Chương 1

CÁC KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT

1.1. GIẢI TÍCH TỔ HỢP

1.1.1. Các quy tắc đếm

1. Quy tắc cộng:

Nếu một công việc được chia ra làm k trường hợp để thực hiện: trường hợp 1 có n_1 cách thực hiện xong công việc; trường hợp 2 có n_2 cách thực hiện xong công việc, . . ., trường hợp k có n_k cách thực hiện xong công việc và không có bất kỳ một cách thực hiện nào ở trường hợp này lại trùng với một cách thực hiện ở trường hợp khác, thì có $n_1 + n_2 + ... + n_k$ cách thực hiện xong công việc.

2. Quy tắc nhân:

Nếu một công việc được chia ra làm k giai đoạn: giai đoạn 1 có n_1 cách thực hiện xong công việc; giai đoạn 2 có n_2 cách thực hiện xong công việc, . . ., giai đoạn k có n_k cách thực hiện xong công việc thì có $n_1 n_2 . . . n_k$ cách thực hiện xong công việc.

3. Tích Descartes:

- Cho hai tập hợp A và B. Tích Descartes của A và B, ký hiệu là $A \times B$ là tập hợp tất cả các cặp (có thứ tự) (a; b) với $a \in A, b \in B$, nghĩa là:

$$A \times B = \{(a, b) / a \in A; b \in B\}$$

Nếu A và B là hai tập hữu hạn thì số phần tử của tập hợp $A \times B$ là $|A \times B| = |A|.|B|$.

- Tương tự, tích Descartes của k tập hợp $A_1, A_2, . . ., A_k$, ký hiệu là $A_1 \times A_2 \times . . . \times A_k$ là tập hợp tất cả các bộ có thứ tự $(a_1, a_2, . . ., a_k)$ trong đó $a_i \in A_i$, mọi $i = 1, 2, . . ., k$.

$$A_1 \times A_2 \times . . . \times A_k = \{(a_1, a_2, . . ., a_k) / a_i \in A_i, i = 1, 2, . . ., k\}$$

Nếu $A_1, A_2, . . ., A_k$ là k tập hữu hạn thì số phần tử của tập hợp $A_1 \times A_2 \times . . . \times A_k$ là $|A_1 \times A_2 \times ... \times A_k| = |A_1| \times |A_2| \times . . . \times |A_k|$.

Ký hiệu $A^k = \underbrace{A \times A \times . . . \times A}_{k \text{ lần}}$.

1.1.2. Chỉnh hợp lặp

Cho A là một tập hợp có n phần tử, nỗi phần tử của tập A^k được gọi là một chỉnh hợp lặp n chập k.

Số các chỉnh hợp lặp n chập k là

$$\boxed{\mathrm{F}_n^k = n^k}$$

Ví dụ 1.1.

a. Có bao nhiêu cách sắp xếp 5 quyển sách vào 3 ngăn?

b. Một đề thi trắc nghiệm khách quan gồm 10 câu, mỗi câu có 4 phương án trả lời. Hỏi bài thi có tất cả bao nhiêu phương án trả lời?

Giải:

a. Mỗi cách sắp xếp 5 quyển sách vào 3 ngăn là một chỉnh hợp lặp 3 chập 5 của (mỗi lần xếp 1 quyển sách vào 1 ngăn xem như chọn 1 ngăn trong 3 ngăn, do có 5 quyển sách nên việc chọn ngăn được tiến hành 5 lần).

Vậy số cách sắp xếp là $\mathrm{F}_3^5 = 3^5 = 243$.

b. Mỗi phương án trả lời bài thi là một chỉnh hợp lặp 4 chập 10, nên số các phương án trả lời bài thi đó là $\mathrm{F}_4^{10} = 4^{10}$.

1.1.3. Chỉnh hợp (không lặp)

Mỗi phần tử của A^k có thành phần đôi một khác nhau được gọi là một chỉnh hợp n chập k ($k \le n$).

Số các chỉnh hợp (không lặp) n chập k là

$$\boxed{\mathrm{A}_n^k = n(n-1)...(n-k+1) = \frac{n!}{(n-k)!}}$$

Quy ước: $0! = 1$.

Ví dụ 1.2. Cho 6 chữ số 2, 3, 4, 5, 6,7. Hỏi

a. Có bao nhiêu số tự nhiên gồm 3 chữ số được thành lập từ 6 chữ số này?

b. Có bao nhiêu số tự nhiên gồm 3 chữ số khác nhau được thành lập từ 6 chữ số này?

c. Có bao nhiêu số tự nhiên gồm 3 chữ số khác nhau và chia hết cho 5 được thành lập từ 6 chữ số này?

Giải:

a. Mỗi số gồm 3 chữ số thành lập từ 6 chữ số này là một chỉnh hợp lặp 6 chập 3. Vậy, số các số gồm 3 chữ số lập từ 6 chữ số này là $F_6^3 = 6^3 = 216$.

b. Số các số có 3 chữ số khác nhau lập thành từ 6 chữ số này là số các chỉnh hợp 6 chập 3 là $A_6^3 = \dfrac{6!}{3!} = 4.5.6 = 120$.

c. Số chia hết cho 5 được thành lập từ 6 chữ số này phải có tận cùng là chữ số 5. Do đó, mỗi cách thành lập một số có 3 chữ số khác nhau và chia hết cho 5 là một cách thành lập một số có 2 chữ số khác nhau từ 5 chữ số còn lại là 2, 3, 4, 6, 7.

Vậy, số các số có 3 chữ số khác nhau và chia hết cho 5 thành lập từ 6 chữ số này là $A_5^2 = \dfrac{5!}{3!} = 20$.

Ví dụ 1.3. Có bao nhiêu từ gồm 2 hoặc 3 mẫu tự khác nhau được thành lập từ 6 mẫu tự của từ "MATRIX" (các từ này có thể có nghĩa hoặc không có nghĩa)?

Giải: Số các từ gồm 2 mẫu tự khác nhau là số chỉnh hợp 6 chập 2, $A_6^2 = \dfrac{6!}{4!} = 5.6 = 30$. Số các từ gồm 3 mẫu tự khác nhau là số chỉnh hợp 6 chập 3, $A_6^3 = \dfrac{6!}{3!} = 4.5.6 = 120$.

Vậy, số các từ gồm 2 hoặc 3 mẫu tự khác nhau thành lập từ 6 mẫu tự của từ "MATRIX" là $A_6^2 + A_6^3 = 30 + 120 = 150$.

1.1.4. Hoán vị

Một chỉnh hợp (không lặp) *n* chập *n* được gọi là một hoán vị của *n* phần tử.

Số các hoán vị của *n* phần tử là

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

Ví dụ 1.4.

Một đoàn khách du lịch dự định đi tham quan 7 địa điểm khác nhau A, B, C, D, E, G, H của thành phố Huế. Họ đi tham quan theo một lộ trình nào đó, chẳng hạn là B → A → C → E → D → H → G. Như vậy, mỗi lộ trình là một hoán vị của tập hợp gồm 7 phần tử là {A, B, C, D, E, G, H} nên đoàn khách có tất cả 7! lộ trình để lựa chọn.

Ví dụ 1.5.

a. Một bàn gồm 4 sinh viên. Hỏi có bao nhiêu cách sắp xếp chỗ ngồi cho 4 sinh viên đó?

b. Có bao nhiêu cách sắp xếp 3 nam và 4 nữ thành một hàng ngang sao cho nam và nữ đứng xen kẽ nhau?

Giải:

a. Số cách sắp xếp là số hoán vị của 4 phần tử, $P_4 = 4! = 24$.

b. Để 3 nam và 4 nữ đứng xen kẽ nhau thì bắt đầu của hàng ngang đó phải là nữ và chỉ có 1 cách sắp xếp vị trí như vậy:

Nữ	Nam	Nữ	Nam	Nữ	Nam	Nữ

Trong đó, số cách sắp xếp vị trí cho 3 nam là $P_3 = 3!$ và số cách sắp xếp vị trí cho 4 nữ là $P_4 = 4!$. Vậy, có tất cả 3!.4! = 144 cách sắp xếp vị trí cho 3 nam và 4 nữ.

1.1.5. Tổ hợp

Mỗi tập con gồm *k* phần tử của một tập hợp gồm *n* phần tử được gọi là một tổ hợp *n* chập *k* ($k \le n$).

Số các tổ hợp *n* chập *k*, ký hiệu là C_n^k và được tính bởi công thức

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

Các tính chất:

a. $C_n^{n-k} = C_n^k,\ k = \overline{0,n}$

b. $C_{n+1}^k = C_n^k + C_n^{k-1},\ k = \overline{1,n}$ (Hằng đẳng thức Pascal).

Nhận xét: Hai tổ hợp khác nhau khi có ít nhất một phần tử khác nhau. Tổ hợp khác chỉnh hợp ở việc không lưu ý đến thứ tự sắp xếp của các phần tử.

Ví dụ 1.6.

a. Mỗi đề thi gồm 3 câu hỏi lấy trong 25 câu hỏi cho trước. Hỏi có thể lập được bao nhiêu đề thi khác nhau?

b. Một đa giác lồi có *n* cạnh thì có bao nhiêu đường chéo?

Giải:

a. Số đoạn thẳng có 2 đầu mút là 2 đỉnh của đa giác lồi *n* đỉnh chính bằng số tổ hợp *n* chập 2, tức là C_n^2. Do đó, số đường chéo của đa giác là C_n^2 - *n*.

b. Số đề thi có thể lập nên là $C_{25}^3 = \dfrac{25!}{3!22!} = \dfrac{25.24.23}{1.2.3} = 2300$.

Ví dụ 1.7. Một hội đồng gồm 5 nam và 4 nữ được tuyển vào một ban quản trị gồm 4 người.

a. Hỏi có bao nhiêu cách tuyển chọn?

b. Hỏi có bao nhiêu cách tuyển chọn nếu có thêm điều kiện ban quản trị phải có ít nhất một nam và một nữ?

Giải:

a. Số cách tuyển chọn là số tổ hợp 9 chập 4, $C_9^4 = \dfrac{9!}{4!5!} = 126$.

b. Số cách chọn ban quản trị là số cách chọn 1 tập {*x*, *y*, z, *t*) trong đó *x* là 1 người được chọn từ 5 nam, *y* là 1 người được chọn từ 4 nữ và *z*, *t* là 2 người được chọn từ 7 người còn lại. Như vậy, số ban quản trị có thể có là $5.4.C_7^2 = 420$.

1.1.6. Nhị thức Newton

Chúng ta đã biết một số hằng đẳng thức đơn giản như

$$a + b = a^1 + b^1$$

$$(a + b)^2 = a^2 + 2ab + b^2$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

Newton đã chứng minh được công thức tổng quát sau (*nhị thức Newton*):

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

1.2. PHÉP THỬ - BIẾN CỐ NGẪU NHIÊN

Trong thực tế, chúng ta gặp nhiều thí nghiệm, phép quan sát mà khi thực hiện thì kết quả không dự đoán trước được và sự kiện chúng ta quan tâm có thể xảy ra hoặc không xảy ra. Lý thuyết xác suất gọi những phép thử như vậy là những phép thử ngẫu nhiên và các sự kiện xảy ra hay không xảy ra đó là các biến cố ngẫu nhiên.

1.2.1. Các khái niệm

Khi tiến hành gieo một con xúc sắc, gieo một đồng xu, gieo thí điểm một loại hạt giống, lấy một sản phẩm từ một lô hàng để kiểm tra hoặc quan sát trạng thái hoạt động của máy móc . . . người ta nói chúng ta đã làm một phép thử.

- Tiến hành một phép thử: là thực hiện một tập hợp các điều kiện xác định nào đó.

Tất nhiên là khi tiến hành phép thử là chúng ta nhằm nghiên cứu một sự kiện hoặc một hiện tượng nào đó. Chẳng hạn, khi "gieo thí điểm một loại hạt giống", chúng ta quan tâm đến sự kiện "nảy mầm" của hạt giống.

- Biến cố: là một kết quả của phép thử, còn được gọi là một sự kiện.

Ví dụ 1.8. Thực hiện phép thử T "gieo 1 con xúc sắc cân đối và đồng chất".

T kết thúc bởi 1 trong 6 kết quả cụ thể: ω_i = "xuất hiện mặt i chấm", $i = 1, 2, \ldots, 6$.

Khi đó, ω_i $(i = 1, 2, \ldots, 6)$ là các biến cố của T.

Ngoài ra, các kết quả phức hợp : U = "số chấm xuất hiện ≤ 6"

V = "số chấm xuất hiện > 6"

A = "số chấm xuất hiện là chẵn"

B = "số chấm xuất hiện là lẻ"

cũng là các biến cố của T.

Khi nói đến biến cố thì phải gắn liền với một phép thử. Tùy theo tính chất xuất hiện của các biến cố trong phép thử, ta có thể phân loại các biến cố:

- Phân loại:

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

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

+ Biến cố ngẫu nhiên, ký hiệu là A, B, C, . . . hay ω, là biến cố có thể xảy ra hoặc không xảy ra khi phép thử thực hiện.

- Biến cố sơ cấp: là kết quả cụ thể của phép thử, có thể hiểu đó là biến cố nhỏ nhất không thể phân chia được nữa.

Ví dụ 1.9. Trong ví dụ 1.8, các biến cố $\omega_1, \ldots, \omega_6$ gọi là các biến cố sơ cấp và U, V, A, B là các biến cố nhưng không phải là các biến cố sơ cấp.

- Không gian mẫu hay không gian các biến cố sơ cấp của một phép thử, ký hiệu Ω, là tập hợp tất cả các biến cố sơ cấp của phép thử. Mỗi biến cố A của phép thử được xem là một tập con của Ω.

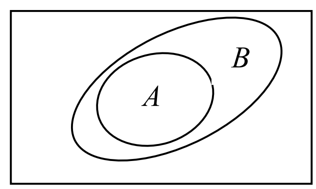
Ví dụ 1.10. Trong phép thử "gieo đồng thời 2 đồng xu cân đối và đồng chất", ta có Ω = {(SS), (SN), (NN), (NS)}.

Nếu gọi A là biến cố "xuất hiện 2 mặt giống nhau" và B là biến cố "xuất hiện 2 mặt khác nhau" thì A và B không phải là biến cố sơ cấp, vì biến cố A xảy ra khi (SS) hay (NN) xảy ra và biến cố B xảy ra khi (SN) hay (NS) xảy ra.

1.2.2. Các phép toán – Quan hệ giữa các biến cố

1. Quan hệ "kéo theo"

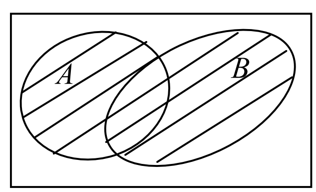
- Biến cố A gọi là kéo theo biến cố B, ký hiệu $A \subset B$, nếu và chỉ nếu A xảy ra thì kéo theo B cũng xảy ra. Mô tả hình học của quan hệ này có thể hình dung A là tập con của B.



- Nếu ω là một biến cố sơ cấp kéo theo biến cố A thì ω được gọi là một biến cố sơ cấp thuận lợi cho A.

2. Tổng của các biến cố

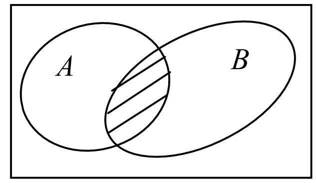
- Tổng của 2 biến cố A và B, ký hiệu $A \cup B$, là biến cố xảy ra khi ít nhất một trong hai biến cố A hay B xảy ra.



- Tổng của n biến cố $A_1, A_2, \ldots, A_n$ ($n \geq 2$), ký hiệu $\bigcup_{i=1}^{n} A_i$, là biến cố xảy ra khi ít nhất một trong n biến cố đó xảy ra.

3. Tích của các biến cố

- Tích của 2 biến cố A và B, ký hiệu $A \cap B$ (hay AB), là biến cố xảy ra khi và chỉ khi cả hai biến cố đó cùng xảy ra.

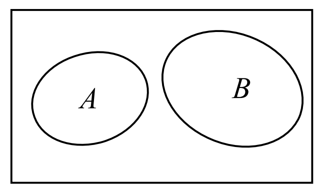


- Tích của n biến cố $A_1, A_2, \ldots, A_n$ ($n \geq 2$), kí hiệu $\bigcap_{i=1}^{n} A_i$ (hay $\prod_{i=1}^{n} A_i$), là biến cố xảy ra khi và chỉ khi n biến cố đó cùng xảy ra.

Ví dụ 1.11. Hai xạ thủ bắn mỗi người một viên đạn vào cùng một bia. Gọi A, B tương ứng là các sự kiện xạ thủ 1 và xạ thủ 2 bắn trúng bia. Có thể mô tả các biến cố $A \cup B$, $A \cap B$ như sau: $A \cup B$ là biến cố "có ít nhất một viên đạn trúng bia" và $A \cap B$ là biến cố "có hai viên đạn trúng bia".

4. Biến cố xung khắc

- Hai biến cố A và B được gọi là xung khắc nếu A và B không đồng thời xảy ra trong cùng một phép thử, ký hiệu $A \cap B = \varnothing$.



Lưu ý, trong trường hợp A và B xung khắc thì tổng của hai biến cố A và B còn được ký hiệu là $A + B$.

- Nhóm n biến cố $\{A_1, A_2, \ldots, A_n\}$ ($n \geq 2$) được gọi là xung khắc từng đôi nếu bất kỳ hai biến cố nào trong nhóm cũng xung khắc với nhau, nghĩa là $A_i \cap Aj = \varnothing$ mọi $i \neq j$.

5. Nhóm đầy đủ các biến cố

Nhóm n biến cố $\{A_1, A_2, \ldots, A_n\}$ ($n \geq 2$) được gọi là một nhóm đầy đủ các biến cố nếu có một và chỉ một biến cố trong n biến cố đó xảy ra khi phép thử thực hiện, nghĩa là

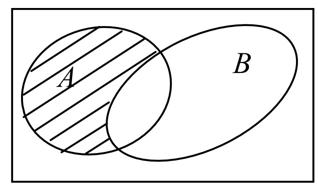
$$A_i \cap A_j = \varnothing \ (i \neq j) \text{ và } \sum_{i=1}^{n} A_i = \Omega.$$

Ví dụ 1.12. Một xí nghiệp dược phẩm có 3 cơ sở cùng sản xuất một loại thuốc A để điều trị bệnh khớp. Lấy ngẫu nhiên một sản phẩm của xí nghiệp để sử dụng. Gọi A_i = "Sản phẩm do cơ sở i sản xuất" (i = 1, 2, 3). Khi đó, $\{A_1, A_2, A_3\}$ có phải là một nhóm đầy đủ các biến cố hay không?

Giải: Lấy ngẫu nhiên một sản phẩm của xí nghiệp thì chỉ xảy ra một trong 3 trường hợp: sản phẩm do cơ sở 1 sản xuất; sản phẩm do cơ sở 2 sản xuất hoặc sản phẩm do cơ sở 3 sản xuất. Nói cách khác, ta có $A_i \cap A_j = \varnothing$ $(i \neq j)$ và $\sum_{i=1}^{n} A_i = \Omega$, rõ ràng $\{A_1, A_2, A_3\}$ là một nhóm đầy đủ các biến cố.

6. Hiệu của hai biến cố

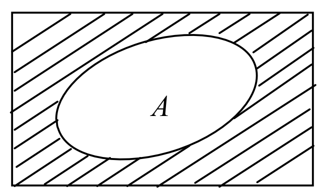
- Hiệu của 2 biến cố, ký hiệu $A \setminus B$, là một biến cố xảy ra khi và chỉ khi A xảy ra và B không xảy ra, nghĩa là $A \setminus B = A \cap \overline{B}$.



7. Biến cố đối lập

- Biến cố đối lập của biến cố A, ký hiệu $\overline{A}$, là biến cố xảy ra khi và chỉ khi A không xảy ra.

Như vậy, $\overline{A} = \Omega \setminus A$ hay $\begin{cases} A + \overline{A} = \Omega \\ A \cap \overline{A} = \varnothing. \end{cases}$



1.3. ĐỊNH NGHĨA XÁC SUẤT

Việc xảy ra hay không xảy ra của một biến cố ngẫu nhiên ta không thể dự đoán trước được. Tuy nhiên bằng trực quan có thể nhận thấy các biến cố ngẫu nhiên khác nhau thường có khả năng xảy ra khác nhau, một số biến cố thường hay xảy ra, một số khác lại thường ít xảy ra. Từ đó nảy sinh vấn đề tìm cách "đo lường" khả năng xuất hiện của mỗi biến cố và khái niệm xác suất ra đời. Xác suất của biến cố là một số không âm, đặc trưng cho khả năng khách quan xảy ra của biến cố đó.

1.3.1. Định nghĩa xác suất theo quan điểm đồng khả năng

Ta xét ví dụ sau: Thực hiện phép thử "gieo đồng thời 2 đồng xu cân đối và đồng chất". Khi đó, $\Omega = \{SS, NN, SN, NS\}$.

Nếu đồng xu được chế tạo cân đối thì các mặt của đồng xu có cùng khả năng xuất hiện nên khả năng xảy ra của 4 kết quả này là ngang nhau, ta nói phép thử có 4 biến cố sơ cấp đồng khả năng.

Gọi A là biến cố "xuất hiện 2 mặt khác nhau", nghĩa là xuất hiện một sấp (S) và một ngửa (N). Khi đó, có 2 trường hợp thuận lợi cho A hay có 2 biến cố sơ cấp thuận lợi cho A. Như vậy, khả năng để biến cố A xảy ra là $\frac{2}{4} = \frac{1}{2}$.

1. Định nghĩa 1.1. (*Theo quan điểm đồng khả năng*)

Nếu trong một phép thử có tất cả n biến cố sơ cấp đồng khả năng, trong đó có m biến cố sơ cấp thuận lợi cho A, thì xác suất của A, ký hiệu $\mathrm{P}(A)$ xác định bởi:

$$\mathrm{P}(A) = \frac{\text{Số biến cố sơ cấp thuận lợi cho } A}{\text{Số biến cố sơ cấp đồng khả năng}} = \frac{m}{n}$$

2. Tính chất 1.1.

1) $0 \le \mathrm{P}(A) \le 1$.
2) $\mathrm{P}(\Omega) = 1$, $\mathrm{P}(\varnothing) = 0$.
3) Nếu A và B xung khắc thì $\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B)$.

Chứng minh:

1) Với A là biến cố bất kỳ thì $0 \le m \le n$. Từ đó suy ra $0 \le \frac{m}{n} \le 1$ hay $0 \le \mathrm{P}(A) \le 1$.

2) Nếu Ω là biến cố chắc chắn của phép thử thì số biến cố sơ cấp thuận lợi cho Ω bằng số biến cố sơ cấp đồng khả năng của phép thử. Do đó, $\mathrm{P}(\Omega) = 1$

3) Nếu A và B là hai biến cố xung khắc thì sẽ không có biến cố sơ cấp nào thuận lợi cho cả A và B. Nói cách khác, số biến cố sơ cấp thuận lợi cho biến cố tổng $A + B$ bằng tổng số biến cố sơ cấp thuận lợi cho A và số biến cố sơ cấp thuận lợi cho B: $m = m_1 + m_2$ trong đó m_1 là số biến cố sơ cấp thuận lợi cho A và m_2 là số biến cố sơ cấp thuận lợi cho B. Từ đó suy ra

$$\mathrm{P}(A + B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = \mathrm{P}(A) + \mathrm{P}(B).$$

Hệ quả 1.1. $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$.

Ví dụ 1.13. Một hộp đựng 8 bi xanh và 4 bi trắng. Ta lấy ngẫu nhiên 3 bi, hãy tính xác suất để lấy được 1 bi xanh và 2 bi trắng?

Giải: Ta lấy ngẫu nhiên 3 bi từ một hộp chứa 12 bi thì sẽ có C_{12}^{3} trường hợp đồng khả năng hay số biến cố sơ cấp đồng khả năng là $n = C_{12}^{3} = 220$.

Gọi A là biến cố "lấy được 1 bi xanh và 2 bi trắng" thì sẽ có $C_8^1.C_4^2$ trường hợp thuận lợi cho A hay số biến cố sơ cấp thuận lợi cho A là $m = C_8^1.C_4^2 = 8.3 = 24$.

Vậy, xác suất để A xảy ra là $P(A) = \dfrac{m}{n} = \dfrac{24}{220} \approx 0{,}109$.

Nhận xét: Định nghĩa xác suất theo quan điểm đồng khả năng đòi hỏi các biến cố sơ cấp phải có tính đồng khả năng. Thường thì tính đồng khả năng được suy ra từ tính đối xứng, chẳng hạn: khi gieo một xúc sắc hay gieo một đồng xu, chúng ta có thể giả thiết nó là cân đối và đồng chất.Tuy nhiên, đây là điều kiện lý tưởng mà trên thực tế thì không có lý do nào để đảm bảo được điều đó. Hơn nữa, những bài toán trong đó ta có thể đưa ra giả thuyết về tính đối xứng thường rất hiếm gặp trong thực tế. Để khắc phục hạn chế của định nghĩa này, người ta đưa ra định nghĩa xác suất theo quan điểm thống kê.

1.3.2. Định nghĩa xác suất theo quan điểm thống kê

Định nghĩa theo thống kê dựa trên tần suất xuất hiện của biến cố trong một lớp các phép thử.

1. Tần suất xuất hiện của biến cố

Giả sử tiến hành n phép thử cùng loại và trong mỗi phép thử ta quan tâm đến sự xuất hiện của biến cố A nào đó.

Xét ví dụ: Kiểm tra ngẫu nhiên lần lượt 100 sản phẩm do một nhà máy sản xuất, người ta phát hiện thấy có 5 phế phẩm. Gọi A = "xuất hiện phế phẩm".

Khi đó, tỉ số $\dfrac{5}{100}$ gọi là tần suất xuất hiện của A trong 100 lần thử, ký hiệu:

$$f_{100}(A) = \frac{5}{100} = 0{,}05\,.$$

Một cách tổng quát, ta có định nghĩa sau:

Định nghĩa 1.2. Tần suất xuất hiện của A trong n lần thử, ký hiệu là $f_n(A)$, được xác định bởi:

$$f_n(A) = \frac{k}{n} = \frac{\text{Số lần xuất hiện của } A}{\text{Số phép thử}}$$

Nhận xét: Khi số phép thử n thay đổi thì tần suất xuất hiện biến cố cũng thay đổi. Người ta nhận thấy nếu n nhỏ thì tần suất có sự dao động rất lớn. Tuy nhiên, nếu n khá lớn thì tần suất xuất hiện biến cố thể hiện tính ổn định khá rõ ràng.

Ví dụ 1.14. Nghiên cứu khả năng xuất hiện mặt sấp khi gieo một đồng xu, Buffon và K.Pearson đã tiến hành gieo một đồng xu nhiều lần liên tiếp và thu được kết quả sau

Người làm thí nghiệm	Số lần gieo	Số lần xuất hiện mặt sấp (S)	Tần suất
Buffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

Quan sát bảng kết quả cho thấy tần suất xuất hiện mặt sấp của đồng xu trong n lần tung ổn định dần về giá trị 0,5.

2. Định nghĩa xác suất (*Theo quan điểm thống kê*)

Định nghĩa 1.3. Khi số lần thực hiện phép thử n khá lớn, nếu tần suất của biến cố A ổn định dần về một giá trị p nào đó thì ta nói A ổn định ngẫu nhiên và số p được gọi là xác suất của biến cố A, ký hiệu P(A).

Từ định nghĩa xác suất theo quan điểm thống kê, ta có thể xấp xỉ P(A) với $f_n(A)$ khi n khá lớn.

$$\boxed{\mathrm{P}(A) \approx f_n(A) \text{ khi } n \text{ khá lớn.}}$$

1.3.3. Định nghĩa xác suất theo quan điểm hình học

Để khắc phục hạn chế của định nghĩa xác suất đòi hỏi phép thử phải có hữu hạn các biến cố sơ cấp đồng khả năng, người ta đưa ra định nghĩa xác suất theo quan điểm hình học. Định nghĩa này mở rộng cho trường hợp phép thử có vô hạn các biến cố sơ cấp đồng khả năng.

Định nghĩa 1.4. (*Theo quan điểm hình học*)

Xét phép thử ngẫu nhiên mà không gian các biến cố sơ cấp Ω được biểu diễn bằng một miền hình học H nào đó trong không gian $\mathbb{R}$, $\mathbb{R}^2$ hay $\mathbb{R}^3$ và tập các biến cố sơ cấp thuận lợi cho biến cố A được biểu diễn bởi một miền $G \subset H$.

Khi đó, nếu mọi điểm của Ω là đồng khả năng thì xác suất của biến cố A là

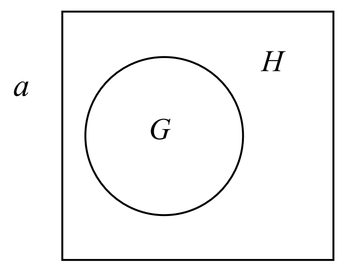
$$\mathrm{P}(A) = \frac{\text{Độ đo miền } G}{\text{Độ đo miền } H}$$

$H \subset \mathbb{R}$: độ đo là độ dài.

$H \subset \mathbb{R}^2$: độ đo là diện tích.

$H \subset \mathbb{R}^3$: độ đo là thể tích.

Ví dụ 1.15. Gieo 1 chấm điểm một cách ngẫu nhiên vào mảnh vải hình vuông H cạnh a, trong đó có một hình tròn G bán kính $r = \frac{a}{4}$. Tìm xác suất để chấm điểm rơi vào trong hình tròn?



Giải:

Khi phép thử "gieo một chấm điểm vào mảnh vải hình vuông H" thực hiện thì sẽ có vô hạn các trường hợp có thể xảy ra. Tập các biến cố sơ cấp đồng khả năng có thể có được biểu diễn bởi miền hình vuông H có cạnh a.

Gọi A là biến cố "chấm điểm rơi vào trong hình tròn G". Khi đó, các biến cố sơ cấp thuận lợi cho A cũng không thể xác định cụ thể. Tuy nhiên, ta nhận thấy rằng tập các biến cố sơ cấp thuận lợi cho A được biểu diễn bởi miền hình tròn G có bán kính $r = \frac{a}{4}$.

Theo định nghĩa xác suất, ta có xác suất để A xảy ra là

$$\mathrm{P}(A) = \frac{\text{Diện tích } G}{\text{Diện tích } H} = \frac{\pi r^2}{a^2}$$

hay $\mathrm{P}(A) = \dfrac{\pi \frac{a^2}{16}}{a^2} = \dfrac{\pi}{16}$.

Những biến cố có xác suất càng gần 1 thì rất dễ xảy ra và được xem là hầu như chắc chắn. Ngược lại, những biến cố có xác suất rất nhỏ thì khả năng xảy ra rất ít và được xem là hầu như không thể xảy ra. Việc quy định một mức xác suất như thế nào để có thể xem là hầu như chắc chắn hay hầu như không chắc chắn tùy thuộc vào từng bài toán cụ thể. Chẳng hạn nếu xác suất máy bay rơi là 0,01 thì xác suất đó chưa thể xem là nhỏ, hầu như không thể xảy ra. Nhưng nếu xác suất một chuyến tàu khởi hành chậm là 0,01 thì có thể xem xác suất này là rất nhỏ.

1.4. CÁC QUY TẮC TÍNH XÁC SUẤT

1.4.1. Định lý cộng xác suất

Từ định nghĩa xác suất, ta đã suy ra được tính chất cơ bản sau của xác suất:

+ Nếu A, B là hai biến cố xung khắc thì

$$P(A + B) = P(A) + P(B). \qquad (1.1)$$

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

$$P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i). \qquad (1.2)$$

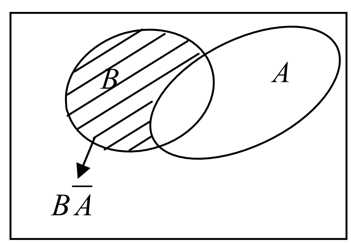
Định lý 1.1. (*Định lý cộng xác suất*)

Nếu A, B là hai biến cố bất kỳ thì $P(A \cup B) = P(A) + P(B) - P(AB)$. (1.3)

Chứng minh:

Trường hợp A, B là hai biến cố bất kỳ, biến cố tổng của chúng có thể biểu diễn thành tổng của 2 biến cố xung khắc là $A \cup B = A + B\overline{A}$. Khi đó, $P(A \cup B) = P(A) + P(B\overline{A})$.

Mặt khác, $B = B(A+\overline{A}) = BA + B\overline{A}$ nên $P(B) = P(BA) + P(B\overline{A})$, suy ra $P(B\overline{A}) = P(B) - P(BA)$. Từ đó, ta có $P(A \cup B) = P(A) + P(B) - P(AB)$.



Mở rộng:

- Nếu A, B, C là ba biến cố bất kỳ thì

  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

- Nếu $A_1, A_2, \ldots, A_n$ là một nhóm đầy đủ các biến cố thì $\sum_{i=1}^{n} P(A_i) = 1$.

Ví dụ 1.16. Trong một lớp học, tỉ lệ học sinh đạt điểm giỏi môn Toán là 10%, tỉ lệ học sinh đạt điểm giỏi môn Anh là 9% và giỏi cả hai môn là 5%. Chọn ngẫu nhiên một học sinh trong lớp. Tính xác suất để học sinh đó không đạt điểm giỏi cả môn Toán lẫn môn Anh?

Giải: Gọi A là biến cố "Sinh viên đó đạt điểm giỏi môn Toán", B là biến cố "Sinh viên đó đạt điểm giỏi môn Anh". Theo giả thiết thì

$$P(A) = 0{,}01,\ P(B) = 0{,}09 \text{ và } P(AB) = 0{,}05$$

Gọi C là biến cố "Sinh viên đó không đạt điểm giỏi cả môn Toán lẫn môn Anh" thì $\overline{C}$ là biến cố "Sinh viên đó đạt điểm giỏi môn Toán hoặc môn Anh" hay $\overline{C} = A \cup B$. Theo định lý cộng xác suất (1.3), ta có

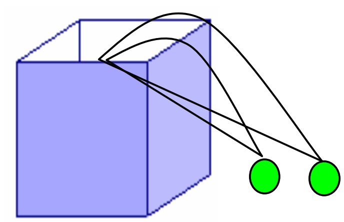
$$P(\overline{C}) = P(A \cup B) = P(A) + P(B) - P(AB) = 0{,}01 + 0{,}09 - 0{,}05 = 0{,}14.$$

Suy ra, $$P(C) = 1 - P(\overline{C}) = 1 - 0{,}14 = 0{,}86.$$

1.4.2. Định lý nhân xác suất

1. Xác suất có điều kiện

Xét ví dụ: Một hộp chứa 10 viên bi giống nhau, trong đó có 6 bi xanh và 4 bi trắng. Người thứ 1 lấy ngẫu nhiên 1 bi (không trả lại vào hộp). Tiếp đó, người thứ 2 lấy 1 bi. Tính xác suất để người thứ 2 lấy được bi xanh nếu biết người thứ 1 đã lấy được bi xanh?



Giải:

Gọi A là biến cố "Người thứ 1 lấy được bi xanh"

B là biến cố "Người thứ 2 lấy được bi xanh". Khi đó, xác suất $P(B)$ sẽ phụ thuộc vào việc A xảy ra hay không xảy ra.

+ Nếu A đã xảy ra thì xác suất của B là $\frac{5}{9}$, ký hiệu $P(B|A) = \frac{5}{9}$.

+ Nếu A không xảy ra thì xác suất của B là $\frac{6}{9}$, ký hiệu $P(B|\overline{A}) = \frac{6}{9}$.

Như vậy, việc xảy ra hay không xảy ra của A đã ảnh hưởng đến khả năng xảy ra của B. Xác suất của B trong điều kiện A đã xảy ra được gọi là xác suất có điều kiện của B trong điều kiện A đã xảy ra, ký hiệu $P(B|A)$.

Định nghĩa 1.5. Giả sử Ω là không gian các biến cố sơ cấp và B là một biến cố ngẫu nhiên của phép thử. Nếu $P(B) > 0$ thì xác suất có điều kiện của biến cố A với điều kiện B đã xảy ra, ký hiệu $P(A|B)$, được xác định bởi:

$$\mathrm{P}(A\,|\,B) = \frac{\mathrm{P}(AB)}{\mathrm{P}(B)}. \qquad (1.4)$$

Tính chất 1.2.

1) $0 \le \mathrm{P}(A\,|\,B) \le 1$.

2) $\mathrm{P}(\Omega\,|\,B) = 1$.

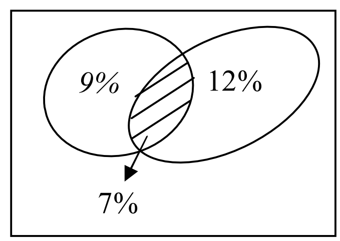
3) $\mathrm{P}(B\,|\,B) = 1$.

4) Nếu $A \cap C = \varnothing$ thì $\mathrm{P}((A+C)\,|\,B) = \mathrm{P}(A\,|\,B) + \mathrm{P}(C\,|\,B)$.

5) $\mathrm{P}(\overline{A}\,|\,B) = 1 - \mathrm{P}(A\,|\,B)$.

Ví dụ 1.17. Trong một vùng dân cư, tỉ lệ người mắc bệnh tim là 9%, mắc bệnh huyết áp là 12%, mắc cả bệnh tim và huyết áp là 7%. Chọn ngẫu nhiên một người dân trong vùng, biết người đó mắc bệnh tim, tìm xác suất để người đó không mắc bệnh huyết áp?

Giải:



Gọi A là biến cố "người đó bị mắc bệnh tim"

B là biến cố "người đó bị mắc bệnh huyết áp".

Theo giả thiết, ta có

$$\mathrm{P}(A) = 0{,}09\ ,\ \mathrm{P}(B) = 0{,}12 \text{ và } \mathrm{P}(AB) = 0{,}07$$

Khi đó, $\mathrm{P}(B\,|\,A)$ là xác suất người chọn ra bị mắc bệnh huyết áp biết người đó đã mắc bệnh tim. Vậy thì $\mathrm{P}(\overline{B}\,|\,A)$ chính là xác suất người chọn ra không bị mắc bệnh huyết áp biết người đó đã mắc bệnh tim.

Theo công thức xác suất có điều kiện (1.4)

$$\mathrm{P}(B\,|\,A) = \frac{\mathrm{P}(BA)}{\mathrm{P}(A)} = \frac{0{,}07}{0{,}09} \approx 0{,}667.$$

Do đó, $\mathrm{P}(\overline{B}\,|\,A) = 1 - \mathrm{P}(B\,|\,A) = 1 - 0{,}667 \approx 0{,}333$.

Ví dụ 1.18. Để xét hiệu quả của một loại Vaccine, người ta điều tra tình hình mắc bệnh trên 1000 người dân có và không tiêm phòng loại Vaccine này. Số liệu thu được như sau:

	Mắc bệnh	Không mắc bệnh	Tổng số
Có tiêm phòng (A)	12	188	200
Không tiêm phòng (B)	288	512	800

Tổng số	300	700	1000

a. Chọn ngẫu nhiên một người trong số 1000 người này, biết người đó có tiêm phòng. Xét xem khả năng người đó mắc bệnh là bao nhiêu?

b. Chọn ngẫu nhiên một người trong số 1000 người đó, biết người đó thuộc nhóm không tiêm phòng. Xét xem khả năng người đó không mắc bệnh là bao nhiêu?

Giải:

a. Gọi A là biến cố "người đó có tiêm phòng"; B là biến cố "người đó không tiêm phòng"; C là biến cố "người đó bị mắc bệnh" và D là biến cố "người đó không bị mắc bệnh".

Khi đó, $\mathrm{P}(A) \approx \dfrac{200}{1000} = 0{,}2$ $\qquad$ $\mathrm{P}(B) \approx \dfrac{800}{1000} = 0{,}8$

$\mathrm{P}(C) \approx \dfrac{300}{1000} = 0{,}3$ $\qquad$ $\mathrm{P}(D) \approx \dfrac{700}{1000} = 0{,}7.$

AC là biến cố "người đó có tiêm phòng và bị mắc bệnh"

AD là biến cố "người đó có tiêm phòng và không bị mắc bệnh"

BC là biến cố "người đó không tiêm phòng và mắc bệnh"

BD là biến cố "người đó không tiêm phòng và không bị mắc bệnh".

$\mathrm{P(AC)} \approx \dfrac{12}{1000} = 0{,}012$ $\qquad$ $\mathrm{P(AD)} \approx \dfrac{188}{1000} = 0{,}188$

$\mathrm{P(BC)} \approx \dfrac{288}{1000} = 0.288$ $\qquad$ $\mathrm{P(BD)} \approx \dfrac{512}{1000} = 0{,}512.$

$\mathrm{P}(C \mid A)$ là xác suất người được chọn bị mắc bệnh biết người đó có tiêm phòng, ta có

$$\mathrm{P}(C \mid A) = \frac{\mathrm{P}(AC)}{\mathrm{P}(A)} \approx \frac{12}{200} = 0{,}06.$$

b. $\mathrm{P}(C \mid B)$ là biến cố người được chọn không bị mắc bệnh biết người đó thuộc nhóm không tiêm phòng, ta có

$$\mathrm{P}(C \mid B) = \frac{\mathrm{P}(BC)}{\mathrm{P}(B)} \approx \frac{288}{800} = 0{,}36.$$

2. Tính độc lập của các biến cố

Định nghĩa 1.6. Hai biến cố A và B được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra của biến cố này không ảnh hưởng đến khả năng xảy ra của biến cố kia và ngược lại. Nói cách khác, hai biến cố A và B được gọi là độc lập với nhau nếu $\mathrm{P}(B \mid A) = \mathrm{P}(B)$ hay $\mathrm{P}(A \mid B) = \mathrm{P}(A)$.

Mở rộng khái niệm độc lập cho n biến cố, ta có

- Hệ $A_1, A_2, \ldots, A_n$ gọi là độc lập từng đôi nếu và chỉ nếu bất kỳ hai biến cố nào trong nhóm cũng độc lập với nhau, nghĩa là $\mathrm{P}(A_i \mid A_k) = \mathrm{P}(A_i)$ $(i \neq k)$.

- Hệ $A_1, A_2, \ldots, A_n$ gọi là độc lập trong toàn thể nếu và chỉ nếu bất kỳ biến cố nào trong nhóm cũng độc lập với tích một số bất kỳ các biến cố trong $(n-1)$ biến cố còn lại. Điều này có nghĩa là mọi dãy $(i_1, i_2, \ldots, i_k) \subset (1, 2, \ldots, \mathrm{n})$,

$$\mathrm{P}(A_j \mid A_{i_1} \ldots A_{i_k}) = \mathrm{P}(A_j),\ j \notin \{i_1, i_2, \ldots, i_k\}.$$

3. Định lý nhân xác suất

Định lý 1.2. (Định lý nhân xác suất)

a. Với A, B là 2 biến cố bất kỳ, ta có $\mathrm{P}(AB) = \mathrm{P}(A \mid B).\mathrm{P}(B)$ với $\mathrm{P}(B) > 0$. (1.5)

b. Với $A_1, A_2, \ldots, A_n$ là n biến cố bất kỳ, ta có:

$$\mathrm{P}(A_1 \ldots A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2 \mid A_1)\mathrm{P}(A_3 \mid A_1 A_2) \ldots \mathrm{P}(A_n \mid A_1 A_2 \ldots A_{n-1}) \quad (1.6)$$

Từ định nghĩa về tính độc lập của các biến cố và định lý nhân xác suất, ta suy ra:

c. Hai biến cố A và B độc lập với nhau nếu và chỉ nếu $\mathrm{P}(AB) = \mathrm{P}(A).\mathrm{P}(B)$.

d. Nếu $\{A_1, A_I, \ldots, A_n\}$ độc lập trong toàn thể thì $\mathrm{P}(A_1 A_2 \ldots A_n) = \mathrm{P}(A_1) \ldots \mathrm{P}(A_n)$.

Ví dụ 1.19. Hai xạ thủ bắn mỗi người một viên vào cùng một mục tiêu. Xác suất trúng đích của xạ thủ 1 là 0,7; của xạ thủ 2 là 0,6. Tính xác suất để mục tiêu bị trúng đạn?

Giải: Gọi A là biến cố "xạ thủ 1 bắn trúng mục tiêu"

B là biến cố "xạ thủ 2 bắn trúng mục tiêu"

H là biến cố "mục tiêu bị trúng đạn".

Lúc đó, $\mathrm{P}(A) = 0{,}7$, $\mathrm{P}(B) = 0{,}6$ và $H = A \cup B$. Theo công thức cộng xác suất,

$$\mathrm{P}(H) = \mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(AB).$$

Mặt khác, do hai biến cố A và B độc lập với nhau nên $\mathrm{P}(AB) = \mathrm{P}(A).\mathrm{P}(B) = 0{,}7.0{,}6 = 0{,}42$.

Suy ra, xác suất để mục tiêu bị trúng đạn là

$$\mathrm{P}(H) = 0{,}7 + 0{,}6 - 0{,}42 = 0{,}88.$$

Ví dụ 1.20. Một người bắn liên tiếp vào một mục tiêu cho đến khi có một phát đạn trúng mục tiêu thì ngưng bắn. Biết rằng xác suất trúng mục tiêu của mỗi lần bắn là như nhau và bằng 0,6. Tính xác suất sao cho khi bắn đến phát thứ tư thì ngưng bắn.

Giải: Gọi A_i là biến cố "phát thứ i trúng mục tiêu", $i = 1, 2, 3, \ldots, n$.

A là biến cố "bắn đến phát thứ tư thì ngưng"

Ta có, $A = \overline{A_1}\ \overline{A_2}\ \overline{A_3}\ A_4$. Theo định lý nhân xác suất thì:

$$\mathrm{P}(A) = \mathrm{P}(\overline{A_1})\,\mathrm{P}(\overline{A_2} \mid \overline{A_1})\ \mathrm{P}(\overline{A_3} \mid \overline{A_1}.\overline{A_2})\ \mathrm{P}(A_4 \mid \overline{A_1}.\overline{A_2}.\overline{A_3}).$$

Trong đó, $\mathrm{P}(\overline{A_1}) = 1 - 0{,}6 = 0{,}4$.

Mặt khác, do $\overline{A_2} \subset \overline{A_1}$; $\overline{A_3} \subset \overline{A_2} \subset \overline{A_1}$ nên $\overline{A_1}\ \overline{A_2} = \overline{A_2}$ và $\overline{A_1}\ \overline{A_2}\ \overline{A_3} = \overline{A_3}$. Từ đó

suy ra $\mathrm{P}(\overline{A_2} \mid \overline{A_1}) = 1 - 0{,}6$ $\quad \mathrm{P}(\overline{A_3} \mid \overline{A_1}.\overline{A_2}) = \mathrm{P}(\overline{A_3} \mid \overline{A_2}) = 1 - 0{,}6$

$$\mathrm{P}(A_4 \mid \overline{A_1}.\overline{A_2}.\overline{A_3}) = \mathrm{P}(A_4 \mid \overline{A_3}) = 0{,}6.$$

Vậy, $\quad \mathrm{P}(A) = 0{,}4.0{,}4.0{,}4.0{,}6 = 0{,}0384.$

1.4.3. Công thức xác suất đầy đủ - Công thức Bayes

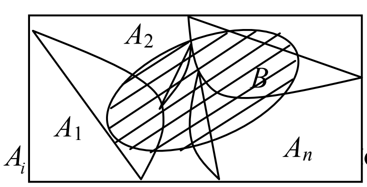
1. Công thức xác suất đầy đủ

Định lý 1.3. Giả sử $\{A_1, A_2, \ldots, A_n\}$ là một nhóm đầy đủ các biến cố, $\mathrm{P}(A_i) > 0 \quad (i = \overline{1, n})$ và B là biến cố bất kỳ trong cùng phép thử. Khi đó,ta có

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(B \mid A_i).\mathrm{P}(A_i) \tag{1.7}$$

Chứng minh: Cho $\{A_1, A_2, \ldots, A_n\}$ là nhóm đầy đủ các biến cố và B là một biến cố bất kỳ của phép thử, ta có $B = BA_1 + \ldots + BA_n$, ở đây $\{BA_1, \ldots, BA_n\}$ xung khắc từng đôi nên theo công thức cộng xác suất

$\mathrm{P}(B) = \mathrm{P}(BA_1) + \ldots + \mathrm{P}(BA_n)$.



Hơn nữa, theo công thức nhân xác suất $\mathrm{P}(BA_i) = \mathrm{P}(B \mid A_i$ ó,

ta có $\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(B \mid A_i).\mathrm{P}(A_i)$.

Công thức này được gọi là công thức xác suất đầy đủ, nó cho phép tính xác suất của biến cố B đối với toàn nhóm biến cố đầy đủ $A_1, A_2, \ldots, A_n$.

2. Công thức xác suất Bayes

Cho $\{A_1, A_2, \ldots, A_n\}$ là một nhóm đầy đủ các biến cố với $P(A_i) > 0$ và B là một biến cố bất kỳ trong cùng phép thử, $P(B) > 0$.

Theo công thức nhân xác suất: $P(A_k.B) = P(A_k \mid B).P(B) \qquad (P(B) > 0)$

$$P(A_k.B) = P(B \mid A_k).P(A_k) \qquad (P(A_k) > 0)$$

Suy ra, $P(A_k \mid B) = \dfrac{P(B \mid A_k).P(A_k)}{P(B)}$ thay $P(B)$ bằng công thức xác suất đầy đủ, ta có

$$P(A_k \mid B) = \frac{P(B \mid A_k).P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i).P(A_i)}.$$

Ta có định lý sau

Định lý 1.4. Cho $\{A_1, A_2, \ldots, A_n\}$ là một nhóm đầy đủ các biến cố với $P(A_i) > 0$ và B là một biến cố bất kỳ trong cùng phép thử, $P(B) > 0$. Khi đó,

$$P(A_k \mid B) = \frac{P(B \mid A_k).P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i).P(A_i)},\ k = 1, 2, \ldots, n \qquad (1.8)$$

Công thức này được gọi là công thức xác suất Bayes. Các xác suất $P(A_1)$, $P(A_2)$, . . ., $P(A_n)$ được xác định trước khi phép thử được tiến hành, được gọi là các xác suất tiên nghiệm. Các xác suất $P(A_k \mid B)$ được xác định sau khi phép thử được tiến hành và biến cố B đã xảy ra, được gọi là các xác suất hậu nghiệm.

Ví dụ 1.21. Một trại chăn nuôi nhận 50 con giống từ ba cơ sở, trong đó 15 con thuộc cơ sở 1; 10 con thuộc cơ sở 2 và 25 con thuộc cơ sở 3. Tỉ lệ con giống không đạt tiêu chuẩn của mỗi cơ sở tương ứng là 16%, 15% và 12%.

a. Hãy xác định tỉ lệ con giống đạt tiêu chuẩn của cả lô con giống?

b. Kiểm tra ngẫu nhiên 1 con từ trại chăn nuôi này thấy không đạt tiêu chuẩn. Hãy xét xem trách nhiệm thuộc về cơ sở nào là lớn hơn?

Giải:

a. Khi thực hiện phép thử "kiểm tra một con giống của trại chăn nuôi" thì có một và chỉ một trong 3 biến cố sau xảy ra:

A_i = "con giống thuộc cơ sở i" $i = 1, 2, 3$. Khi đó, $\{A_1, A_2, A_3\}$ là một nhóm đầy đủ các biến cố.

Gọi B là biến cố "con giống không đạt tiêu chuẩn". Theo giả thiết, ta có

$$P(A_1) = 0{,}3 \qquad P(A_2) = 0{,}2 \qquad P(A_3) = 0{,}5$$

$$P(B \mid A_1) = 0{,}16 \qquad P(B \mid A_2) = 0{,}15 \qquad P(B \mid A_3) = 0{,}12 .$$

Cần tính $P(\overline{B})$?

Theo công thức xác suất đầy đủ, ta có $P(B) = \sum_{i=1}^{3} P(B \mid A_i)P(A_i)$ = 0,3 . 0,16 + 0,2 . 0,15 + 0,5 . 0,12 = 0,048 + 0,03 + 0,06 = 0,138. Từ đó suy ra, xác suất để một con giống đạt tiêu chuẩn là $P(\overline{B})$ = 1 – P(B) = 1 - 0,138 = 0,862.

Vậy, tỉ lệ con giống đạt tiêu chuẩn của cả lô con giống là 86,2%.

b. B đã xảy ra, cần tính và so sánh $P(A_1 \mid B)$, $P(A_2 \mid B)$, $P(A_3 \mid B)$?

Áp dụng công thức xác suất Bayes

$$P(A_1 \mid B) = \frac{P(B \mid A_1).P(A_1)}{P(B)} = \frac{0{,}3.0{,}16}{0{,}138} = \frac{0{,}048}{0{,}138}$$

$$P(A_2 \mid B) = \frac{P(B \mid A_2).P(A_2)}{P(B)} = \frac{0{,}2.0{,}15}{0{,}138} = \frac{0{,}03}{0{,}138}$$

$$P(A_3 \mid B) = \frac{P(B \mid A_3).P(A_3)}{P(B)} = \frac{0{,}5.0{,}12}{0{,}138} = \frac{0{,}06}{0{,}138}.$$

Ta thấy $P(A_3 \mid B)$ là lớn nhất nên nhiều khả năng nhất là con giống không đạt tiêu chuẩn đó được nhận về từ cơ sở 3. Nghĩa là, trách nhiệm thuộc về cơ sở 3 là lớn hơn.

1.5. DÃY PHÉP THỬ BERNOULLI

Định nghĩa 1.7. Một phép thử trong đó biến cố A xảy ra với xác suất p và không xảy ra với xác suất $q = 1 - p$ được gọi là phép thử Bernoulli. Tiến hành lặp lại phép thử Bernoulli n lần độc lập nhau, ta có dãy n phép thử Bernoulli hay còn gọi là lược đồ Bernoulli.

Chẳng hạn, tỷ lệ nảy mầm của một loại hạt giống là 75%, người ta gieo thí điểm 10 hạt. Đó là dãy 10 phép thử Bernoulli.

1.5.1. Công thức Bernoulli

Bài toán đặt ra: Tìm xác suất để trong n phép thử Bernoulli, biến cố A xuất hiện k lần.

Định lý 1.5. Xác suất để trong n phép thử Bernoulli, biến cố A xuất hiện đúng k lần, ký hiệu $P_n(k)$, được tính theo công thức:

$$P_n(k) = C_n^k p^k q^{n-k} \quad \text{với } k = \overline{0,n}. \tag{1.9}$$

Chứng minh:

Gọi B = "trong n phép thử, A xuất hiện k lần".

B xảy ra theo nhiều phương thức khác nhau, trong đó việc xảy ra của A đúng k lần và $\overline{A}$ đúng $n-k$ lần có thể diễn ra theo các trình tự khác nhau. Nói cách khác, B là tổng của các biến cố xung khắc có dạng sau

$$B = A_1A_2...A_k\overline{A_{k+1}}...\overline{A_n} + ... + \overline{A_1}\,\overline{A_2}...\overline{A_{n-k}}A_{n-k+1}...A_n$$

Mỗi biến cố thành phần của B là một cách chọn k phép thử trong đó A xảy ra từ n vị trí, suy ra tổng số các biến cố thành phần của B là C_n^k.

Mặt khác, A_i và A_j độc lập với nhau nên mỗi biến cố thành phần của B có xác suất là p^kq^{n-k}.

Hơn nữa, các biến cố thành phần của B là các biến cố xung khắc từng đôi nên $\mathrm{P}(B) = \mathrm{C}_n^k p^k q^{n-k}$.

Hệ quả 1.2. Xác suất để trong n phép thử Bernoulli có từ k_1 đến k_2 lần xuất hiện biến cố A, kí hiệu $\mathrm{P}_n(k_1,k_2)$, được tính theo công thức $\mathrm{P}_n(k_1,k_2) = \sum_{k=k_1}^{k_2} \mathrm{C}_n^k p^k q^{n-k}$.

Ví dụ 1.22. Xác suất thành công của một thí nghiệm sinh học là 0,7. Một nhóm gồm 5 sinh viên cùng tiến hành thí nghiệm trên một cách độc lập nhau. Tính xác suất để trong 5 thí nghiệm:

a. Có đúng 3 thí nghiệm thành công.

b. Có từ 2 đến 4 thí nghiệm thành công.

c. Có ít nhất 1 thí nghiệm thành công.

Giải:

a. Gọi A là biến cố "thí nghiệm thành công". Khi đó, $\mathrm{P}(A) = p = 0{,}7$.

Xác suất để có đúng 3 thí nghiệm thành công được tính theo công thức Bernoulli là

$$\mathrm{P}_5(3;0,7) = \mathrm{C}_5^3(0,7)^3.(0,3)^2 = 0{,}3087.$$

b. Xác suất để có từ 2 đến 4 thí nghiệm thành công là $\mathrm{P}_5(2,4) = \mathrm{C}_5^2(0,7)^2.(0,3)^3 + \mathrm{C}_5^3(0,7)^3.(0,3)^2 + \mathrm{C}_5^4(0,7)^4.(0,3)^1 = 0{,}80115.$

c. Gọi B là biến cố "có ít nhất 1 thí nghiệm thành công". Khi đó,

$\overline{B}$ là biến cố "không có thí nghiệm nào thành công".

Ta có, $P(\overline{B}) = P_5(0;0,7) = C_5^0(0,7)^0.(0,3)^5 = 0,00243$.

Suy ra, $P(B) = 1 - P(\overline{B}) = 0,99757$.

1.5.2. Số có khả năng nhất

Bài toán đặt ra: Khi gieo thí điểm 10 hạt giống, khả năng lớn nhất là có bao nhiêu hạt nảy mầm. Theo công thức Bernoulli, có thể tính được xác suất để trong 10 hạt được gieo có k hạt nảy mầm (k = 0, 1, 2, ..., 10). Trong các xác suất này sẽ tồn tại số lớn nhất và k_0 ứng với xác suất lớn nhất sẽ là số hay xảy ra nhất.

Định nghĩa 1.8. Cho một dãy n phép thử Bernoulli (độc lập), số k_0 mà ứng với nó $P_n(k_0)$ lớn nhất được gọi là số có khả năng nhất:

$$P_n(k_0) = \max_{0\le k\le n} P(k).$$

Quy tắc tìm số có khả năng nhất:

a. Nếu $(n + 1)p$ là số nguyên thì $k_0 = (n + 1)p$ và $k_0 = (n + 1)p - 1$

b. Nếu $(n + 1)p$ là số thập phân thì k_0 là số nguyên lớn nhất thỏa mãn

$$k_0 < (n + 1)p$$

Chứng minh: Xét sự biến thiên của $P_n(k)$ theo k để tìm k_0.

Ta lập tỉ số $\dfrac{P_n(k+1)}{P_n(k)} = \dfrac{C_n^{k+1} p^{k+1}.q^{n-k-1}}{C_n^k p^k.q^{n-k}} = \dfrac{(n-k).p}{(k+1).q}$.

Từ đây suy ra rằng, $P_n(k+1) > P_n(k)$ khi $k < np - q$ hay $k < (n+1)p - 1$; $P_n(k+1) = P_n(k)$ khi $k = np - q$ hay $k = (n+1)p - 1$; $P_n(k+1) < P_n(k)$ khi $k > np - q$ hay $k > (n+1)p - 1$.

Ta thấy khi k tăng từ 0 đến n, hàm $P_n(k)$ thoạt tiên tăng, sau đó đạt cực đại, rồi giảm dần. Do vậy,

a. Nếu $(n + 1)p$ là số nguyên thì $(n+1)p - 1$ cũng là số nguyên, khi đó $P_n(k)$ đạt cực đại tại 2 giá trị của k là $k_0 = (n + 1)p$ và $k_0 = (n + 1)p - 1$.

b. Nếu $(n + 1)p$ là số thập phân thì $(n+1)p - 1$ cũng là số thập phân, khi đó $P_n(k)$ đạt cực đại tại $k = k_0$ với k_0 là số nguyên thỏa mãn $P_n(k_0) \ge P_n(k_0 - 1) \Leftrightarrow k_0 - 1 \le (n+1)p - 1$ hay $k_0 \le (n+1)p$. Do k là số nguyên nên $k_0 < (n+1)p$.

Ví dụ 1.23. Tỉ lệ nảy mầm của một loại hạt giống là 75%, người ta gieo thí điểm 10 hạt giống loại này.

a. Tính xác suất để có 9 hạt nảy mầm?

b. Tìm số hạt giống nảy mầm có khả năng nhất?

Giải:

a. Việc gieo mỗi hạt giống là một phép thử, ta có 10 phép thử độc lập. Trong mỗi phép thử chỉ có 2 trường hợp: Hoặc hạt giống nảy mầm, hoặc hạt giống không nảy mầm. Xác suất để một hạt giống nảy mầm là 0,75. Như vậy, bài toán thỏa mãn lược đồ Bernoulli với $n = 10$ và $p = 0,75$.

Do đó, xác suất để có đúng 9 hạt nảy mầm được tính theo công thức Bernoulli là

$$P_{10}(9) = C_{10}^{9}(0,75)^{9}.(0,25)^{1}.$$

b. Ta có $(n + 1)p = 10.0,75 = 7,5$.

Gọi k_0 là số hạt giống nảy mầm có khả năng nhất thì k_0 là số nguyên lớn nhất thỏa $k_0 < 7,5$ nên $k_0 = 7$.

Vậy, số hạt giống nảy mầm có khả năng nhất là $k_0 = 7$ hạt.

BÀI TẬP CHƯƠNG 1

Bài 1. Một hộp gồm 5 bi trắng và 3 bi xanh. Lấy từ hộp ra 2 bi, có 3 cách lấy:

Cách 1: lấy ngẫu nhiên đồng thời 2 bi.

Cách 2: lấy ngẫu nhiên lần lượt (không hoàn lại) 2 bi.

Cách 3: lấy ngẫu nhiên lần lượt có hoàn lại 2 bi.

Xét theo mỗi cách lấy:

a) Có bao nhiêu cách lấy 2 bi?

b) Có bao nhiêu cách lấy 2 bi trắng?

c) Có bao nhiêu cách lấy 1 bi trắng và 1 bi xanh?

Bài 2. Có bao nhiêu cách sắp xếp k quả cầu khác nhau vào n hộp khác nhau?

Bài 3. Có bao nhiêu cách phân phối 15 sản phẩm cho 3 cơ sở sao cho cơ sở 1 có 2 sản phẩm, cơ sở 2 có 3 sản phẩm và cơ sở 3 có 10 sản phẩm?

Bài 4. Một lớp học có 50 sinh viên trong đó có 30 là nam. Có bao nhiêu cách chọn ra một ban cán sự gồm 4 sinh viên nếu:

a) Có đúng 2 nam

b) Có nhiều nhất 2 nam

c) Không có nam

d) Có ít nhất 1 nam.

Bài 5.

a) Có bao nhiêu cách sắp xếp 5 sinh viên và 2 giáo viên ngồi trên một chiếc ghế dài sao cho 2 giáo viên luôn ngồi cạnh nhau?

b) Có bao nhiêu cách sắp xếp 5 sinh viên và 2 giáo viên ngồi quanh một chiếc bàn tròn sao cho 2 giáo viên luôn ngồi cạnh nhau?

Bài 6. Lấy ngẫu nhiên trong một lô hàng ra 4 sản phẩm để kiểm tra, quan tâm đến số phế phẩm trong 4 sản phẩm đó.

a) Xác định các biến cố sơ cấp, không gian mẫu.

b) Biểu diễn các biến cố sau theo các biến cố sơ cấp

A = "có nhiều nhất 1 phế phẩm"

B = "có ít nhất 1 phế phẩm"

C = "có ít nhất 2 phế phẩm".

c) Chỉ ra các biến cố xung khắc, đối lập trong các biến cố thu được.

Bài 7. Ba xạ thủ cùng bắn, mỗi người 1 viên vào cùng một bia. Gọi A, B, C là các sự kiện các xạ thủ tương ứng bắn trúng bia.

a) Nhóm các biến cố $\{A, B, C\}$ có xung khắc từng đôi hay không?

b) Hãy biểu diễn các sự kiện sau theo A, B, C:

D = "có ít nhất 1 viên trúng bia"

E = "cả 3 viên đều trúng bia"

F = "chỉ 1 viên trúng bia"

G = "không viên nào trúng bia".

Bài 8. Một phòng điều trị có 3 bệnh nhân nặng cùng nhập viện một lúc. Gọi A_1, A_2, A_3 là các sự kiện các bệnh nhân tương ứng cần cấp cứu trong vòng 1giờ. Hãy biểu diễn các sự kiện sau:

a) Có 2 bệnh nhân cần cấp cứu trong vòng 1 giờ.

b) Có ít nhất một bệnh nhân cần cấp cứu trong vòng 1 giờ.

Bài 9. Có hai hộp đựng bi xanh, bi đỏ và bi trắng: Hộp 1 chứa 3 bi trắng, 2 bi đỏ và 4 bi xanh; Hộp 2 chứa 4 bi trắng, 4 bi đỏ và 3 bi xanh. Lấy ngẫu nhiên ở mỗi hộp ra 1 bi. Gọi

T_1, T_2; D_1, D_2 và X_1, X_2 lần lượt là các biến cố lấy được bi trắng, bi đỏ, bi xanh từ hộp 1 và hộp 2.

a) Hãy biểu diễn các biến cố sau theo các biến cố sơ cấp:

A = "Lấy được 2 bi cùng màu"

B = "2 bi lấy ra không có bi xanh"

C = "2 bi lấy ra ít nhất có 1 bi xanh".

b) Có bao nhiêu biến cố sơ cấp thuận lợi cho biến cố D = "Lấy được 2 bi khác màu"?

Bài 10. Có thể xem xác suất sinh con trai là bao nhiêu nếu theo dõi 88200 trẻ sơ sinh trong một Quốc gia thấy có 45600 bé trai.

Bài 11. (Bài toán gặp nhau) Hai người X, Y thỏa thuận gặp nhau ở một địa điểm hẹn trước. Mỗi người đến đó độc lập với nhau tại một thời điểm ngẫu nhiên từ 8 giờ đến 9 giờ và quy ước người nào đến trước thì chờ trong 20 phút, nếu người kia không đến thì sẽ bỏ đi. Tìm xác suất để hai người gặp nhau.

Bài 12. (Bài toán cây kim Buffon) Trên mặt phẳng có kẻ vô số các đường thẳng song song và cách nhau một khoảng $2a$. Tung ngẫu nhiên một cây kim có chiều dài $2l$ ($l < a$). Tính xác suất để cây kim cắt một trong các đường thẳng bất kỳ.

Bài 13. Phát hành 100 vé số, trong đó có 5 vé trúng thưởng. Một người mua 3 vé. Tính xác suất để có đúng 1 vé trúng thưởng?

Bài 14. Một hộp đựng 4 viên bi trắng và 6 viên bi đỏ, lấy ngẫu nhiên không hoàn lại lần lượt 2 viên. Tính xác suất để lấy được:

a) 2 viên trắng

b) ít nhất 1 viên đỏ

c) viên thứ 2 đỏ.

Bài 15. Một lô thỏ có 48 thỏ có gen dị hợp tử Xt, 16 thỏ có gen dị hợp tử XX, trong đó X là gen màu xám (gen trội) và t là gen màu trắng (gen lặn). Bắt ngẫu nhiên từng con một ra 2 thỏ.

a) Tìm xác suất để 2 thỏ cùng gen.

b) Giả sử 2 thỏ bắt được có một thỏ đực và một thỏ cái. Cặp thỏ này sinh được 4 thỏ xám. Tìm xác suất để cặp thỏ bố mẹ cùng gen Xt, cùng gen XX.

Bài 16. Có 3 người cùng chơi bóng rổ, mỗi người ném một quả. Xác suất ném trúng rổ của mỗi người lần lượt là 0,5; 0,6; 0,7. Tính xác suất để:

a) Cả 3 người đều ném trúng rổ

b) Có ít nhất 1 người ném trúng rổ

c) Có ít nhất 1 người ném không trúng rổ

d) Có đúng 2 người ném trúng rổ.

Bài 17. Hãy chứng minh các khẳng định trong mở rộng của định lý cộng xác suất. Hãy viết công thức tính xác suất của tổng 4 biến cố A, B, C, D.

Bài 18.

a) Hệ n biến cố độc lập từng đôi thì có độc lập trong toàn thể không?

b) Chứng minh nếu A và B độc lập thì A và $\overline{B}$; $\overline{A}$ và B; $\overline{A}$ và $\overline{B}$ cũng độc lập.

c) Chứng minh nếu A và B_1 độc lập, A và B_2 độc lập, B_1 và B_2 xung khắc, thì A và (B_1+B_2) độc lập.

Bài 19. Nếu hệ $\{A_1, A_2, \ldots, A_n\}$ là độc lập từng đôi thì có thể khẳng định được rằng $P(A_1A_2...A_n) = P(A_1)...P(A_n)$ hay không?

Bài 20.

a) Chứng minh rằng nếu A và B xung khắc và có xác suất dương thì A và B không thể độc lập với nhau. Cho ví dụ minh họa.

b) Có phải "nếu A và B không xung khắc thì chúng luôn độc lập với nhau" hay không?

Bài 21. Điều tra sở thích xem ti vi của các cặp vợ chồng trẻ trong một vùng, ta có các kết quả sau: 50% các cô vợ thích xem phim Hàn Quốc, tỉ lệ này ở các ông chồng là 30%. Song nếu thấy vợ thích xem thì xác suất để các ông chồng thích xem cùng là 40%. Gặp ngẫu nhiên một cặp vợ chồng. Tính xác suất để:

a) Hai người cùng thích xem

b) Ít nhất một người thích xem

c) Cả hai người đều không thích xem

d) Nếu thấy chồng thích xem thì vợ cũng xem

e) Vợ không xem nhưng chồng vẫn xem.

Bài 22. Một công nhân đứng 3 máy, biết các máy hoạt động độc lập với nhau. Xác suất để trong thời gian $t = 5$ năm máy 1, máy 2, máy 3 không bị hỏng tương ứng là 0,7; 0,8 và 0,9. Tìm xác suất để ít nhất 1 trong 3 máy không bị hỏng trong khoảng thời gian đó.

Bài 23. Một người có 3 con gà mái, 2 con gà trống nhốt chung trong cùng một lồng. Một người đến mua, người bán gà bắt ngẫu nhiên ra một con. Người mua chấp nhận mua con đó.

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

b) Người thứ hai đến mua, người bán gà lại bắt ngẫu nhiên ra một con. Tìm xác suất để người thứ hai mua được gà trống.

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

Bài 24. Để dập tắt nạn sâu bệnh hại lúa, đội bảo vệ thực vật của Hợp tác xã đã tiến hành phun thuốc 3 lần liên tiếp trong một tuần. Xác suất sâu bị chết sau lần phun thứ 1 là 0,5. Nếu sâu sống sót thì khả năng bị chết sau lần phun thứ 2 là 0,7. Tương tự, sau lần phun thứ 3 là 0,9. Tìm xác suất sâu bị chết sau đợt phun thuốc.

Bài 25. Có 12 hộp thuốc trong đó có 3 hộp đã quá hạn sử dụng, được chia làm 3 gói mỗi gói 4 hộp. Tính xác suất để trong mỗi gói đều có hộp thuốc đã quá hạn.

Bài 26. Trong điều trị bệnh lao có hiện tượng kháng thuốc. Gọi A là hiện tượng "kháng INH của vi khuẩn lao", B là hiện tượng "kháng PAS của vi khuẩn lao" và C là hiện tượng "kháng Streptomycin của vi khuẩn lao".

Qua theo dõi ta biết khả năng kháng INH của vi khuẩn lao là 20%, nghĩa là $P(A) = 0,2$. Tương tự, $P(B) = 0,4$ và $P(C) = 0,3$. Việc kháng các loại thuốc khác nhau là độc lập với nhau.

Nếu phối hợp cả 3 loại thuốc trên thì khả năng khỏi bệnh là bao nhiêu.

Bài 27. Một hộp kín đựng 3 quả cầu đỏ và 4 quả cầu xanh. Tính xác suất để khi chia hộp cầu đó một cách ngẫu nhiên thành 3 phần bằng nhau thì:

a) Cả 3 quả cầu đỏ cùng ở trong một phần.

b) Mỗi một phần có một quả cầu đỏ.

Bài 28. Một dự án được triển khai ở một vùng nông thôn. Trong đó, có 40% số hộ vay vốn dự án. Sau một chu kỳ sản xuất, kết quả điều tra cho thấy 70% số hộ vay vốn được nâng cao thu nhập. Tỉ lệ này đối với hộ không vay vốn là 30%. Tính tỉ lệ hộ có nâng cao thu nhập của toàn vùng?

Bài 29. Bắn 3 viên đạn vào cùng một mục tiêu một cách độc lập. Xác suất trúng mục tiêu của mỗi viên là 0,2; 0,3 và 0,5. Nếu có 1 viên trúng thì mục tiêu bị phá hủy với xác suất

0,4. Nếu có từ 2 viên trở lên trúng thì mục tiêu chắc chắn bị phá hủy. Tìm xác suất để mục tiêu bị phá hủy khi bắn 3 viên đạn trên?

Bài 30. Một thiết bị gồm ba linh kiện loại 1, 2, 3; chúng chiếm tương ứng 35%, 25%, 40% tổng số linh kiện của thiết bị. Tỉ lệ bị hỏng sau một khoảng thời gian hoạt động của các loại linh kiện tương ứng là 15%, 25% và 5%. Thiết bị đang hoạt động bỗng nhiên có một linh kiện bị hỏng. Tính xem linh kiện loại nào có nhiều khả năng bị hỏng nhất.

Bài 31. Một lô hạt giống được phân thành ba loại: loại 1 chiếm 2/3 số hạt của cả lô; loại 2 chiếm 1/4; còn lại là loại 3. Tỉ lệ nảy mầm tương ứng của mỗi loại là 80%, 60% và 40%. Hãy xác định tỉ lệ nảy mầm chung của lô hạt giống?

Bài 32. Trong một trạm cấp cứu bỏng có 80% bệnh nhân bỏng do nóng, 20% bệnh nhân bỏng do hóa chất. Loại bỏng do nóng có 30% bị biến chứng, loại bỏng do hóa chất có 50% bị biến chứng.

a) Từ tập hồ sơ bệnh nhân, người ta chọn ngẫu nhiên ra một bệnh án. Tìm xác suất để gặp một bệnh án của bệnh nhân bị biến chứng.

b) Rút ngẫu nhiên được một bệnh án của bệnh nhân bị biến chứng. Tìm xác suất để bệnh án đó là của bệnh nhân bị biến chứng do bỏng gây ra.

Bài 33. Tỉ lệ người dân nghiện thuốc lá là 30%, biết rằng tỉ lệ người bị viêm họng trong số người nghiện thuốc lá là 60%; còn tỉ lệ người bị viêm họng trong số người không hút thuốc lá là 40%.

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

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

Bài 34. Có hai chuồng thỏ thí nghiệm: chuồng 1 có 12 thỏ trắng và 3 thỏ nâu; chuồng 2 có 16 thỏ trắng và 4 thỏ nâu. Tình cờ một con thỏ từ chuồng 2 nhảy sang chuồng 1. Từ chuồng 1, người ta bắt ngẫu nhiên một con. Tính xác suất để thỏ bắt được là thỏ trắng.

Bài 35. Trong một vùng dân cư có tỉ lệ nam: nữ là 9: 11, một nạn dịch truyền nhiễm xuất hiện trong vùng với khả năng mắc bệnh ở nam giới là 6% và ở nữ giới là 2%.

a) Khả năng gặp một người trong vùng bị mắc bệnh là bao nhiêu.

b) Chọn ngẫu nhiên một người trong vùng thì gặp phải người mắc bệnh. Xét xem khả năng người được chọn đó là nam giới cao hơn hay nữ giới cao hơn.

Bài 36. Ta biết rằng một cặp sinh đôi có thể là sinh đôi thật (do một trứng sinh ra), trong trường hợp đó chúng cùng giới hoặc có thể là giả sinh đôi (do hai trứng sinh ra), trong trường hợp này xác suất để chúng cùng giới là 0,5. Ta giả thiết rằng đã biết xác suất của một cặp sịnh đôi là sinh đôi thật trong một họ nào đó là p.

a) Tìm xác suất để một cặp sinh đôi là sinh đôi thật biết rằng chúng cùng giới.

b) Tìm xác suất để một cặp sinh đôi là sinh đôi giả biết rằng chúng khác giới.

Bài 37. Tỉ lệ cha mắt đen và con mắt đen là 0,05; cha mắt đen và con mắt xanh là 0,079; cha mắt xanh và con mắt đen là 0,089; cha mắt xanh và con mắt xanh là 0,782.

a) Tìm khả năng con mắt xanh biết rằng cha mắt xanh.

b) Tìm khả năng con mắt không đen biết rằng cha mắt đen.

Bài 38. Một người ốm vào bệnh viện, bác sỹ chẩn đoán sơ bộ người này có thể bị mắc bệnh A với xác suất 70%, mắc bệnh B với xác suất 30%. Để có thêm thông tin chẩn đoán, bác sỹ đã cho tiến hành xét nghiệm sinh hóa. Sau 3 lần xét nghiệm thấy có một lần dương tính, biết rằng khả năng dương tính của mỗi lần xét nghiệm đối với bệnh A là 10%, đối với bệnh B là 30%. Hãy cho biết nên chẩn đoán bệnh nhân mắc bệnh nào.

Bài 39. Có hai chuồng thỏ: chuồng 1 có 3 thỏ trắng và 3 thỏ nâu; chuồng thứ 2 có 6 thỏ trắng và 4 thỏ nâu. Bắt ngẫu nhiên 4 con thỏ từ chuồng thứ nhất nhốt sang chuồng thứ hai. Sau đó, bắt ngẫu nhiên ở chuồng thứ hai ra 1 con thỏ. Tính xác suất để bắt được thỏ nâu từ chuồng thứ hai.

Bài 40. Theo dõi kết quả điều tra về bệnh lao, tỉ lệ người bị lao ở một vùng nọ là 0,001. Tìm xác suất để khi khám cho 10 người:

a) Không ai bị lao

b) Có 5 người bị lao

c) Ít nhất 1 người bị lao

d) Tìm số người bị lao có khả năng nhất.

Bài 41. Xác suất để mỗi con lợn khi tiêm phòng bằng một loại vaccine được miễn dịch là 0,9. Người ta tiêm phòng cho 40 con. Tìm số lợn được miễn dịch có khả năng nhất?

Bài 42. Một lô hạt giống với tỉ lệ hạt lép là 5%. Ta phải lấy một mẫu cỡ bao nhiêu sao cho xác suất để ít nhất có 1 hạt lép không nhỏ hơn 95%.

Bài 43. Một bác sỹ có tiếng về chữa một loại bệnh với xác suất chữa khỏi bệnh là 0,8. Có người nói rằng cứ 5 người đến chữa thì chắc chắn có 4 người khỏi bệnh. Điều đó có đúng không? Tại sao?

Bài 44. Một người bắn liên tiếp 5 viên đạn vào một mục tiêu với xác suất trúng đích của mỗi viên là 0,2. Để phá hủy mục tiêu cần từ 3 viên trúng mục tiêu trở lên. Tính xác suất để mục tiêu bị tiêu diệt.

Bài 45. Một lô thuốc đóng chai có tỉ lệ phế phẩm là 10%. Lấy ngẫu nhiên 10 chai. Tìm xác suất để trong số chai thuốc lấy ra:

a) Có đúng 4 phế phẩm

b) Có ít nhất 8 phế phẩm.

Bài 46. Trong một kho rượu số lượng chai rượu loại A và loại B là bằng nhau. Người ta lấy ngẫu nhiên một chai rượu trong kho và đưa cho 4 người sành rượu nếm thử để xác định xem đây là rượu loại nào. Giả sử mỗi người có khả năng đoán đúng là 80%. Có 3 người kết luận chai rượu thuộc loại A và một người kết luận chai rượu thuộc loại B. Vậy chai rượu được chọn thuộc loại A với xác suất bằng bao nhiêu.

CHƯƠNG II:

BIẾN NGẪU NHIÊN VÀ QUY LUẬT PHÂN PHỐI XÁC SUẤT

2.1. Biến ngẫu nhiên

2.1.1. Khái niệm

a) Khái niệm: Biến ngẫu nhiên là một đại lượng nhận các giá trị của nó với xác suất tương ứng nào đó.

- Biến ngẫu nhiên thường biểu thị giá trị kết quả của một phép thử ngẫu nhiên.

- Biến ngẫu nhiên thường được ký hiệu $X, Y, Z, ...$ hay $\xi, \eta, \delta ...$

Ví dụ. - Tung một con xúc xắc cân đối, đồng chất. Gọi X là số chấm xuất hiện trên mặt con xúc xắc thì X là một biến ngẫu nhiên.

- Đo ngẫu nhiên chiều cao của 1 sinh viên. Gọi ξ là chiều cao của sinh viên đó thì ξ là một biến ngẫu nhiên.

b) Phân loại biến ngẫu nhiên:

- Biến ngẫu nhiên rời rạc

- Biến ngẫu nhiên liên tục

2.1.2. Biến ngẫu nhiên rời rạc và bảng phân phối xác suất

a) Định nghĩa: Biến ngẫu nhiên được gọi là rời rạc nếu tập các giá trị của nó là hữu hạn hoặc đếm được.

Ví dụ. - Số chấm xuất hiện trên mặt con xúc xắc khi tung.

- Số tai nạn giao thông trong một ngày ở một vùng.

b) Bảng phân phối xác suất

- Giả sử biến ngẫu nhiên X nhận các giá trị $x_1, x_2, ..., x_n, ...$ với xác suất tương ứng $p_i = P(X = x_i),\ i = 1, 2...$ thì nó có thể được mô tả bằng bảng phân phối xác suất như sau:

Bảng 2.1: Bảng phân phối xác suất của biến ngẫu nhiên rời rạc

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

trong đó $\sum_i p_i = 1$.

Ví dụ. Một chuồng có 10 con thỏ: 7 thỏ trắng, 3 thỏ nâu. Bắt ngẫu nhiên từ chuồng ra 3 con thỏ. Gọi X là số thỏ trắng bắt ra được. Lập bảng phân phối xác suất của X và tính các xác suất $P(0 < X \le 2); P(|X| \le 1)$.

2.1.3. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

a) Định nghĩa: Biến ngẫu nhiên được gọi là liên tục nếu tập các giá trị của nó là một khoảng (hay một đoạn) trên trục số thực.

Ví dụ. - Nhiệt độ không khí ở mỗi thời điểm nào đó.

- Chiều cao của con người.
- Thời gian sống của một loại cây trồng.

b) <u>Hàm mật độ xác suất</u>

- Để mô tả biến ngẫu nhiên liên tục ta dùng khái niệm hàm mật độ.
- Hàm $f(x)$ được gọi là hàm mật độ của biến ngẫu nhiên X nào đó nếu thỏa mãn 2 điều kiện:

  i. $f(x) \geq 0, \quad \forall x \in R$.

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

- Khi đó, xác suất để X thuộc vào khoảng $[x_1, x_2)$ được xác định như sau:

$$P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} f(x)dx$$

c) <u>Ý nghĩa của hàm mật độ</u>

- Từ định nghĩa của hàm mật độ, ta có $P(x \leq X < x + \Delta x) \approx f(x)\Delta x$.
- Xác suất để X nhận giá trị thuộc lân cận khá bé $(x, x + \Delta x)$ gần như tỷ lệ với $f(x)$.

Ví dụ. - Hàm $f(x)$ xác định bởi

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

gọi là hàm mật độ của phân phối đều trên $[a,b]$.

- Tuổi thọ (tính bằng tháng) của một loại côn trùng là một biến ngẫu nhiên liên tục có hàm mật độ

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

Xác định k và tính xác suất để loại côn trùng trên chết trước khi được 1 tháng tuổi.

2.1.4. Tính độc lập giữa hai biến ngẫu nhiên

Hai biến ngẫu nhiên X, Y được gọi là độc lập nếu mọi biến cố liên quan đến X đều độc lập với biến cố bất kỳ liên quan đến Y, tức là

$$P\{X \in (x_1, x_2); Y \in (y_1, y_2)\} = P\{X \in (x_1, x_2)\}.P\{Y \in (y_1, y_2)\}$$

Ví dụ. Có 2 chuồng gà: chuồng I có 5 gà trống, 3 gà mái; chuồng II có 4 gà trống, 6 gà mái. Gọi X là số gà trống trong 1 con gà bắt ra từ chuồng I, Y là số gà trống trong 2 con gà bắt ra từ chuồng II. Lập bảng phân phối xác suất của $X + Y$.

2.2. Hàm phân phối xác suất

2.2.1. Định nghĩa

a) <u>Định nghĩa</u>: Hàm phân phối của biến ngẫu nhiên X tại $x \in R$ xác định như sau:

$$F_X(x) = P(X < x)$$

b) <u>Ý nghĩa</u>:

- $F_X(x)$ là xác suất để biến ngẫu nhiên nhận giá trị bên trái x
- Hàm phân phối phản ánh mức độ tập trung xác suất về bên trái điểm x.

Khi đó:

- Nếu X là biến ngẫu nhiên rời rạc nhận các giá trị $x_1, x_2, ..., x_n, ...$ với xác suất tương ứng $p_i = P(X = x_i),\ i = 1, 2, ...$ thì

$$F_X(x) = \sum_{i:x_i < x} P(X = x_i) = \sum_{i:x_i < x} p_i$$

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

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

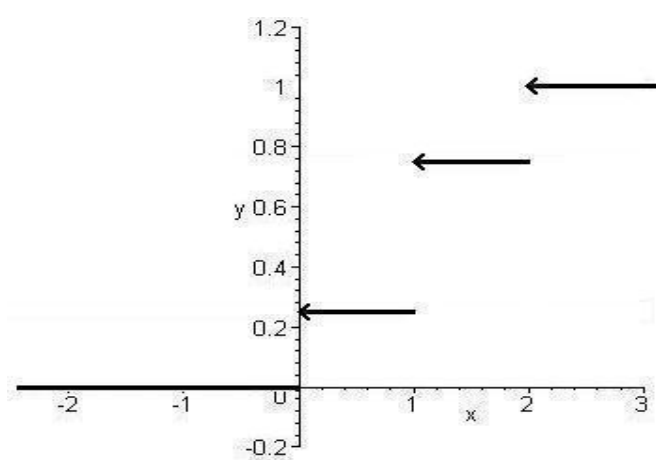
- Trong trường hợp không cần đề cập đến biến ngẫu nhiên, ta có thể ký hiệu $F(x)$ là hàm phân phối của biến ngẫu nhiên X nào đó.

Ví dụ. Gieo đồng thời 2 đồng xu cân đối đồng chất. Gọi X là số lần xuất hiện mặt sấp.

- Hàm phân phối $F(x)$ là:

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

- Đồ thị $F(x)$:



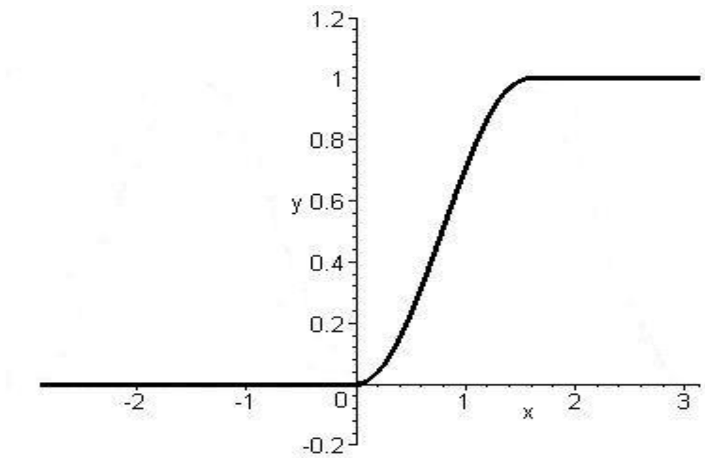
Ví dụ. Cho biến ngẫu nhiên X có hàm mật độ:

$$f(x)=\begin{cases}0 & \text{nÕu } x\notin[0,\pi/2]\\ \sin 2x & \text{nÕu } x\in[0,\pi/2]\end{cases}$$

- Hàm phân phối $F(x)$ là:

$$F(x)=\begin{cases}0 & \text{nÕu } x\le 0\\ \dfrac{1-\cos 2x}{2} & \text{nÕu } 0<x\le \pi/2\\ 1 & \text{nÕu } \pi/2<x\end{cases}$$

- Đồ thị $F(x)$:



2.2.2. Tính chất

a) Hàm phân phối có các tính chất sau:

i. $0\le F(x)\le 1,\ \forall x\in R$.

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

iii. Nếu $x_1<x_2$ thì $F(x_1)\le F(x_2)$ (tính không giảm).

iv. $F(x)$ liên tục trái tại mọi điểm.

v. Với $a<b$ thì $P(a\le X<b)=F(b)-F(a)$.

b) Nhận xét:

- Nếu hàm mật độ $f(x)$ liên tục tại $x=x_0$ thì $F'(x_0)=f(x_0)$.

- Nếu hàm phân phối của X liên tục tại x_0 thì $P(X=x_0)=0$.

2.3. Các tham số đặc trưng của biến ngẫu nhiên

2.3.1. Kỳ vọng toán

a) <u>Định nghĩa</u>: Kỳ vọng của biến ngẫu nhiên X, ký hiệu $E(X)$, được xác định bởi:

$$E(X)=\begin{cases}\sum_i x_i p_i & \text{nếu } X \text{ rời rạc với } P(X=x_i)=p_i, i=1,2,\dots\\ \int_{-\infty}^{\infty} xf(x)dx & \text{nếu } X \text{ liên tục với hàm mật độ } f(x)\end{cases}$$

trong trường hợp các đại lượng tổng chuỗi và tích phân trên hữu hạn.

b) <u>Ý nghĩa</u>:

- Kỳ vọng của biến ngẫu nhiên là giá trị trung bình (theo xác suất) của biến ngẫu nhiên.
- Kỳ vọng phản ánh giá trị trọng tâm của phân phối xác suất của biến ngẫu nhiên.

c) <u>Tính chất</u>:

i. $EC=C, C$ là hằng số.

ii. $E(CX)=CE(X)$.

iii. $E(X\pm Y)=E(X)\pm E(Y)$.

iv. Nếu X,Y độc lập thì $E(XY)=E(X).E(Y)$.

v. Cho hàm số g. Khi đó

vi. $$Eg(X)=\begin{cases}\sum_i g(x_i)p_i & \text{nếu } X \text{ rời rạc với } P(X=x_i)=p_i, i=1,2,\dots\\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{nếu } X \text{ liên tục với hàm mật độ } f(x)\end{cases}$$

Ví dụ. Cho biến ngẫu nhiên X có bảng phân phối xác suất sau:

X	1	2	3	4
P	1/7	2/7	3/7	1/7

Tính $E(X); E(X^2-3X+2)$.

Ví dụ. Cho hai biến ngẫu nhiên X,Y độc lập, cùng có hàm mật độ

$$f(x)=\begin{cases}2e^{-2x} & \text{nếu } x\in[0,2]\\ 0 & \text{nếu } x\notin[0,2]\end{cases}$$

Tính $E(X); E(XY); E(X^3); E(X+2Y-3)$.

2.3.2. Trung vị (Median)

a) <u>Định nghĩa</u>: Trung vị của biến ngẫu nhiên X là đại lượng, ký hiệu $med(X)$, được xác định như sau:

$$P(X<med(X))\le\frac{1}{2}\le P(X\le med(X))$$

b) <u>Ý nghĩa</u>: Trung vị là điểm phân đôi phân phối xác suất thành hai phần bằng nhau.

c) <u>Nhận xét</u>:

- Nếu hàm phân phối $F(x)$ của X liên tục thì $med(X)$ là nghiệm của phương trình

$$F(x) = 1/2\,.$$

- Một biến ngẫu nhiên có thể có một hoặc nhiều giá trị trung vị.

Ví dụ. Tìm các trung vị của các ví dụ trong phần 2.3.1.

2.3.3. Mode

a) <u>Định nghĩa</u>:

- Nếu X là biến ngẫu nhiên rời rạc thì Mode của X, ký hiệu $Mod(X)$, là giá trị của X mà tại đó xác suất tương ứng là lớn nhất.

- Nếu X biến ngẫu nhiên liên tục có hàm mật độ $f(x)$ thì $Mod(X)$ là giá trị x_0 làm cực đại hàm $f(x)$.

b) <u>Ý nghĩa</u>: $Mod(X)$ là giá trị của biến ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận của nó.

Ví dụ. Tìm các Mode của các ví dụ trong phần 2.3.1.

c) <u>Nhận xét</u>:

- Một biến ngẫu nhiên có thể có một hoặc nhiều giá trị Mode.

- Nếu phân phối của biến ngẫu nhiên là đối xứng hoặc gần đối xứng thì dùng kỳ vọng để đặc trưng là tốt nhất.

- Nếu phân phối của biến ngẫu nhiên là quá lệch thì dùng trung vị hoặc Mode để đặc trưng là tốt nhất.

- Nếu phân phối của biến ngẫu nhiên là đối xứng và có một Mode thì 3 đặc trưng: kỳ vọng, trung vị, Mode là trùng nhau.

2.3.4. Phương sai

a) <u>Định nghĩa</u>: Phương sai của biến ngẫu nhiên X, ký hiệu $Var(X)$, là đại lượng không âm được xác định bởi

$$Var(X) = E(X - EX)^2 = EX^2 - (EX)^2$$

trong đó

$$EX^2 = \begin{cases} \sum_i x_i^2 p_i & \text{nÕu } X \text{ rêi r¹c víi } P(X = x_i) = p_i,\, i = 1,2,\ldots \\ \int_{-\infty}^{\infty} x^2 f(x)dx & \text{nÕu } X \text{ liªn tôc víi hµm mËt ®é } f(x) \end{cases}$$

b) <u>Ý nghĩa</u>:

- Phương sai của biến ngẫu nhiên X dùng để đo mức độ phân tán của các giá trị của X xung quanh giá trị trung bình EX của nó.

- $Var(X)$ nhỏ thì mức độ phân tán nhỏ, độ tập trung lớn.

- $Var(X)$ càng lớn thì độ phân tán càng cao.

c) <u>Tính chất</u>:

i. $Var(C) = 0, C$ là hằng số.

ii. $Var(CX) = C^2 Var(X)$.

iii. Nếu X, Y độc lập thì $Var(X \pm Y) = Var(X) + Var(Y)$.

Ví dụ. Tính phương sai của các ví dụ trong phần 2.3.1.

2.3.5. Độ lệch chuẩn

a) Định nghĩa: Độ lệch chuẩn của biến ngẫu nhiên, ký hiệu $\sigma(X)$, được xác định bởi:

$$\sigma(X) = \sqrt{Var(X)}$$

b) Ý nghĩa: Độ lệch chuẩn dùng để đánh giá mức độ phân tán các giá trị của biến ngẫu nhiên theo đơn vị của biến ngẫu ngẫu nhiên.

2.3.6. Moment

a) Định nghĩa:

- Moment cấp k của biến ngẫu nhiên X là số $\mu_k = E(X^k)$.

- Moment trung tâm cấp k của biến ngẫu nhiên X là số $m_k = E[X - E(X)]^k$.

b) Nhận xét:

- Moment cấp 1 của X là kỳ vọng của X.

- Moment trung tâm cấp 2 của X là phương sai của X.

2.4. Một số quy luật phân phối xác suất thông dụng

2.4.1. Phân phối 0-1 (Bernoulli)

a) Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối 0-1 với tham số $p, p \in (0,1)$ nếu

$$P(X = 0) = q;\ P(X = 1) = p, \quad (p + q = 1)$$

Ví dụ. Một loại hạt giống nảy mầm với xác suất p. Số hạt nảy mầm khi gieo 1 hạt có phân phối 0-1.

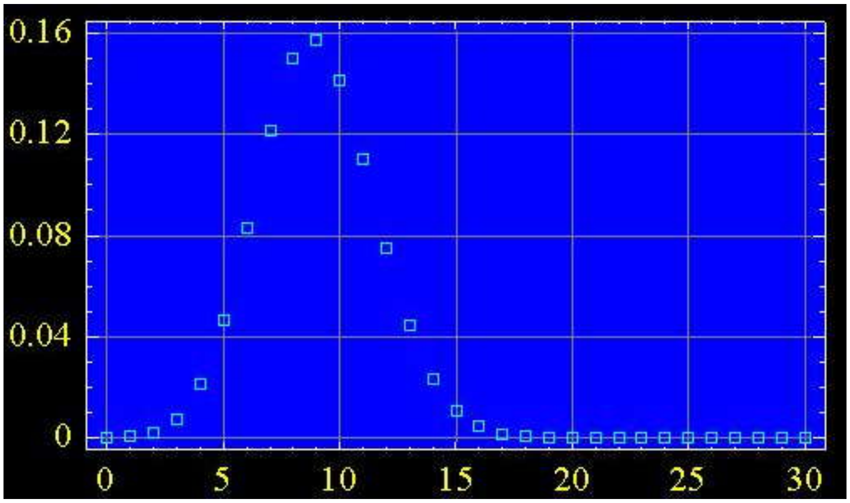
b) Các đặc trưng: $EX = p; Var(X) = pq$.

2.4.2. Phân phối nhị thức

a) Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối nhị thức với tham số $n, p, p \in (0,1)$, ký hiệu $X \sim B(n, p)$, nếu

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \qquad k = 0..n$$

Đồ thị:



Hình 2.1: Đồ thị phân phối nhị thức

Ví dụ. Một xạ thủ bắn trúng mục tiêu với xác suất p. Cho xạ thủ đó bắn 10 viên đạn. Số viên đạn anh ta bắn trúng có phân phối nhị thức $B(10, p)$.

b) <u>Nhận xét</u>:

- Phân phối 0-1 là phân phối nhị thức với $n = 1$.

- Xét dãy n phép thử độc lập Bernoulli với xác suất thành công $p = P(A)$ (A là biến cố quan sát của phép thử). Gọi X là số lần xuất hiện biến cố A trong n phép thử trên thì $X \sim B(n, p)$.

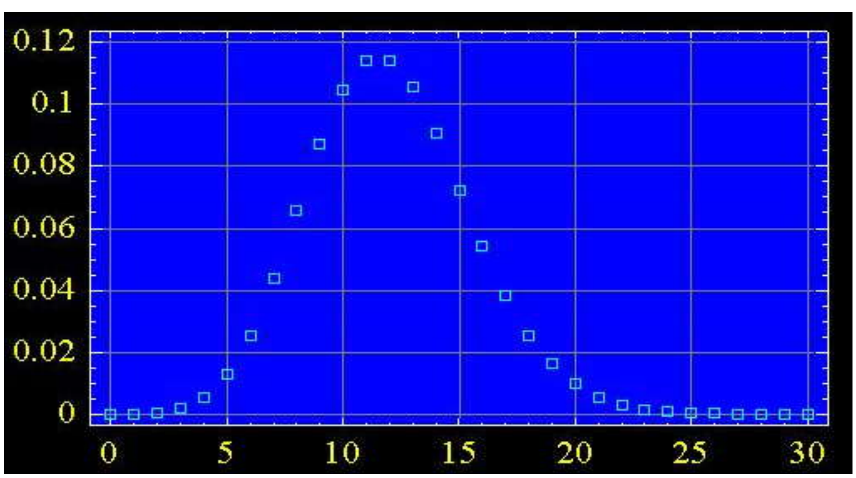
c) <u>Các đặc trưng</u>: $EX = np; Var(X) = np(1-p)$.

2.4.3. Phân phối Poisson

a) <u>Định nghĩa</u>:Biến ngẫu nhiên X được gọi là có phân phối Poisson với tham số $\lambda, \lambda > 0$, ký hiệu $X \sim P(\lambda)$, nếu

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad k = 0, 1, 2...$$

<u>Đồ thị</u>:



Hình 2.2: Đồ thị phân phối Poisson

Ví dụ. - Số cuộc điện thoại nhận được ở một trạm điện thoại trong một phút.

- Số tai nạn giao thông xảy ra trong một ngày.

- Số máy bị hỏng trong một năm, số lỗi in trong một tập sách ...

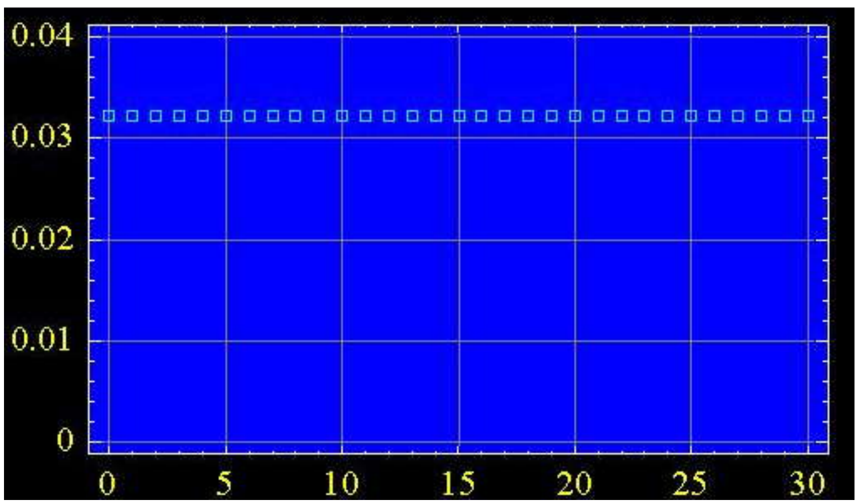
b) <u>Các đặc trưng</u>: $EX = Var(X) = \lambda$.

2.4.4. Phân phối đều

a) <u>Định nghĩa</u>:Biến ngẫu nhiên X được gọi là có phân phối đều (rời rạc) trên tập $\{1,2,...,n\}$, ký hiệu $U(n)$, nếu

$$P(X=k)=\frac{1}{n}, \quad k=1,2,...n$$

<u>Đồ thị</u>:



Hình 2.3: Đồ thị phân phối đều.

Ví dụ. Số chấm xuất hiện khi tung một con xúc xắc.

b) <u>Các đặc trưng</u>: $EX=\dfrac{n+1}{2}; Var(X)=\dfrac{n^2-1}{12}$.

c) <u>Định nghĩa</u>: Biến ngẫu nhiên X gọi là có phân phối đều (liên tục) trên $[a,b]$, ký hiệu $U[a,b]$, nếu có hàm mật độ dạng

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

Ví dụ. Thời điểm đến trạm của xe buýt trong thời gian từ 7h đến 7h30.

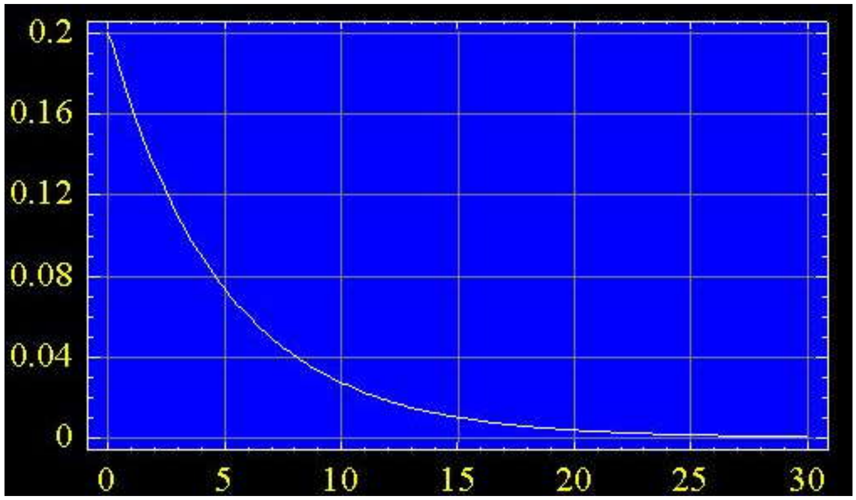
d) <u>Các đặc trưng</u>: $EX=\dfrac{(a+b)}{2}; Var(X)=\dfrac{(a-b)^2}{12}$.

2.4.5. Phân phối mũ

a) <u>Định nghĩa</u>: Biến ngẫu nhiên X được gọi là có phân phối mũ với tham số $\lambda, \lambda>0$, ký hiệu $X \sim Exp(\lambda)$, nếu có hàm mật độ dạng

$$f(x)=\begin{cases}0 & \text{nếu } x\le 0,\\ \lambda e^{-\lambda x} & \text{nếu } x>0\end{cases}$$

Đồ thị:



Hình 2.4: Đồ thị phân phối mũ.

Ví dụ. - Thời gian một khách hàng cần phục vụ ở hiệu uốn tóc;

- Thời gian sống của bóng đèn điện

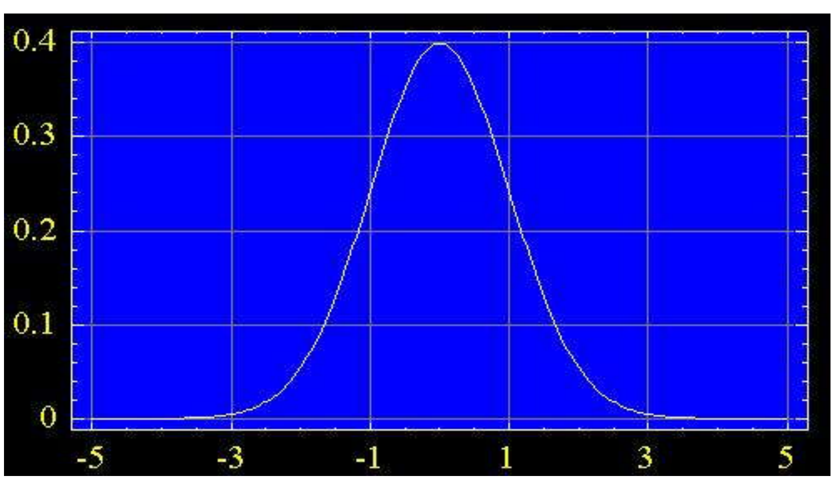
b) Các đặc trưng: $EX=\dfrac{1}{\lambda};Var(X)=\dfrac{1}{\lambda^2}$.

2.4.6. Phân phối chuẩn

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

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

Đồ thị:



Hình 2.5: Đồ thị phân phối chuẩn.

Ví dụ. - Trọng lượng và chiều cao của người lớn, chiều cao của cây ...

- Điểm thi của các thí sinh.

- Hàm lượng và khoáng vật trong đất, đá ...

b) Chú ý: - Trường hợp $\mu = 0, \sigma = 1$ thì $N(0,1)$ gọi là phân phối chuẩn tắc. Đây là phân phối quan trọng trong lý thuyết xác suất.

- Hàm phân phối chuẩn $N(0,1)$ được ký hiệu là

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

còn gọi là hàm Laplace.

- Nếu $X \sim N(\mu, \sigma^2)$, bằng phép đổi biến

$$Y = \frac{X - \mu}{\sigma}$$

thì $Y \sim N(0,1)$. Khi đó

$$P(a \le X < b) = P(\tfrac{a-\mu}{\sigma} \le \tfrac{X-\mu}{\sigma} < \tfrac{b-\mu}{\sigma}) = \Phi(\tfrac{b-\mu}{\sigma}) - \Phi(\tfrac{a-\mu}{\sigma})$$

c) Các đặc trưng: $EX = \mu; Var(X) = \sigma^2$.

2.5. Vectơ ngẫu nhiên 2 chiều

2.5.1. Khái niệm

Giả sử $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên. Khi đó:

- Bộ $(X_1, X_2, ..., X_n)$ được gọi là vectơ ngẫu nhiên n chiều

- $X_1, X_2, ..., X_n$ được gọi là các thành phần của vectơ ngẫu nhiên.

Trường hợp $n = 2$, ta được vectơ ngẫu nhiên 2 chiều.

Ví dụ. Gọi X, Y tương ứng là chiều cao, cân nặng của con người thì (X, Y) là một vectơ ngẫu nhiên 2 chiều.

2.5.2. Vectơ ngẫu nhiên 2 chiều rời rạc và bảng phân phối xác xuất đồng thời

a) Định nghĩa: Vectơ ngẫu nhiên 2 chiều được gọi là rời rạc nếu các thành phần của nó là các biến ngẫu nhiên rời rạc.

b) Bảng phân phối xác suất đồng thời:

- Giả sử biến ngẫu nhiên X nhận các giá trị $x_1, x_2, ...x_n$ với xác suất tương ứng $p_i = P(X = x_i), i = 1..n$

− Biến ngẫu nhiên Y nhận các giá trị $y_1, y_2, ..., y_m$ với xác suất tương ứng $q_j = P(Y = y_j), j = 1..m$.

- Đặt $p_{ij} = P(X = x_i, Y = y_j)$: gọi là xác suất đồng thời để X nhận giá trị x_i và Y nhận giá trị y_j với $i = 1..n, j = 1..m$.

Bảng phân phối xác suất đồng thời của vectơ ngẫu nhiên (X,Y)

Bảng 2.2: Bảng phân phối xác suất đồng thời của vectơ ngẫu nhiên 2 chiều.

X \ Y	y_1	y_2	...	y_j	...	y_m
x_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2m}
$\vdots$	...	...	...	...	...	...
x_i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{im}
$\vdots$	...	...	...	...	...	...
x_n	p_{n1}	p_{n2}	...	p_{nj}	...	p_{nm}

trong đó

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1;\ \sum_{j=1}^{m} p_{ij} = p_i;\ \sum_{i=1}^{n} p_{ij} = q_j$$

Ví dụ. Tung 1 đồng xu và 1 con xúc xắc cân đối, đồng chất. Gọi X là số lần xuất hiện mặt sấp của đồng xu, Y là số chấm xuất hiện trên con xúc xắc. Ta có bảng phân phối xác suất của vectơ ngẫu nhiên (X,Y) như sau:

X \ Y	1	2	3	4	5	6	p_i
0 (N)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
1 (S)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
q_j	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

c) <u>Chú ý</u>: Nếu X,Y độc lập thì $P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j)$

2.5.3. Hàm phân phối xác suất đồng thời

a) <u>Định nghĩa</u>: Hàm phân phối xác suất của vectơ ngẫu nhiên (X,Y), ký hiệu $F_{X,Y}(x,y)$ được xác định như sau:

$$F_{X,Y}(x,y) = P(X < x, Y < y)$$

b) <u>Tính chất</u>

i. $0 \le F_{X,Y}(x,y) \le 1,\ (x,y) \in R^2$.

ii. $\lim\limits_{x\to+\infty} F_{X,Y}(x,y) = F_Y(y);\ \lim\limits_{y\to+\infty} F_{X,Y}(x,y) = F_X(x)$.

$\lim\limits_{x\to-\infty} F_{X,Y}(x,y) = \lim\limits_{y\to-\infty} F_{X,Y}(x,y) = 0;\ \lim\limits_{\substack{x\to+\infty \\ y\to+\infty}} F_{X,Y}(x,y) = 1$.

iii. Nếu X,Y độc lập thì $F_{X,Y}(x,y)=F_X(x).F_Y(y)$.

- Trường hợp vectơ ngẫu nhiên (X,Y) là rời rạc thì

$$F_{X,Y}(x,y)=\sum_{i:x_i<x}\sum_{j:y_j<y}P(X=x_i,Y=y_j)$$

Ví dụ. Tìm hàm phân phối xác suất đồng thời ở ví dụ phần 2.5.2.

2.5.4. Hàm mật độ xác suất đồng thời

a) <u>Định nghĩa</u>: Hàm mật độ xác suất đồng thời của vectơ ngẫu nhiên (X,Y), ký hiệu $f_{X,Y}(x,y)$, nếu thỏa 2 điều kiện:

i. $f_{X,Y}(x,y)\geq 0,\ (x,y)\in R^2$,

ii. $\iint_{R^2} f_{X,Y}(x,y)dxdy=1$.

Khi đó:

- Hàm phân phối đồng thời $F_{X,Y}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)dxdy$.

- Hàm mật độ xác suất của các biến ngẫu nhiên thành phần:

$$f_X(x)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy;\quad f_Y(y)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

Ví dụ. Hàm mật độ chuẩn 2 chiều

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

trong đó $\mu_1=EX,\sigma_1^2=Var(X),\mu_2=EY,\sigma_2^2=Var(Y),\rho$ là hệ số tương quan của X,Y.

2.5.5. Các đặc trưng của vectơ ngẫu nhiên 2 chiều

2.5.5.1. Vectơ kỳ vọng

- Giả sử vectơ ngẫu nhiên $Z=(X,Y)$ với các biến ngẫu nhiên thành phần X,Y tồn tại kỳ vọng.

- Vectơ kỳ vọng của Z là $EZ=(EX,EY)$.

2.5.5.2. Hiệp phương sai (covariance)

a) <u>Định nghĩa</u>: Hiệp phương sai (hay covariance) của 2 biến ngẫu nhiên X,Y được xác định:

$$cov(X,Y)=E[(X-EX)(Y-EY)]$$

nếu kỳ vọng ở vế phải tồn tại.

b) <u>Nhận xét</u>: Nếu tồn tại các kỳ vọng $EX,EY,E(XY)$ thì

$$cov(X,Y)=E(XY)-EX.EY$$

Ví dụ. Tìm hiệp phương sai của 2 biến ngẫu nhiên ở ví dụ phần 2.5.2.

2.5.5.3. Hệ số tương quan

a) Định nghĩa: Hệ số tương quan của 2 biến ngẫu nhiên X,Y được xác định:

$$\rho(X,Y)=\frac{cov(X,Y)}{\sigma(X).\sigma(Y)}$$

nếu các đại lượng ở vế phải tồn tại.

b) Ý nghĩa:

- Hệ số tương quan đo mức độ tương quan, phụ thuộc tuyến tính giữa 2 biến ngẫu nhiên.
- Nếu $|\rho(X,Y)|$ càng lớn thì sự phụ thuộc càng chặt chẽ.
- $\rho(X,Y)>0$ gọi là tương quan thuận, $\rho(X,Y)<0$ gọi là tương quan nghịch.

c) Tính chất:

i. $|\rho(X,Y)|\le 1$.

ii. Nếu X,Y độc lập thì $\rho(X,Y)=0$.

iii. $|\rho(X,Y)|=1 \Leftrightarrow \exists\, a,b: Y=aX+b$.

Ví dụ. Tìm hệ số tương quan của 2 biến ngẫu nhiên ở ví dụ phần 2.5.2.

2.6. Luật số lớn

2.6.1. Luật số lớn Tchebyshev

Giả sử dãy các biến ngẫu nhiên $X_1,X_2,...,X_n,...$ thỏa các điều kiện sau:

- Độc lập đôi một, tức là X_i,X_j độc lập với $i\neq j$.
- Kỳ vọng hữu hạn và phương sai giới nội đều, tức là $Var(X_n)\le C,\forall n$

Đặt $S_n=\sum_n X_n$ thì

$$\lim_{n\to\infty}P\{|\frac{S_n-E(S_n)}{n}|>\varepsilon\}=0,\ \forall\varepsilon>0$$

2.6.2. Luật số lớn Khintchin

Giả sử dãy biến ngẫu nhiên $X_1,X_2,...,X_n,...$ thỏa các điều kiện sau:

- Độc lập đôi một,
- Có cùng phân phối và kỳ vọng hữu hạn.

Đặt $S_n=\sum_n X_i,\ \mu=EX_i$ thì

$$\lim_{n\to\infty}P\{|\frac{S_n}{n}-\mu|>\varepsilon\}=0,\ \forall\varepsilon>0$$

2.6.3. Luật số lớn Bernoulli

Xét dãy phép thử độc lập Bernoulli với biến cố quan sát A. Gọi f_n là tần số xuất hiện biến cố A trong n phép thử và p là xác suất xuất hiện biến cố A trong mỗi phép thử thì

$$\lim_{n\to\infty} P\{|f_n - p| > \varepsilon\} = 0, \ \forall \varepsilon > 0$$

2.7. Các định lý giới hạn

2.7.1. Định lý giới hạn địa phương Moivre - Laplace

- Xét n phép thử độc lập Bernoulli với biến cố quan sát A có xác suất $p = P(A)$.
- $P_n(k)$ là xác suất có k lần xuất hiện biến cố A trong n phép thử. Khi đó

$$\lim_{n\to\infty} \frac{\sqrt{npq}\ P_n(k)}{f(x_k)} = 1$$

trong đó

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \ x_k = \frac{k-np}{\sqrt{npq}}, \ p+q=1$$

- Công thức xấp xỉ: Khi n đủ lớn thì

$$P_n(k) \approx \frac{1}{\sqrt{npq}} f(x_k)$$

Ví dụ. Gieo 1800 lần một đồng xu cân đối đồng chất. Tính xác suất để có 10 lần đồng xu xuất hiện mặt sấp.

2.7.2. Định lý giới hạn trung tâm

Giả sử dãy biến ngẫu nhiên $X_1, X_2, ...$ độc lập, cùng phân phối có phương sai hữu hạn khác 0. Khi đó

$$\lim_{n\to\infty} P\{\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) < x\} = \Phi(x), \ \forall x \in R$$

trong đó

$$\mu = EX_1, \sigma^2 = Var(X_1), \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$ lμ hμm Laplace

Khi đó, xét dãy các biến ngẫu nhiên (X_i) độc lập, cùng phân phối 0-1, tham số p. Theo định lý giới hạn trung tâm, với $X = \sum_i X_i$, ta được

$$\lim_{n\to\infty} P(\frac{X-np}{\sqrt{npq}} < x) = \Phi(x), \ \forall x \in R$$

Công thức xấp xỉ: Khi n đủ lớn thì

$$P(a < X < b) = P(\frac{a-np}{\sqrt{npq}} < \frac{X-np}{\sqrt{npq}} < \frac{b-np}{\sqrt{npq}}) \approx \Phi(\frac{b-np}{\sqrt{npq}}) - \Phi(\frac{a-np}{\sqrt{npq}})$$

Ví dụ. Gieo 3200 lần một đồng xu cân đối đồng chất. Gọi X là số lần xuất hiện mặt sấp. Tính $P(1600 < X < 1800)$.

2.7.3. Định lý Poisson

Giả sử dãy biến ngẫu nhiên $X_n \sim B(n, p_n)$ độc lập và $\lim\limits_{n\to\infty} np_n = \lambda$. Khi đó

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad (k = 0,1,2...)$$

Khi đó, nếu biến ngẫu nhiên $X \sim B(n, p)$ với n khá lớn, p khá bé thì theo định lý Poisson, ta có công thức xấp xỉ:

$$P(X = k) \approx \frac{e^{-\lambda}\lambda^k}{k!} \text{ ví i } \lambda = np$$

Ví dụ. Trong số 500 trang sách của một cuốn sách có 10 lỗi in. Tìm xác suất sao cho khi lấy ngẫu nhiên một trang sách thì có đúng 2 lỗi in.

Bài tập chương 2

1) Cho hàm số

$$f(x)=\begin{cases} A\cos x & \text{nÕu } |x|\le \pi/2 \\ 0 & \text{nÕu } |x|> \pi/2 \end{cases}$$

a) Tìm A để f là hàm mật độ của biến ngẫu nhiên X nào đó.

b) Xác định hàm phân phối F(x).

c) Tính các đặc trưng: E(X), Var(X).

2) Cho hàm số

$$f(x)=\begin{cases} 0 & \text{nÕu } x\le 0 \text{ hoÆc } x>2 \\ ax^2 & \text{nÕu } 0<x<1 \\ a(2-x)^2 & \text{nÕu } 1\le x<2 \end{cases}$$

a) Tìm a để f là hàm mật độ.

b) Xác định hàm phân phối F(x) tương ứng.

3) Cho hàm số

$$f(x)=\begin{cases} Ax^2e^{-2x} & \text{nÕu } x\ge 0 \\ 0 & \text{nÕu } x<0 \end{cases}$$

a) Tìm A để f là hàm mật độ của biến ngẫu nhiên X nào đó.

b) Xác định hàm phân phối F(x) .

c) Tính các đặc trưng: E(X), Var(X).

4) Một hộp có 3 bi trắng, 4 bi đen. Lấy ngẫu nhiên lần lượt từng quả cho đến khi nào lấy được bi trắng. Gọi X là số quả lấy được.

a) Tìm bảng phân phối xác suất của X, hàm phân phối F(x).

b) Tính $E(X); Var(X); E(3X+2); E(4X^2-X-3); Var(1-3X)$.

c) Tính $P(1<X\le 4); P(|X|<3); P(X^2-4X\le 3)$.

5) Một hộp chứa 10 viên bi đỏ, 6 viên bi xanh. Lấy ngẫu nhiên 3 viên. Gọi X là số bi đỏ lấy ra.

a) Tìm bảng phân phối xác suất của X và các đặc trưng E(X), Var(X).

b) Giả sử lấy mỗi viên bi đỏ được 5 điểm, mỗi viên bi xanh được 8 điểm. Gọi Y là số điểm tổng cộng cho 3 viên bi lấy ra. Lập bảng phân phối xác suất của Y và $E(X), Var(X)$

6) Hộp I có 1 bi trắng, 4 bi đỏ; hộp II có 4 bi trắng. Lấy ngẫu nhiên 2 bi từ hộp I sang hộp II. Sau đó lấy từ hộp II 3 bi bỏ vào hộp I. Gọi X_1, X_2 là số bi trắng ở hộp I, II sau 2 lần chuyển bi. Tìm phân phối xác suất của X_1, X_2.

7) Cho 2 biến ngẫu nhiên độc lập X, Y có bảng phân phối xác suất như sau:

X	1	2	3
P	0,2	0,3	0,5

Y	2	4
P	0,4	0,6

a) Tìm phân phối xác suất của $Z = X + Y$ và tính $E(Z), Var(Z)$.

b) Tìm phân phối xác suất của $W = X.Y$ và tính $E(W), Var(W)$.

8) Gieo 3 con xúc xắc cân đối, đồng chất. Gọi X là số điểm trên 3 con xúc xắc. Tính $E(X), Var(X)$.

9) Một xấp bài gồm 7 quân đỏ và 3 quân đen. Trong một ván chơi, người chơi rút ngẫu nhiên 1 quân bài. Nếu được quân đen thưởng 20 đồng, nếu được quân đỏ bị phạt 8 đồng. Gọi X là số tiền người chơi chơi trong 10 ván. Tính số tiền trung bình mà người chơi thu được.

10) Có 3 chuồng gà: chuồng I có 4 mái, 2 trống; chuồng II có 2 trống, 3 mái; chuồng III để trống.

a) Từ mỗi chuồng bắt ngẫu nhiên 1 con gà. Sau đó dồn hết số gà còn lại vào chuồng III. Từ chuồng III bắt ngẫu nhiên ra 2 con gà. Gọi X là số gà trống bị bắt ra. Tính $E(X), Var(X)$.

b) Từ chuồng I bắt ngẫu nhiên 2 con gà bỏ vào chuồng II. Sau đó từ chuồng II bắt ngẫu nhiên ra 2 con gà. Gọi Y là số gà trống còn lại trong chuồng II. Tính $E(Y), Var(Y), P(5 < Y < 8)$.

11) Có 2 chuồng thỏ: chuồng I có 5 con (2 trắng, 3 nâu); chuồng II có 6 con (4 trắng, 2 nâu). Ngẫu nhiên có 2 thỏ từ chuồng II chạy sang chuồng I. Sau đó, người ta bắt từ chuồng II ra 3 con. Gọi X là số thỏ trắng được bắt ra. Lập bảng phân phối xác suất của X. Tính các đặc trưng $E(X), Var(X)$.

12) Gieo 3 hạt giống khác nhau I, II, III với xác suất nảy mầm của mỗi loại tương ứng là 0,8; 0,9; 0,85. Giả sử rằng sự nảy mầm của mỗi hạt là độc lập lẫn nhau. Gọi X là số hạt nảy mầm sau khi gieo. Lập bảng phân phối xác suất của X và tính các đặc trưng $E(X), Var(X)$.

13) Một thiết bị máy móc gồm 3 bộ phận I, II, III hoạt động độc lập với nhau. Xác suất để các bộ phận bị hỏng trong khoảng thời gian 2 năm lần lượt là 0,3; 0,15; 0,2. Gọi X là số bộ phận bị hỏng của thiết bị trên trong khoảng thời gian 2 năm. Lập bảngphân phối xác suất của X và tính các đặc trưng $E(X), Var(X)$.

14) Có 2 chuồng gà: chuồng I có 6 con (4 gà trống, 2 gà mái); chuồng II có 5 con (1 gà trống, 4 gà mái). Từ mỗi chuồng bắt ngẫu nhiên ra 2 con gà. Gọi X là số gà trống trong 4 con gà bắt ra. Lập bảng phân phối xác suất của X và tính các đặc trưng $E(X), Var(X)$.

15) Gọi X là số khách hàng vào 1 của hàng trong khoảng thời gian 1 giờ. Giả sử X có phân phối Poisson với $\lambda = 6,5$. Tính

a) Trung bình 1 giờ có bao nhiêu khách hàng đến cửa hàng.

b) Tính xác suất để số khách hàng vào cửa hàng từ 4 đến 8 khách hàng.

c) Tính xác suất để có ít nhất 3 khách vào cửa hàng.

d) Khả năng vắng khách trong thời gian 1 giờ là bao nhiêu.

16) Gọi X là trọng lượng (kg) của con gà công nghiệp 4 tháng tuổi. Giả sử $X \sim N(3,5;0,25)$.

a) Trọng lượng trung bình của mỗi con gà là bao nhiêu.

b) Tính tỷ lệ đàn gà có trọng lượng trên 2kg.

c) Gọi Y là số con gà nặng trên 2kg trong 100 con bắt ra. Lập bảng phân phối xác suất của Y.

17) Giả sử $X \sim N(3;0,09), Y \sim Exp(2)$ và X, Y độc lập. Tìm

a) $E(2X+3Y-5), Var(X-5Y+3)$.

b) $E(X^2-4Y^2+3XY-Y+2)$.

18) Có 2 hộp: hộp I đựng 6 bi (1 bi số 1, 2 bi số 2, 3 bi số 3); hộp II đựng 8 bi (2 bi số 1, 5 bi số 2, 1 bi số 3). Rút ngẫu nhiên mỗi hộp 1 bi. Gọi X, Y là số ghi trên bi rút ra từ hộp I, II tương ứng.

a) Lập bảng phân phối xác suất của vectơ ngẫu nhiên (X,Y).

b) Tính hệ số tương quan $\rho(X,Y)$.

CHƯƠNG 3

LÝ THUYẾT MẪU

3.1. Mẫu ngẫu nhiên

3.1.1. Định nghĩa

- Để nghiên cứu đại lượng ngẫu nhiên X của một tổng thể, ta quan sát một cách độc lập (cân, đong, đo, đếm, ...) n lần về X. Gọi X_i là giá trị quan sát ở lần thứ *i* (*i*=1..n) thì (X_i) là các biến ngẫu nhiên độc lập, cùng phân phối như X. Khi đó, mẫu $(X_1, X_2, \cdots, X_n)$ được gọi là mẫu ngẫu nhiên sinh ra từ X, n gọi là cỡ mẫu hay số lần quan sát.
- Gọi x_i là 1 kết quả cụ thể của X_i. Khi đó $(x_1, x_2, \cdots, x_n)$ là một giá trị cụ thể mà mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ nhận và được gọi là mẫu thực nghiệm. Dựa trên đó, ta sẽ dùng các phương pháp và kết quả thống kê toán học để phân tích và rút ra những kết luận cần thiết.

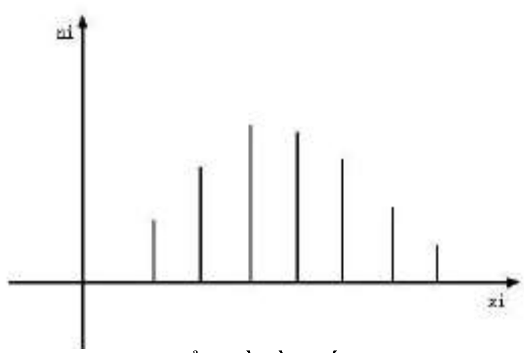
3.1.2. Bảng tần số mẫu, đa giác tần suất, tổ chức đồ

3.1.2.1. Bảng phân phối tần số thực nghiệm:

Giả sử mẫu thực nghiệm $(x_1, x_2, \cdots, x_n)$ có k giá trị khác nhau, ký hiệu $x_{(1)} < x_{(2)} < \cdots < x_{(k)}$. Gọi n_i là số các quan sát trong mẫu có giá trị x_i Khi đó ta có bảng tần số mẫu như sau

$X_{(i)}$	$x_{(1)}$	$x_{(2)}$	$\cdots$	$x_{(k)}$
n_i	n_1	n_2	$\cdots$	n_k

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



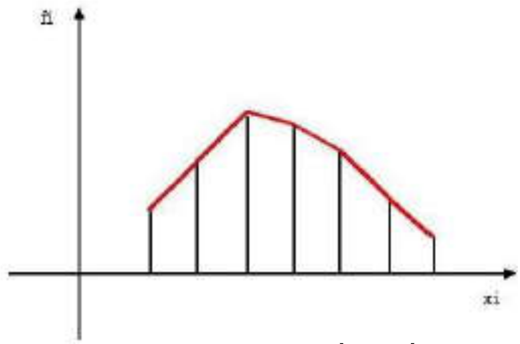
Hình 3.1. Biểu đồ tần số thực nghiệm

3.1.2.2. Bảng phân phối tần suất

$X_{(i)}$	$x_{(1)}$	$x_{(2)}$	$\cdots$	$x_{(k)}$
f_i	f_1	f_2	$\cdots$	f_k

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

<u>Đa giác tần suất</u>: Trên hệ trục tọa độ, biễu diễn các điểm có tọa độ (x_i, f_i). Đồ thị gồm các đường thẳng nối các điểm đó gọi là đa giác tần suất.



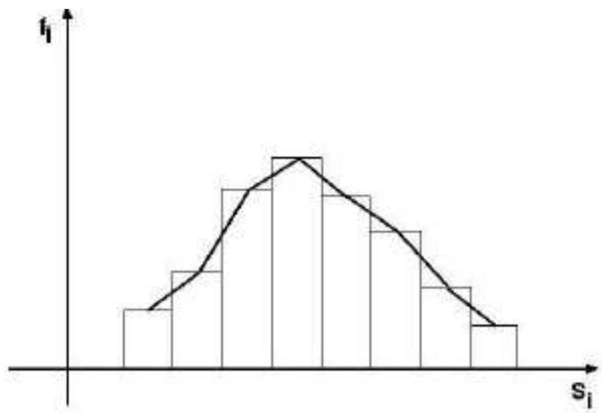
Hình 3.2. Đa giác tần suất

- Trong trường hợp số k lớn, ta có thể chia các giá trị của mẫu thành các khoảng con bằng nhau $S_i = (a_{i-1}, a_i](i = 1..k)$ với $a_i - a_{i-1} = h$ (hằng số). Gọi n_i là số các quan sát trong mẫu có giá trị rơi vào khoảng S_i. Khi đó ta có bảng tần số mẫu như sau

X_i	S_1	S_2	$\cdots$	S_k
n_i	n_1	n_2	$\cdots$	n_k

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

<u>Tổ chức đồ</u>: Trên hệ trục tọa độ, biểu đồ cột có chiều rộng h, chiều cao tương ứng mỗi cột là $f_i = \frac{n_i}{h.n}$ được gọi là tổ chức đồ.



Hình 3.3. Tổ chức đồ

- Đa giác tần suất và tổ chức đồ cho ta dáng điệu của hàm mật độ của biến ngẫu nhiên X.

3.1.3. Phân phối thực nghiệm

- Giả sử mẫu thực nghiệm $(x_1, x_2, \cdots, x_n)$ sinh từ X. Ta xây dựng hàm

$$F_n(x) = \frac{\text{Sè c¸c gi¸ trÞ } x_i < x}{n}$$

được gọi là hàm phân phối thực nghiệm.

- Định lý Glivenco. Giả sử F(x) là hàm phân phối của biến ngẫu nhiên X mà ta cần tìm. $F_n(x)$ là hàm phân phối thực nghiệm nhận được từ mẫu ngẫu nhiên cỡ n sinh ra từ X. Khi đó

$$P\{\sup_{x \in R} |F_n(x) - F(x)| \to 0 \text{ khi } n \to \infty\} = 1$$

Như vậy hàm phân phối thực nghiệm là một xấp xỉ của hàm phân phối lý thuyết. Với n cố định, hàm phân phối thực nghiệm cho ta hình ảnh hình học về phân phối lý thuyết cần tìm.

3.2. Các đặc trưng mẫu

- Giả sử $(X_1, X_2, \cdots, X_n)$ là mẫu ngẫu nhiên sinh ra từ X có $EX = \mu, DX = \sigma^2$.

3.2.1. Thống kê

- Một hàm $T = T(X_1, X_2, \cdots, X_n)$ được gọi là một thống kê.

Ví dụ: $T(X_1, X_2, \cdots, X_n) = \max_n \{X_1, X_2, \cdots, X_n\}$ là một thống kê.

3.2.2. Kỳ vọng mẫu (Trung bình mẫu)

- Thống kê, ký hiệu $\overline{X}$, xác định bởi

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

gọi là kỳ vọng mẫu.

- Kỳ vọng mẫu là một biến ngẫu nhiên có các đặc trưng

$$E(\overline{X}) = \mu; \quad D(\overline{X}) = \frac{\sigma^2}{n}$$

3.2.3. Phương sai mẫu

- Thống kê, ký hiệu $\widehat{S}^2$, xác định bởi

$$\widehat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

gọi là phương sai mẫu.

- Dùng các phép biến đổi, ta được

$$\widehat{S}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

- Phương sai mẫu cũng là một biến ngẫu nhiên và $E\widehat{S}^2 = \dfrac{n-1}{n}\sigma^2$.

- Thống kê

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}\widehat{S}^2$$

gọi là phương sai mẫu điều chỉnh.

3.2.4. Độ lệch tiêu chuẩn mẫu

- Thống kê $\widehat{S} = \sqrt{\widehat{S}^2}$ được gọi là độ lệch tiêu chuẩn mẫu.

- Thống kê $S = \sqrt{S^2}$ được gọi là độ lệch tiêu chuẩn mẫu điều chỉnh.

3.2.5. Cách tính $\overline{X}, \widehat{S}^2$

- Giả sử mẫu thực nghiệm $(x_1, x_2, \cdots, x_n)$ có bảng tần số mẫu:

X_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

- Khi đó

$$n=\sum_{i=1}^{k} n_i, \qquad \overline{X}=\frac{1}{n}\sum_{i=1}^{k} n_i x_i, \qquad \widehat{S}^2=\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2-\overline{X}^2$$

- Nếu các giá trị x_i cách đều nhau thì ta có thể đơn giản tính toán bằng phép đổi biến sau

$$u_i=\frac{x_i-x_0}{h} \quad (i=1..k)$$

trong đó x_0 là giá trị chọn nào đó, h là khoảng cách đều giữa các x_i. Khi đó

$$\overline{X}=\frac{h}{n}\sum_{i=1}^{k} n_i u_i+x_0=h\overline{u}+x_0, \qquad \widehat{S}^2=h^2\left[\frac{1}{n}\sum_{i=1}^{k} n_i u_i^2-\overline{u}^2\right]=h^2.s_u^2$$

- Chú ý: Nếu bảng tần số của mẫu cho dưới dạng khoảng thì ta chọn giá trị đại diện x_i cho mỗi khoảng là giá trị trung bình của mỗi khoảng.
- Ví dụ: Tại một trang trại, điều tra trọng lượng của một loại trái cây, ta thu được kết quả sau:

Trọng lượng (g)	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75
Số trái	3	20	30	44	105	30	13	3	2

Tính $\overline{X}, S^2$.

Thực hiện đổi biến: $u_i=\frac{x_i-x_0}{h}$ với $x_0=52,5; h=5$.

Ta lập bảng tính như sau:

X_i	n_i	u_i	n_iu_i	$n_iu_i^2$
32,5	3	−4	−12	48
37,5	20	−3	−60	180
42,5	30	−2	−60	120
47,5	44	−1	−44	44
52,5	105	0	0	0
57,5	30	1	30	30
62,5	13	2	26	52
67,5	3	3	9	27
72,5	2	4	8	32
$\sum$	250		−103	533

Khi đó:

$$n=250; \quad \overline{u}=\frac{-103}{250}=-0,412; \quad s_u^2=\frac{533}{250}-(-0,412)^2=1,962$$

Suy ra

$$\overline{X}=h\overline{u}+x_0=50,44; \quad \widehat{S}^2=h^2 s_u^2=49,06; \quad S^2=\frac{n}{n-1}\widehat{S}^2=49,25$$

3.3. Một số phân phối trong thống kê

3.3.1. Phân phối χ^2

- Cho dãy biến ngẫu nhiên (X_i) độc lập, cùng phân phối chuẩn $N(0,1)$. Khi đó biến ngẫu nhiên

$$\chi^2(n) = X_1^2 + X_2^2 + \cdots + X_n^2$$

gọi là có phân phối khi-bình phương với bậc tự do n.

- Hàm mật độ của phân phối $\chi^2(n)$ có dạng

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

trong đó $\Gamma(p) = \int_0^{+\infty} x^{p-1}e^{-x}dx, p > 0$ là hàm Gamma.

3.3.2. Phân phối Student t

- Giả sử $U \sim N(0,1), V \sim \chi^2(n)$ độc lập nhau. Khi đó phân phối của biến ngẫu nhiên

$$T = \frac{U}{\sqrt{V/n}}$$

gọi là phân phối Student t với n bậc tự do.

Hàm mật độ của phân phối Student t có dạng

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

3.3.3. Phân phối Fisher F

- Giả sử $U \sim \chi^2(m), V \sim \chi^2(n)$ độc lập nhau. Khi đó phân phối của biến ngẫu nhiên

$$F = \frac{U/m}{V/n}$$

gọi là phân phối F với m,n bậc tự do.

3.4. Phân phối của $\overline{X}, S_n^2$

- Giả sử mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ sinh ra từ phân phối chuẩn $N(\mu, \sigma^2)$. Khi đó

- $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ hay $\frac{(\overline{X}-\mu)}{\sigma}\sqrt{n} \sim N(0;1)$.
- $\frac{\overline{X}-\mu}{\hat{S}}\sqrt{n-1}$ hoặc $\frac{\overline{X}-\mu}{S}\sqrt{n}$ có phân phối Student với n bậc tự do.
- $\frac{n\hat{S}^2}{\sigma^2} \sim \chi^2(n-1)$ hay $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
- $\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$

- Nếu mẫu $(X_1, X_2, \cdots, X_n)$ sinh từ $N(\mu_1, \sigma_1^2)$ và mẫu $(Y_1, Y_2, \cdots, Y_m)$ sinh từ $N(\mu_2, \sigma_2^2)$ thì

- $\overline{X} - \overline{Y} : \; N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$.
- $\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}\sqrt{\dfrac{n+m}{nm}}}$ có phân phối Student với m+n-2 bậc tự do.
- $\dfrac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} : \; F(n-1, m-1)$ bậc tự do.
- $\dfrac{\sum_{i=1}^{n}(X_i - \mu_1)^2/\sigma_1^2}{\sum_{i=1}^{n}(Y_i - \mu_2)^2/\sigma_2^2} \cdot \dfrac{m}{n} : \; F(n, m)$?bậc tự do.

- Chú ý: Nếu mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ sinh ra từ một biến ngẫu nhiên bất kỳ có phương sai hữu hạn khác không thì theo định lý giới hạn trung tâm, khi n lớn thì trung bình mẫu $\overline{X}$ có phân phối tiệm cận chuẩn. Do đó các kết quả trên vẫn đúng đối với mẫu ngẫu nhiên sinh ra từ một biến ngẫu nhiên bất kỳ.

Bài tập

1. Đo độ dài của 36 chi tiết được chọn ngẫu nhiên của một loại sản phẩm, ta được bảng số liệu sau:

39 43 41 41 40 41 43 42 41 39 40 42
44 42 42 41 41 42 43 40 41 41 42 43
39 40 41 39 40 42 40 43 39 41 40 40

a) Lập bảng tần số thực nghiệm và hàm phân phối thực nghiệm.

b) Tìm các đặc trưng mẫu $\overline{X}, S^2$.

2. Quan sát chiều cao của 1 loại cây dầu sau 6 tháng tuổi ở khu trồng trọt A, ta thu được bảng số liệu sau:

Chiều cao X (cm)	$30-35$	$35-40$	$40-45$	$45-50$	$50-55$	$55-60$	$60-65$
Số cây n_i	5	11	44	59	50	27	4

a) Lập bảng tần số thực nghiệm và hàm phân phối thực nghiệm.

b) Tìm các đặc trưng mẫu $\overline{X}, S^2$.

3. Kết quả đo chiều cao của sinh viên trường Đại học A cho ở bảng số liệu sau:

Chiều cao X (cm)	$150-155$	$155-160$	$160-165$	$165-170$	$170-175$	$175-180$
Số sinh viên n_i	32	74	120	57	15	2

Tìm các đặc trưng mẫu $\overline{X}, S^2$.

4. Kết quả khảo sát về hàm lượng Vitamin C của một loại trái cây cho ở bảng số liệu sau:

Hàm lượng X (%)	$6-7$	$7-8$	$8-9$	$9-10$	$10-11$	$11-12$	$12-13$
Số trái n_i	15	22	28	42	25	10	8

Tìm hàm lượng Vitamin C trung bình của trái cây.

5. Có thể nói gì về mẫu có độ lệch tiêu chuẩn mẫu bằng 0?

CHƯƠNG 4:

ƯỚC LƯỢNG THAM SỐ.

4.1. Các khái niệm cơ bản về ước lượng điểm và ước lượng khoảng.

Lập thống kê $\hat{\theta}(X_1, X_2, ..., X_n)$ để ước lượng cho tham số θ.

4.1.1. Định nghĩa: Thống kê $\hat{\theta}(X_1, X_2, ..., X_n)$ ước lượng cho tham số θ được gọi là hàm ước lượng của θ.

Với mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì $\hat{\theta}((x_1, x_2, ..., x_n)$ là 1 điểm trên trục số R, điểm đó dùng thay cho θ. $\hat{\theta}(X_1, X_2, ..., X_n)$ được gọi là ước lượng điểm của θ.

4.1.2. Định nghĩa: Khoảng tin cậy $(\hat{\theta}_1, \hat{\theta}_2)$ của tham số số θ với độ tin cậy $1-\alpha$ là 1 khoảng $(\hat{\theta}_1, \hat{\theta}_2)$ với $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, ..., X_n)$; $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, ..., X_n)$ sao cho $P\{\theta_1(X_1, X_2, ..., X_n)\} \leq \theta \leq \hat{\theta}_2(X_1, X_2, ..., X_n)\} = 1 - \alpha$.

4.2. Một số loại ước lượng điểm

4.2.1. Ước lượng không chệch

a. Hàm ước lượng $\hat{\theta}$ $(X_1, X_2, ..., X_n)$ được gọi là ước lượng không chệch của θ nếu $E\hat{\theta}(X_1, X_2, ..., X_n) = \theta$.

b. Giả sử $\hat{\theta}(X_1, X_2, ..., X_n)$ là hàm ước lượng của θ

Nếu $\lim\limits_{n\to\infty}\left[E\hat{\theta}(X_1,...X_n) - \theta\right] = 0$ thì $\hat{\theta}$ được gọi là tiệm cận không chệch của θ.

Độ chệch của một ước lượng được đo bằng giá trị $BS = \left|E(\hat{\theta}) - \theta\right|$.

Giữa những ước lượng chệch thì ta chọn ước lượng có BS nhỏ nhất.

<u>Ví dụ</u>:

1. Trung bình mẫu $\overline{X}$ là ước lượng không chệch của $EX = \mu$..

Thật vậy: $EX = EX_i = \mu$; $\quad E\overline{X} = E\left(\frac{1}{n}\sum X_i\right) = \mu$

2. $VarX = E(X-EX)^2 = \sigma^2$

$S^2 = \frac{1}{n-1}\hat{S}^2$, S^2 là ước lượng không chệch cho σ^2

Thật vậy: $ES^2 = \dfrac{n}{n-1}E\hat{S}^2 = \sigma^2$

3. $E(f) = P$. Tần suất mẫu là ước lượng không chệch cho xác suất P.

4.2.2. Ước lượng vững

$\hat{\theta}$ được gọi là ước lượng vững cho θ nếu ∀ε>0: $\lim\limits_{n\to\infty} P\left\{\left|\hat{\theta}(X_1,...,X_n)-\theta\right|<\varepsilon\right\} = 1$.

Ký hiệu $\hat{\theta} \xrightarrow{P} \theta$

Theo luật số lớn Chebyshev và luật số lớn Bernoulli thì $\overline{X}$ là ước lượng vững của $EX = \mu$; tần suất mẫu f là ước lượng vững của xác suất P.

Định lý : Nếu $\hat{\theta}$ $(X_1, X_2, ..., X_n)$ là ước lượng tiệm cận không chệch của θ và $\lim\limits_{n\to\infty}$ var $[\hat{\theta}(X_1, X_2, ..., X_n)] = 0$ thì $\hat{\theta}$ $(X_1, X_2, ..., X_n)$ là ước lượng vững của θ.

Chứng minh:

Theo bất đẳng thức chebyshev $P\left\{\left|\hat{\theta} - E\hat{\theta}\right| < \varepsilon\right\} \geq 1 - \dfrac{\text{var}\,\hat{\theta}}{\varepsilon^2}$

$\Rightarrow \lim P\left\{\left|\hat{\theta} - \theta\right| < \varepsilon\right\} = 1$ vì var $\hat{\theta} \to 0$

4.2.3. Ước lượng hiệu quả (ước lượng không chệch có phương sai tối thiểu)

Định nghĩa 1: Giả sử $\hat{\theta}$ $(X_1, X_2, ..., X_n)$ và $T(X_1, X_2, ..., X_n)$ là các ước lượng không chệch của θ.

Nếu var $(\hat{\theta}) \leq$ var T, ∀ T là ước lượng bất kỳ khác thì $\hat{\theta}$ ước lượng không chệch có phương sai tối thiểu (hay ước lượng hiệu quả của θ).

Định nghĩa 2: (Lượng thông tin Fisher)

a) Giả sử biến ngẫu nhiên rời rạc X có hàm xác suất $P(X=x) = P(x,\theta)$ phụ thuộc θ. Lượng thông tin Fisher về tham số θ chứa trong một quan sát được định nghĩa như sau : $I(\theta) = E\left(\dfrac{\partial LnP(x,\theta)}{\partial\theta}\right)^2 = \sum\left[\dfrac{\partial LnP(x,\theta)}{\partial\theta}\right]^2 P(x,\theta)$

b) Nếu X là biến ngẫu nhiên liên tục với hàm mật độ xác suất $P(x,\theta)$ thì

$$I(\theta) = E\left(\frac{\partial LnP(x,\theta)}{\partial \theta}\right)^2 = \int_{-\infty}^{-\infty}\left[\frac{\partial LnP(x,\theta)}{\partial \theta}\right]^2 P(x,\theta)dx$$

Định lý: Nếu mẫu (X_1, X_2, ..., X_n) lấy từ biến ngẫu nhiên X có hàm mật độ xác suất P(x,θ) và $\hat{\theta}$ là một ước lượng không chệch của θ thì:

$$Var(\hat{\theta}) \geq \frac{1}{nE\left[\frac{\partial LnP(x,\theta)}{\partial \theta}\right]^2} = \frac{1}{nI(\theta)}$$ (bất đẳng thức Cramer - Rao)

Ý nghĩa: Không tồn tại một ước lượng nào có phương sai bằng 0 và cận dưới của phương sai các ước lượng là $\frac{1}{nI(\theta)}$.

4.3. Phương pháp tìm ước lượng điểm

4.3.1. Phương pháp ước lượng hợp lý cực đại

Giả sử X là biến ngẫu nhiên có hàm mật độ xác suất p(x,θ), trong đó p đã biết còn θ chưa biết. Để ước lượng θ, ta lấy mẫu (X_1, X_2, ..., X_n) và lập hàm

$$L(\theta) = p(X_1, \theta).\ p(X_2,\theta)...p(X_n,\theta)$$

Tìm $\hat{\theta}$ (X_1, X_2, ..., X_n) sao cho : $L(\hat{\theta}) \geq L(\theta), \forall \theta \in T$ (1)

Hàm L(θ) được gọi là hàm hợp lý của tham số θ; giá trị của hàm hợp lý là mật độ xác suất tại điểm (x_1, x_2, ..., x_n). Giá trị của thống kê $\hat{\theta}$ tại điểm đó được gọi là ước lượng hợp lý cực đại của θ nếu ứng với giá trị này của θ, hàm hợp lý L(θ) đạt cực đại (nghĩa là mật độ xác suất tại (x_1, x_2, ..., x_n) đạt cực đại). Do hàm L(θ) và LnL(θ) cùng đạt cực đại tại 1 điểm $\hat{\theta}$ nên người ta tìm cực đại của LnL bằng cách giải nghiệm của phương trình:

$\frac{dLnL}{d\theta} = 0$; giả sử nghiệm là $\theta = \hat{\theta} = f(x_1, x_2, ..., x_n)$.

Xét $\frac{d^2LnL}{d\theta^2}$. Nếu tại $\theta = \hat{\theta}$, $\frac{d^2LnL}{d\theta^2} < 0$ thì L(θ) đạt cực đại tại $\hat{\theta}$.

Khi đó $\hat{\theta} = f(x_1, x_2, ..., x_n)$ là ước lượng điểm hợp lý tối đa cần tìm của θ.

Ví dụ: Cho X ~ N(μ,σ^2); giả sử đã biết σ^2 ,tìm ước lượng hợp lý cực đại của μ.

$$L(\theta) = L(\mu) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}$$

$$\text{LnL}(\theta) = \text{LnL}(\mu) = \text{nLn}\frac{1}{\sigma\sqrt{2n}} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i - \mu\right)^2$$

$$\frac{d\text{LnL}(\theta)}{d\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \mu\right)$$

$$\frac{d\text{LnL}(\theta)}{d\theta} = 0 \Leftrightarrow \sum_{i=1}^{n}\left(X_i - \mu\right) = 0$$

$$\Leftrightarrow \mu = \frac{\sum_{1}^{n} X_i}{n} = \overline{X}$$

Xét đạo hàm cấp hai $\frac{d^2\text{LnL}(\theta)}{d\theta^2} = -\frac{n}{\sigma^2} < 0$

$\Rightarrow$ LnL(μ) đạt cực đại tại $\overline{X}$.

Vậy $\overline{X}$ là ước lượng hợp lý cực đại của μ.

Kết quả trên có thể mở rộng cho trường hợp tham số θ nhiều chiều, chẳng hạn

$\theta = (\theta_1, \theta_2)$. Để tìm $\hat{\theta}$, ta giải hệ $\left\{\frac{\partial \text{LnL}(\theta)}{\partial \theta_i} = 0,\ i = \overline{1,2}\right.$

Nghiệm của hệ: $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ là ước lượng hợp lý cực đại của θ nếu thỏa mãn (1).

<u>Ví dụ</u>: Cho $X \sim N(\mu, \sigma^2)$, trong đó μ, σ^2 : chưa biết. Tìm ước lượng hợp lý cực đại hợp lý cho (μ, σ^2).

$$\frac{\partial}{\partial \mu}\text{LnL}(\theta) = \frac{1}{\partial^2}\left(\sum X_i - n\mu\right) = 0 \quad (3)$$

$$\frac{\partial \text{LnL}(\theta)}{\partial \sigma^2} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^3} = 0 \quad (4)$$

$$\Leftrightarrow \begin{cases} \mu = \overline{X} \\ \sigma^2 = \hat{S}^2 \end{cases}$$

Ta có: $\text{LnL}(\theta) = nLn\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$

$\Rightarrow \text{LnL}(\mu,\sigma^2) = -\text{nLn}\sigma - \frac{n}{2}\text{Ln}(2\pi) - \frac{n\hat{S}^2 + n\left(\bar{X}-\mu\right)^2}{2\sigma^2}$

$\text{LnL}(\bar{X},\hat{S}^2) = -\text{nLn}\hat{S} - \frac{n}{2}\text{Ln}(2\pi) - \frac{n}{2}$

Để $\text{LnLn}(\mu,\sigma^2) \le \text{LnLn}(\bar{X},\hat{S}^2) \Leftrightarrow -\text{nLn}\sigma - \frac{n\hat{S}^2 + n(\bar{X}-\mu)^2}{2\sigma^2} \le -\text{nLn}\hat{S} - \frac{n}{2}$

$\Leftrightarrow 0 \le \left[\left(\frac{\hat{S}^2}{2\sigma^2} - \frac{1}{2}\right) - \text{Ln}\frac{\hat{S}}{\sigma}\right] + \frac{(\bar{X}-\mu)^2}{2\sigma^2}$ (5)

(5) luôn luôn đúng vì $\frac{1}{2}(x^2-1) \ge \text{Ln}x, \forall x>0$.

Vậy $\bar{X}$ và $\hat{S}^2$ là ước lượng hợp lý cực đại của μ và σ^2

4.3.2. Phương pháp Moment

Xét mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ lấy từ biến ngẫu nhiên X. Giả sử X có phân phối $F(x,\theta_1, \theta_2,... \theta_m)$ và ta cần ước lượng m tham số này.

Các moment của X lần lượt là:

$EX = g_1(\theta_1, \theta_2,... \theta_m)$

$EX^2 = g_2(\theta_1, \theta_2,... \theta_m)$

....

$EX^m = g_m(\theta_1, \theta_2,... \theta_m)$

Gọi các moment mẫu cấp k là $\frac{1}{n}\sum_{i=1}^{n} X_i^k$, $k = \overline{1,m}$.

Ước lượng $\hat{\theta} = (\hat{\theta}_1, ...,\hat{\theta}_k)$ là nghiệm của hệ phương trình:

$$g_k(\theta_1, \theta_2,... \theta_m) = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \ k = \overline{1,m}.$$

Hàm $\hat{\theta}$ được gọi là ước lượng của θ bằng phương pháp moment.

<u>Ví dụ</u>: Cho $X \sim N(\mu,\sigma^2)$ với μ,σ^2 chưa biết. Hãyước lượng μ,σ^2 bằng phương pháp moment.

Ta có $EX = \mu$

$EX^2 = \sigma^2 + \mu^2$

hệ phương trình :

$$\begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} X_i = \mu \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \end{cases}$$

có nghiệm là:

$$\hat{\mu} = \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \hat{S}^2$$

Vậy, bằng phương pháp moment, ta tìm được: ước lượng của μ là $\overline{X}$; ước lượng của σ^2 là $\hat{S}^2$

4.4. Ước lượng khoảng.

4.4.1. Khái niệm

Mục đích của bài toán là tìm khoảng tin cậy ($\hat{\theta}_1$, $\hat{\theta}_2$) của tham số θ với độ tin cậy 1- α.

Với mẫu ngẫu nhiên ($X_1, X_2, ..., X_n$) ta xây dựng thống kê $K = f(X_1, X_2, ..., X_n, \theta)$ sao cho quy luật phân phối xác suất của K hoàn toàn được xác định.

Với độ tin cậy 1 - α, và với α_1, α_2 sao cho: $\alpha_1 + \alpha_2 = \alpha$, ta tìm được các giá trị $k_{1-\alpha_1}, k_{\alpha_2}$ sao cho :

$$P(K > k_{\alpha_2}) = \alpha_2$$

$$P(K < k_{1-\alpha_1}) = \alpha_1$$

$$\Rightarrow P(k_{1-\alpha_1} < K < k_{\alpha_2}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$$

Từ khoảng tin cậy ($k_{1-\alpha_1}; k_{\alpha_2}$) của K ta suy ra khoảng tin cậy (θ_1, $\hat{\theta}_2$) của θ sao cho $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$.

Trên một phép thử với mẫu ngẫu nhiên ($X_1, X_2, ..., X_n$), ta nhận được mẫu cụ thể ($x_1, x_2, ..., x_n$) và các giá trị của $\hat{\theta}_1$, $\hat{\theta}_2$ ứng với mẫu cụ thể này. Từ đó nhận được khoảng tin cậy của θ ứng với mẫu thu được.

4.4.2. Ước lượng kỳ vọng toán μ của biến ngẫu nhiên phân phối chuẩn.

Giả sử X là biến ngẫu nhiên phân phối chuẩn $N(\mu,\sigma^2)$. Ta ước lượng μ theo 2 trường hợp sau:

1) Đã biết phương sai σ^2.

Ta dùng thống kê K $= \dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}$, theo chương 3, ta đã biết K phân phối theo quy luật chuẩn N(0;1)

Với : $\alpha_1 + \alpha_2 = \alpha$, ta tìm được $u_{1-\alpha_1}$, u_{α_2} sao cho:

$$P(U > u_{\alpha_2}) = \alpha_2 \text{ và } P(U < u_{1-\alpha_1}) = \alpha_1$$

Do tính đối xứng của phân phối N(0;1) nên: $u_{1-\alpha_1} = -\, u_{\alpha_1}$.

Từ đó suy ra: $P(-u_{\alpha_1} < U < u_{\alpha_2}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$

$$\Leftrightarrow P\left(-u_{\alpha_1} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < u_{\alpha_2}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\overline{X} - u_{\alpha_2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{\alpha_1}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Khoảng tin cậy $1 - \alpha$ dành cho μ :

$$\overline{X} - u_{\alpha_2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{\alpha_1}\frac{\sigma}{\sqrt{n}}$$

* Với $\alpha_1 = \alpha_2 = \alpha/2$, ta có khoảng tin cậy đối xứng:

$$\overline{X} - u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

* Với $\alpha_1 = 0$, $\alpha_2 = \alpha$, ta có khoảng tin cậy phía phải:

$$\overline{X} - u_{\alpha}\frac{\sigma}{\sqrt{n}} < \mu < +\infty$$

* Với $\alpha_1 = \alpha$, $\alpha_2 = 0$, ta có khoảng tin cậy phía trái:

$$-\infty < \mu < \overline{X} + u_{\alpha}\frac{\sigma}{\sqrt{n}}$$

2) Chưa biết phương sai σ^2:

Xét thống kê $K = \dfrac{\overline{X}-\mu}{S}\sqrt{n}$, K phân phối theo quy luật Student với số bậc tự do k = n-1.

Với $\alpha = \alpha_1 + \alpha_2$, ta tìm các giá trị $t_{1-\alpha_1}, t_{\alpha_2}$ sao cho:

$P(K > t_{\alpha_2}) = \alpha_2$

$P(K < t_{1-\alpha_1}) = P(K < -t_{\alpha_1}) = \alpha_1$.

$\Rightarrow P(-t_{\alpha_1} < K < t_{\alpha_2}) = 1-\alpha$

$\Leftrightarrow P\left(-t_{\alpha_1} < \dfrac{\overline{X}-\mu}{S}\sqrt{n} < t_{\alpha_2}\right) = 1-\alpha$

$\Rightarrow$ khoảng tin cậy dành cho μ: $\overline{X} - t_{\alpha_2}\dfrac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha_1}\dfrac{S}{\sqrt{n}}$

Với $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$, ta có khoảng tin cậy đối xứng:

$$\overline{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$

* Với $\alpha_1 = 0$, $\alpha_2 = \alpha$, ta có khoảng tin cậy phía phải:

$$\left(\overline{X} - t_{\alpha}\frac{S}{\sqrt{n}}\ ; +\infty\right)$$

* Với $\alpha_1 = \alpha$, $\alpha_2 = 0$, ta có khoảng tin cậy phía trái:

$$\left(-\infty; \overline{X} + t_{\alpha}\frac{S}{\sqrt{n}}\right)$$

4.4.3. Ước lượng khoảng tin cậy phương sai của biến ngẫu nhiên phân phối chuẩn (trường hợp chưa biết μ)

Cho biến ngẫu nhiên $X \sim N(\mu, \sigma^2)$. Ta xây dựng khoảng tin cậy cho tham số σ^2.

Xét thống kê $K = \dfrac{(n-1)S^2}{\sigma^2}$ phân phối χ^2 với (n-1) bậc tự do.

Với $\alpha_1 + \alpha_2 = \alpha$, tìm được các giá trị $\chi^2_{1-\alpha_1}$; $\chi^2_{\alpha_2}$ sao cho:

$P(K > \chi^2_{\alpha_2}) = \alpha_2$

$P(K > \chi^2_{1-\alpha_1}) = 1 - \alpha_1$

$\Leftrightarrow P\left(\chi^2_{1-\alpha_1} < K < \chi^2_{\alpha_2}\right) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$.

$\Leftrightarrow P\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha_2}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{1-\alpha_1}}\right) = 1 - \alpha$.

Khoảng tin cậy của σ^2: $\dfrac{(n-1)S^2}{\chi^2_{\alpha_2}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{1-\alpha_1}}$

* $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ $\qquad \dfrac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}}$

* $\alpha_1 = 0; \alpha_2 = \alpha$ $\qquad \left(\dfrac{(n-1)S^2}{\chi^2_{\alpha}};\ +\infty\right)$

* $\alpha_1 = \alpha$; $\alpha_2 = 0$ $\qquad \left(0;\ \dfrac{(n-1)S^2}{\chi^2_{1-\alpha}}\right)$

4.4.4. Ước lượngkhoảng tin cậy xác suất p của biến ngẫu nhiên phân phối theo quy luật 0 - 1:

Ta có: tần suất $f = \dfrac{m}{n} = \dfrac{\sum X_i}{n}$ có kì vọng $Ef = p$, phương sai $Df = \dfrac{p(1-p)}{n}$ ($Df = \dfrac{\sum DX_i}{n^2} = \dfrac{npq}{n^2}$).

Ta đã biết :

$$K = \frac{f-p}{\sqrt{\dfrac{p(1-p)}{n}}}$$ có phân phối tiệm cận chuẩn N(0;1) khi $n \to \infty$.

Trong thực tế:

* Nếu n lớn, p nhỏ thì f phân phối xấp xỉ quy luật Poisson.

* Nếu n lớn, p không nhỏ song thỏa mãn điều kiện:

$$n > 5 \text{ và } \frac{\left|\sqrt{\dfrac{p}{1-p}} - \sqrt{\dfrac{1-p}{p}}\right|}{\sqrt{n}} < 0{,}3.$$

thì K = $\dfrac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ phân phối xấp xỉ chuẩn N(0;1).

Với $\alpha_1 = \alpha_2 = \alpha / 2$, ta tìm được P ($U>U_{\alpha/2}$) = α / 2

$$P(U_{1-\alpha/2} < U < U_{\alpha/2}) = P(-U_{\alpha/2} < U < U_{\alpha/2}) = 1 - \alpha$$

$$\Leftrightarrow P\left(-U_{\alpha/2} < \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}} < U_{\alpha/2}\right) = 1-\alpha.$$

Việc giải p là phức tạp, nên trong thực tế nếu $n \geq 100$, ta chọn thống kê:

K = $\dfrac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}}$. Từ đó tìm được khoảng tin cậy đối xứng dành cho p:

$$f - u_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} < p < f + u_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}$$

Ví dụ: Gieo 1000 hạt giống thấy có 765 hạt nẩy mầm. Với độ tin cậy 95%, hãy ước lượng khoảng tin cậy cho tỉ lệ nẩy mầm của loại hạt giống nầy.

BÀI TẬP

4.1. Theo dõi năng suất của 100 điểm trồng giống lúa A ở địa phương B, người ta thu được kết quả sau:

Năng suất (tạ/ ha)	Số điểm (n_i)
25 - 30	3
30 - 35	5
35 - 40	12
40 - 45	18
45 - 50	26
50 - 55	20
55 - 60	8
60 - 65	5
65 - 70	3

1) Với độ tin cậy 98%, hãy ước lượng khoảng tin cậy cho tỉ lệ điểm trồng lúa có năng suất cao hơn 50tạ/ha của địa phương B.
2) Với độ tin cậy 95%, hãy ước lượng khoảng tin cậy cho năng suất trung bình của giống lúa A ở địa phương B.

4.2. Với số liệu trong bài tập 3.2 chương III, hãy:

1) Ước lượng khoảng tin cậy cho tỉ lệ cây dầu có chiều cao lớn hơn 50 cm của khu trồng trọt A với độ tin cậy 95%. Nếu muốn cho sai số của ước lượng không vượt quá 0,01 thì ta phải tăng kích thước mẫu lên bao nhiêu?

2) Với độ tin cậy 99%, hãy ước lượng khoảng tin cậy cho chiều cao trung bình của cây dầu sau 6 tháng tuổi ở khu trồng trọt A.

4.3. Tại một khu rừng nguyên sinh, người ta đeo vòng cho 2000 con chim rồi thả vào rừng. Sau một thời gian, bắt lại 500 con thì thấy có 92 con có đeo vòng. Với độ tin cậy 98%, hãy ước lượng số lượng chim trong khu rừng.

4.4. Đường kính của 176 chi tiết máy do một máy sản xuất được cho trên bảng sau:

Đường kính (mm)	Số chi tiết máy
9,5	5
9,6	8
9,7	15
9,8	22
9,9	34
10,0	38
10,1	26
10,2	15
10,3	7
10,4	6

Những chi tiết có đường kính từ 9,8mm đến 10,2mm được xem là đạt tiêu chuẩn kỹ thuật.

1) Ước lượng tỉ lệ và ước lượng trung bình đường kính của những chi tiết đạt tiêu chuẩn với cùng độ tin cậy 99%.
2) Để sai số ước lượng đường kính trung bình của các chi tiết đạt tiêu chuẩn là 0,02mm và sai số ước lượng tỉ lệ chi tiết đạt tiêu chuẩn là 0,05 với cùng độ tin cậy 99% thì cần phải đo thêm bao nhiêu chi tiết nữa.

4.5. Người ta dùng một phương pháp đo để đo 5 lần đường kính của một chi tiết máy, kết quả như sau: 4,89; 5,22; 5,15; 4,98 và 4,92mm. Với độ tin cậy 95%, hãy ước lượng khoảng tin cậy cho phương sai của phương pháp đo này.

NỘI DUNG CHI TIẾT

Chương 1: Các khái niệm cơ bản về xác suất.

1.1. Giải tích tổ hợp.
1.2. Biến cố ngẫu nhiên.
1.3. Xác suất của biến cố ngẫu nhiên.
1.4. Các định lý cơ bản về xác suất.

Chương 2: Biến ngẫu nhiên và qui luật phân phối xác suất.

2.1. Biến ngẫu nhiên.
2.2. Hàm phân phối xác suất.
2.3. Các đặc trưng của biến ngẫu nhiên.
2.4. Một số qui luật phân phối xác suất thông dụng.
2.5. Vectơ ngẫu nhiên 2 chiều.
2.6. Luật số lớn.
2.7. Các định lí giới hạn.

Chương 3: Lý thuyết mẫu.

3.1. Mẫu ngẫu nhiên.
3.2. Các đặc trưng của mẫu.
3.3. Một số phân phối trong thống kê.

Chương 4: Ước lượng tham số.

4.1. Các khái niệm.
4.2. Một số ước lượng điểm
4.3. Các phương pháp tìm hàm ước lượng.
4.4. Ước lượng bằng khoảng tin cậy.

Chương 5: Kiểm định giả thiết

5.1. Các khái niệm.
5.2. Kiểm định giả thiết về kì vọng toán của biến ngẫu nhiên phân phối chuẩn.
5.3. Kiểm định giả thiết về xác suất p của biến ngẫu nhiên phân phối theo qui luật 0-1.
5.4. Kiểm định giả thiết về phương sai của biến ngẫu nhiên phân phối chuẩn.
5.5. So sánh hai kì vọng toán.
5.6. So sánh hai xác suất.
5.7. So sánh hai phương sai.
5.8. Kiểm định giả thiết về qui luật phân phối xác suất của biến ngẫu nhiên.
5.9. Kiểm định giả thiết về tính độc lập.

Chương 6: Tương quan và hồi qui.

6.1. Các khái niệm.
6.2. Phân tích hồi qui tuyến tính đơn.

CHƯƠNG 5 :

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

5.1. CÁC KHÁI NIỆM CƠ BẢN

5.1.1. Giả thiết thống kê

Giả thiết thống kê là mệnh đề về dạng phân phối $F(x, \theta)$ của biến ngẫu nhiên X; về tham số θ của X hoặc về tính độc lập của các biến ngẫu nhiên.

Giả thiết thống kê đưa ra được ký hiệu là H_0.

Giả thiết đối lập của H_0 được ký hiệu là H_1.

Vấn đề đặt ra là hãy kiểm tra xem chấp nhận được hay không chấp nhận được giả thiết H_0. Việc kiểm định này được dựa vào thông tin của mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ nên được gọi là kiểm định thống kê.

5.1.2. Phương pháp chung

Gọi $\mathcal{U}$ là không gian các giá trị của mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$. Ta chia $\mathcal{U}$ thành 2 phần W và $\overline{W}$ sao cho $W \cap \overline{W} = \varnothing$ và $W \cup \overline{W} = \mathcal{U}$. Sau đó, ta chọn quyết định theo quy tắc sau:

+ Nếu $(X_1, X_2, \ldots, X_n) \in W$ thì ta bác bỏ giả thiết H_0 và chấp nhận đối thiết H_1.

+ Nếu $(X_1, X_2, \ldots, X_n) \notin W$ thì ta chấp nhận giả thiết H_0.

Miền W được gọi là miền tiêu chuẩn. Vấn đề là ta chọn miền W như thế nào?

Với phương pháp kiểm định nêu trên, ta có thể mắc hai loại sai lầm sau:

1) Sai lầm loại 1: Bác bỏ giả thiết H_0 trong khi H_0 đúng. Xác suất mắc sai lầm loại này được ký hiệu là $P(W/H_0)$.

2) Sai lầm loại 2: Chấp nhận giả thiết H_0 trong khi H_0 sai. Xác suất mắc sai lầm loại này được ký hiệu là $P(\overline{W}/H_1)$.

Ta có,
$$P(W/H_0) + P(\overline{W}/H_0) = 1$$
$$P(W/H_1) + P(\overline{W}/H_1) = 1.$$

Miền tiêu chuẩn W làm cực tiểu $P(W/H_0)$ nhưng chưa chắc đã làm cực tiểu $P(\overline{W}/H_1)$, nên ta không thể cùng một lúc làm cực tiểu cả 2 sai lầm. Thông thường, ta xác định miền W sao cho:

$$\begin{cases} P(W/H_0) = \alpha \ \ (\alpha \text{ cho trước khá bé}) \\ P(\overline{W}/H_1) = \beta \longrightarrow \min. \end{cases}$$

Giá trị α được gọi là mức ý nghĩa của tiêu chuẩn.

5.1.3. Thống kê kiểm định

Từ mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ của biến ngẫu nhiên X, ta lập thống kê

$K = f(X_1, X_2, \ldots, X_n, \theta)$, θ là tham số liên quan đến giả thiết kiểm định.

Điều kiện của thống kê K là nếu giả thiết H_0 đúng thì quy luật phân phối xác suất của K hoàn toàn xác định. Thống kê K được gọi là thống kê kiểm định.

5.1.4. Quy tắc kiểm định giả thiết thống kê

+ Với thống kê kiểm định K đã biết quy luật phân phối xác suất, ta xác định miền W_α sao cho $P(K \in W_\alpha / H_0) = \alpha$ với α là một xác suất khá bé cho trước.

W_α được gọi là miền bác bỏ giả thiết.

+ Trên một phép thử cụ thể, ta nhận được mẫu cụ thể $(x_1, x_2, \ldots, x_n)$.

Gọi $K_0 = f(x_1, x_2, \ldots, x_n, \theta)$.

Nếu $K_0 \in W_\alpha$: ta bác bỏ H_0, thừa nhận H_1.

Nếu $K_0 \notin W_\alpha$: ta nói chưa có cơ sở bác bỏ H_0.

5.2. KIỂM ĐỊNH GIẢ THIẾT VỀ KỲ VỌNG TOÁN CỦA BIẾN NGẪU NHIÊN PHÂN PHỐI CHUẨN

Giả sử X là biến ngẫu nhiên phân phối chuẩn $N(\mu; \sigma^2)$. Ta dựa vào mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ để kiểm định giả thiết H_0: $\mu = \mu_0$.

1. Trường hợp đã biết phương sai σ^2

Thống kê kiểm định $K = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$.

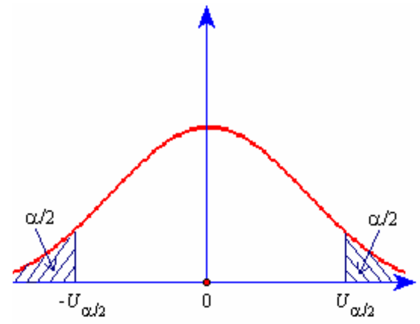
Nếu giả thiết H_0 đúng thì $K = \dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$ phân phối theo quy luật chuẩn $N(0; 1)$.

a. Kiểm định giả thiết H_0: $\mu = \mu_0$ với đối thiết H_1: $\mu \neq \mu_0$

Với xác suất α cho trước, ta tìm được giá trị $U_{\frac{\alpha}{2}}$ sao cho P(K>$U_{\frac{\alpha}{2}}$) = $\frac{\alpha}{2}$.

Miền bác bỏ giả thiết $W_\alpha = (-\infty;\ -U_{\frac{\alpha}{2}}) \cup (U_{\frac{\alpha}{2}};\ +\infty)$

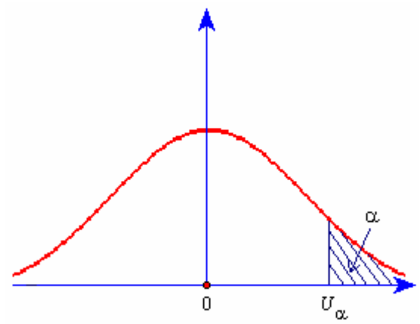
.



b. Kiểm định giả thiết H_0: $\mu = \mu_0$ với đối thiết H_1: $\mu > \mu_0$

Với α cho trước, ta tìm được U_α sao cho P($K > U_\alpha$) = P($K \in W_\alpha / H_0$) = α.

Miền bác bỏ giả thiết $W_\alpha = (U_\alpha;\ +\infty)$.

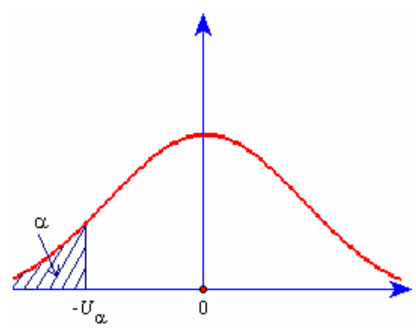


c. Kiểm định giả thiết H_0: $\mu = \mu_0$ với đối thiết H_1: $\mu < \mu_0$

Với α cho trước, ta tìm được $U_{1-\alpha} = -U_\alpha$ sao cho

P($K \in W_\alpha / H_0$) = P($K < U_{1-\alpha}$) = P($K < -U_\alpha$) = α.

Miền bác bỏ giả thiết $W_\alpha = (-\infty;\ -U_\alpha)$.



Ví dụ 5.1. Theo hợp đồng giữa nhà cung cấp súc vật thí nghiệm với bộ phận giáo tài, thỏ cung cấp để làm thí nghiệm phải nặng trên 1 kg. Cân ngẫu nhiên 50 con trong trại thỏ, tính được cân nặng trung bình 1,1 kg. Hỏi với mức ý nghĩa α = 0,05, thỏ ở trại đã có thể cung cấp làm thí nghiệm được chưa? Giả sử cân nặng của thỏ là biến ngẫu nhiên phân phối chuẩn với là σ = 0,1.

Giải: Gọi X là trọng lượng của thỏ. $X \approx N(\mu;\ \sigma^2)$.

Bài toán kiểm định giả thiết của μ của biến ngẫu nhiên phân phối chuẩn trong trường hợp đã biết phương sai.

Kiểm định giả thiết H_0: $\mu = 1$ với đối thiết H_1: $\mu > 1$.

Thống kê kiểm định $K = \dfrac{(\overline{X}-1)\sqrt{50}}{0,1}$.

Nếu giả thiết H_0 đúng thì $K = \dfrac{(\overline{X}-\mu)\sqrt{50}}{0,1}$ phân phối theo quy luật chuẩn N(0; 1).

Với $\alpha = 0{,}05$, miền bác bỏ giả thiết là $W_\alpha = (u_\alpha;\ +\infty) = (u_{0,05};\ +\infty) = (1{,}645;\ +\infty)$.

Giá trị quan sát $K_0 = \dfrac{(1{,}1-1)\sqrt{50}}{0{,}1} \approx 7{,}07$.

Ta có, $K_0 \in W_\alpha$ nên bác bỏ giả thiết H_0, chấp nhận H_1. Từ đó kết luận rằng thỏ ở trại đã có thể cung cấp để làm thí nghiệm.

2. Trường hợp chưa biết phương sai σ^2

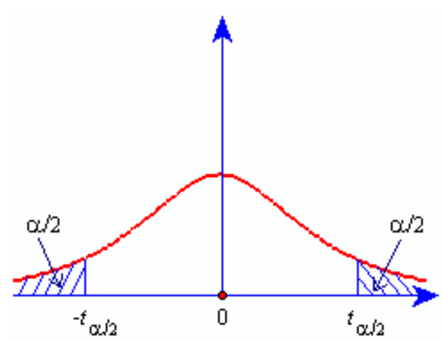
Thống kê kiểm định $K = \dfrac{(\overline{X}-\mu_0)\sqrt{n}}{S}$.

Nếu giả thiết H_0 đúng thì $K = \dfrac{(\overline{X}-\mu)\sqrt{n}}{S}$ phân phối theo quy luật Student với số bậc tự do $(n-1)$.

a. Kiểm định giả thiết H_0: $\mu = \mu_0$ với đối thiết H_1: $\mu \neq \mu_0$

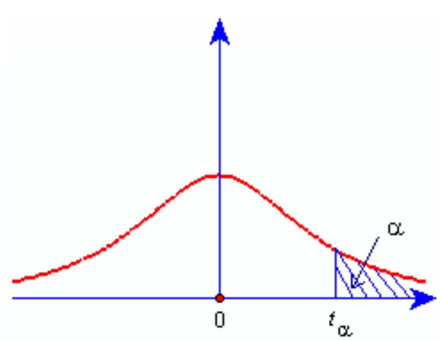
$P(K \in W_\alpha/H_0) = P(K < -t_{\frac{\alpha}{2}}) + P(K > t_{\frac{\alpha}{2}}) = \alpha$.

Miền bác bỏ giả thiết $W_\alpha = (-\infty;\ -t_{\frac{\alpha}{2}}) \cup (t_{\frac{\alpha}{2}};\ +\infty)$.



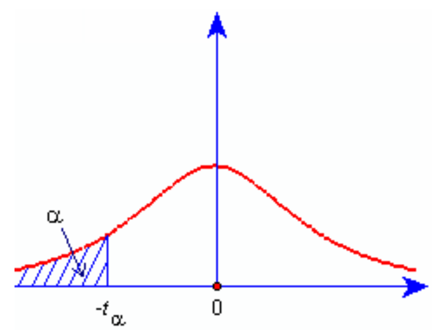
b. Kiểm định giả thiết H_0: $\mu = \mu_0$ với đối thiết H_1: $\mu > \mu_0$

$P(K \in W_\alpha/H_0) = P(K > t_\alpha) = \alpha$. Miền bác bỏ giả thiết $W_\alpha = (t_\alpha;\ +\infty)$.



c. Kiểm định giả thiết H_0: $\mu = \mu_0$ với đối thiết H_1: $\mu < \mu_0$

$P(K \in W_\alpha/H_0) = P(K < -t_\alpha) = \alpha$. Miền bác bỏ giả thiết $W_\alpha = (-\infty; -t_\alpha)$.



Ví dụ 5.2. Trong điều kiện chăn nuôi bình thường, lượng sữa trung bình của một con bò là 14 kg / 1ngày. Nghi ngờ điều kiện chăn nuôi bò kém đi làm cho lượng sữa giảm xuống, người ta đã điều tra ngẫu nhiên 25 conbò và tính được lượng sữa trung bình của một con trong 1 ngày là 12,5 kg và độ lệch chuẩn mẫu là 2,5 kg . Với mức ý nghĩa α = 0,05, hãy kết luận về điều nghi ngờ nói trên. Giả thiết lượng sữa bò là biến ngẫu nhiên phân phối chuẩn.

Giải: Gọi X là lượng sữa bò.

Bài toán kiểm định giả thiết của μ của biến ngẫu nhiên phân phối chuẩn trong trường hợp chưa biết phương sai.

Kiểm định giả thiết H_0: $\mu = 14$ với đối thiết H_1: $\mu < 14$.

Thống kê kiểm định $K = \dfrac{(\overline{X} - 14)\sqrt{25}}{S}$.

Nếu giả thiết H_0 đúng thì $K = \dfrac{(\overline{X} - \mu)\sqrt{25}}{S}$ phân phối theo quy luật Student với số bậc tự do là 24.

Với α = 0,05, miền bác bỏ giả thiết là $W_\alpha = (-\infty; -t_\alpha^{(24)}) = (-\infty; -1{,}711)$.

Từ mẫu cụ thể ta có, giá trị quan sát $K_0 = \dfrac{(12{,}5 - 14)\sqrt{25}}{2{,}5}$ = -3. Ta có, $K_0 \in W_\alpha$ nên bác bỏ giả thiết H_0, chấp nhận H_1. Từ đó có thể kết luận rằng điều nghi ngờ nói trên là đúng.

5.3. KIỂM ĐỊNH GIẢ THIẾT VỀ XÁC SUẤT *P* CỦA BIẾN NGẪU NHIÊN PHÂN PHỐI THEO QUY LUẬT 0_1

Giả sử (X_1, X_2, . . ., X_n) là mẫu ngẫu nhiên lập từ biến ngẫu nhiên X phân phối theo quy luật 0_1 với tham số p.

Khi đó $\sum_{i=1}^{n} X_i$ phân phối theo quy luật nhị thức B(n, p) và tần suất $f = \dfrac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$ có phân phối tiệm cận chuẩn.

Ta kiểm định giả thiết H_0: $p = p_0$ trong trường hợp n khá lớn và p không quá bé:

Xét thống kê $$K = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$$

Nếu H_0 đúng thì $K = \dfrac{(f - p)\sqrt{n}}{\sqrt{p(1-p)}}$ phân phối theo quy luật N(0; 1).

1) Kiểm định giả thiết H_0: $p = p_0$ với đối thiết H_1: $p \neq p_0$

$$\mathrm{P}(K \in W_\alpha \mid H_0) = \mathrm{P}(|K| > U_{\frac{\alpha}{2}}) = \alpha$$

$$W_\alpha = (-\infty; -U_{\frac{\alpha}{2}}) \cup (U_{\frac{\alpha}{2}}; +\infty)$$

2) Kiểm định giả thiết H_0: $p = p_0$ với đối thiết H_1: $p > p_0$

$$\mathrm{P}(K \in W_\alpha \mid H_0) = \mathrm{P}(K > U_\alpha) = \alpha$$

$$W_\alpha = (U_\alpha; +\infty)$$

3) Kiểm định giả thiết H_0: $p = p_0$ với đối thiết H_1: $p < p_0$

$$\mathrm{P}(K \in W_\alpha \mid H_0) = \mathrm{P}(K < -U_\alpha) = \alpha$$

$$W_\alpha = (-\infty; -U_\alpha)$$

Ví dụ 5.3. Một tổng kết của thư viện cho biết có 75% sinh viên thường xuyên đến thư viện đọc sách tham khảo. Một mẫu thăm dò cho thấy trong 200 sinh viên có 162 sinh viên thường xuyên đến phòng đọc. Với mức ý nghĩa 5% hãy kiểm định báo cáo tổng kết của thư viện.

Giải: Kiểm định giả thiết H_0: p = 0,75 với đối thiết H_1: $p \neq$ 0,75

Xét thống kê $$K = \frac{(f - 0,75)\sqrt{n}}{\sqrt{0,75(1-0,75)}}$$

Nếu H_0 đúng thì K phân phối theo quy luật N(0; 1).

Với α = 0,05, miền bác bỏ giả thiết $W_\alpha = (-\infty; -u_{0,025}) \cup (u_{0,025}; +\infty)$ = $(-\infty$; -1,96) $\cup$ (1,96; $+\infty$).

Giá trị quan sát $K_0 = \dfrac{(\frac{162}{200} - 0,75)\sqrt{200}}{\sqrt{0,75.0,25}}$ ≈ 2,29. Ta có, $K_0 \in W_\alpha$ nên bác bỏ giả thiết H_0, chấp nhận H_1: p $\neq$ 0,75.

5.4. KIỂM ĐỊNH GIẢ THIẾT VỀ PHƯƠNG SAI CỦA BIẾN NGẪU NHIÊN PHÂN PHỐI CHUẨN

Giả sử biến ngẫu nhiên phân phối theo quy luật chuẩn N(μ, σ^2).

Để kiểm định giả thiết H_0: $\sigma^2 = \sigma_0^2$, ta chọn thống kê kiểm định $K = \dfrac{(n-1)S^2}{\sigma_0^2}$.

Nếu giả thiết H_0 đúng thì $K = \dfrac{(n-1)S^2}{\sigma^2}$ phân phối theo quy luật χ^2 với (n-1) bậc tự do.

1) Kiểm định giả thiết H_0: $\sigma^2 = \sigma_0^2$ với đối thiết H_1: $\sigma^2 \neq \sigma_0^2$

$$\mathrm{P}(K \in W_\alpha \mid H_0) = \mathrm{P}(K > \chi^2_{\frac{\alpha}{2}}) + \mathrm{P}(K < \chi^2_{1-\frac{\alpha}{2}}) = \alpha$$

$$W_\alpha = (0; \chi^2_{1-\frac{\alpha}{2}}) \cup (\chi^2_{\frac{\alpha}{2}}; +\infty)$$

2) H_0: $\sigma^2 = \sigma_0^2$, H_1: $\sigma^2 > \sigma_0^2$

$$W_\alpha = (\chi^2_\alpha; +\infty)$$

3) H_0: $\sigma^2 = \sigma_0^2$, H_1: $\sigma^2 < \sigma_0^2$

$$W_\alpha = (0; \chi^2_{1-\alpha})$$

Ví dụ 5.4. Để kiểm tra độ chính xác của một máy, người ta đo ngẫu nhiên kích thước của 15 chi tiết máy do máy đó sản xuất và tính được s^2 =14,6. Với mức ý nghĩa α = 0,01, hãy kết luận máy móc có hoạt động bình thường không, biết rằng kích thước của chi tiết máy là biến ngẫu nhiên phân phối chuẩn có dung sai theo thiết kế là σ^2 = 12.

Giải: Gọi X là kích thước của chi tiết máy, theo giả thiết X tuân theo quy luật phân phối chuẩn.

Kiểm định giả thiết H_0: $\sigma^2 = 12$ với đối thiết H_1: $\sigma^2 > 12$

Chọn thống kê kiểm định $K = \dfrac{14.S^2}{12}$.

Nếu giả thiết H_0 đúng thì $K = \dfrac{14.S^2}{\sigma^2}$ phân phối theo quy luật χ^2 với 14 bậc tự do.

Với $\alpha = 0{,}01$, ta có $W_\alpha = (\chi_\alpha^{2(14)};+\infty) = (29{,}14;\ +\infty)$.

Giá trị quan sát $K_0 = \dfrac{14.S^2}{12} = \dfrac{14.14{,}6}{12} = 17{,}033$. Suy ra, $K_0 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thiết H_0 hay có thể kết luận máy móc vẫn hoạt động bình thường.

5.5. KIỂM ĐỊNH GIẢ THIẾT VỀ 2 KỲ VỌNG TOÁN CỦA 2 BIẾN NGẪU NHIÊN PHÂN PHỐI CHUẨN

Giả sử 2 biến ngẫu nhiên *X, Y* cùng phân phối theo quy luật chuẩn.

$$X \sim \mathrm{N}(\mu_1, \sigma_1^2)$$

$$Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$$

Ta kiểm định giả thiết H_0: $\mu_1 = \mu_2$ dựa trên mẫu ngẫu nhiên ($X_1, X_2, ..., X_{n1}$) lấy từ biến X và mẫu ngẫu nhiên ($Y_1, Y_2, ..., Y_{n2}$) lấy từ biến Y.

5. 5. 1. Trường hợp đã biết các phương sai σ_1^2, σ_2^2.

Thống kê kiểm định $K = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Với giả thiết H_0 đúng: $K = \dfrac{(\overline{X} - \overline{Y})}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ phân phối N(0; 1).

1) Kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 \neq \mu_2$

Miền bác bỏ giả thiết mức α:

$$W_\alpha = (-\infty; -U_{\frac{\alpha}{2}}) \cup (U_{\frac{\alpha}{2}}; +\infty)$$

2) Kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 > \mu_2$

$W_\alpha = (U_\alpha; +\infty)$

3) Kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 < \mu_2$

$W_\alpha = (-\infty; -U_\alpha)$

5. 5. 2. Trường hợp chưa biết các phương sai σ_1^2, σ_2^2 và có cơ sở để giả thiết $\sigma_1^2 = \sigma_2^2$

Thống kê kiểm định $K = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ với $S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

Nếu giả thiết H_0 đúng thì $K = \dfrac{(\overline{X} - \overline{Y})}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ phân phối theo quy luật Student với ($n_1 + n_2 - 2$) bậc tự do.

1) Kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 \neq \mu_2$

Miền bác bỏ giả thiết mức α:

$W_\alpha = (-\infty; -t_{\frac{\alpha}{2}}) \cup (t_{\frac{\alpha}{2}}; +\infty)$

2) Kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 > \mu_2$

$W_\alpha = (t_\alpha; +\infty)$

3) Kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 < \mu_2$

$W_\alpha = (-\infty; -t_\alpha)$

5. 5. 3. Trường hợp chưa biết các phương sai σ_1^2, σ_2^2 và không có cơ sở để giả thiết $\sigma_1^2 = \sigma_2^2$

Thống kê kiểm định $K = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

Nếu giả thiết H_0 đúng thì $K = \dfrac{(\overline{X} - \overline{Y})}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ phân phối theo quy luật Student với số bậc tự do

$$k = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)c^2 + (1 - c)^2 (n_1 - 1)} \text{ với } c = \frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Miền bác bỏ mức α được xác định như sau:

1) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$

$$W_\alpha = (-\infty; -t_{\frac{\alpha}{2}}) \cup (t_{\frac{\alpha}{2}}; +\infty)$$

2) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 > \mu_2$

$$W_\alpha = (t_\alpha; +\infty)$$

3) H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 < \mu_2$

$$W_\alpha = (-\infty; -t_\alpha)$$

Ví dụ 5.5. Người ta cân trẻ sơ sinh ở 2 khu vực thành thị và nông thôn, thu được bảng kết quả sau

Khu vực	Số trẻ được cân	Trọng lượng trung bình	Phương sai mẫu
Nông thôn	$n_1 = 2500$	$\overline{x}_1 = 3,0$	2
Thành thị	$n_2 = 500$	$\overline{x}_2 = 3,1$	3

Với mức ý nghĩa 0,01 có thể coi trọng lượng trung bình của trẻ sơ sinh ở hai khu vực là bằng nhau được không?

Giải: Gọi trọng lượng trẻ sơ sinh ở nông thôn và thành thị tương ứng là X_1 và X_2. Lúc đó, trọng lượng trẻ sơ sinh trung bình chính là μ_1 và μ_2.

Đây là bài toán kiểm định giả thiết H_0: $\mu_1 = \mu_2$ với đối thiết H_1: $\mu_1 \neq \mu_2$, trong trường hợp chưa biết phương sai σ_1^2, σ_2^2 và không có cơ sở để giả thiết $\sigma_1^2 = \sigma_2^2$.

Thống kê kiểm định $K = \dfrac{(\overline{X_1} - \overline{X_2})}{\sqrt{\dfrac{S_1^2}{2500} + \dfrac{S_2^2}{500}}}$

Do $n_1 > 30$, $n_2 > 30$ nên nếu giả thiết H_0 đúng thì K có phân phối xấp xỉ N(0; 1).

Với $\alpha = 0,01$, ta có $W_\alpha = (-\infty; -u_{0,005}) \cup (u_{0,005}; +\infty) = (-\infty; -2,576) \cup (2,576; +\infty)$.

Giá trị quan sát $K_0 = \dfrac{(3,0-3,1)}{\sqrt{\dfrac{2}{2500}+\dfrac{3}{500}}} \approx$ -1,21.

Vậy, $K_0 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thiết H_0 hay có thể coi trọng lượng của trẻ sơ sinh ở nông thôn và thành thị là như nhau.

5.6. SO SÁNH 2 XÁC SUẤT

Giả sử X, Y là 2 biến ngẫu nhiên phân phối theo quy luật 0_1 với các tham số p_1, p_2.

Ta kiểm định giả thiết H_0: $p_1 = p_2$ dựa trên 2 mẫu ngẫu nhiên ($X_1, X_2, \ldots, X_{n_1}$) lấy từ biến X và ($Y_1, Y_2, \ldots, Y_{n_2}$) lấy từ biến Y.

Để kiểm định giả thiết H_0, ta dùng thống kê

$$K = \frac{(f_1 - f_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

K phân phối xấp xỉ chuẩn N(0; 1) với $n_1 > 30$, $n_2 > 30$. Nếu giả thiết H_0 đúng ($p_1 = p_2 = p$) thì:

$$K = \frac{(f_1 - f_2)}{\sqrt{p(1-p)(\dfrac{1}{n_1} + \dfrac{1}{n_2})}}$$

vì p chưa biết nên p được thay bằng $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

$K = \dfrac{(f_1 - f_2)}{\sqrt{f(1-f)(\dfrac{1}{n_1} + \dfrac{1}{n_2})}}$ phân phối xấp xỉ chuẩn N(0; 1) với $n_1 > 30$, $n_2 > 30$.

1) Kiểm định giả thiết H_0: $p_1 = p_2$ với đối thiết H_1: $p_1 \neq p_2$

Miền bác bỏ giả thiết mức α:

$$W_\alpha = (-\infty; -U_{\frac{\alpha}{2}}) \cup (U_{\frac{\alpha}{2}}; +\infty)$$

2) Kiểm định giả thiết H_0: $p_1 = p_2$ với đối thiết H_1: $p_1 > p_2$

$$W_\alpha = (U_\alpha; +\infty)$$

3) Kiểm định giả thiết H_0: $p_1 = p_2$ với đối thiết H_1: $p_1 < p_2$

$W_\alpha = (-\infty; -U_\alpha)$

Ví dụ 5.6. Có 2 loại thuốc A và B cùng điều trị một loại bệnh nào đó. Qua theo dõi trên số bệnh nhân có dùng 2 loại thuốc này thu được các số liệu sau:

Loại thuốc	Số bệnh nhân dùng thuốc	Số bệnh nhân khỏi bệnh
A	160	120
B	56	40

Với mức ý nghĩa α = 0,05, hỏi tác dụng của 2 loại thuốc trên trong việc chữa bệnh có như nhau không?

Giải: Gọi p_1, p_2 tương ứng là tỉ lệ khỏi bệnh khi dùng thuốc A và B.

Đây là bài toán kiểm định giả thiết H_0: $p_1 = p_2$ với đối thiết H_1: $p_1 \neq p_2$ với n_1, n_2 khá lớn.

Thống kê kiểm định $K = \dfrac{(f_1 - f_2)}{\sqrt{f(1-f)(\dfrac{1}{n_1} + \dfrac{1}{n_2})}}$ với $f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

K phân phối xấp xỉ chuẩn N(0; 1).

Với α = 0,05, ta có $W_\alpha = (-\infty; -u_{0,025}) \cup (u_{0,025}; +\infty) = (-\infty; -1,96) \cup (1,96; +\infty)$.

Với 2 mẫu cụ thể ta có

$$f_1 = \frac{120}{160} = 0,75 \qquad f_2 = \frac{40}{56} \approx 0,714$$

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} \approx \frac{120 + 40}{160 + 56} \approx 0,741.$$

Tính được $K_0 \approx$ 0,529.

Vậy, $K_0 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thiết H_0 , ta chấp nhận công dụng của 2 loại thuốc A, B đối với việc điều trị bệnh nào đó là như nhau.

5.7. SO SÁNH 2 PHƯƠNG SAI

Giả sử biến ngẫu nhiên X phân phối chuẩn $N(\mu, \sigma_1^2)$

biến ngẫu nhiên Y phân phối chuẩn $N(\mu, \sigma_2^2)$

Ta kiểm định giả thiết H_0: $\sigma_1^2 = \sigma_2^2$ dựa trên 2 mẫu ngẫu nhiên ($X_1, X_2, \ldots, X_{n_1}$) lấy từ biến X và ($Y_1, Y_2, \ldots, Y_{n_2}$) lấy từ biến Y.

Thống kê kiểm định $K = \dfrac{S_1^2.\sigma_1^2}{S_2^2.\sigma_2^2}$ nếu $S_1^2 > S_2^2$

Nếu H_0 đúng thì $K = \dfrac{S_1^2}{S_2^2}$ ($S_1^2 > S_2^2$) phân phối theo quy luật Fisher với (n_1 - 1) và (n_2–1) bậc tự do.

Miền bác bỏ giả thiết mức α được xác định như sau:

1) H_0: $\sigma_1^2 = \sigma_2^2$; H_1: $\sigma_1^2 \neq \sigma_2^2$

$W_\alpha = (0; f_{1-\frac{\alpha}{2}}) \cup (f_{\frac{\alpha}{2}}; +\infty)$.

2) H_0: $\sigma_1^2 = \sigma_2^2$; H_1: $\sigma_1^2 > \sigma_2^2$

$W_\alpha = (f_\alpha; +\infty)$.

Ví dụ 5.7. Có 2 giống lúa có năng suất trung bình xấp xỉ như nhau song mức độ phân tán về năng suất có thể khác nhau. Để kiểm tra điều đó, người ta gặt mẫu tại 2 vùng trồng 2 giống lúa đó và thu được kết quả sau:

Giống lúa	Số điểm gặt	Phương sai
A	$n_1 = 41$	$s_1^2 = 11,41$
B	$n_2 = 30$	$s_2^2 = 6,52$

Với mức ý nghĩa α = 0,05 hãy kết luận về vấn đề trên, biết năng suất lúa là biến ngẫu nhiên phân phối chuẩn.

Giải: Gọi X, Y là năng suất của 2 giống lúa A và B, X và Y có phân phối chuẩn.

Kiểm định giả thiết H_0: $\sigma_1^2 = \sigma_2^2$ với đối thiết H_1: $\sigma_1^2 \neq \sigma_2^2$

Nếu H_0 đúng thì thống kê kiểm định có dạng $K = \dfrac{S_1^2}{S_2^2}$ ($S_1^2 > S_2^2$) phân phối theo quy luật Fisher với 40 và 29 bậc tự do.

Do α = 0,05 nên $f_{\frac{\alpha}{2}}^{(40,29)} = f_{0,025}^{(40,29)} = 2,03$ và $f_{1-\frac{\alpha}{2}}^{(40,29)} = \dfrac{1}{f_{\frac{\alpha}{2}}^{(29,40)}} = \dfrac{1}{1,94}$ =0,52.

Vậy, miền bác bỏ là $W_\alpha = (0;\ 0,52) \cup (2,03;\ +\infty)$.

Giá trị quan sát $K_{qs} = \dfrac{11,41}{6,52} \approx 1,75$.

$K_{qs} \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thiết H_0 hay độ phân tán của năng suất hai giống lúa nói trên là như nhau.

5.8. KIỂM ĐỊNH GIẢ THIẾT VỀ QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

Xét giả thiết H_0: biến ngẫu nhiên X có quy luật phân phối xác suất xác định bởi hàm phân phối $F(x)$.

Dựa vào mẫu thực nghiệm để xét xem giả thiết H_0 đúng không? Nói cách khác, các số liệu thực nghiệm ($x_1, x_2, \ldots, x_n$) có phù hợp với giả thiết lý thuyết phân phối xác suất của X là hàm $F(x)$ hay không?

5. 8. 1. Trường hợp X là biến ngẫu nhiên rời rạc

Giả sử mẫu thực nghiệm ($x_1, x_2, \ldots, x_n$) có bảng phân phối tần số thực nghiệm:

x_i	x_1	x_2	. . . . . .	x_k
n_i	n_1	n_2	. . . . . .	n_k

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

Nếu H_0 đúng, phân phối xác suất của X được xác định, ta tính được các xác suất lý thuyết $P(x = x_i) = P_i,\ i = \overline{1,k}$

Từ đó suy ra tần số lý thuyết của giá trị x_i là $np_i,\ i = \overline{1,k}$

Thống kê kiểm định $K = \sum_{i=1}^{k} \dfrac{(n_i - np_i)^2}{np_i}$ có phân phối χ^2 với $(k - 1)$ bậc tự do.

Miền bác bỏ giả thiết mức α: $(\chi_\alpha^2, +\infty)$.

5. 8. 2. Trường hợp X là biến ngẫu nhiên liên tục

Từ mẫu cụ thể ($x_1, x_2, \ldots, x_n$), ta có bảng phân phối tần số thực nghiệm dạng ghép lớp:

x_{i-1} - x_i	x_0 - x_1	x_1 - x_2	. . . . . .	x_{k-1} - x_k
n_i	n_1	n_2	. . . . . .	n_k

Nếu H_0 đúng, ta tính được các xác suất: $P(x_{i-1} < x < x_i) = P_i,\ i = \overline{1,k}$

Khi đó, tần số lý thuyết của lớp thứ i là $np_i,\ i = \overline{1,k}$

Ta kiểm định giả thiết H_0 bằng cách thống kê kiểm định $K = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$

Ví dụ 5.8. Kiểm định giả thiết H_0: X phân phối chuẩn $N(\mu, \sigma_2^2)$.

Khi đó $P(x_{i-1} < x < x_i) = \Phi(\frac{x_i - \mu}{\sigma}) - \Phi(\frac{x_{i-1} - \mu}{\sigma})$.

Thông thường ta chưa biết μ và σ nên chúng được thay bằng các ước lượng hợp lý tối đa tương ứng là $\overline{x}$ và $\hat{s}$.

$$P_i \approx \Phi(\frac{x_i - \overline{x}}{\hat{s}}) - \Phi(\frac{x_{i-1} - \overline{x}}{\hat{s}})$$

Nhận xét: Tiêu chuẩn χ^2 được sử dụng tốt khi n đủ lớn ($n > 50$) và $n_i \geq 5,\ \forall i = \overline{1,k}$.

5.9. KIỂM ĐỊNH GIẢ THIẾT VỀ TÍNH ĐỘC LẬP

Xét vectơ ngẫu nhiên 2 chiều (X, Y). Ta kiểm định giả thiết H_0: X và Y độc lập với nhau dựa vào mẫu ngẫu nhiên gồm n cặp quan sát từ vectơ ngẫu nhiên (X,Y)

$$(x_1, y_1), (x_2, y_2), (x_3, y_3)$$

Phân phối tần số thực nghiệm của mẫu được trình bày trên bảng hai chiều như sau:

X_i \ Y_i	y_1	y_2	...	y_j	...	y_s	Σ
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1s}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2.}$
. . .	. . .	.	.	.	.	.	.
x_i	n_{i1}	n_{i2}		n_{ij}	...	n_{is}	$n_{i.}$
. . .	. . .	.	.	.	.	.	.
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rs}	$n_{r.}$
Σ	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.s}$	n

Với X khá lớn thì theo định nghĩa thống kê về xác suất, ta có:

$$\mathrm{P}\{X = x_i; Y = y_j\} \approx \frac{n_{ij}}{n}, \qquad i = \overline{1,r},\ j = \overline{1,s}$$

$$\mathrm{P}(X = x_i) \approx \frac{n_{i.}}{n}$$

$$\mathrm{P}(Y = y_j) \approx \frac{n_{.j}}{n}$$

Nếu H_0 đúng, nghĩa là X, Y độc lập thì

$$\frac{n_{ij}}{n} = \frac{n_{i.}}{n}.\frac{n_{.j}}{n}$$

Thống kê kiểm định được chọn là:

$$K = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - \frac{n_{i.}.n_{.j}}{n}\right)^2}{\frac{n_{i.}.n_{.j}}{n}} = n\left(\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{n_{ij}^2}{n_{i.}.n_{.j}} - 1\right)$$

Nếu giả thiết H_0 đúng và n khá lớn thì thống kê K phân phối theo quy luật χ^2 với

$(r-1)(s-1)$ bậc tự do.

Miền bác bỏ giả thiết mức α: $W_\alpha = (\chi^2_\alpha; +\infty)$.

BÀI TẬP

5.1. **M**ột công ty tuyên bố rằng tuổi thọ trung bình của một loại sản phẩm do công ty sản xuất là 160 giờ **. B**iết rằng tuổi thọ của loại sản phẩm này là biến ngẫu nhiên phân phối chuẩn với độ lêch chuẩn $\sigma = 8$ giờ. Kiểm tra ngẫu nhiên 45 sản phẩm của công ty, ta tính được tuổi thọ trung bình là 152 giờ. Với mức ý nghĩa $\alpha = 0{,}02$, hãy kiểm định tuyên bố của công ty có đúng không?

5. 2. Trong năm trước trọng lượng trung bình trước khi xuất chuồng của bò ở một trại chăn nuôi là 380 kg. Năm nay, người ta áp dụng thử một chế độ chăn nuôi mới với hy vọng là bò sẽ tăng trọng nhanh hơn. Sau một thời gian áp dụng thử, người ta lấy ngẫu nhiên 50 con bò trước khi xuất chuồng đem cân và tính được trọng lượng trung bình của chúng là 390kg. Vậy, với mức ý nghĩa $\alpha = 0{,}01$ có thể cho rằng trọng lượng trung bình của bò trước khi xuất chuồng đã tăng lên hay không? Giả thiết trọng lượng của bò là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 35,2 kg.

5. 3. Nhịp mạch trung bình của nam thanh niên là 72 lần /phút. Kiểm tra 84 nam thanh niên làm việc trong hầm lò thấy nhịp mạch trung bình của họ là 75 lần/ phút với độ lệch mẫu là 9. Hãy xét xem làm việc trong hầm lò có làm tăng nhịp mạch hay không? ($\alpha = 0{,}04$)

5. 4. Trọng lượng đóng bao của các bao gạo trong kho là bién ngẫu nhiên phân phối chuẩn với trọng lượng trung bình theo quy định là 50 kg. Nghi ngờ bị đóng thiếu, người ta đem cân ngẫu nhiên 25 bao và thu được các số liệu sau:

Trọng lượng bao (kg)	Số bao tương ứng
48,0 – 48,5	2
48,5 – 49	5
49,0 – 49,5	10
49,5 – 50,0	6
50,0 – 50,5	2
	$n = 25$

Với mức ý nghĩa $\alpha = 0{,}05$ hãy kết luận về điều nghi ngờ nói trên.

5.5. Đối với người Việt Nam, lượng huyết sắc tố trung bình là 138 g/l. Khám cho 120 công nhân ở nhà máy có tiếp xúc hóa chất thấy huyết sắc tố trung bình là 125,8 g/l và s=15 g/l. Từ kết quả trên có thể kết luận lượng sắc tố trung bình của công nhân nhà máy này thấp hơn mức chung hay không ($\alpha = 0{,}01$)?

5. 6. Năng suất trung bình của một giống lúa là 50 tạ / ha. Sau thời gian dài canh tác, người ta nghi ngờ giống lúa đó bị thoái hóa, năng suất giảm.Dựa vào mẫu kích thước bằng 25 tìm được năng suất trung bình từ mẫu là 48,7 tạ / ha và độ lệch chuẩn mẫu là 4,2 tạ / ha. Hãy kết luận về điều nghi ngờ nói trên với ý nghĩa là $\alpha = 0{,}05$. Cho biết năng suất của giống lúa đó là biến ngẫu nhiên tuân theo quy luật chuẩn.

5. 7. Một kho hạt giống có tỉ lệ nảy mầm xác định $p_0 = 0{,}9$. Ngẫu nhiên có một thiết bị bị hỏng làm thay đổi điều kiện bên trong của kho. Để có thông tin về tỉ lệ nảy mầm mới của kho hạt giống, người ta làm thí nghiệm 400 hạt thấy có 290 hạt nảy mầm. Hỏi tỉ lệ nảy mầm của kho hạt giống có còn giữ nguyên hay không ($\alpha = 0{,}05$)?

5. 8. Trên lý thuyết, hàm lượng dầu của một loại cây bằng 0,5. Sau một thời gian chăm bón, lấy ngẫu nhiên một mẫu gồm 100 cây, thấy có tỉ lệ dầu là 0,7. Với mức ý nghĩa 0,01 có thể kết luận tỉ lệ dầu của loại cây đó đã tăng lên?

5. 9. Tỉ lệ phế phẩm của một nhà máy là 10%. Sau khi cải tiến kỹ thuật, điều tra 400 sản phẩm thấy có 32 phế phẩm. Với α = 0,01 hãy kết luận về việc cải tiến kỹ thuật có làm giảm tỉ lệ phế phẩm hay không?

5. 10. Một nhà máy sản xuất thuốc chữa dị ứng thực phẩm tuyên bố rằng 90% người dùng thuốc thấy thuốc có tác dụng trong vòng 8g. Kiểm tra 200 người bị dị ứng thức phẩm thì thấy trong vòng 8g thuốc làm giảm bớt dị ứng đối với 160 người. Hãy kiểm định xem lời tuyên bố trên của nhà sản xuất có đúng hay không với mức ý nghĩa α = 0,01.

5. 11. Chủ một hãng sản xuất cho biết độ lệch chuẩn của sai số đo của thiết bị là 5 đơn vị. Người ta kiểm tra 19 lần đo cùng một đối tượng thì thấy s^2 = 33. Với α = 0,05, có kết luận gì về ý kiến của chủ hãng?

5.13. Điều tra trên một mẫu gồm 130 thanh niên lứa tuổi 18 ở vùng dân cư A ta tính được chiều cao trung bình là 169 cm; độ lệch chuẩn 8cm. Trên một mẫu 80 thanh niên lứa tuổi 18 của vùng dân cư B thì chiều cao trung bình là 165cm; độ lệch chuẩn 9cm. Với mức ý nghĩa α = 0,05, có thể nói chiều cao trung bình của thanh niên vùng A là cao hơn vùng B hay không?

5.14. Gieo một loại hạt giống theo 2 phương pháp khác nhau:

Phương pháp I: gieo 100 hạt thấy có 82 hạt nẩy mầm.

Phương pháp II: gieo 120 hạt thấy có 92 hạt nẩy mầm.

Với mức ý nghĩa α = 0,01 có thể nói tỉ lệ nẩy mầm của 2 phương pháp là như nhau hay không?

CHƯƠNG VI

TƯƠNG QUAN VÀ HỒI QUY

6.1. Các khái niệm

6.1.1. Hệ số tương quan mẫu

- Giả sử ta có mẫu ngẫu nhiên $\{(X_1,Y_1);\cdots;(X_n,Y_n)\}$ là những quan sát đồng thời về hai biến ngẫu nhiên X,Y. Hệ số tương quan mẫu của X,Y được xác định bởi:

$$R_{X,Y}=\frac{\overline{XY}-\overline{X}.\overline{Y}}{S_X.S_Y}$$

trong đó

$$\overline{XY}=\frac{1}{n}\sum_{i=1}^{n}X_iY_i \qquad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad \overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$$

$$S_X^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2 \qquad S_Y^2=\frac{1}{n}\sum_{i=1}^{n}Y_i^2-\overline{Y}^2$$

Ý nghĩa: Hệ số tương quan được dùng để đo mức độ phụ thuộc tuyến tính giữa 2 biến ngẫu nhiên. Nếu $|r|$ càng lớn (càng gần 1) thì sự phụ thuộc tuyến tính càng rõ (khả năng xảy ra phụ thuộc tuyến tính giữa X, Y càng cao). Nếu $|r|$ càng gần 0 thì sự phụ thuộc tuyến tính càng yếu (khả năng xẩy ra phụ thuộc tuyến tính giữa X, Y càng thấp).

- Đối với một mẫu cụ thể và nhiều số liệu, ta thực hiện các bước tính hệ số tương quan mẫu như sau:

  + Gọi r là số các giá trị khác nhau của mẫu $(x_1,x_2,\cdots,x_n)$ với $x_{(1)}<x_{(2)}<\cdots<x_{(r)}$.

  + Gọi s là số các giá trị khác nhau của mẫu $(y_1,y_2,\cdots,y_n)$ với $y_{(1)}<y_{(2)}<\cdots<y_{(s)}$.

  + Gọi n_{ij} là số các quan sát từ mẫu ghép có giá trị $(x_{(i)},y_{(j)})$ $(i=1..r, j=1..s)$.

  + Tính các giá trị $n_{i.}=\sum_{j=1}^{s}n_{ij}$; $n_{.j}=\sum_{i=1}^{r}n_{ij}$; $n=\sum_{i=1}^{r}\sum_{j=1}^{s}n_{ij}$

- Khi đó

$$\overline{XY}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}n_{ij}x_iy_i \qquad \overline{X}=\frac{1}{n}\sum_{i=1}^{r}n_{i.}x_i \qquad \overline{Y}=\frac{1}{n}\sum_{j=1}^{s}n_{.j}y_j$$

$$S_X^2=\frac{1}{n}\sum_{i=1}^{r}n_{i.}x_i^2-\overline{X}^2 \qquad S_Y^2=\frac{1}{n}\sum_{j=1}^{s}n_{.j}y_j^2-\overline{Y}^2$$

6.1.2. Kỳ vọng có điều kiện

- Giả sử có vectơ ngẫu nhiên 2 chiều (X,Y)

- Kỳ vọng có điều kiện của biến ngẫu nhiên Y với điều kiện $X=x$, ký hiệu $E(Y/x)$, xác định bởi:

  + Nếu Y là rời rạc thì

$$E(Y/x)=\sum_{i=1}^{m} y_j P(X=x, Y=y_j)$$

+ Nếu Y liên tục thì

$$E(Y/x)=\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dy$$

6.1.3. Hàm hồi quy

- Hàm hồi quy của Y đối với X là $f(x)=E(Y/x)$.
- Hàm hồi quy của X đối với Y là $g(y)=E(X/y)$.

Ý nghĩa: Hàm hồi quy của Y đối với X có thể dùng để dự báo kì vọng giá trị của Y với một giá trị nào đó của X.

6.2. Hồi quy tuyến tính đơn

- Hai biến ngẫu nhiên X, Y gọi là có hồi quy tuyến tính nếu hàm hồi quy của Y đối với X có dạng:

$$y=E(Y/x)=ax+b$$

6.2.1. ước lượng các hệ số a,b

- Giả sử ta có mẫu cụ thể $\{(x_1,y_1);\cdots;(x_n,y_n)\}$ là những quan sát đồng thời về 2 biến ngẫu nhiên X, Y. Dựa vào phương pháp bình phương bé nhất, tức là tìm a, b sao cho hàm

$$F(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2$$

đạt cực tiểu. Khi đó ta có các ước lượng sau:

$$a=R_{X,Y}\cdot\frac{S_Y}{S_X} \qquad b=\overline{Y}-a\overline{X}$$

Bài tập chương 6

6.1. Số vi khuẩn X (triệu) sinh sản sau thời gian Y (giờ) được ghi lại trong bảng sau qua một thí nghiệm

X	0	1	2	3	4	5	6	7	8
Y	30	32	35	40	48	52	58	62	69

a) Xác định hệ số tương quan giữa X và Y.
b) Tìm đường hồi quy tuyến tính của X theo Y . Dự báo về số vi khuẩn có sau 10 giờ.

6.2. Đo chiều cao X (cm) và trọng lượng Y (kg) của 100 học sinh, ta được kết quả sau:

Y X	145 - 150	150 - 155	155 - 160	160 - 165	165 - 170
35 - 40	3				
40 - 45	5	10			
45 - 50		14	20	6	
50 - 55			15	12	5
55 - 60				6	4

c) Xác định hệ số tương quan giữa chiều cao và cân nặng.
d) Tìm đường hồi quy tuyến tính của chiều cao theo cân nặng.

6.3. Theo dõi lượng phân bón X (kg/ha) và năng suất lúa Y (tấn/ha) của 100 ha lúa ở một vùng, ta thu được kết quả sau:

Y X	120	140	160	180	200
2,2	2				
2,6	5	3			
3,0		11	8	4	
3,4			15	17	
3,8			10	6	7
4,2					12

e) Xác định hệ số tương quan giữa lượng phân bón và năng suất lúa.
f) Tìm đường hồi quy tuyến tính của năng suất lúa theo lượng phân bón.

6.4. Đo chiều cao X (m) và đường kính Y (cm) của một loại cây, ta được kết quả sau:

Y X	6	8	10	12	14
30	2	17	9	3	
35		10	17	9	
40		3	24	16	13
45			6	24	12
50			2	11	22

g) Xác định hệ số tương quan giữa chiều cao và đường kính của cây.
h) Tìm đường hồi quy tuyến tính của đường kính theo chiều cao.