các tham số hoặc qui luật phân phối xác suất hoặc tính độc lập của các ĐLNN.

+ Việc tìm ra kết luận là chấp nhận hoặc bác bỏ một gi thiết gọi là kiểm định giả thiết thống kê.

+ Giả thiết cần kiểm định gọi là giả thiết không, ký hiệu là H_0 (hoặc H). Mệnh đề đối lập với H_0 được gọi là giả thiết đối và ký hiệu là H_1 (hoặc $\overline{H}$).

Chẳng hạn: $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$. Kiểm định với giả thiết đối dạng này gọi là kiểm định giả thiết hai phía.

Còn nếu kiểm định với giả thiết đối $H_1 : \theta > \theta_0$ (hoặc $H_1 : \theta < \theta_0$) thì gọi là kiểm định một phía.

b. Mức ý nghĩa, miền bác bỏ.

Xét ĐLNN X trên một tổng thể. Giả sử cần kiểm định giả thiết $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$ (chẳng hạn).

Lập mẫu ngẫu nhiên $(X_1, ..., X_n)$.

Chọn tiêu chuẩn kiểm định $Z = f(X_1, ..., X_n, \theta_0)$ sao cho nếu H_0 đúng thì ta sẽ xác định được qui luật phân phối của Z và tính được giá trị của Z ứng với mỗi mẫu cụ thể.

Giả sử H_0 là đúng => Chọn được miền W_α sao cho $P(Z \in W_\alpha) = \alpha$, với $\alpha > 0$ rất bé (biến cố $Z \in W_\alpha$ là rất khó xảy ra).

Với mẫu cụ thể, ta tính giá trị cụ thể của Z là z.

+Nếu $Z \in W_\alpha$ thì ta bác bỏ giả thiết H_0, thừa nhận H_1.

+Nếu $Z \notin W_\alpha$ thì ta tạm chấp nhận gi thiết H_0. Miền W_α gọi là miền bác bỏ gi thiết H_0.

Trị α gọi là mức ý nghĩa của kiểm định. Thường chọn $\alpha \in [1\%; 5\%]$.

c. Sai lầm loại 1 và sai lầm loại 2.

+ Sai lầm loại 1 là sai lầm mắc phải khi ta bác bỏ H_0 nhưng trong thực tế H_0 là đúng. Xác suất mắc sai lầm loại 1 là $P(Z \in W_\alpha) = \alpha$ (chính là mức ý nghĩa α).

+Sai lầm loại 2 là sai lầm mắc phải khi ta chấp nhận H_0 nhưng trong thực tế H_0 là sai.

Khi kiểm định giả thiết thống kê, ta phải hạn chế khả năng mắc phải hai loại sai lầm trên. Thường ta làm như sau: ấn định trước mức ý nghĩa α (cũng chính là xác suất mắc sai lầm loại 1).

Tìm miền bác bỏ W_α sao cho xác suất mắc sai lầm loại 2 là nhỏ nhất.

Các miền bác bỏ W_α mà ta sử dụng trong các kiểm định sẽ nói ở sau đều thỏa yêu cầu này.

4.5.2 Kiểm định giả thiết về trung bình tổng thể

Xét ĐLNN X trên một tổng thể. Giả sử trung bình tổng thể là $E(X) = \mu$ và ta chưa biết μ.

+ Ta cần kiểm định giả thiết $H_0 : \mu = m_0$ và giả thiết đối $H_1 : \mu \neq m_0$ với mức ý nghĩa α.

Lập mẫu ngẫu nhiên $(X_1, ..., X_n)$.

$n > 30$ (**hoặc** $n < 30$ **nhưng X có phân phối chuẩn) biết** σ^2 Xét $Z = \frac{\overline{X}-m_o}{\sigma/\sqrt{n}}$ Gs H_0 đúng=> $Z \in N(0,1)$. Với α đủ nhỏ, chọn được z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ (Biến cố $(\|Z\| > z_\alpha)$ là biến cố rất khó xãy ra). => Miền bác bỏ là $W = Z : \|Z\| > z_\alpha$ với z_α ở trên.	$n \geq 30$ **và chưa biết phưng sai** σ^2**.** Xét $Z = \frac{\overline{X}-m_o}{S/\sqrt{n}}$ Gs H_0 đúng=> $Z \in N(0,1)$. Với α đủ nhỏ, chọn được z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ (Biến cố $(\|Z\| > z_\alpha)$ là biến cố rất khó xãy ra) => Miền bác bỏ là $W = Z : \|Z\| > z_\alpha$ với z_α ở trên.	$n < 30$, **chưa biết phưng sai** σ^2 **và X có phân phối chuẩn.** Xét $Z = \frac{\overline{X}-m_o}{\sigma/\sqrt{n}}$ Gs H_0 đúng=> T có phân phối Student với bậc tự do $n-1$. Với α đủ nhỏ, chọn được t_α sao cho $P(\|T\| > t_\alpha) = \alpha$ (Biến cố $(\|T\| > t_\alpha)$ là biến cố rất khó xãy ra). => Miền bác bỏ là $W = Z : \|T\| > t_\alpha$ với t_α trên.
$n > 30$ (**hoặc** $n < 30$ **nhưng X có phân phối chuẩn) biết** σ^2 **Qui tắc quyết định** +Chọn mẫu cụ thể kích thước n. + tính giá trị $z = \frac{\overline{x}-m_o}{\sigma/\sqrt{n}}$. +Tra bảng phân phối chuẩn tắc tìm giá trị tới hạn z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ +$\|z\| > z_\alpha$ => bác bỏ giả thiết H_0; $\|z\| \leq z_\alpha$ => tạm chấp nhận H_0.	$n \geq 30$ **và chưa biết phưng sai** σ^2**.** **Qui tắc quyết định** +Chọn mẫu cụ thể kích thước n. + tính giá trị $z = \frac{\overline{x}-m_o}{s/\sqrt{n}}$. +Tra bảng phân phối chuẩn tắc tìm giá trị tới hạn z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ +$\|z\| > z_\alpha$ => bác bỏ giả thiết H_0; $\|z\| \leq z_\alpha$ => tạm chấp nhận Ho.	$n < 30$, **chưa biết phưng sai** σ^2 **và X có phân phối chuẩn.** **Qui tắc quyết định** +Chọn mẫu cụ thể kích thước n. + tính giá trị $t = \frac{\overline{x}-m_o}{s/\sqrt{n}}$. +Tra bảng ph/phối Student với bậc tự do $n-1$, tìm giá trị tới hạn t_α sao $P(\|T\| > t_\alpha) = \alpha$. + $\|t\| > t_\alpha$ => bác bỏ giả thiết H_0. $\|t\| \leq t$ => tạm chấp nhận Ho.

Ta cần kiểm định giả thiết $H_0 : \mu = m_0$ và giả thiết đối $H_1 : \mu > m_0$ với mức ý nghĩa α

Qui tắc quyết định +Tính z như trên. +Tìm giá trị tới hạn z'_α sao cho $P(Z > z'_\alpha) = \alpha$ + $z > z'_\alpha$ =>bác bỏ H_0 $z \leq z'_\alpha$ =>Tạm chấp nhận H_0	**Qui tắc quyết định** +Tính z như trên. +Tìm giá trị tới hạn z'_α sao cho $P(Z > z'_\alpha) = \alpha$ + $z > z'_\alpha$ =>bác bỏ H_0 $z \leq z'_\alpha$ =>Tạm chấp nhận H_0	**Qui tắc quyết định** +Tính t như trên. +Tìm giá trị tới hạn t'_α sao cho $P(T > t'_\alpha) = \alpha$ +$t > t'_\alpha$ =>bác bỏ H_0 $t \leq t'_\alpha$ =>tạm chấp nhận H_0

Ta cần kiểm định giả thiết $H_0 : \mu = m_0$ và gi thiết đối $H_1 : \mu < m_0$ với mức ý nghĩa α.

Qui tắc quyết định	**Qui tắc quyết định**	**Qui tắc quyết định**
+Tính z như trên. +Tìm giá trị tới hạn z'_α sao cho $P(Z < -z'_\alpha) = \alpha$ + $z < -z'_\alpha$ =>bác bỏ H_0 $z \geq -z'_\alpha$ =>Tạm chấp nhận H_0	+Tính z như trên. +Tìm giá trị tới hạn z'_α sao cho $P(Z < -z'_\alpha) = \alpha$ + $z < -z'_\alpha$ =>bác bỏ H_0 $z \geq -z'_\alpha$ =>Tạm chấp nhận H_0	+Tính t như trên. +Tìm giá trị tới hạn t'_α sao cho $P(T < -t'_\alpha) = \alpha$ +$t < -t'_\alpha$ =>bác bỏ H_0 $t \geq -t'_\alpha$ =>tạm chấp nhận H_0

Chú ý. + Khi tìm z_α để $P(|Z| > z_\alpha) = \alpha$ (hoặc tìm t_α để $P(|T| > t_\alpha) = \alpha$).

Nếu tra bằng bảng $P(Z > z'_\alpha) = \alpha$ thì thay $z_\alpha = z'_{\alpha/2}$.

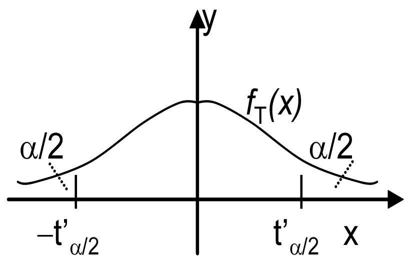
Nếu tra bảng $P(T > z'_\alpha) = \alpha$ thì thay $t_\alpha = t'_{\alpha/2}$.

+ Khi tìm z'_α để $P(Z > z'_\alpha) = \alpha$ (hoặc tìm t'_α để $P(T > t'_\alpha) = \alpha$).

Nếu tra bảng $P(|Z| > z_\alpha) = \alpha$ thì thay $z'_\alpha = z_{2\alpha}$.

Nếu tra bảng $P(|T| > t_\alpha) = \alpha$ thì thay $t'_\alpha = t_{2\alpha}$.

Tóm lại: Có thể thay $z_\alpha = z'_{\alpha/2}; z'_\alpha = z_{2\alpha}$; còn thay $t_\alpha = t'_{\alpha/2}$ và $t'_\alpha = t_{2\alpha}$.



Ví dụ 4.5.1. Trọng lượng các bao hàng do một máy đóng bao sản xuất là ĐLNN phân phối theo qui luật chuẩn với trọng lượng trung bình qui định là 50 Kg. Người ta cân thử 25 bao hàng và tính được $\overline{x} = 49,74$ Kg và $s = 0,5$ Kg. Với mức ý nghĩa 1% hãy kết luận tình hình làm việc của máy đóng bao đó: đóng các bao hàng đúng trọng lượng trung bình qui định không?

Giải. Gọi μ là trọng lượng trung bình thực tế của máy đóng bao sản xuất.

Đặt giả thiết $H_0 : \mu = 50$ và $H_1 : \mu \neq 50$. Do σ^2 chưa biết, ta kiểm định theo qui tắc sau:

Tính $t = \dfrac{\overline{x} - 50}{s/\sqrt{25}} = \dfrac{(49,74 - 50)}{0,5}.5 = -2,6$.

$\alpha = 0,01$. Tra bảng Student 24 bậc tự do với $P(|T| > t_\alpha) = 0,01$, ta có $t_\alpha = 2,797$.

Ta có $|t| < t_\alpha$ => Tạm chấp nhận H_0.

KL: Với mức ý nghĩa 1%, tạm coi máy đã đóng các bao hàng có trọng lương trung bình thực tế đúng qui định.

Ví dụ 4.5.2. Cũng ví dụ như trên nhưng với yêu cầu muốn kiểm tra xem các bao hàng do máy đóng có trọng lượng trung bình có nhỏ hơn qui định không. (Do đã có nghi ngờ).

Giải. Gọi μ là trọng lượng trung bình thực tế của máy đóng bao sản xuất.

Đặt giả thiết $H_0 : \mu = 50$ và $H_1 : \mu < 50$. Do σ^2 chưa biết, ta kiểm định theo qui tắc sau:

$\alpha = 0,01$. Tra bảng Student 24 bậc tự do với $P(T > t'_\alpha) = \alpha = 0,01$, ta có $t'_\alpha = t_{2\alpha} = 2,492$.

Tính $t = \dfrac{49,74 - 50}{0,5}.\sqrt{25} = -2,6$.

Ta có $t = -2,6 < t'_\alpha = -2,492$=> Bác bỏ H_0.

KL: Với mức ý nghĩa $\alpha = 1\%$, các bao hàng do máy đóng có trọng lượng trung bình thực tế nhỏ hơn qui định.

4.5.3 Kiểm định giả thiết về tỉ lệ

Giả sử tỉ lệ các phần tử có tính chất A của tổng thể là p và ta chưa biết p.

Cần kiểm định giả thiết $H_0 : p = p_0$ với mức ý nghĩa α .

Xét ĐLNN X trên tổng thể. X được xác định như sau: $X = 1$ nếu phần tử có tính chất A và X = 0 nếu phần tử không có tính chất A. Lập mẫu ngẫu nhiên $(X_1, ..., X_n)$ với kích thước $n > 30$.

Xét $Z = \dfrac{F_n - p_0}{\sqrt{p_0(1 - p_0)}/\sqrt{n}}$.

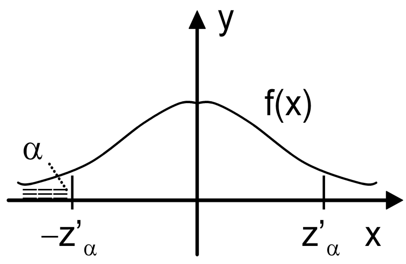
Gs H_0 là đúng. Người ta chứng minh $Z \sim N(0, 1)$. Tương tự kiểm định giá trị trung bình đưa đến.

Chọn mẫu cụ thể kích thước n lớn ($n > 30$).

Qui tắc quyết định

$H_0 : p = p_0; H_1 : p \neq p_0$ +Tính $z = \frac{f-p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$ +Tìm z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$ +$\|z\| > z_\alpha$=> bác bỏ H_0. $\|z\| \leq z'_\alpha$ =>Tạm chấp nhận H_0	$H_0 : p = p_0; H_1 : p > p_0$ +Tính $z = \frac{f-p_o}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$ +Tìm z'_α sao cho $P(Z > z'_\alpha) = \alpha$ +$z > z_\alpha$=> bác bỏ H_0. + $z \leq z'_\alpha$ =>Tạm chấp nhận H_0	$H_0 : p = p_0; H_1 : p < p_0$ +Tính $z = \frac{f-p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$ +Tìm z'_α với $P(Z < z'_\alpha) = \alpha$ +$z < -z_\alpha$=> bác bỏ H_0. $z \geq -z'_\alpha$ =>Tạm chấp nhận H_0

Ví dụ 4.5.3. Tỉ lệ phế phẩm của một dây chuyền sản xuất là 5%. Sau khi tiến hành một cải tiến kỹ thuật người ta kiểm tra ngẫu nhiên 3000 sản phẩm thì thấy có 120 phế phẩm. Với mức ý nghĩa 0,01 hãy kết luận việc cải tiến kỹ thuật có làm giảm tỉ lệ phế phẩm không?



Giải. Gọi p là tỉ lệ phế phẩm mới của dây chuyền sau cải tiến.

Kiểm tra giả thiết $H_0 : p = 0,05, H_1 : p < 0,05$.

Tỉ lệ phế phẩm của mẫu là $120/3000 = 0,04$ Tính $z = \dfrac{f - p_0}{\sqrt{p_o(1-p_0)}/\sqrt{n}} = \dfrac{0,04 - 0,05}{\sqrt{0,05.(1-0,05)}}.\sqrt{3000} = -2,513$.

$\alpha = 0,01$. Tìm z'_α sao: $P(Z < z'_\alpha) = \alpha => z'_\alpha = z_{2\alpha} = 2,326$.

Ta có $z < -2,326 = -z'_\alpha =>$ bác bỏ H_0.

Kết luận: với mức ý nghĩa $\alpha = 1\%$, cải tiến kỹ thuật có làm giảm tỉ lệ phế phẩm.

4.5.4 Kiểm định giả thiết bằng nhau của hai trung bình

Cho hai ĐLNN X và Y độc lập, có cùng phân phối chuẩn với $E(X)$ và $E(Y)$ đều chưa biết.

Cần kiểm định giả thiết $H_0 : E(X) = E(Y)$ với mức ý nghĩa α.

Từ X lập mẫu ngẫu nhiên W_X kích thước n_1, Từ Y lập mẫu ngẫu nhiên W_Y kích thước n_2.

Xét thống kê $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{D(X)}{n_1} + \frac{D(Y)}{n_2}}}$.

Giả sử H_0 là đúng. Người ta chứng minh được $Z \sim N(0,1)$. Lập luận tương tự trên suy ra:

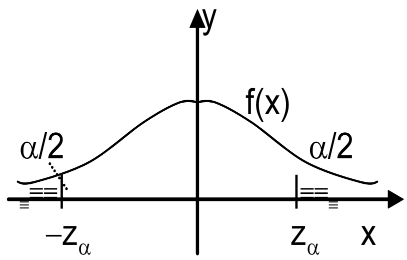
Qui tắc quyết định:

+Lấy mẫu cụ thể kích thước n_1 (đối với X) và mẫu kích thước n_2 (đối với Y).

+Tính $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{D(X)}{n_1} + \frac{D(Y)}{n_2}}}$ (hoặc $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}}$ nếu $D(X), D(Y)$ chưa biết,

$n_1, n_2 > 30$ đủ lớn).

$H_0 : E(X) = E(Y); H_1 : E(X) \neq E(Y)$ +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\lvert Z\rvert > z_\alpha) = \alpha$. $+\lvert z\rvert > z_\alpha =>$ bác bỏ H_0. $\lvert z\rvert \leq z =>$ tạm chấp nhận H_0.	$H_0 : E(X) = E(Y); H_1 : E(X) > E(Y)$ +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\lvert Z\rvert > z'_\alpha) = \alpha$. $+z > z'_\alpha =>$ bác bỏ H_0 $z \leq z'_\alpha =>$tạm chấp nhận H_0.	$H_0 : E(X) = E(Y); H_1 : E(X) < E(Y)$ +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\lvert Z\rvert < -z'_\alpha) = \alpha$. $+z < -z'_\alpha =>$ bác bỏ H_0 $z \geq -z'_\alpha =>$tạm chấp nhận H_0.



Ví dụ 4.5.4. Trong lượng một loại sản phẩm do hai nhà máy sản xuất ra là ĐLNN có phân phối chuẩn và có cùng độ lệch tiêu chuẩn là $\sigma = 1Kg$. Cân thử 25 sản phẩm của nhà máy thứ nhất ta có $\overline{x} = 50Kg$ và cân thử 20 sản phẩm của nhà máy thứ hai ta có $\overline{y} = 50,6Kg$.

Với mức ý nghĩa $\alpha = 0,05$, hãy kết luận xem trọng lượng trung bình của sản phẩm do hai nhà máy sản xuất ra có như nhau không.

Giải. Gọi các trọng lượng sản phẩm do nhà máy thứ nhất và thứ hai sản xuất là X và Y (kg).

X và Y là các ĐLNN có phân phối chuẩn và $D(X) = D(Y) = 1$.

Kiểm định giả thiết $H_0 : E(X) = E(Y); H_1 : E(X) \neq E(Y)$.

$\alpha = 0,05$. Tìm z_α sao : $P(|Z| > z_\alpha) = \alpha$. Ta có $z_\alpha = 1,96$.

Tính $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{D(X)}{n_1} + \frac{D(Y)}{n_2}}} = \dfrac{50 - 50,6}{\sqrt{\frac{1}{25} + \frac{1}{20}}} = -2.$

Ta có $|z| > z_\alpha = 1,96 =>$ bác bỏ H_0.

KL: Với mức ý nghĩa $\alpha = 0,05$, không thể xem trọng lượng trung bình của sản phẩm hai nhà máy là như nhau được.

Ghi chú. TH: $D(X) = D(Y)$ nhưng chưa biết giá trị cụ thể của chúng. Người ta chứng minh được: Nếu H_0 đúng $(E(X) = E(Y)$) thì với

$$S_*^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}$$

có phân phối Student $(n_1 + n_2 - 2)$ bậc tự do.

Dựa vào kết quả trên hãy suy ra qui tắc quyết định khi n_1 hay $n_2 < 30$.

4.5.5 Kiểm định về sự bằng nhau của hai tỉ lệ

Giả sử p_1 và p_2 lần lượt là tỉ lệ các phần tử có tính chất A trong tổng thể thứ nhất và thứ hai.

Cần kiểm định giả thiết $H_0 : p_1 = p_2$ và giả thiết đối $H_1 : p_1 \neq p_2$ với mức ý nghĩa α.

Lấy hai mẫu cụ thể kích thước n_1 và $n_2 (\geq 30)$ đủ lớn lần lượt trong tổng thể thứ nhất và thứ hai.

Qui tắc quyết định

+Tính f_1, f_2 lần lượt là tỉ lệ mẫu có tính chất A của hai mẫu cụ thể.

Và tính p^* là tỉ lệ có tính chất A của mẫu gồm hai mẫu gộp lại ($p^* = \frac{n_1.f_1+n_2.f_2}{n_1+n_2}$).

$$z = \frac{f_1 - f_2}{\sqrt{p*(1-p*).\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$H_0 : p_1 = p_2; H_1 : p_1 \neq p_2$ +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\|Z\| > z_\alpha) = \alpha$. $+\|z\| > z_\alpha =>$ bác bỏ H_0. $\|z\| \leq z =>$ tạm chấp nhận H_0.	$H_0 : p_1 = p_2; H_1 : p_1 > p_2$ +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\|Z\| > z'_\alpha) = \alpha$. $+z > z'_\alpha =>$ bác bỏ H_0 $z \leq z'_\alpha =>$tạm chấp nhận H_0.	$H_0 : p_1 = p_2; H_1 : p_1 < p_2$ +Tra bảng phân phối chuẩn tắc tìm z_α sao cho $P(\|Z\| < -z'_\alpha) = \alpha$. $+z < -z'_\alpha =>$ bác bỏ H_0 $z \geq -z'_\alpha =>$tạm chấp nhận H_0.

Ví dụ 4.5.5. Kiểm tra các sản phẩm chọn ngẫu nhiên do hai nhà máy cùng sản xuất. Ta có bảng sau

Nhà máy	số sản phẩm được kiểm tra	số phế phẩm
A	1000	20
B	900	30

Với mức ý nghĩa $\alpha = 0,05$, có thể coi tỉ lệ phế phẩm 2 nhà máy là như nhau không?

Giải. Gọi p_1 và p_2 là các tỉ lệ phế phẩm của nhà máy thứ nhất và thứ hai.

Xét giả thiết $H_0 : p_1 = p_2$ và $H_1 : p1 \neq p2$.

$\alpha = 0,05$. Tìm z_α sao: $P(|Z| > z_\alpha) = \alpha => z_\alpha = 1,96$.

+Tính tỉ lệ phế phẩm nhà máy thứ nhất $f_1 = 0,02$ và của nhà máy thứ hai là $f_2 = 0,0333$.

$p^* = (20+30)/(1000+900) = 1/38$ và $1 - p^* = 37/38$.

+Tính

$$z = \frac{f_1 - f_2}{\sqrt{p*(1-p*).\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,02 - 0,033}{\sqrt{\frac{1}{38}.\frac{37}{38}\left(\frac{1}{1000} + \frac{1}{900}\right)}} = -1,81$$

+ $|z| < z_\alpha = 1,96 =>$ Tạm chấp nhận H_0.

KL: Với mức ý nghĩa 5%, tạm có thể coi tỉ lệ phế phẩm của hai nhà máy như nhau.

BÀI TẬP CHƯƠNG 4

4.1. Đo chiều cao và đường kính của một loại cây có cùng độ tuổi. Ta được kết quả cho ở bảng sau:

Chiều cao (cm)	104	99	105	103	102	106
Đường kính (cm)	11	13	15	14	12	12

Tính trung bình và phương sai mẫu của chiều cao và đường kính.

4.2. Quan sát về thời gian cần thiết để sản xuất một chi tiết máy, ta thu được các số liệu cho ở bảng sau:

Khoảng thời gian (phút)	Số quan sát	Khoảng thời gian (phút)	Số quan sát
20-25	2	40-45	14
25-30	14	45-50	8
30-35	26	50-55	4
35-40	32		

Tính trung bình mẫu và phương sai mẫu?

4.3. Tiến hành quan sát về hai chỉ tiêu X và Y, ta thu được các số liệu cho ở bảng sau:

X	0.25	0.37	0.44	0.55	0.60	0.62	0.68	0.70
Y	2.57	2.31	2.12	1.92	1.75	1.71	1.60	1.51

X	0.73	0.75	0.92	0.84	0.87	0.88	0.90	0.95	1
Y	1.50	1.41	1.33	1.31	1.25	1.20	1.19	1.15	1

Tính: $\overline{x}$; $\overline{y}$; $\overline{xy}$; s_X^2; s_Y^2.

4.4. Điều tra năng suất của 100 hecta lúa ở một vùng, người ta thu được kết quả cho ở bảng sau:

Năng suất (tạ/ha)	Diện tích (ha)	Năng suất (tạ/ha)	Diện tích (ha)
30-35	7	50-55	20
35-40	12	55-60	8
40-45	18	60-65	5
45-50	27	65-70	3

Tính trung bình mẫu và phương sai mẫu?

4.5. Để nghiên cứu nhu cầu của một loại hàng hóa ở một khu vực, người ta tiến hành kho sát 800 gia đình. Kết quả cho ở bảng dưới đây:

Nhu cầu (kg/tháng)	Số gia đình (n_i)	Nhu cầu (kg/tháng)	Số gia đình (n_i)
30-35	25	55-60	142
35-40	48	60-65	94
40-45	83	65-70	50
45-50	159	70-75	10
50-55	189		

Tính trung bình mẫu và độ lệch chuẩn mẫu?

4.6. Thống kê số hàng hóa bán được mỗi ngày và số ngày bán được lượng hàng tương ứng, ta có bảng số liệu sau:

Lượng hàng bán trong ngày (kg)	số ngày (n_i)	Lượng hàng bán trong ngày (kg)	số ngày (n_i)
100-200	5	400-450	70
200-250	12	450-500	35
250-300	56	500-550	30
300-350	107	550-700	10
350-400	75		

Tính trung bình mẫu và cho biết ý nghĩa thực tế của nó?

4.7. Kho sát về thu nhập và tỉ lệ thu nhập chi cho giáo dục ở 400 hộ gia đình, ta thu được số liệu cho ở bảng sau:

$Y \diagdown X$	10	20	30	40	50
150-250	10	40	20		
250-350		40	60	20	
350-450		20	80	40	
450-550			30	30	10

Trong đó: X là tỉ lệ thu nhập chi cho giáo dục (tính theo %), Y là thu nhập bình quân của một người trong gia đình (đơn vị tính là ngàn đồng/tháng). Tính trung bình mẫu và phương sai mẫu của X và Y?

4.8. Kho sát một mẫu gồm 12 người cho thấy số lần họ xem phim trong 1 năm như sau: 14, 16, 17, 24, 20, 32, 18, 29, 31, 15, 35. Tìm khoảng tin cậy 95% cho số lần trung bình mà một người tới rạp chiếu bóng trong 1 năm, biết số lần xem phim của một người là đại lượng có phân phối chuẩn.

4.9. Để định mức thời gian gia công một chi tiết máy người ta theo dõi ngẫu nhiên thời gian gia công 25 chi tiết và thu được bảng số liệu sau đây:

Thời gian (phút)	14	16	18	20	24
Số chi tiết	2	6	11	4	2

Với mức tin cậy 0.95, hãy ước lượng thời gian gia công trung bình tối đa với loại chi tiết trên. Cho biết thời gian gia công chi tiết đó tuân theo quy luật chuẩn.

4.10. Năng suất ngô của một vùng A được báo cáo lên qua 25 điểm thu hoạch là:

Năng suất (tạ/ha)	7	9	11	13	17
Số điểm thu hoạch	2	7	12	3	1

Với mức tin cậy là 0.95, hãy tính năng suất trung bình ngô tối thiểu của vùng này. Cho biết năng suất ngô tuân theo quy luật chuẩn.

4.11. Chọn ngẫu nhiên 36 công nhân của một xí nghiệp thì thấy lưng trung bình là 380.000 đ/1 tháng. Giả sử lương công nhân tuân theo quy luật chuẩn, với độ lệch chuẩn $\sigma = 14$ ngàn đồng. Với độ tin cậy 95%, hãy ước lượng mức lưng trung bình của công nhân trong toàn xí nghiệp trên.

4.12. Điều tra năng suất lúa trên diện tích 100 hécta trồng lúa của một vùng, ta thu được bảng số liệu sau:

Năng suất (tạ/ha) (x_i)	41	44	45	46	48	52	54
Số ha có năng suất tương ứng (n_i)	10	20	30	15	10	10	5

1) Lập bảng tính trung bình mẫu $\overline{x}$ và phương sai mẫu s^2 ứng với số liệu trên.

2) Ước lượng năng suất lúa trung bình của vùng đó với độ tin cậy 95%?

3) Ước lượng năng suất lúa trung bình tối đa và tối thiểu của vùng đó với độ tin cậy 90%?

4) Những thửa ruộng có năng suất từ 48 tạ/ha trở lên là những thửa có năng suất cao.

a) Hãy ước lượng năng suất lúa trung bình của những thửa có năng suất cao trong vùng đó với độ tin cậy 99%.

b) Hãy ước lượng tỷ lệ diện tích có năng suất cao trong vùng đó với độ tin cậy 95%.

4.13. Một trại chăn nuôi A xuất chuồng 200 con lợn, trọng lượng hơi tính bằng (Kg), ta có bảng số liệu sau:

Trọng lượng (x_i)	45-55	55-65	65-75	75-85	85-95	95-105
Số lợn (n_i)	10	30	45	80	30	5

a) Lập bảng tính trung bình mẫu và phương sai mẫu ứng với bộ số liệu trên.

b) Ước lượng trọng lượng lợn (hơi) trung bình của trại chăn nuôi A với độ tin cậy 95%.

c) Nếu chủ trại chăn nuôi A thông báo trọng lượng (hơi) trung bình của trại chăn nuôi là trên 80 kg thì có chấp nhận được không với mức ý nghĩa 10%.

d) Lợn có trọng lượng từ 85 kg trở lên được gọi là lợn chóng lớn, hãy ước lượng tỷ lệ lợn chóng lớn của trại chăn nuôi trên với mức ý nghĩa 99%.

4.14. Trọng lượng một loại chi tiết là một biến số ngẫu nhiên quy theo quy luật chuẩn với độ lệch tiêu chuẩn là 1.2 kg. Phải chọn ít nhất bao nhiêu chi tiết để điều tra, nếu muốn độ chính xác của ước lượng không vượt quá 0.3 và mức tin cậy của ước lượng là 0.95.

4.15. Chiều dài của một loại sản phẩm A do một máy tự động sản xuất là một biến số ngẫu nhiên quy theo quy luật chuẩn với độ lệch tiêu chuẩn là 3cm. Phải chọn ít nhất bao nhiêu chi tiết để đo, nếu muốn độ dài khoảng tin cậy không vượt quá 0,6 và mức tin cậy của ước lượng là 0,99.

4.16. Để điều tra số cá trong một hồ, người ta đánh bắt 1000 con cá, đánh dấu rồi th xuống hồ. Lần sau bắt lại 200 con thì được 40 con đánh dấu. Với độ tin cậy 95% hãy:

a) Ước lượng tỷ lệ cá được đánh dấu trong hồ.

b) Ước lượng số cá trong hồ.

4.17. Gieo 400 hạt giống thì có 20 hạt giống không ny mầm. Tỷ lệ hạt giống không ny mầm tối đa là bao nhiêu? Yêu cầu kết luận với mức tin cậy 0.95.

4.18. Cơ quan cảnh sát giao thông kiểm tra hệ thống phanh của 40 chiếc xe tải trên đường quốc lộ. Họ phát hiện ra 14 chiếc có phanh chưa đảm bo an toàn. Tìm khoảng tin cậy 98% cho tỉ lệ xe tải có phanh chưa an toàn.

4.19. Trong đợt vận động bầu cử tổng thống người ta phỏng vấn ngẫu nhiên 1600 cử tri thì được biết 960 người trong số đó bỏ phiếu cho ứng cử viên A. Với mức tin cậy 0.99, ứng cử viên A chiếm được tối thiểu bao nhiêu phần trăm số phiếu.

4.20. Tỷ lệ nảy mầm của một loại hạt giống trong mẫu là 90%. Cần ước lượng tỷ lệ nảy mầm của loại hạt giống đó với mức tin cậy 0.95. Với độ dài khoảng tin cậy đối xứng không vượt quá 0.02, thì phải gieo bao nhiêu hạt?

4.21. Để xác định tỷ lệ người mắc chứng bướu cổ do thiếu hụt iode ở một khu vực dân cư nào đó, cần khám bao nhiêu người? Nếu muốn cho khong ước lượng có độ chính xác không vượt quá 0.04 với mức tin cậy 0.95.

4.22. Giả sử điểm trung bình môn toán của 100 thí sinh thi vào ĐHNT là 5, với độ lệch chuẩn mẫu hiệu chỉnh là 2.5.

a)Ước lượng điểm trung bình môn toán của toàn thể sinh viên với độ tin cậy 95%.

b)Với độ chính xác là 0.25 điểm, hãy xác định độ tin cậy.

4.23. Tuổi thọ của một loại bóng đèn được biết theo quy luật chuẩn với độ lệch chuẩn 100 giờ.

a) Chọn ngẫu nhiên 100 bóng để thử nghiệm, thấy mỗi bóng tuổi thọ trung bình là 1000 giờ.

Hãy ước lượng tuổi thọ trung bình của bóng đèn xí nghiệp A sản xuất với độ tin cậy 95%.

b) Với độ chính xác 15 giờ, hãy xác định độ tin cậy.

c) Với độ chính xác 25 giờ và độ tin cậy 95% thì cần thử nghiệm bao nhiêu bóng.

4.24. 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 (độ chính xác) 3%, hãy xác định độ tin cậy.

4.25. Điều tra năng suất lúa ở 1 một vùng trong vụ hè thu năm 2007, người ta thu được bảng số liệu sau:

Năng suất (tấn/ha)	3-4	4-4.5	4.5-5	5-5.5	5.5-6	6-6.5	6.5-7	7-8
Diện tích (ha)	5	10	26	30	28	12	6	4

a) Những thửa ruộng có năng suất trên 6 tấn/ha là những thửa ruộng có năng suất cao. Ước lượng diện tích lúa có năng suất cao vùng này với độ tin cậy 95%. Biết diện tích lúa gieo trồng ở vùng này là 8000 ha.

(HD: Tính khoảng $(p_1; p_2)$ sau đó nhân với 8000 ha).

a) Điều tra sơ bộ vụ hè thu năm 2006 thì thấy năng suất lúa ở vùng này là 5 tấn/ha và phương sai mẫu hiệu chỉnh là 0,49. Vụ hè thu năm 2007 người ta áp dụng 1 biện pháp kỹ thuật mới.

Hãy xét về tác dụng của biện pháp kỹ thuật mới này với mức ý nghĩa 5%.

Tài liệu tham khảo

[1] Nguyễn Văn Hộ, *Xác suất thống kê*, NXB Giáo Dục, 2006.

[2] Phạm Văn Kiều, *Giáo trình xác suất và thống kê*, NXB Giáo Dục, 2007.

[3] Đào Hữu Hồ, *Xác suất thống kê*, NXB Đại Học Quốc Gia, 2001

[4] Lê Văn Tiến, *Lý thuyết xác suất và thống kê toán học*, NXB Nông Nghiệp, 1999.

[5] Nguyễn Đình Hiền, *Giáo trình xác suất thống kê*, NXB Đại học Sư phạm, 2003.