XÁC SUẤT – THỐNG KÊ

ĐẠI HỌC TÔN ĐỨC THẮNG
TON DUC THANG UNIVERSITY

ĐỀ CƯƠNG:

- Chương 1: KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT.
- Chương 2: BIẾN NGẪU NHIÊN VÀ CÁC QUY LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG.
- Chương 3: LÝ THUYẾT MẪU.
- Chương 4: ƯỚC LƯỢNG THAM SỐ THỐNG KÊ.
- Chương 5: KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ.

Bổ túc về giải tích tổ hợp
Quy tắc cộng -1

Quy tắc cộng: Giả sử một công việc nào đó được chia làm 2 trường hợp:

Trường hợp 1 có n_1 cách thực hiện, trường hợp 2 có n_2 cách thực hiện. Và không có một cách thực hiện nào ở trường hợp này trùng ở trường hợp kia. **Khi đó, số cách thực hiện công việc đã cho là**: $\boldsymbol{n = n_1 + n_2}$ cách

Quy tắc cộng -2

Giả sử một công *việc nào đó được chia làm k trường hợp:*

- TH1: có n_1 cách thực hiện
- TH2: có n_2 cách thực hiện
- . . .
- THk: có n_k cách thực hiện
- Vậy sẽ có $\boldsymbol{n = n_1 + n_2 + \ldots + n_k}$ cách thực hiện xong công việc.

Quy tắc cộng -3

Ví dụ 1:

Trong một lớp học có 50 nam sinh viên và 40 nữ sinh viên. Giảng viên muốn gọi một sinh viên để kiểm tra bài. Có bao nhiêu cách gọi một sinh viên?

Ví dụ 2:

Một hộp đựng sản phẩm, trong đó có 5 phế phẩm và 15 chính phẩm. Lấy một sản phẩm từ hộp. Có bao nhiêu cách lấy một sản phẩm?

Quy tắc nhân -1

Qui tắc nhân:

Giả sử một công việc nào đó được chia làm 2 giai đoạn:

Giai đoạn 1 có n_1 cách thực hiện,

Giai đoạn 2 có n_2 cách thực hiện.

Khi đó, số cách thực hiện công việc đã cho là:

$$\boldsymbol{n = n_1 \times n_2} \text{ cách}$$

Quy tắc nhân -2

Giả sử một công *việc nào đó được chia làm k giai đoạn:*

- Giai đoạn 1: Có $n_1$cách thực hiện
- Giai đoạn 2: Có n_2 cách thực hiện
- . . .
- Giai đoạn k: Có n_k cách thực hiện
- Vậy sẽ có $n = n_1 \times n_2 \times \cdots \times n_k$ cách thực hiện xong công việc.

Các ví dụ quy tắc nhân -1

Ví dụ 1:

Một hộp đựng sản phẩm, trong đó có 5 phế phẩm và 15 chính phẩm. Lấy ngẫu nhiên 2 sản phẩm từ hộp. Có bao nhiêu cách lấy 2 sản phẩm trong đó có một phế phẩm và một chính phẩm?

Các ví dụ quy tắc nhân -2

Ví dụ 3: Cho tập số A = {1,2,3,4,5,6,7,8,9} .

a. Có bao nhiêu số gồm 5 chữ số?

b. Có bao nhiêu số gồm 5 chữ số khác nhau?

c. Có bao nhiêu số chẵn gồm 5 chữ số?

d. Có bao nhiêu số chẵn gồm 5 chữ số khác nhau?

e. Các số được lấy từ tập A.

Hoán vị -1

Cho tập gồm n phần tử. Mỗi cách sắp xếp n phần tử vào n vị trí gọi là một hoán vị. Số cách sắp xếp n phần tử vào n vị trí chính là số hoán vị. Số hoán vị n phần tử được ký hiệu là: $P_n = n!$

- $n!$ đọc là n giai thừa.
- Trong đó $n! = 1 \times 2 \times 3 \times \ldots \times n$ và $0! = 1$.

Hoán vị -2

Ví dụ 1: Có bao nhiêu số gồm 5 chữ số khác nhau được lấy từ tập { 7, 3, 4, 5,6 }?

Ví dụ 2: Có 5 người. Hỏi có bao nhiêu cách sắp xếp 5 người này thành:

a. Một hàng dài

b. Bàn tròn có đánh số.

Chỉnh hợp -1

Cho tập gồm n phần tử. Một nhóm gồm k phần tử lấy từ tập và sắp xếp theo một thứ tự nào đó gọi là một chỉnh hợp chập k của n phần tử. Số chỉnh hợp chập k của n phần tử được ký hiệu là A_n^k và được xác định theo công thức sau:

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

Chỉnh hợp -2

Một lớp học có 30 sinh viên trong đó có 20 sinh viên nam. Có bao nhiêu cách chọn ra một ban cán sự gồm 1 lớp trưởng, một lớp phó và một thủ quỹ. Nếu:

a. Chọn bất kỳ

b. Lớp trưởng là nam

c. Toàn là nữ

d. Ít nhất một nam

e. Nhiều nhất một nam.

Tổ hợp -1

Cho một tập gồm n phần tử. Một nhóm gồm k phần tử lấy ra từ tập trên không kể thứ tự gọi là một tổ hợp chập k của n phần tử ($k \leq n$). Số cách chọn k phần tử từ n phần tử trên gọi là số tổ hợp n chập k. Ký hiệu là C_n^k .

- **Công thức**: $C_n^k = \frac{n!}{(n-k)!k!}$

Tổ hợp -2

Ví dụ 1: Trong một hội nghị, các đại biểu bắt tay nhau từng đôi một.

a. Nếu có 12 đại biểu thì có bao nhiêu cái bắt tay?

b. Nếu đếm được 190 cái bắt tay thì hội nghị có bao nhiêu đại biểu?

Tổ hợp -3

Ví dụ 2:

Một hộp đựng 10 sản phẩm, trong đó có 3 hỏng. Lấy ngẫu nhiên 3 sản phẩm ra để kiểm tra. Có bao nhiêu cách lấy ra 3 sản phẩm, biết rằng:

a. 3 sản phẩm cùng loại

b. Có đúng một sản phẩm hỏng

c. Ít nhất một sản phẩm hỏng

d. Nhiều nhất 1 sản phẩm hỏng

I. Phép thử, không gian mẫu và biến cố

1. Khái niệm phép thử, không gian mẫu và biến cố

- *Phép thử* là việc thực hiện những điều kiện đã đặt ra để nghiên cứu một sự vật hay hiện tượng nào đó.
- *Không gian mẫu, ký hiệu Ω,* là tập tất cả các kết quả có thể có của phép thử. Tập *$A \subset \Omega$, A gọi là biến cố*.

Các loại biến cố

a. Biến cố chắc chắn

Là biến cố nhất định xảy ra khi thực hiện phép thử, ký hiệu là Ω.

- *Ví dụ 1:* Tung con súc sắc.

Gọi Ω là biến cố ''xuất hiện mặt có số chấm nhỏ hơn 7''. Khi đó Ω là một biến cố chắc chắn.

Ví dụ 2:

Một hộp đựng 10 quả cầu trắng. Từ hộp lấy ngẫu nhiên 1 quả cầu.

Gọi Ω là biến cố lấy được quả cầu trắng.

Khi đó Ω là biến cố chắn chắn.

b. Biến cố không thể (ký hiệu Φ)

Là biến cố nhất định không xảy ra khi thực hiện phép thử.

- *Ví dụ*:

Một hộp đựng 10 viên bi, trong đó có 4 bi đỏ, 5 bi xanh và còn lại là bi trắng. Chọn ngẫu nhiên từ hộp ra 2 bi.

Gọi **Φ** là biến cố lấy ra 2 bi trắng.

Khi đó **Φ** là biến cố không thể.

c. Biến cố ngẫu nhiên

Là biến cố có thể xảy ra hoặc không xảy ra khi thực hiện phép thử. Thường dùng các chữ cái A, B, C.,. . . để ký hiệu biến cố ngẫu nhiên.

Ví dụ 1:

Tung con súc sắc. Gọi A là biến cố xuất hiện mặt một chấm. Khi đó A là biến cố ngẫu nhiên.

Ví dụ 2 :

Lấy một sản phẩm từ hộp đựng sản phẩm để kiểm tra. Gọi B là biến cố lấy được chính phẩm. Khi đó, B được gọi là biến cố ngẫu nhiên.

d. Biến cố tổng

Biến cố C được gọi là biến cố tổng của hai biến cố A và B, ký hiệu là $C = A \cup B$ $(C = A + B)$.

Biến cố C xảy ra khi và chỉ khi có ít nhất một trong hai biến cố trên xảy ra.

- ***Tương tự, có thể mở rộng cho tổng của n biến cố:***

Ví dụ 1:

Có 2 xạ thủ cùng bắn vào một tấm bia

Gọi A là biến cố người thứ I bắn trúng

B là biến cố người thứ II bắn trúng

C là biến cố tấm bia trúng đạn

Khi đó $C = A \cup B$?

Ví dụ 2:

Có hai hộp đựng thuốc, mỗi hộp đều có 1 số lọ hỏng. Từ mỗi hộp lấy ra 1 lọ.

Gọi A là biến cố lấy được lọ hỏng từ hộp I

Gọi B là biến cố lấy được lọ hỏng từ hộp II

Khi đó biến cố C được gọi như thế nào để có $C = A \cup B$?

e. Biến cố tích

Biến cố C được gọi là biến cố tích của hai biến cố A và B, ký hiệu là $C = A \cap B$ $(hoặc\ C = AB)$

$Biến\ cố$ C xảy ra khi và chỉ khi A và B đồng thời xảy ra.

- *Tương tự, có thể mở rộng cho tích n biến cố*.

Ví dụ 1:

Có 2 xạ thủ cùng bắn vào một tấm bia.

Gọi A là biến cố người thứ nhất bắn không trúng

Gọi B là biến cố người thứ 2 bắn không trúng

Gọi C là biến cố tấm bia không trúng đạn.

- Khi đó ta có $C = A \cap B$.

Ví dụ 2:

Có hai hộp đựng thuốc, mỗi hộp đều có 1 số lọ hỏng. Từ mỗi hộp lấy ra 1 lọ.

Gọi A là biến cố lấy được lọ hỏng từ hộp I ;

Gọi B là biến cố lấy được lọ hỏng từ hộp II

Gọi C là biến cố 2 lọ lấy ra đểu hỏng

Khi đó $C = A \cap B$

f. Biến cố xung khắc

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

Ví dụ 1:

Tung con súc sắc.

Gọi $A_i \, (i = \overline{1,6})$ là biến cố mặt có i chấm xuất hiện. Khi đó các cặp A_i và A_j xung khắc nhau với i khác j.

Ví dụ 2:

Một hộp đựng 10 viên bi, trong đó có 4 bi đỏ, 5 bi xanh và còn lại là bi trắng. Chọn ngẫu nhiên từ hộp ra 2 bi.

Gọi A là biến cố lấy được 2 bi đỏ

Gọi B là biến cố lấy được 2 bi xanh

Gọi C là biến cố lấy được 1 bi đỏ, 1 xanh.

Khi đó các biến cố xung khắc từng đôi.

Ví dụ 3:

- Có 2 hộp đựng bi như trên. Từ mỗi hộp lấy ra 1 bi.

  Gọi A là biến cố lấy được bi đỏ từ hộp I

  Gọi B là biến cố lấy được bi đỏ từ hộp II

  Hỏi A và B có xung khắc không?

g. Biến cố đối lập

Biến cố "không xảy ra biến cố A", ký hiệu là $\bar{A}$, gọi là biến cố đối lập với biến cố A.

Như vậy, trong phép thử nhất định có một và chỉ một biến cố A hoặc $\bar{A}$ xảy ra, tức là $A \cap \bar{A} = \emptyset$ và $A \cup \bar{A} = \Omega$

Ví dụ 1:

Một hộp đựng 10 sản phẩm, trong đó có 3 phế phẩm. Lấy ngẫu nhiên từ hộp 1 sản phẩm. Gọi A là biến cố lấy được phế phẩm.

Khi đó, biến cố $\bar{A}$ là biến cố . . .

Ví dụ 2:

Một hộp đựng 10 viên bi, trong đó có 4 bi đỏ, 5 bi xanh và còn lại là bi trắng. Lấy ngẫu nhiên từ hộp ra 5 bi.

Gọi A là biến cố trong 5 bi lấy ra có 4 bi đỏ.

Khi đó biến cố $\bar{A}$ là biến cố . . .

Lưu ý: Hai biến cố đối lập thì xung khắc nhưng ngược lại thì chưa chắc đúng.

Ví dụ 1: Tung một con súc sắc.

- Tìm 2 biến cố đối lập.
- Tìm 2 biến cố xung khắc nhưng không đối lập.

Ví dụ 2: Có 3 thí sinh đi thi môn XSTK.

Gọi A_i $(i = \overline{1,3})$ là biến cố thí sinh thứ i thi đậu.

Hãy biểu diễn các biến cố sau đây qua các biến cố A_i

a. Có 3 thí sinh thi đậu
b. Có 2 thí sinh thi đậu
c. Có đúng một thí sinh thi đậu
d. Có ít nhất một thí sinh thi đậu
e. Không có thí sinh thi đậu

h. Biến cố đồng khả năng và biến cố sơ cấp

Biến cố đồng khả năng:

Là các biến cố có khả năng khách quan xảy ra như nhau.

Biến cố sơ cấp:

Là biến cố không thể biểu diễn thành tổng (tích) các biến cố khác.

- Ví dụ: Tung con súc sắc.

k. Nhóm biến cố đầy đủ

Nhóm biến cố gọi là đầy đủ nếu chúng xung khắc từng đôi và có một trong các biến cố đó chắc chắn xảy ra.Tức là:

Nhóm $A_1, A_2, \ldots, A_n$ là *nhóm biến cố đầy đủ* khi và chỉ

khi $\begin{cases} A_i \cap A_j = \Phi \ (xung\ khắc) \\ A_1 \cup A_2 \cup \ldots \cup A_n = \Omega \ (biến\ cố\ chắc\ chắn) \end{cases}$

Ví dụ 1:

Sản phẩm X do 3 nhà máy sản xuất ra. Người tiêu dùng mua ngẫu nhiên 1 sản phẩm.

Gọi $A_i(i = \overline{1,3})$ là biến cố mua được sản phẩm do nhà máy i sản xuất.

Khi đó, nhóm các biến cố $A_i(i = \overline{1,3})$ là nhóm biến cố đầy đủ.

Ví dụ 2:

Từ 3 hộp đựng sản phẩm, chọn ngẫu nhiên một hộp. Từ đó lấy ra một sản phẩm để kiểm tra.

Trường hợp 1. Nếu gọi $A_i(i = \overline{1,3}\ )$ là biến cố chọn được hộp thứ i.

Trường hợp 2: Nếu gọi $B_i(i = \overline{1,3}\ )$ là biến cố lấy được chính phẩm từ hộp i.

- Hỏi nhóm biến cố nào là nhóm biến cố đầy đủ?

II. Các định nghĩa xác suất :

1. Định nghĩa xác suất theo lối cổ điển :

Nếu phép thử có n biến cố sơ cấp đồng khả năng xảy ra, trong đó có m biến cố sơ cấp thuận lợi cho biến cố A. Khi đó, xác suất của biến cố A , ký hiệu $P(A)$ được định nghĩa :

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

Ví dụ 1 :

Một lớp học có 35 học sinh, trong đó gồm 10 học sinh giỏi và còn lại là học sinh khá. Chọn ngẫu nhiên 2 học sinh từ lớp đó. Tính xác suất:

a. Hai học sinh đều giỏi

b. Hai học sinh đều khá

c. Một học sinh giỏi và một học sinh khá.

Ví dụ 2 :

Một thí sinh đi thi chỉ làm được 18 câu trong tổng số 30 câu hỏi. Đề thi có 5 câu được lấy từ 30 câu hỏi trên. Nếu thí sinh làm được từ 4 câu trở lên thì sẽ đậu. Ngược lại thì rớt. Tính xác suất để thí sinh đó:

a. Làm được 5 câu

b. Thí sinh đó sẽ thi đậu

c. Thí sinh đó thi rớt.

Ví dụ 3:

Một hộp đựng 15 viên bi, trong đó có 4 bi đỏ, 5 bi xanh và còn lại là bi trắng. Chọn ngẫu nhiên từ hộp ra 2 bi. Tính xác suất để được 2 bi cùng màu.

Hạn chế của định nghĩa xác suất theo lối cổ điển:

+ Chỉ xét cho trường hợp hữu hạn biến cố sơ cấp đồng khả năng;

+ Đồng khả năng không phải lúc nào cũng thỏa.

2.Định nghĩa xác suất bằng tần suất

Nếu ta lặp lại phép thử n lần, ta thấy biến cố A xảy ra m lần thì m gọi là tần số của biến cố A. Còn $\frac{m}{n}$ được gọi là tần suất của biến cố A trong n phép thử.

Khi n đủ lớn, tần suất dao động quanh một số không đổi p, thì số này gọi là xác suất của biến cố A.

Tức là, khi n đủ lớn thì $P(A) \approx \frac{m}{n}$

Ví dụ :

Kiểm tra lần lượt 100 sản phẩm, thấy có 10 phế phẩm. Gọi A là biến cố lấy được phế phẩm. Vậy xác suất lấy được phế phẩm là:

$$P(A) \approx \frac{m}{n} = \frac{10}{100}$$

3.Tính chất của xác suất

a. $0 \leq P(A) \leq 1, \quad \forall A$

b. $P(\Omega) = 1; P(\emptyset) = 0$

4.Ý nghĩa của xác suất

III. Các công thức tính xác suất

1. Công thức cộng:

a. Công thức 1:

Nếu A và B là hai biến cố xung khắc thì:

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

- ***Có thể mở rộng cho trường hợp n biến cố:***

$$P(A_1 \cup A_2 \cup \cdots A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

nếu các biến cố xung khắc từng đôi.

Ví dụ 1:

Một hộp đựng 15 viên bi, trong đó có 4 bi đỏ, 5 bi xanh và còn lại là bi trắng. Chọn ngẫu nhiên từ hộp ra 2 bi. Tính xác suất để được 2 bi cùng màu.

Ví dụ 2:

Một lô hàng chứa 15 sản phẩm gồm 10 sản phẩm tốt và 5 sản phẩm xấu. Chọn ngẫu nhiên từ lô hàng ra 4 sản phẩm. Tính xác suất để trong 4 sản phẩm chọn ra có ít nhất 1 sản phẩm xấu.

b. Hệ quả

Với A là một biến cố bất kỳ, ta có:

$$P(\bar{A}) = 1 - \mathrm{P(A)}$$

Ví dụ:

Dựa vào hệ quả, hãy giải lại ví dụ 2:

Một lô hàng chứa 15 sản phẩm gồm 10 sản phẩm tốt và 5 sản phẩm xấu. Chọn ngẫu nhiên từ lô hàng ra 4 sản phẩm.

Tính xác suất để trong 4 sản phẩm chọn ra có ít nhất 1 sản phẩm xấu.

c. Công thức cộng xác suất

Nếu A và B là hai biến cố không xung khắc thì:

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

Mở rộng cho trường hợp 3 biến cố:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

Ví dụ:

Một lớp học có 100 sinh viên, trong đó có 60 sinh viên giỏi Toán, 70 sinh viên giỏi Anh văn và 40 sinh viên giỏi cả Toán và Anh văn. Chọn ngẫu nhiên một sinh viên của lớp.

Tính xác suất để chọn được sinh viên giỏi ít nhất một trong hai môn Toán hoặc Anh văn.

2. Công thức xác suất có điều kiện và công thức nhân

a. Công thức XS có điều kiện

Xác suất có điều kiện của biến cố A khi biết biến cố B đã xảy ra, ký hiệu là P(A/B), là xác suất của biến cố A nhưng được tính trong trường hợp biến cố B đã xảy ra rồi và được xác định như sau:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Ví dụ:

Một hộp có 10 quả cầu, trong đó có 4 quả cầu đỏ và 6 quả cầu xanh. Lần thứ nhất lấy ra 2 quả cầu, 2 quả cầu lấy ra không bỏ trở lại hộp. Lần thứ 2 cũng lấy ra 2 quả cầu.

Tính xác suất để lần thứ hai lấy được 2 quả cầu đỏ biết rằng lần thứ nhất lấy được 2 quả cầu xanh.

b.Công thức nhân :

Công thức nhân thứ 1:

Từ công thức xác suất có điều kiện, ta có:

Với A và B là hai biến cố bất kỳ, thì

$$P(A \cap B) = P(\mathrm{A}).P(B/A) = P(B).P(A/B)$$

Mở rộng cho trường hợp n biến cố:

$$P(A_1 \cap A_2 \ldots \cap A_n)$$

$$= P(A_1)P(A_2/A_1) \ldots P(A_\mathrm{n}/A_1 A_2 \ldots A_{\mathrm{n}-1})$$

Ví dụ 1:

Có một hộp đựng 10 sản phẩm, trong đó có 3 sản phẩm hỏng. Lấy ngẫu nhiên lần lượt từ hộp 2 sản phẩm để kiểm tra (không hoàn lại).

a. Tính xác suất để cả 2 đều là hỏng;

b. Tính xác suất để lần thứ nhất hỏng, lần thứ hai không hỏng;

c. Có một sản phẩm hỏng.

Ví dụ 2:

Một lô hàng chứa 10 sản phẩm, trong đó có 2 sản phẩm xấu và 8 sản phẩm tốt. Khách hàng kiểm tra bằng cách lấy ra từng sản phẩm cho đến khi nào được sản phẩm xấu thứ hai thì dừng lại.

a. Tính xác suất để khách hàng dừng lại ở lần kiểm tra thứ 3;

b. Biết rằng khách hàng dừng lại ở lần kiểm tra thứ 3, tính xác suất để lần thứ I gặp sản phẩm xấu.

Công thức nhân thứ 2

Tính độc lập:

Hai biến cố A và B được gọi là độc lập nhau nếu biến cố B có xảy ra hay không cũng không ảnh hưởng đến xác suất của biến cố A , tức là P(A/B)= P(A).

Công thức nhân thứ 2:

Nếu hai biến cố A và B độc lập nhau thì:

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

Mở rộng cho trường hợp n biến cố độc lập từng đôi:

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

Ví dụ 1:

Hai xạ thủ cùng bắn vào một mục tiêu. Mỗi người bắn một viên đạn. Xác suất người thứ nhất bắn trúng là 0,8, người thứ hai bắn trúng là 0,7.

Tính xác suất để tấm bia trúng đạn.

Ví dụ 2:

Có hai hộp đựng thuốc. Hộp thứ nhất có 10 lọ thuốc, trong đó có 2 lọ hỏng. Hộp thứ hai 15 lọ, trong đó có 3 lọ hỏng. Từ mỗi hộp lấy ra một lọ. Hãy tính xác suất để :

a. Có 2 lọ hỏng

b. Không có lọ hỏng nào;

c. Có đúng một lọ hỏng;

d. Ít nhất một lọ hỏng.

Ví dụ 3:

Có 2 hộp đựng thuốc, hộp I có 10 lọ, trong đó có 3 hỏng; hộp II có 12 lọ, trong đó có 4 hỏng. Từ mỗi hộp lấy ra 2 lọ.

a. Tính xác suất để có 4 hỏng;

b. Tính xác suất để có 4 lọ cùng loại;

c. Tính xác suất để có đúng 2 lọ hỏng;

d. Tính xác suất có ít nhất 1 lọ hỏng.

3.Công thức xác suất đầy đủ và Bayes

a. Công thức xác suất đầy đủ (toàn phần)

Cho $A_1, A_2, \dots, A_n$ là nhóm biến cố đầy đủ, và A là một biến cố bất kỳ có thể xảy ra trong phép thử. Khi đó ta có công thức xác suất của biến cố A:

$$P(A)$$
$$= P(A_1)P(A/A_1) + P(A_2)P(A/A_2) + \cdots + P(A_n)P(A/A_n)$$

gọi là công thức xác suất toàn phần (đầy đủ).

b. Công thức xác suất Bayes

$$P(A_i/A) = \frac{P(A_i)P(A/A_i)}{P(A)}$$

Ví dụ 1:

Có 12 hộp sản phẩm cùng loại, mỗi hộp chứa nhiều sản phẩm, trong đó có 4 hộp sản phẩm do xí nghiệp I sản xuất, 5 hộp sản phẩm do xí nghiệp II sản xuất và còn lại là do xí nghiệp III sản xuất. Biết rằng tỷ lệ sản phẩm tốt của hộp do XN I, II và III sản xuất lần lượt là 60%, 70% và 75%.

Lấy ngẫu nhiên một hộp và từ đó lấy ngẫu nhiên một sản phẩm.

a. Tính xác suất để lấy được sản phẩm tốt;

b. Giả sử lấy được sản phẩm tốt, theo bạn sản phẩm đó thuộc xí nghiệp nào?

Ví dụ 2:

Có 2 hộp đựng thuốc, hộp I có 10 lọ, trong đó có 3 hỏng; hộp II có 12 lọ, trong đó có 4 hỏng. Người ta chọn 1 hộp ngẫu nhiên bằng cách gieo con súc sắc. Nếu mặt một chấm xuất hiện thì chọn hộp I, ngược lại chọn hộp II.Từ hộp được chọn, lấy ra 1 lọ.

a. Tính xác suất để được lọ tốt.

b. Giả sử lấy được lọ tốt, tính xác suất để lọ tốt lấy từ hộp I.

Ví dụ 3:

Có 3 hộp đựng sản phẩm, hộp I có 10 sản phẩm, trong đó có 2 hỏng; hộp II có 15 sản phẩm, trong đó có 3 hỏng; hộp III có 20 sản phẩm trong đó có 5 hỏng. Từ mỗi hộp lấy ra 1 sản phẩm.

a. Tính xác suất để có 2 sản phẩm hỏng.

b. Giả sử lấy ra có 2 sản phẩm hỏng. Tính xác suất để sản phẩm lấy từ hộp I là tốt.

Ví dụ 4:

Một nhà máy sản xuất bóng đèn gồm 3 máy. Máy I sản xuất 25% bóng; máy II sản 35% và máy III sản xuất 40% bóng đèn. Tỷ lệ bóng hỏng do mỗi máy sản xuất ra lần lượt là 3%, 2% và 1%. Một người mua một bóng đèn do nhà máy đó sản xuất.

a. Tính xác suất để mua được bóng tốt;
b. Biết rằng bóng mua được là bóng hỏng. Tính xác suất để bóng đó do nhà máy III sản xuất.

Ví dụ 5:

Có 2 chuồng thỏ. Chuồng I có 5 con thỏ đen và 10 con thỏ trắng. Chuồng II có 3 con thỏ đen và 7 con thỏ trắng. Từ chuồng I bắt một con thỏ cho vào chuồng II và sau đó lại bắt ngẫu nhiên một con thỏ ở chuồng II .

a. Tính xác suất để từ chuồng II bắt ra con thỏ trắng.

b. Giả sử từ chuồng II bắt được con thỏ trắng. Tính xác suất đề từ chuồng I bắt ra con thỏ trắng.

I. Biến ngẫu nhiên (BNN)

1. Định nghĩa

Xét phép thử T và tập tất cả các biến cố sơ cấp, ký hiệu Ω. Giả sử, ứng với mỗi biến cố sơ cấp $\omega \in \Omega$, ta liên kết một số thực $X(\omega) \in R$, thì X được gọi là một biến ngẫu nhiên.

Hay X là một ánh xạ: $\Omega \rightarrow R$

$$\omega \epsilon \Omega \rightarrow X(\omega)$$

Ví dụ :

Trong một cuộc xổ số được tiến hành để giúp quỹ từ thiện địa phương, 10.000 vé số sẽ được bán với giá 5USD mỗi vé. Giải thưởng là một chiếc xe ôtô trị giá 12.000USD. Giả sử một người mua hai vé số. Gọi X là khoản lợi (gain) của người mua. Khi đó, X là biến ngẫu nhiên.

2.Phân loại

– Biến ngẫu nhiên mà chỉ nhận hữu hạn hoặc vô hạn đếm được các giá trị “cách quãng” nhau gọi là biến *ngẫu nhiên rời rạc*, tức là $X(\Omega)$ có dạng:

$X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ hoặc $X(\Omega) = \{x_1, x_2, \ldots, x_n, \ldots\}$

– Biến ngẫu nhiên mà giá trị có thể có của nó lấp đầy một khoảng nào đó trên trục số gọi là *biến ngẫu nhiên liên tục*.

– Thường dùng các chữ cái X, Y, Z, .. . để ký hiệu biến ngẫu nhiên.

Ví dụ 1:

Lấy 5 sản phẩm từ hộp có 15 sản phẩm, trong đó có 3 sản phẩm hỏng. Gọi X là số sản phẩm hỏng trong 5 sản phẩm lấy ra.

Khi đó, X là BNN rời rạc.

Ví dụ 2:

Quan sát số người vào mua hàng tại một cửa hàng trong một ngày. Gọi Y là " số người vào mua hàng tại cửa hàng trên trong một ngày". Khi đó, Y là biến ngẫu nhiên rời rạc.

Ví dụ 3:

Gọi X là khoảng cách từ điểm chạm của viên đạn đến tấm bia.

Khi đó, X là biến ngẫu nhiên . . .

Ví dụ 4:

Hoặc gọi Y là chiều cao của sinh viên; T là tuổi thọ của con người. Khi đó, Y, T các là biến ngẫu nhiên .

II. Quy luật phân phối xác suất của biến ngẫu nhiên

1. Biến ngẫu nhiên rời rạc

Để xác định một biến ngẫu nhiên rời rạc X, người ta cần xác định các giá trị $x_i, i = 1,2,\ldots$ mà biến ngẫu nhiên X có thể nhận được và đồng thời xác định xác suất để X nhận giá trị này là bao nhiêu, tức là xác định $P(X = x_i)$ với $i = 1,2,\ldots$

a. Bảng phân phối xác suất

Giả sử tập giá trị biến ngẫu nhiên rời rạc X có dạng $X = \{x_1, x_2, \ldots, x_n\}$ với $p_i = P(X = x_i)$ và được biểu diễn dưới dạng bảng như sau:

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

Bảng trên được gọi là bảng phân phối xác suất của biến ngẫu nhiên X nếu:

$$0 \leq p_i \leq 1 \text{ và } \sum_{i=1}^{n} p_i = 1$$

Ví dụ 1:

Một hộp có 10 sản phẩm trong đó có 6 chính phẩm. Lấy ngẫu nhiên 2 sản phẩm. Xây dựng bảng phân phối xác suất của số chính phẩm trong 2 sản phẩm lấy ra.

Ví dụ 2:

Trong một cuộc xổ số được tiến hành để giúp quỹ từ thiện địa phương, 10.000 vé số sẽ được bán với giá 5USD mỗi vé. Giải thưởng là một chiếc xe ôtô trị giá 12.000USD. Giả sử một người mua hai vé số. Gọi X là khoản lợi (số tiền thu được) của người mua. Lập bảng phân phối xác suất của X.

Ví dụ 3:

Có 3 xạ thủ cùng bắn vào 1 tấm bia. Mỗi người bắn 1 viên đạn. Xác suất ba xạ thủ bắn trúng lần lượt là 0,5 ; 0,6 ; 0,7. Gọi X là số viên đạn bắn trúng. Lập bảng phân phối xác suất của X.

Ví dụ 4:

Trong một lô 10 sản phẩm có 7 sản phẩm loại A. Lấy ngẫu nhiên cùng lúc 5 sản phẩm để kiểm tra. Gọi X là số sản phẩm loại A gặp khi kiểm tra.

Lập bảng phân phối xác suất của X.

Ví dụ 5:

Một hộp đựng 10 lọ thuốc, trong đó có 4 lọ hỏng. Kiểm tra từng lọ (không hoàn lại) cho đến khi gặp lọ tốt thì dừng. Gọi X là số lần kiểm tra (số lọ lấy ra).

Lập bảng phân phối xác suất của X.

Ví dụ 6:

Có 2 kiện hàng. Kiện I có 12 sản phẩm trong đó có 10 sản phẩm loại 1, kiện II có 15 sản phẩm trong đó có 12 sản phẩm loại 1.

a. Từ mỗi kiện lấy ra 2 sản phẩm, nếu 2 sản phẩm đều là loại 1 thì mua kiện đó. Gọi X là số kiện được mua. Lập bảng phân phối xác suất của X.

b. Từ mỗi kiện lấy ra 1 sản phẩm. Gọi Y là số sản phẩm loại 1. Lập bảng phân phối xác suất của Y.

c. Từ 2 kiện trên, chọn ngẫu nhiên một kiện. Từ kiện đó lấy ngẫu nhiên 2 sản phẩm, nếu cả hai đều là loại 1 thì mua kiện đó. Gọi Z là số kiện được mua. Hãy lập bảng phân phối xác suất của Z.

d. Từ kiện I, lấy ngẫu nhiên 2 sản phẩm và bỏ vào kiện II. Gọi T là số sản phẩm loại 1 có trong kiện II sau khi bỏ 2 sản phẩm từ kiện I vào. Hãy lập bảng phân phối xác suất của T

b. Hàm mật độ xác suất

i. Định nghĩa

Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất. Khi đó, hàm số $f: R \rightarrow R$ *xác định bởi*

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

được gọi là hàm mật độ của X,

ii. Tính chất:

$$+\ 0 \leq f(\mathrm{x}) \leq 1, \forall x \in R$$

$$+\ \sum_x f(x) = 1.$$

c. Hàm phân phối xác suất

i. Định nghĩa

Cho biến ngẫu nhiên rời rạc X có hàm mật độ xác suất f. Khi đó, hàm số $F: R \to R$ *xác định bởi*

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

được gọi là hàm phân phối xác suất của X.

ii. Tính chất của hàm phân phối:

- Tính chất 1: $0 \leq F(x) \leq 1, với\ mọi\ x\ thuộc\ R$
- Tính chất 2: $F(-\infty) = 0$ và $F(+\infty) = 1$
- Tính chất 3: $F\ là\ hàm\ tăng$
- Tính chất 4: F liên tục bên phải tại mọi x thuộc R.
- Tính chất 5: $P(a < X \leq b) = F(b) - F(a)$

- Nếu $F(x) = P(X < x)$ xét cho X rời rạc
- P(a$\leq X < b) = F(b) - F(a)$
- P(a$\leq X \leq b) = F(b) - F(a) + P(X = b)$

2. Biến ngẫu nhiên liên tục

a. Hàm mật độ xác suất của BNN liên tục X

Định nghĩa:

Hàm số $f: R \to R$ được gọi là hàm mật độ xác suất của biến ngẫu nhiên liên tục X nếu f(x) thỏa:

i. $f(\mathrm{x}) \geq 0, \forall x \in R$ và $\int_{-\infty}^{+\infty} f(x)dx = 1$

ii. $P(a \leq X \leq b) = \int_a^b f(x)dx$ với mọi a,b thuộc R.

b. Hàm phân phối xác suất của BNN liên tục X

i. Định nghĩa:

Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất f. Khi đó, hàm $số\ F: R \rightarrow R\ xác\ định\ bởi$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \quad \text{với } t \in R$$

được gọi là hàm phân phối xác suất của X.

ii. Tính chất: Lưu ý: $P(a \leq X \leq b) = F(b) - F(a)$.

iii. Mối liên hệ giữa hàm mật độ và hàm phân phối:

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

Các ví dụ

- Ví dụ 1:

Cho biến ngẫu nhiên liên tục X có hàm mật độ:

$$f(x) = \begin{cases} 0 & với \quad x < 1 \\ \dfrac{k}{x^2} & với \quad x \geq 1 \end{cases}$$

a. Hãy xác định hệ số k;

b. Tìm hàm phân phối $F(x)$;

c. Tính $P(2 < X < 3)$;

d. Tính xác suất để trong 4 phép thử độc lập, biến ngẫu nhiên X đều không lấy giá trị trong khoảng (2; 3).

Đáp số:

a. k=1. Dựa vào tính chất hàm mật độ và công thức tính tích phân suy rộng.

b. $F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & với \quad x < 1 \\ \frac{x-1}{x} & với \quad x \geq 1 \end{cases}$

c. $P(2 < X < 3) = \frac{1}{6}$

d. $(\frac{5}{6})^4$

Ví dụ 2:

- Cho biến ngẫu nhiên liên tục X có hàm phân phối:

$$F(x) = \begin{cases} 0 & với & x \le 0 \\ kx^2 & với & 0 < x \le 1 \\ 1 & với & x > 1 \end{cases}$$

a. Xác định hệ số k

b. Tìm hàm mật độ $f(x)$.

Đáp số:

a. $k = 1$. Dựa vào tính liên tục của hàm phân phối tại $x = 1$.

b. Kết quả của $f(x)$ chính là đạo hàm của $F(x)$.

III.Các đặc trưng của biến ngẫu nhiên

1. Biến ngẫu nhiên rời rạc

a. Kỳ vọng, ký hiệu E(X)

i. Định nghĩa:

Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất, khi đó E(X) được xác định:

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

ii. Tính chất của kỳ vọng:

- Tính chất 1: $E(a) = a$, với a là hằng số
- Tính chất 2: $E(aX) = a.E(X)$, với a là hằng số
- Tính chất 3: $E(X \mp Y) = E(X) \mp E(Y)$

iii. Ý nghĩa kỳ vọng:

- Là giá trị trung bình của biến ngẫu nhiên. Phản ánh giá trị trung tâm của phân phối xác suất của biến ngẫu nhiên.

- Trong sản xuất hay kinh doanh, nếu cần chọn phương án cho năng suất cao hay lợi nhuận cao ta chọn phương án cho năng suất hay lợi nhuận có kỳ vọng cao.

Ví dụ 1:

Trong một cuộc xổ số được tiến hành để giúp quỹ từ thiện địa phương, 10.000 vé số sẽ được bán với giá 5 USD mỗi vé. Giải thưởng là một chiếc xe ôtô trị giá 12.000USD. Nếu anh/chị mua hai vé số, thì khoản lợi (gain) trung bình của anh/chị là bao nhiêu?

Ví dụ 2:

- Một trò chơi chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen và 4 bi trắng.

a. Nếu chọn được 1 bi trắng thì sẽ được thưởng 2 USD. Hãy xác định số tiền nhận được trung bình sau một lần chơi.

b. Nếu chọn được 1 bi trắng thì được thưởng 2 USD và chọn được một bi đen thì được thưởng 3USD. Hãy tính số tiền nhận được trung bình sau 1 lần chơi.

b. Phương sai của biến ngẫu nhiên rời rạc:

i. Định nghĩa

Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất, khi đó phương sai của X, ký hiệu V(X) hay D(X) được xác định như sau:

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

Trong đó: $EX^2 = \sum_{i=1}^{n} x_i^2 p_i$

ii. Tính chất của phương sai:

- Tính chất 1: D(a) $= 0$, với a là hằng số
- Tính chất 2: D$(aX) = a^2 D(X)$, với a là hằng số
- Tính chất 3: D$(X \mp Y) = D(X) + D(Y)$ nếu X và Y độc lập.

iii. Ý nghĩa của phương sai

- Phương sai đặc trưng cho độ phân tán của biến ngẫu nhiên quanh giá trị trung bình, nghĩa là phương sai càng nhỏ thì độ phân tán nhỏ vì vậy độ tập trung lớn, ngược lại phương sai lớn thì độ phân tán lớn, vì vậy độ tập trung nhỏ.

- Trong kỹ thuật, phương sai đặc trưng cho độ sai số của thiết bị. Trong kinh doanh nó đặc trưng cho độ rủi ro của các quyết định.

Ví dụ:

Theo thống kê, một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm cho sinh mạng cho một năm với số tiền chi trả là 1000 USD, còn tiền đóng là 10 USD.

a. Hỏi lợi nhuận trung bình của công ty là bao nhiêu? Giải thích ý nghĩa kết quả?

b. Hãy tính phương sai của lợi nhuận của công ty.

- Gọi X là lợi nhuận của công ty

a. Lợi nhuận trung bình là $E(X) = 2$. Điều đó cho thấy công ty bảo hiểm có thể làm ăn có lãi.

b. Phương sai của X: $D(X) = 7936$. Phương sai khá lớn. Điều đó cho thấy mặc dù kinh doanh bảo hiểm có lãi nhưng rủi ro khá lớn.

c. Mốt của biến ngẫu nhiên rời rạc

i. Định nghĩa:

Mốt của biến ngẫu nhiên rời rạc X là giá trị của X ứng với xác suất lớn nhất, ký hiệu là M_0.

ii. Ví dụ:

Tìm mốt của X, với X có bảng phân phối trên.

iii. Tính chất:

Mốt của biến ngẫu nhiên rời rạc X không duy nhất.

2. Biến ngẫu nhiên liên tục

a. Kỳ vọng của biến ngẫu nhiên liên tục:

i. Định nghĩa:

Cho X là biến ngẫu nhiên liên tục có hàm mật độ $f(x)$. Khi đó, kỳ vọng của X được xác định:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

ii. Tính chất:

Tương tự như trường hợp rời rạc.

Ví dụ:

- Tuổi thọ của một loại côn trùng nào đó là biến ngẫu nhiên X (đơn vị: tháng) với hàm mật độ như sau:

$$f(x) = \begin{cases} kx^2(4-x) & nếu \quad 0 \le x \le 4 \\ 0 & nếu \quad ngược\ lại \end{cases}$$

a. Tìm k?

b. Tìm hàm phân phối xác suất

c. Tính tuổi thọ trung bình của loài côn trùng trên.

d. Tính xác suất để loài côn trùng trên có thời gian sống trên 5 tháng.

b. Phương sai của biến ngẫu nhiên liên tục:

i. Định nghĩa:

Cho X là biến ngẫu nhiên liên tục có hàm mật độ $f(x)$. Khi đó, phương sai của X được xác định:

$$D(X) = EX^2 - (EX)^2$$

trong đó $EX^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx$

ii. Ví dụ:

Tìm phương sai của tuổi thọ loài côn trùng trên.

c. Mốt của biến ngẫu nhiên liên tục:

Định nghĩa: Cho biến ngẫu nhiên liên tục X. Khi đó, $modX$ là giá trị của X mà tại đó hàm mật độ xác suất đạt giá trị max.

d. Độ lệch tiêu chuẩn của biến ngẫu nhiên

Cho BNN X, độ lệch tiêu chuẩn của BNN X được xác định là $\delta = \sqrt{D(X)}$.

IV. Một số quy luật phân phối xác suất thông dụng
1. Phân phối Nhị thức
a. Định nghĩa thí nghiệm Nhị thức

Một thí nghiệm nhị thức là thí nghiệm có các đặc trưng sau:

- Thí nghiệm đó có n lần thử giống nhau;
- Mỗi lần thử tạo ra hai kết quả;
- Xác suất thành công trong một lần thử là bằng p và xác suất này không đổi qua các lần thử;
- Các phép thử là độc lập nhau.

- Khi đó, xác suất để trong n lần thử có k lần thử thành công được tính theo công thức:

$P(k) = C_n^k p^k (1-p)^{n-k}$: công thức Becnuli.

b. Định nghĩa phân phối Nhị thức

Giả sử thí nghiệm Nhị thức có n lần thử giống nhau, xác suất để một lần thử "thành công" là p. Gọi X là số lần thử "thành công" trong n lần thử. Khi đó,

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

Một BNN X có xác suất được tính theo (*) được gọi là BNN có phân phối Nhị thức với 2 tham số n, p. Ký hiệu là $X \sim B(n, p)$.

Ví dụ 1:

Xác suất để một con gà mái đẻ trứng trong ngày là 0,6. Một gia đình nuôi 5 con gà mái. Tính xác suất để:

a. Trong ngày có 2 con đẻ;

b. Có nhiều nhất 1 con đẻ.

Ví dụ 2:

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

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

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

Ví dụ 3:

Trong một vùng dân cư có 65% gia đình có máy giặt, chọn ngẫu nhiên 12 gia đình.

Gọi X là số gia đình có máy giặt trong số 12 gia đình này.

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

Ví dụ 4:

Một nhà máy sản xuất với tỷ lệ phế phẩm là 7%. Quan sát ngẫu nhiên 10 sản phẩm. Tính xác suất để có:

i. Có đúng một phế phẩm,

ii. Có ít nhất một phế phẩm,

iii. Có nhiều nhất một phế phẩm.

Ví dụ 5:

Một lô thuốc gồm 10 lọ trong đó có 2 lọ thuốc hỏng. Lấy ngẫu nhiên 5 lọ từ lô thuốc, có hoàn lại.

Gọi X là số lọ hỏng trong 5 lọ lấy ra.

Tính xác suất để có 2 lọ hỏng.

c. Các đặc trưng của phân phối Nhị thức

Mệnh đề:

Giả sử $X \sim B(n, p)$. Khi đó:

$$+ \ E(X) = np$$

$$+ D(X) = npq \text{ với } q = 1 - p$$

và $np - q \leq modX \leq np - q + 1$

Ví dụ 6:

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

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

b. Mỗi năm nhân viên này đi bán 300 ngày. Gọi Y là số ngày bán được hàng trong năm. Tìm số ngày bán được hàng nhiều khả năng nhất trong năm.

2. Phân phối Poisson

a. Định nghĩa:

Biến ngẫu nhiên X được gọi là có phân phối Poisson với tham số μ, ký hiệu là $X \sim P(\mu)$ nếu xác suất để (X = k) được tính theo công thức:

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

Nếu $X{\sim}B(n,p)$ và khi n khá lớn và p khá nhỏ thì:

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \approx e^{-\mu} \times \frac{\mu^k}{k!}$$

2.Phân phối Poisson

Hay, nếu $X \sim B(n,p)$, trong đó p nhỏ và n lớn thì X xem như có phân phối Poisson $X \sim P(\mu)$, $\mu = np$.

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

Ví dụ 7:

Khi tiêm truyền một loại huyết thanh, trung bình có một trường hợp phản ứng trên 1000 trường hợp. Dùng loại huyết thanh này tiêm cho 2000 người. Tính xác suất để:

a. có 3 trường hợp phản ứng,

b. có nhiều nhất 3 trường hợp phản ứng

c. có nhiều hơn 3 trường hợp phản ứng.

b. Mệnh đề

Cho $X \sim P(\mu)$. Ta có:

i. Trung bình:

$$E(X) = \mu.$$

ii. Phương sai:

$$D(X) = \mu.$$

Ví dụ 8:

Một trung tâm bưu điện nhận được trung bình 3 cuộc điện thoại trong mỗi phút.

Tính xác suất để trung tâm này nhận được:

a. 1 cuộc,

b. 2 cuộc,

c. 3 cuộc gọi trong mỗi phút.

Biết rằng số cuộc gọi trong mỗi phút có phân phối Poisson.

3. Phân phối chuẩn (phân phối Gauss):

a. Phân phối chuẩn hóa:

i. Định nghĩa:

Biến ngẫu nhiên X được gọi là có phân phối chuẩn hóa (tắc), ký hiệu là $X \sim N(0,1)$ nếu hàm mật độ xác suất có dạng:

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

ii. Mệnh đề:

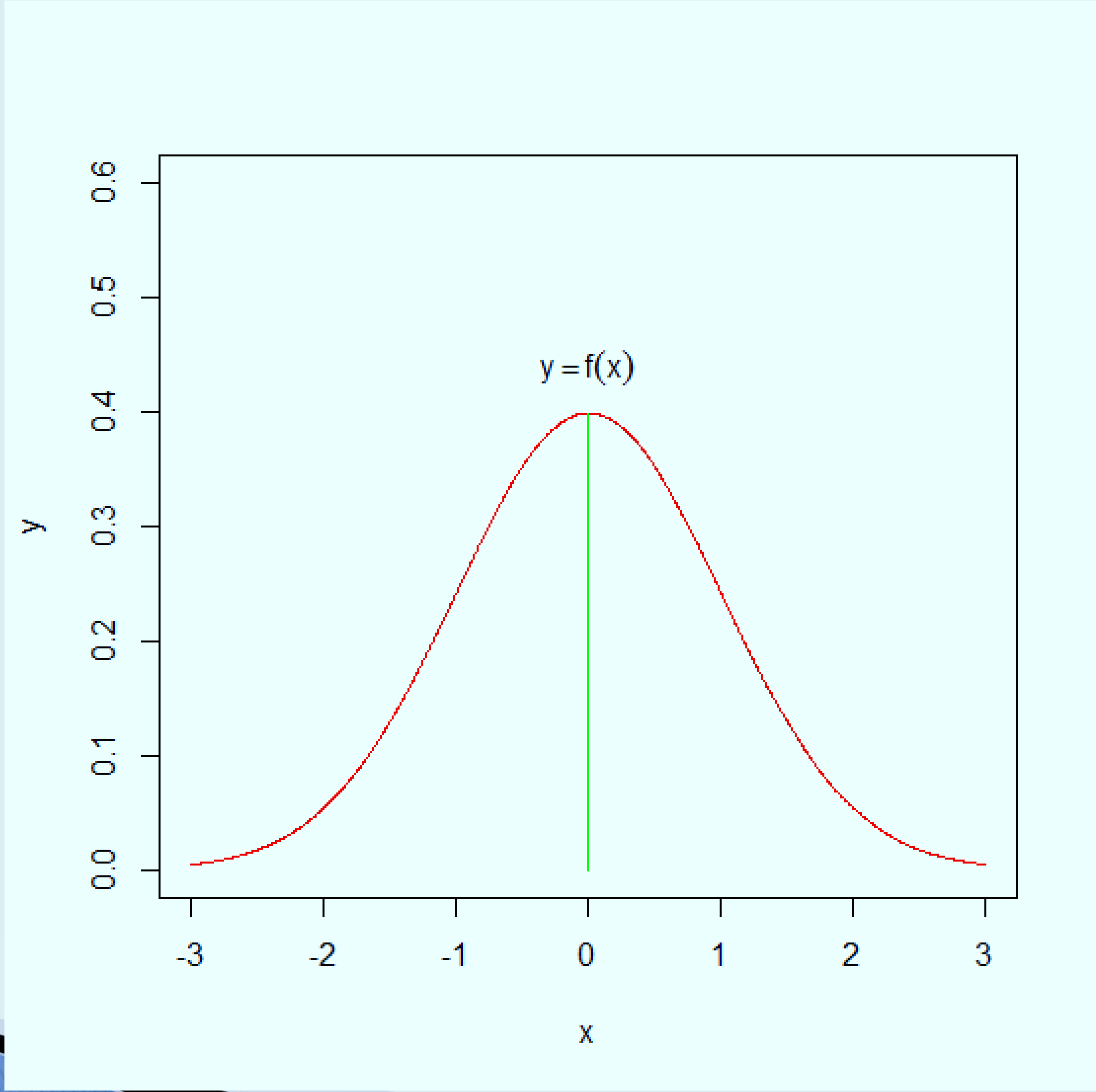
Cho biến ngẫu nhiên $X \sim N(0,1)$. Ta có:

$i.\ E(X) = 0$

$ii.\ D(X) = 1$

iii. $modX = 0$

Đồ thị hàm mật độ y=f(x)=$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$



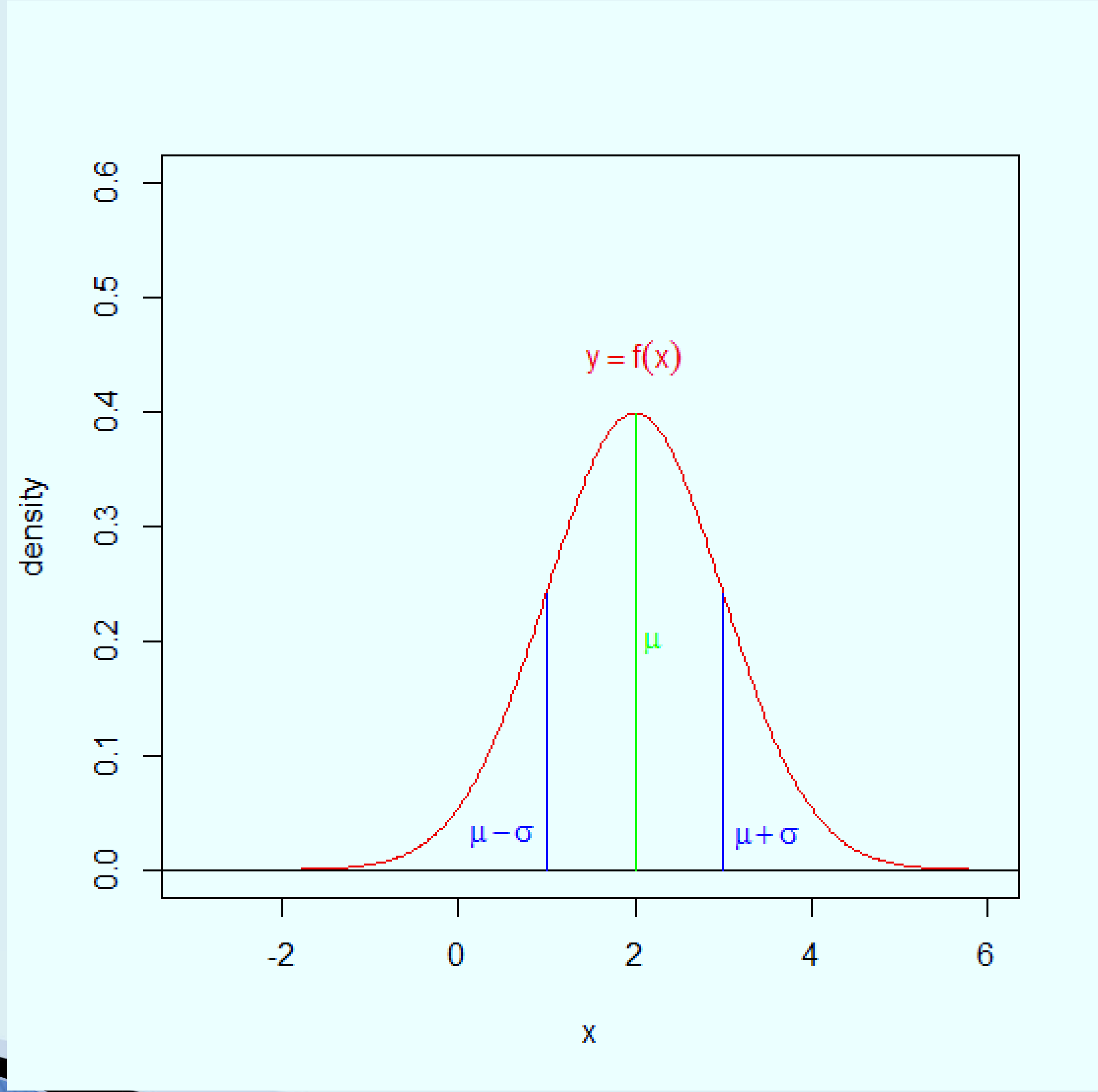
b. Phân phối chuẩn (trường hợp tổng quát)

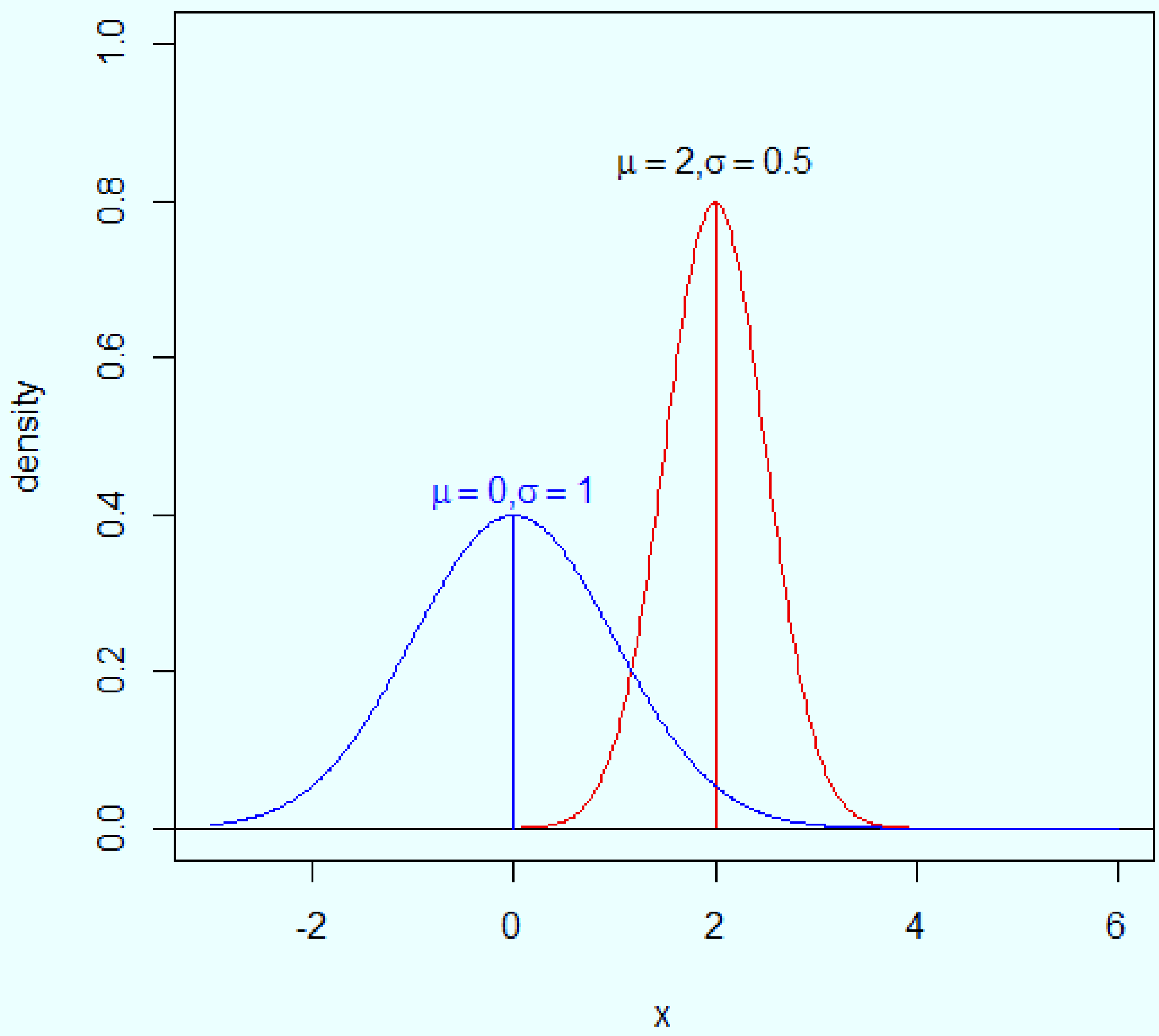
i. Định nghĩa:

Biến ngẫu nhiên X gọi là có phân phối chuẩn, ký hiệu là $X \sim N(\mu, \sigma^2)$ nếu hàm mật độ xác suất của X là:

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

Đồ thị hàm mật độ $y=f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$





μ = 2,σ = 0.5
μ = 0,σ = 1
density
0.0
0.2
0.4
0.6
0.8
1.0
-2
0
2
4
6
x

ii. Mệnh đề:

Cho biến ngẫu nhiên $X \sim N(\mu, \sigma^2)$. Khi đó:

$$+ E(X) = \mu$$

$$+ D(X) = \sigma^2$$

$$+ \ ModX = \mu$$

$$+ \ N\text{ếu}\ Y = \frac{X-\mu}{\sigma}\ th\text{ì}\ \mathrm{Y} \sim N(0,1\ ).$$

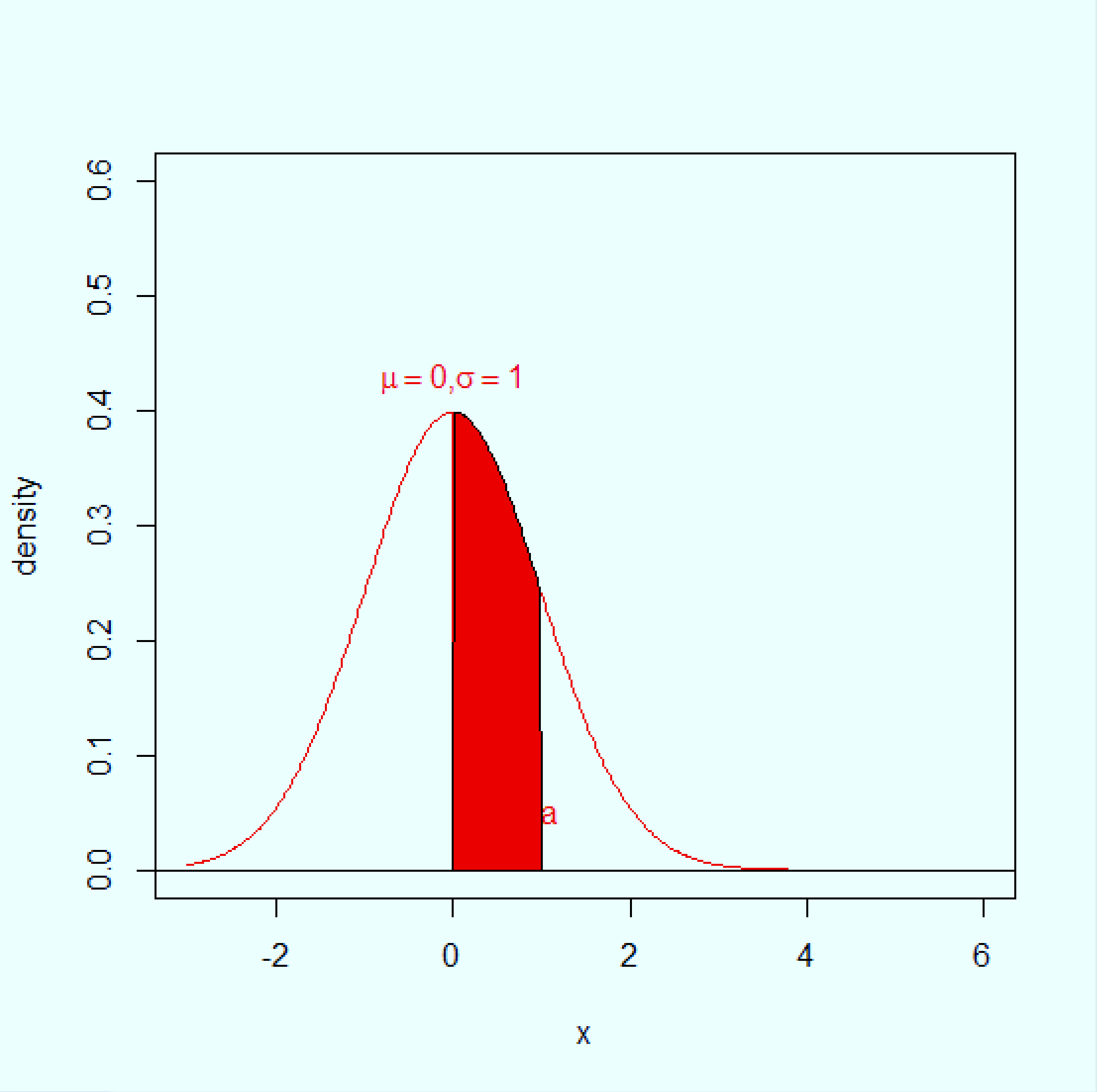
c. Các công thức tính xác suất

Công thức 1: Cho $X \sim N(0;1)$ thì ta có:

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

với $\varphi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ được gọi là tích phân Laplace (tra bảng phụ lục).

$$\varphi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$: tích phân Laplace



c. Các công thức tính xác suất

Công thức 1: Cho $X \sim N(0;1)$ thì ta có:

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

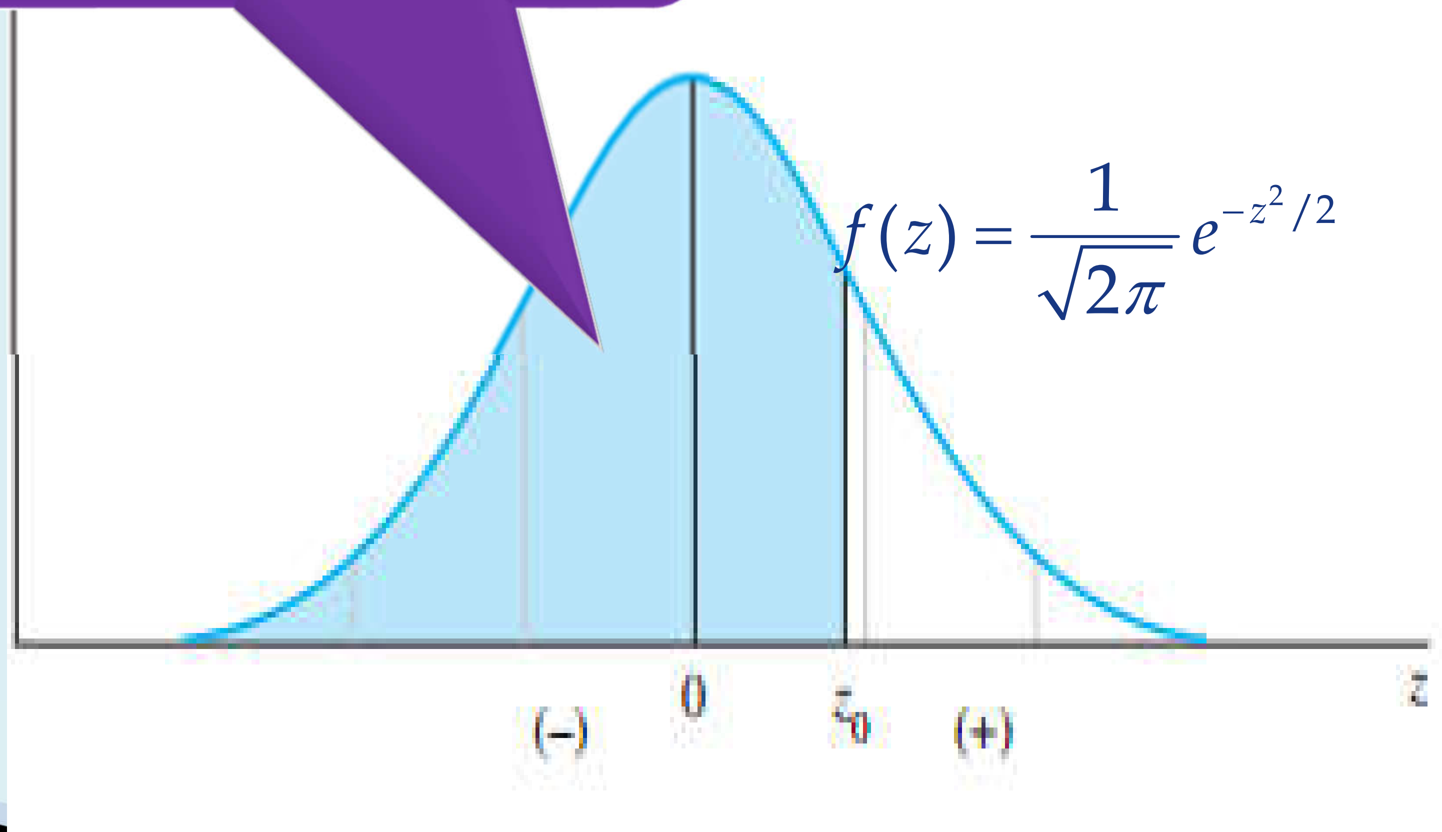
với $\Phi(x) = \mathrm{P}(\mathrm{X} < \mathrm{x}) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

được gọi là xác suất tích lũy – cumulative probability (tra bảng 1).

Giả sử $Z \sim N(0, 1)$

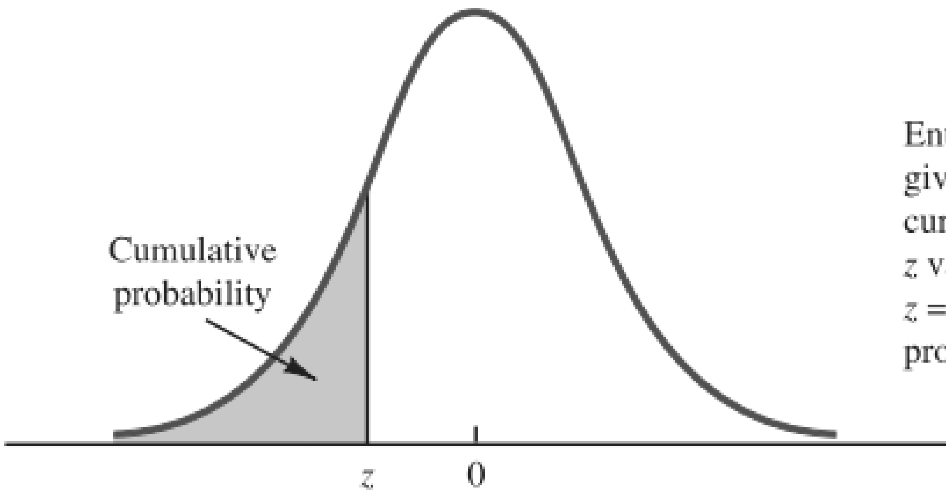
Cumulate probability,

$\Phi(z_0) = P(Z \le z_o) = \int_{-\infty}^{z_0} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$



$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$: **Xác suất tích lũy – cumulative probability**

TABLE 1 CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION



Entries in the table give the area under the curve to the left of the z value. For example, for $z = -.85$, the cumulative probability is .1977.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048

Standard Normal Distribution -5

- **Example:** Cumulative Probability Table for the Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.[illegible]264	.8289	.8315	.8340	.8365	.8389
.	.	.	.	.		.	.	.	.	.

$P(Z \leq 0.83)$

Công thức 2:

Giả sử $X \sim N(\mu; \sigma^2)$, khi đó:

$$P(a < X < b) = \varphi\left(\frac{b-\mu}{\sigma}\right) - \varphi\left(\frac{a-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Từ đó ta có quy tắc $\boldsymbol{k-\sigma}$:

$$P(|X-\mu| < k\sigma) = 2\varphi(k)$$

Với $\varphi(x)$ tra trong bảng tích phân Laplace và $\Phi(x)$ tra trong bảng xác suất tích lũy (bảng 1).

Lưu ý: Hàm tích phân Laplace

- *Trong bảng tích phân hàm Laplace, chỉ có*

  $x \geq 0$ *và nếu* $x \geq 5$ *thì* $\varphi(x) = 0{,}5$.

–*Hàm* $\varphi(x)$ *là hàm lẻ, tức là* $\varphi(-x) = -\varphi(x)$.

Ví dụ 1:

Nghiên cứu chiều cao của những người trưởng thành, người ta nhận thấy rằng chiều cao tuân theo quy luật phân phối chuẩn với trung bình là 175cm và độ lệch tiêu chuẩn là 4cm. Hãy tính:

a. Tỷ lệ người trưởng thành có tầm vóc trên 180cm,

b. Tỷ lệ người trưởng thành có chiều cao từ 166cm đến 177cm,

c. Tìm h_0, nếu biết rằng 33% người trưởng thành có tầm vóc dưới mức h_0.

Ví dụ 2:

Trọng lượng X (tính bằng gam) một loại trái cây có phân phối $N(\mu;\sigma^2)$, với $\mu = 500(gam)$ và $\sigma^2 = 16(gam^2)$.

Trái cây thu hoạch được đem phân loại trọng lượng như sau:

a. loại 1 : trên 505 gam,

b. loại 2 : từ 495 đến 505 gam,

c. loại 3 : dưới 495 gam.

Tính tỷ lệ mỗi loại.

e. Định lý Moirve – Laplace địa phương

Cho $X \sim B(n,p)$. Giả sử n khá lớn và p không quá gần 0 cũng không quá gần 1.

Khi đó có thể xấp xỉ bằng đại lượng ngẫu nhiên X có phân phối chuẩn, tức là $X \sim N(\mu, \sigma^2)$ với $\mu = np, \sigma^2 = npq$ và $q = 1 - p$.

Khi đó, ta có:

a. $P(X=k) \approx \frac{1}{\sqrt{npq}} f(\frac{k-np}{\sqrt{npq}})$

- trong đó $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ là hàm mật độ Gauss.

  Lưu ý: Hàm mật độ Gauss là hàm chẵn.

b. $P(k_1 \leq X \leq k_2) \approx \varphi\left(\frac{k_2-\mu}{\sigma}\right) - \varphi\left(\frac{k_1-\mu}{\sigma}\right)$

trong đó $\varphi(x)$ là hàm tích phân Laplace.

Ví dụ 3:

Một nhà máy sản xuất ra một loại sản phẩm loại A với xác suất 0.485. Tính xác suất sao cho trong 200 sản phẩm do máy sản xuất ra có ít nhất 95 sản phẩm loại A.

Ví dụ 4:

Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên có phân phối chuẩn với trung bình 50kg và phương sai là 100. Sản phẩm được gọi là sản phẩm loại A nếu có trọng lượng từ 45kg đến 70 kg.

a. Tính xác suất để gặp sản phẩm loại A.

b. Kiểm tra 100sp. Tính xác suất để:

+ có ít nhất 60 sản phẩm loại A;

+ Có đúng 60 sản phẩm loại A.

4. Phân phối Chi – bình phương

a. Định nghĩa

Xét $Z_1, Z_2, \ldots, Z_n$ là n biến ngẫu nhiên có phân phối chuẩn hóa, tức là $Z_i \sim N(0,1)$ với $i = 1,..,n$. $Z_1, Z_2, \ldots, Z_n$ độc lập với nhau.

Đặt

$$\boxed{Y = \sum_{i=1}^{n} Z_i^2 = Z_1^2 + Z_2^2 + \ldots + Z_n^2}$$

Đại lượng ngẫu nhiên Y gọi là có phân phối Chi – bình phương với n bậc tự do.

Ký hiệu: $Y \sim \chi^2(n)$

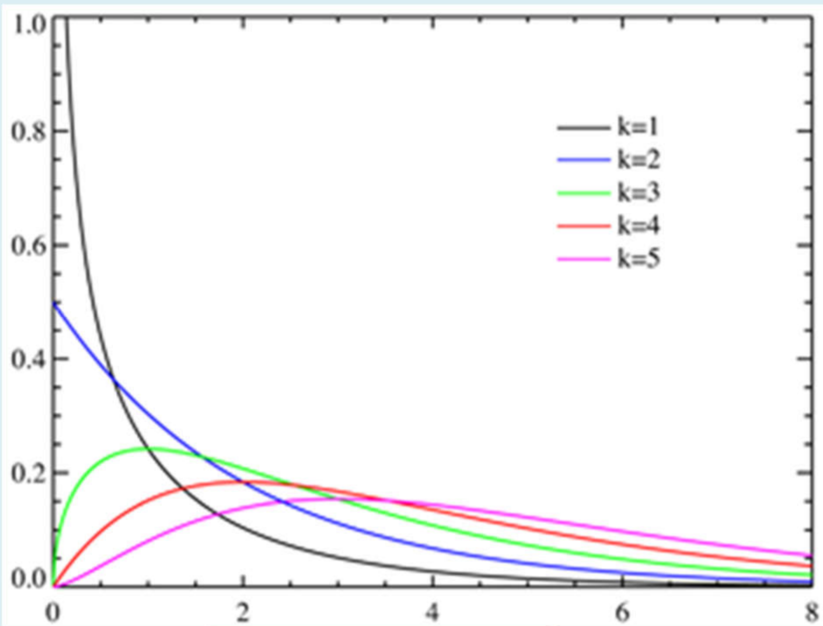
4. Phân phối chi- bình phương

b. Định lý

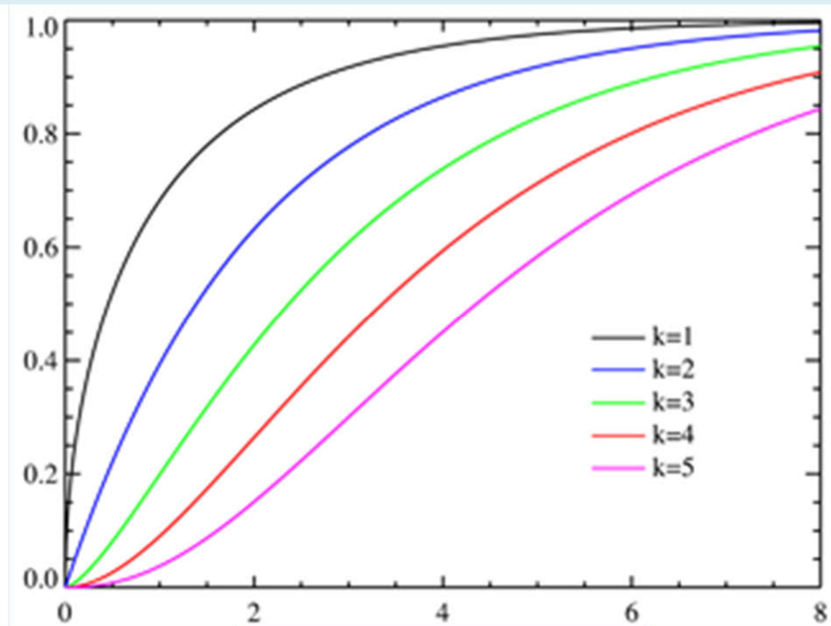
Cho $Y \sim \chi^2(n)$. Khi đó:

$E(Y) = n$; $D(Y) = 2n$

4. Phân phối Chi – bình phương



Hàm mật độ



Hàm phân phối

Hàm mật độ và hàm phân phối của
phân phối Chi – bình phương với các bậc tự do
Khác nhau

5. Phân phối Student

a. Định nghĩa

Xét biến ngẫu nhiên Z ~ N(0,1) và $Y \sim \chi^2(n)$; Z và Y độc lập với nhau.

Đặt

$$\boxed{T = \frac{Z}{\sqrt{Y/n}}}$$

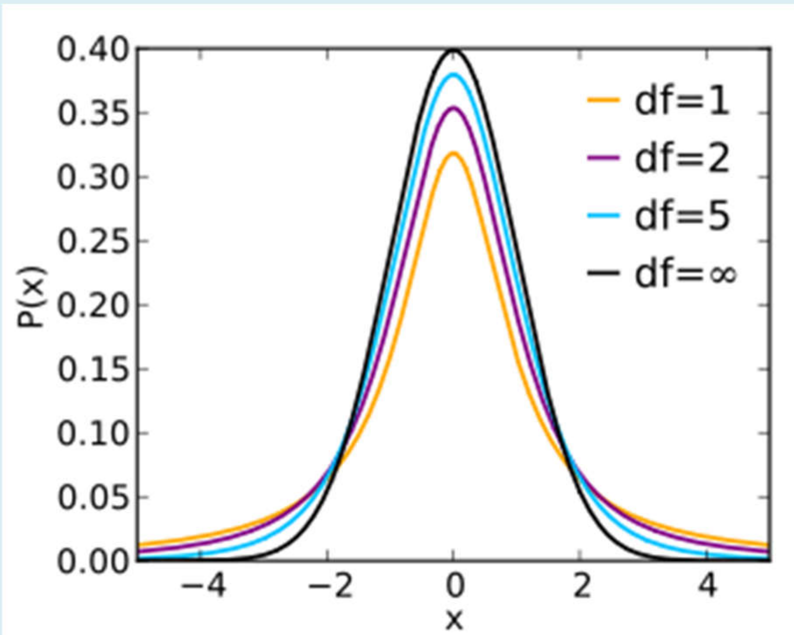
Đại lượng ngẫu nhiên T gọi là có phân phối Student với n bậc tự do. Ký hiệu: $T \sim t(n)$

5. Phân phối Student

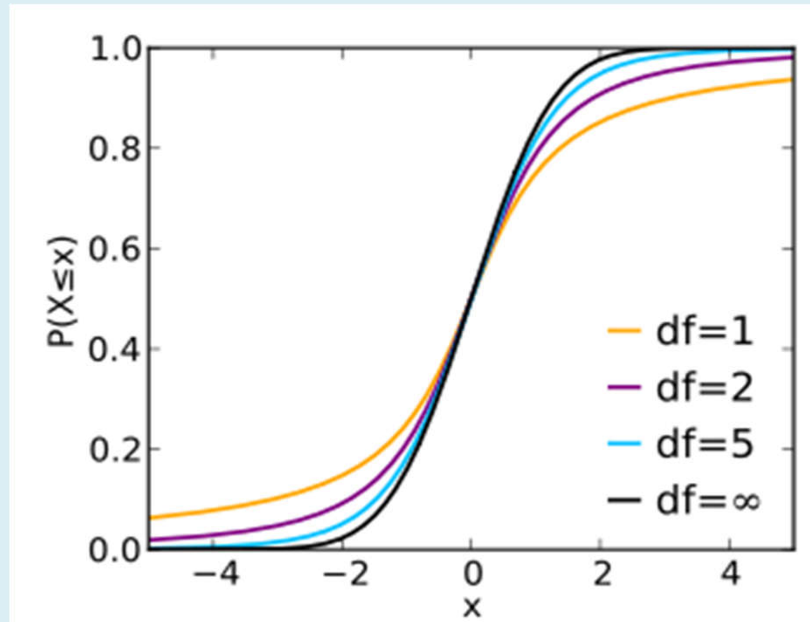
b. Định lý:

Cho $T \sim T(n)$, khi đó $E(T) = 0, D(T) = \frac{n}{n-2}$

Phân phối Student



Hàm mật độ



Hàm phân phối

Hàm mật độ và hàm phân phối của phân phối Student với các bậc tự do Khác nhau

Chương 3: Lý thuyết mẫu

Nội dung:

I. Tổng thể và mẫu ngẫu nhiên

II. Thống kê và một số thống kê thông dụng

III. Phân phối của các đặc trưng mẫu

I. Tổng thể và mẫu ngẫu nhiên

1. Khái niệm tổng thể:

Tập tất cả các phần tử do mục đích nghiên cứu qui định.

Số phần tử tổng thể gọi là kích thước của tổng thể, kí hiệu là N.

Đặc tính (dấu hiệu) mà ta khảo sát thường kí hiệu là X, Y, . .

Ví dụ:

- Số phần tử của tổng thể thường rất lớn nên ta không thể chọn hết những phần tử để thực hiện thí nghiệm vì những lý do sau:
  - Thời gian và kinh phí không cho phép.
  - Có thể làm hư hại các phần tử của tổng thể.

- Vì vậy người ta sẽ chọn một tập con của tổng thể để nghiên cứu, một tập con như vậy gọi là **Mẫu**. Số phần tử của mẫu gọi là cỡ mẫu, kí hiệu n.

2. Mẫu ngẫu nhiên:

Giả sử từ tổng thể lấy ra n phần tử, tạo nên mẫu kích thước n theo phương pháp lấy mẫu giản đơn, tức là mỗi lần lấy ra 1 phần tử, ngẫu nhiên và có hoàn lại.

2. Mẫu ngẫu nhiên:

Gọi X_i là giá trị của dấu hiệu X trên phần tử thứ i của mẫu nên $X_1, X_2, \ldots, X_n$ là các đại lượng ngẫu nhiên *độc lập* và có *cùng qui luật phân phối xác suất của X* và kí hiệu mẫu ngẫu nhiên là $(X_1, X_2, \cdots, X_n)$.

3. Mẫu thực nghiệm:

Khi ta xét mẫu cụ thể, thì mẫu ngẫu nhiên là $(X_1, X_2, \cdots, X_n)$ nhận giá trị là $(x_1, x_2, \cdots, x_n)$ và gọi là mẫu thực nghiệm.

4. Mẫu thực nghiệm:

Khi ta xét mẫu cụ thể, thì mẫu ngẫu nhiên là $(X_1, X_2, \cdots, X_n)$ nhận giá trị là $(x_1, x_2, \cdots, x_n)$ và gọi là mẫu thực nghiệm.

5. Dạng bảng số liệu thường gặp:

Dạng1: Liệt kê dưới dạng: $x_1, x_2, \ldots, x_n$, trong đó mỗi giá trị có thể lặp lại nhiều lần.

Dạng 2: Mẫu dạng điểm:

X	x_1	x_2	. . .	x_k
n_i	n_1	n_2		n_k

- Trong đó $x_1 < x_2 < \ldots < x_k$ và mỗi số liệu x_i xuất hiện n_i lần. n_i là tần số của x_i .

Dạng 3: Mẫu dạng khoảng:

X	$a_1 - b_1$	$a_2 - b_2$	. . .	$a_k - b_k$
n_i	n_1	n_2	. . .	n_k

Khi đó ta đưa mẫu dạng khoảng về mẫu dạng điểm:

với $x_i = \frac{a_i + b_i}{2}$

6. Tham số:

Là một đặc trưng cụ thể của tổng thể, thường kí hiệu θ.

Ví dụ:

Tham số trung bình μ

Tham số phương sai σ^2

II. Thống kê và một số thống kê thông dụng

1.Thống kê

Một hàm của $(X_1, X_2, \cdots, X_n)$, kí hiệu $G = f(X_1, X_2, \cdots, X_n)$ được gọi là một thống kê G.

G là một đại lượng ngẫu nhiên.

2. Một số thống kê thông dụng

a. Trung bình mẫu ngẫu nhiên

i. Định nghĩa:

Cho mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ được xây dựng từ đại lượng ngẫu nhiên gốc X, khi đó

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum X_i$$

là một thống kê.

Trung bình mẫu cụ thể

Nếu xét mẫu cụ thể $(x_1, x_2, \cdots, x_n)$ $thì$ $\overline{\mathrm{X}}$ có giá trị cụ thể là

$$\bar{x} = \frac{1}{n}\left(x_1 + x_2 + \cdots + x_n\right) = \frac{1}{n}\sum x_i$$

và gọi là trung bình mẫu cụ thể.

ii. Tính chất

Nếu biến ngẫu nhiên X có $E(X) = \mu$ và $D(X) = \sigma^2$ thì

$$E(\bar{X}) = \mu \text{ và } D(\bar{X}) = \frac{\sigma^2}{n}.$$

iii. Định lý:

Nếu X có qui luật phân phối $N(\mu; \sigma^2)$ thì $\bar{X}$ có phân phối $N\left(\mu; \frac{\sigma^2}{n}\right)$.

b. Phương sai của mẫu ngẫu nhiên

i. Định nghĩa:

Cho mẫu ngẫu nhiên $(X_1, X_2, \cdots, X_n)$ được xây dựng từ đại lượng ngẫu nhiên gốc X, khi đó phương sai của mẫu ngẫu nhiên, kí hiệu

$$S^2 = \frac{1}{n}\sum\left(X_i - \overline{X}\right)^2$$

là một thống kê.

Phương sai mẫu cụ thể

Nếu xét mẫu cụ thể $(x_1, x_2, \cdots, x_n)$ thì S^2

có giá trị cụ thể là

$$s^2 = \frac{1}{n}\sum\left(x_i - \overline{x}\right)^2$$

và gọi là phương sai mẫu cụ thể.

ii. Tính chất

- Nếu biến ngẫu nhiên X có $E(X) = \mu$ và $D(X) = \sigma^2$ thì

$$E\left(S^2\right) = \frac{n-1}{n}\sigma^2$$

- Nếu đặt $S'^2 = \frac{n}{n-1}S^2$ thì $E\left(S'^2\right) = \sigma^2$

- Và S'^2 gọi là phương sai có điều chỉnh của mẫu ngẫu nhiên.

Chứng minh.

3. Cách tính đặc trưng của mẫu thực nghiệm

a.Tỷ lệ mẫu:

Cho mẫu có kích thước n, trong đó có m phần tử có tính chất A. Khi đó, $f = \frac{m}{n}$ là tỷ lệ mẫu.

b.Trung bình mẫu: $\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{k} x_i n_i}$

c. Phương sai mẫu:

$$\boxed{S^2 = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \overline{x^2} - (\overline{x})^2}$$

Trong đó :

$$\boxed{\overline{x^2} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2}$$

d. Phương sai mẫu hiệu chỉnh, ký hiệu là:

$$\boxed{S'^2 = \frac{n}{n-1} S^2}$$

Độ lệch tiêu chuẩn: $\boxed{S' = \sqrt{S'^2}}$

Các ví dụ

- **Ví dụ 1**:
- Thời gian tự học (giờ/ngày) của 90 sinh viên trường 1 Đại học cho bởi bảng sau:

Thời gian tự học	1	2	3	4	5	6
Số sv	7	8	17	24	20	14

Tính các đặc trưng mẫu: trung bình mẫu và phương sai mẫu có điều chỉnh.

Ví dụ

- **Ví dụ 2:**

Khảo sát chiều cao của 15 sinh viên trong một lớp học:

160, 165, 155, 162, 167, 145, 158, 170, 165, 155, 158, 160, 170, 175, 169.

Tính các đặc trưng mẫu: trung bình mẫu, phương sai mẫu.

III. Phân phối của các đặc trưng mẫu

Định lý 1:

Nếu $X \sim N(\mu, \sigma^2)$ $thì$ $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$\text{và } U = \frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1).$$

Định lý 2:

- Nếu biến ngẫu nhiên X có $E(X)=\mu$ và $D(X)=\sigma^2$

thì
$$\chi_1{}^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$$

- Và

$$\chi_2{}^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{S'}\right)^2\sim\chi^2(n-1)$$

Chứng minh.

Định lý 3:

Nếu $X \sim N(\mu, \sigma^2)$ với σ^2 chưa biết thì

$$T = \frac{\overline{X} - \mu}{S'}\sqrt{n} \sim t(n-1)$$

Khi n>30 thì

$$T = \frac{\overline{X} - \mu}{S'}\sqrt{n} \sim t(n-1) \approx N(0,1)$$

Định lý 4:

Giả sử đại lượng ngẫu nhiên X có phân phối xác suất nào đó, không phải là phân phối chuẩn thì

$$U_1 = \frac{\bar{X}-\mu}{S'}\sqrt{n} \text{ và } U_2 = \frac{\bar{X}-\mu}{\sigma}\sqrt{n}$$

sẽ có phân phối xấp xỉ chuẩn N(0,1) khi n khá lớn (n>30).

Chương 4:

Ước lượng tham số thống kê

I. Ước lượng điểm

- Xét tổng thể có đặc tính X cần khảo sát.
- Biến ngẫu nhiên X có hàm mật độ $f(x)$ phụ thuộc vào tham số θ chưa biết.
- Vấn đề: cần tìm tham số θ.
- $(X_1, X_2, \dots, X_n)$ là một mẫu ngẫu nhiên được chọn từ X.
- Thống kê $\hat{\theta} = h(X_1, X_2, \dots, X_n)$ gọi là một ước lượng của θ.
- Với mẫu thực nghiệm $(x_1, x_2, \dots, x_n)$ ta có một ước lượng cụ thể của θ:

$$\hat{\theta} = h(x_1, x_2, \dots, x_n)$$

TIÊU CHUẨN ƯỚC LƯỢNG KHÔNG CHỆCH

Ước lượng $\hat{\theta}$ gọi là một ước lượng không chệch cho tham số θ nếu:

$$E(\hat{\theta}) = \theta.$$

Nếu $\hat{\theta}$ *là ước lượng chệch của tham số* θ thì $E(\hat{\theta}) - \theta$ gọi là độ chệch của ước lượng.

- Các kết quả dùng cho ước lượng điểm khi cỡ mẫu lớn.
  - Trung bình mẫu $\bar{X}$ là ước lượng không chệch cho μ $(kỳ\ vọng\ tổng\ thể)$.
  - Tỷ lệ mẫu f là ước lượng không chệch của tỷ lệ tổng thể (p).
  - Phương sai mẫu có hiệu chỉnh S'^2 là ước lượng không chệch của phương sai tổng thể σ^2.

Nhận xét:

Ước lượng điểm có nhược điểm là khi cỡ mẫu nhỏ thì ước lượng điểm tìm được sai lệch rất nhiều so với giá trị của đặc trưng cần ước lượng. Do đó, người ta thường dùng phương pháp ước lượng bằng khoảng tin cậy.

II. Ước lượng khoảng – Khoảng tin cậy

- Giả sử tổng thể có đặc trưng θ *chưa biết*. Căn cứ vào mẫu có kích thước n, tìm hai đại lượng θ_1, θ_2 sao cho:

$$P(\theta_1 < \theta < \theta_2) = 1 - \alpha \ (\alpha \text{ đã biết})$$

- Khoảng (θ_1, θ_2) gọi là khoảng tin cậy.
- $(1 - \alpha)$ gọi là độ tin cậy.

1. Ước lượng trung bình của tổng thể

- Giả sử $X \sim N(\mu, \sigma^2)$, trong đó $\mu\ chưa\ biết, ta$ cần ước lượng cho μ.
- Từ tổng thể chọn một mẫu với kích thước n, ta tìm hai số μ_1, μ_2 sao cho:

$$P(\mu_1 < \mu < \mu_2) = 1 - \alpha$$

- Trong thực hành, ta xét 4 trường hợp:

1. Ước lượng trung bình của tổng thể

Trường hợp 1: $n > 30$ $và$ δ^2 đã biết

- Áp dụng qui tắc $k - \delta, ta$ $có$:
- Công thức ước lượng khoảng cho μ $là$:

$$\left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

- Trong đó $\varphi(Z_{\frac{\alpha}{2}}) = \frac{1-\alpha}{2}$ và φ là hàm Laplace.
- $\varepsilon = Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ là độ chính xác (sai số) của ước lượng.

1. Ước lượng trung bình của tổng thể

Với độ tin cậy $1-\alpha$, ta có thể tìm giá trị $Z_{\frac{\alpha}{2}}$ từ bảng phân phối xác suất tích lũy (bảng 1) bằng cách dựa vào công thức:

$$\boxed{\Phi(z_{\frac{\alpha}{2}}) = P(Z < z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}}$$

- Trường hợp 2: $n > 30\ và\ \delta^2$ chưa biết

  Công thức khoảng ước lượng tương tự trường hợp 1 nhưng thay $\sigma\ bằng\ S'$.

- Trường hợp 3: $n \leq 30\ và\ \delta^2$ đã biết

  Công thức khoảng ước lượng tương tự công thức ở trường hợp 1.

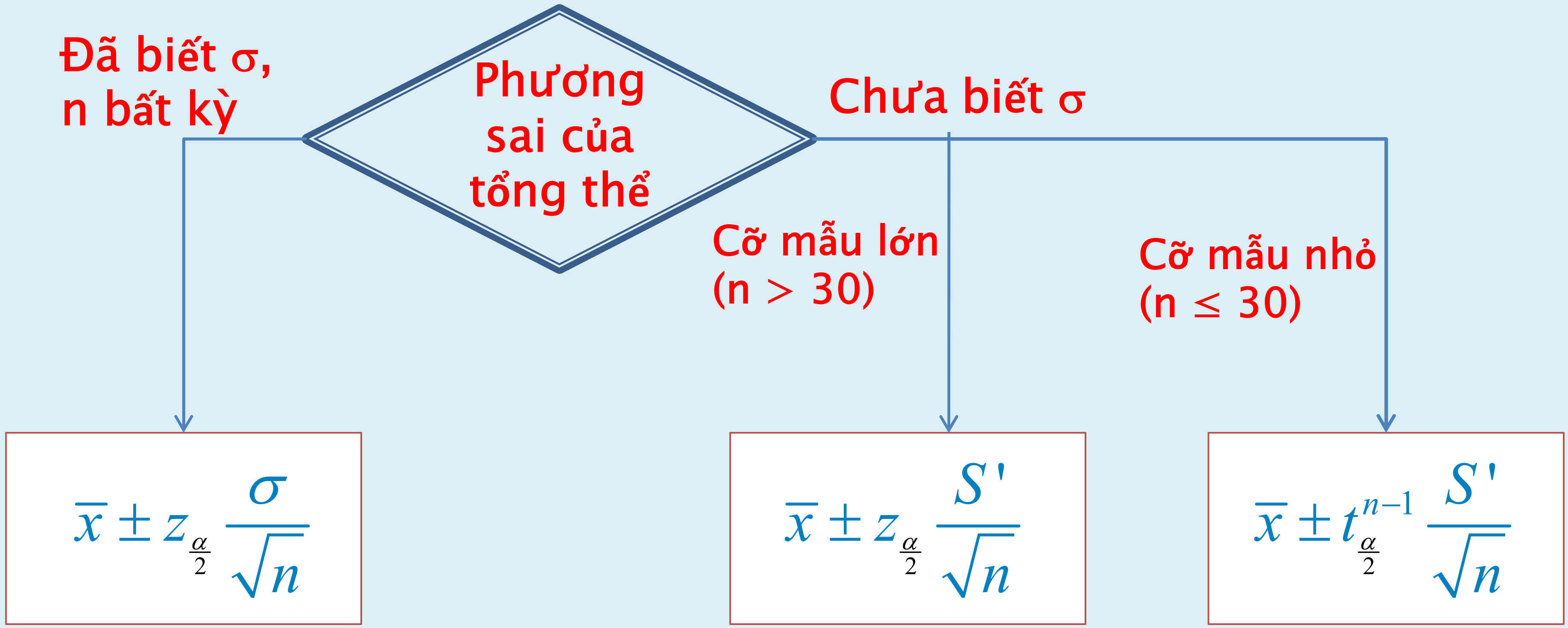
Trường hợp 4: $n \leq 30$ và δ^2 chưa biết

Khi đó công thức khoảng ước lượng cho trung bình:

$$\left(\overline{x} - t_{\frac{\alpha}{2}}^{n-1}\frac{s'}{\sqrt{n}}; \overline{x} + t_{\frac{\alpha}{2}}^{n-1}\frac{s'}{\sqrt{n}}\right)$$

Trong đó $t_{\frac{\alpha}{2}}^{n-1}$ tra bảng phân phối Student (bảng 2).

Công thức khoảng tin cậy cho trung bình TT – TÓM TẮT-



Ví dụ 1:

Cân thử 100 sản phẩm, ta có trọng lượng trung bình của một sản phẩm là 500g. Cho biết trọng lượng của sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn với $\sigma^2 = 40.000$. Hãy ước lượng bằng khoảng tin cậy trọng lượng trung bình của sản phẩm với độ tin cậy 95%.

Ví dụ 2:

Để xác định trọng lượng trung bình của các hộp đựng sản phẩm. Cân thử 15 hộp và có trọng lượng trung bình là 39,8. Biết trọng lượng của các hộp sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn với $S'^2 = 0{,}144$. Hãy ước lượng trọng lượng trung bình của các hộp sản phẩm với độ tin cậy 99%.

2. Ước lượng tỷ lệ của tổng thể

- Giả sử trong tổng thể có tỷ lệ phần tử có tính chất A là p chưa biết. Ta cần ước lượng p bằng khoảng tin cậy. Căn cứ vào mẫu có kích thước n, tỷ lệ mẫu của các phần tử có tính chất A là f. Tìm hai số p_1, p_2 sao cho:

$$P(p_1 < p < p_2) = 1 - \alpha$$

- Áp dụng qui tắc $k - \sigma$ $ta\ có\ công\ thức\ ước\ lượng$:

- Khi $nf \geq 5$ và $n(1-f) \geq 5$, áp dụng qui tắc $k - \sigma$ ta có công thức ước lượng cho p:

$$\left(f - Z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\,;\, f + Z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

- Với $\varphi\left(Z_{\frac{\alpha}{2}}\right) = \frac{1-\alpha}{2}$, tra trong bảng tích phân Laplace.

- $\varepsilon = Z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}$ gọi là độ chính xác (sai số) của ước lượng.

Tìm $Z_{\frac{\alpha}{2}}$ từ bảng 1:

Với độ tin cậy $1-\alpha$, ta có thể tìm giá trị $Z_{\frac{\alpha}{2}}$ từ bảng phân phối xác suất tích lũy (bảng 1) bằng cách dựa vào công thức:

$$\boxed{\Phi(z_{\frac{\alpha}{2}}) = P(Z < z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}}$$

Ví dụ 1:

Một khu dân cư có 2000 hộ gia đình. Để điều tra về nhu cầu tiêu dùng của một loại hàng hóa, người ta theo dõi 100 gia đình thấy có 60 gia đình có nhu cầu hàng này.Với độ tin cậy 95%, hãy ước lượng số gia đình trong khu dân cư trên có nhu cầu về loại hàng hóa trên.

Ví dụ 2:

Trước kỳ bầu cử tổng thống, người ta phỏng vấn ngẫu nhiên 1800 cử tri thì thấy có 1180 người ủng hộ cử tri A. Với độ tin cậy 99%, hỏi ứng cử viên A có thể thu được tối thiểu bao nhiêu phần trăm phiếu bầu?

III. Một số chỉ tiêu chính của bài toán ước lượng

1. Các chỉ tiêu chính của bài toán ước lượng trung bình

TH 1: $n > 30$ và σ^2 đã *biết*:

Độ chính xác: $\varepsilon = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} (*)$

Từ $(*)$ ta suy ra:

- Độ tin cậy:

- $Ta\ có\ Z_{\frac{\alpha}{2}} = \varepsilon \frac{\sqrt{n}}{\sigma}$ suy ra độ tin cậy $(1-\alpha)$ như sau:

III. Một số chỉ tiêu chính của bài toán ước lượng

- *Dựa vào bảng tích phân Laplace:*

  Từ $Z_{\frac{\alpha}{2}} = \varepsilon \frac{\sqrt{n}}{\sigma}$ suy ra độ tin cậy:

$$1 - \alpha = 2\varphi\left(Z_{\frac{\alpha}{2}}\right)$$

- *Dựa vào xác suất tích lũy (bảng 1):*

  Từ $Z_{\frac{\alpha}{2}} = \varepsilon \frac{\sqrt{n}}{\sigma}$ suy ra độ tin cậy $(1-\alpha)$ từ:

$$\boxed{\Phi(z_{\frac{\alpha}{2}}) = P(Z < z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}}$$

- Tìm cỡ mẫu tối thiểu: n= $\frac{\sigma^2}{\varepsilon^2} Z_{\frac{\alpha}{2}}^{2}$
- Trường hợp 2, 3 và 4 tương tự:

  Nếu σ *chưa biết* ta thay σ bởi S'.

2. Các chỉ tiêu chính của bài toán ước lượng tỷ lệ

Độ chính xác: $\varepsilon = Z_{\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}$ $(**)$

Từ $(**)$, ta suy ra:

- Độ tin cậy $(1-\alpha)$: $Z_{\frac{\alpha}{2}} = \varepsilon \sqrt{\frac{n}{f(1-f)}}$

suy ra $1-\alpha$ từ:

$$\boxed{\Phi(z_{\frac{\alpha}{2}}) = P(Z < z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}}$$

- Cỡ mẫu tối thiểu: $n = \left(\frac{Z_{\frac{\alpha}{2}}}{\varepsilon}\right)^2 f(1-f).$

Ví dụ :

- Để khảo sát chỉ tiêu X của một loại sản phẩm của xí nghiệp I, người ta quan sát một mẫu trong kho và có kết quả sau:

X(cm)	11 – 15	15 – 19	19 – 23	23 – 27	27 – 31	31 – 35	35 – 39
Số SP	8	9	20	16	16	13	18

a. Ước lượng giá trị trung bình của chỉ tiêu X của loại sản phẩm trên với độ tin cậy 96%.

b. Nếu ước lượng giá trị trung bình của X với độ chính xác 1,8cm thì sẽ đạt được độ tin cậy là bao nhiêu?

c. Nếu ước lượng giá trị trung bình của X với độ chính xác 1,5cm và độ tin cậy 99% thì phải điều tra thêm ít nhất bao nhiêu sản phẩm nữa?

d. Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được gọi là những sản phẩm loại B. Ước lượng giá trị trung bình của chỉ tiêu X của những sản phẩm loại B với độ tin cậy 98%.

e. Hãy ước lượng tỷ lệ sản phẩm loại B với độ tin cậy 92%. Bảng số liệu trên được chọn ngẫu nhiên từ một kho trong đó có 1000 sản phẩm loại B. Hãy ước lượng số sản phẩm có trong kho với độ tin cậy 92%.

f. Nếu ước lượng tỷ lệ những sản phẩm loại B với độ tin cậy 96% và độ chính xác 8% thì cần phải điều tra thêm bao nhiêu sản phẩm nữa?

g. Giả sử trong kho để lẫn 1000 sản phẩm của xí nghiệp II và trong 100 sản phẩm lấy từ kho có 9 sản phẩm của xí nghiệp II. Hãy ước lượng số sản phẩm của xí nghiệp I có trong kho với độ tin cậy 82%.

Chương 5: KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Nội dung

- I. Một số khái niệm
- II. Kiểm định trung bình
- III. Kiểm định tỷ lệ

I. Một số khái niệm

1. Kiểm định giả thuyết là bài toán đi xác định có nên chấp nhận hay bác bỏ một khẳng định về giá trị của một tham số của tổng thể.

2. Giả thuyết không, gọi tắt là giả thuyết, ký hiệu H_0, là một giả định thăm dò về tham số của tổng thể.

3. Đối thuyết, ký hiệu H_1, là khẳng định có trạng thái đối lập với giả thuyết.

- Đối thuyết là vấn đề người làm kiểm định cần thiết lập.

4. Sai lầm loại I và sai lầm loại II

- Bởi vì kiểm định giả thuyết dựa trên số liệu mẫu, nên có khả năng xảy ra những sai lầm.

- *Sai lầm loại I* là bác bỏ H_0 khi nó đúng.

- Xác suất mắc phải sai lầm loại một khi giả thuyết H_0 đúng bằng một đại lượng gọi là *mức ý nghĩa* của kiểm định, ký hiệu α.

- Những ứng dụng của kiểm định giả thuyết để kiểm soát sai lầm loại một thường được gọi là *kiểm định ý nghĩa.*

Sai lầm loại II

- *Sai lầm loại II* là sai lầm mà ta chấp nhận H_0 khi nó sai.

- Trong kiểm định, để hạn chế gặp phải sai lầm loại II, người ta thường sử dụng khẳng định "không bác bỏ H_0" và không dùng khẳng định "chấp nhận H_0".

II. Kiểm định giả thuyết cho trung bình

- Giả thuyết: luôn có trường hợp "=".

- Tổng quát, một bài toán kiểm định giả thuyết cho kỳ vọng μ sẽ có một trong 3 dạng dưới đây (với μ_0 là giá trị kiểm định của trung bình):

Một phía (Bên trái)	Một phía (Bên phải)	Hai phía
$H_0: \mu \geq \mu_0$	$H_0: \mu \leq \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu \neq \mu_0$

Giả sử tổng thể có trung bình μ chưa biết. Ta cần kiểm tra giả thuyết H_0: $\mu = \mu_0$ (cho trước). Dựa vào mẫu có kích thước n và qui tắc kiểm định, ta chấp nhận hay bác bỏ giả thuyết trên với mức ý nghĩa α.

Trong thực hành, ta chia 4 trường hợp:

- Trường hợp 1: $n > 30, \sigma^2$ đã biết:
- Đặt giả thuyết: H_0: $\mu = \mu_0$

- Tính giá trị $$t = \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}$$
- Tìm $t_{\frac{\alpha}{2}}$ từ công thức $\boxed{\Phi(t_{\frac{\alpha}{2}}) = P(Z < t_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}}$
- Nếu $t > t_{\frac{\alpha}{2}}$, ta bác bỏ giả thuyết
- Ngược lại, ta chấp nhận giả thuyết

- Trường hợp 2: $n > 30, \sigma^2$ *chưa biết*.

  Ta làm như trường hợp 1 nhưng thay σ *bởi* S'.

- Trường hợp 3: $n \leq 30, \sigma$ đã *biết*, X *có phân phối chuẩn*.

  Ta làm như trường hợp 1.

- Trường hợp 4: $n \leq 30, \sigma^2$ *chưa biết*, X *có phân phối chuẩn*.

Khi đó,ta làm tương tự trường hợp 1 nhưng thay S' bởi σ:

Kiểm định $t = \frac{|\bar{x} - \mu_0|}{\frac{S'}{\sqrt{n}}}$ so sánh với $t_{\frac{\alpha}{2}}^{n-1}$(tra trong bảng phân phối Student với (n-1) bậc tự do).

Ví dụ 1: Kem đánh răng P/S

- Kiểm định hai phía cho kỳ vọng: σ chưa biết

Dây chuyền sản xuất kem đánh răng P/S được thiết kế để đóng hộp những tuýp kem có trọng lượng trung bình là 168g. Một mẫu gồm 40 tuýp kem được chọn ngẫu nhiên để kiểm tra định kỳ.

Bộ phận điều khiển dây chuyền phải đảm bảo để trọng lượng trung bình mỗi tuýp kem là 168g; nếu nhiều hơn hoặc ít hơn, dây chuyền phải được điều chỉnh lại.

Ví dụ 1: Kem đánh răng P/S

- Kiểm định hai phía cho kỳ vọng: σ chưa biết

Giả sử trung bình mẫu của 40 tuýp kem là 170g và độ lệch tiêu chuẩn của mẫu S'= 5,6g. Thực hiện kiểm định giả thuyết với mức ý nghĩa 3% để xác định xem dây chuyền sản xuất có vận hành tốt hay không?

Ví dụ 2:

Trọng lượng trung bình khi xuất chuồng ở một trại chăn nuôi trước là 3,3kg/con. Năm nay người ta sử dụng một loại thức ăn mới, cân thử 15 con khi xuất chuồng ta được số liệu như sau:

3,25; 2,5; 4; 3,75; 3,8; 3,9; 4,02; 3,6; 3,8; 3,2; 3,82; 3,4; 3,75; 4; 3,5.

a. Hãy cho kết luận về tác dụng của loại thức ăn này với mức ý nghĩa 5%.

b. Nếu trại chăn nuôi báo cáo trọng lượng trung bình khi xuất chuồng là 3,5kg/con thì có chấp nhận được không với mức ý nghĩa 5%?

III. Kiểm định giả thuyết hai phía cho tỷ lệ

Giả sử tổng thể chia làm hai loại phần tử. Tỉ lệ phần tử có tính chất ta quan tâm là $p\ chưa\ biết$. Ta cần kiểm tra giả thuyết: $H_0: p = p_0$ (cho trước)

Dựa vào mẫu có kích thước n và qui tắc kiểm định, ta bác bỏ hay chấp nhận giả thuyết trên với mức ý nghĩa α.

Quy tắc thực hành:

- Đặt giả thuyết H_0: $p = p_0$, khi np_0 ≥5 và n$(1 - p_0) \geq 5$
- Từ mẫu cụ thể, ta tính tỉ lệ mẫu $f = \frac{m}{n}$ với m là số phần tử có tính ta quan tâm.
- Tính kiểm định: $t = \frac{|f - p_0|}{\sqrt{\frac{p_0 q_0}{n}}}$ với $q_0 = 1 - p_0$
- Tính $t_{\frac{\alpha}{2}}$ từ: $\boxed{\Phi(t_{\frac{\alpha}{2}}) = P(Z < t_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}}$
- Nếu $t > t_{\frac{\alpha}{2}}$, ta bác bỏ giả thuyết.

  Ngược lại, ta chấp nhận giả thuyết.

Ví dụ:

Tỷ lệ phế phẩm của một nhà máy trước đây là 5%. Năm nay nhà máy áp dụng một biện pháp kỹ thuật mới. Để nghiên cứu tác dụng của biện pháp kỹ thuật mới, người ta lấy một mẫu gồm 800 sản phẩm để kiểm tra thì thấy có 24 phế phẩm.

a. Hãy cho kết luận về phương pháp mới này với mức ý nghĩa 5%.
b. Nếu nhà máy báo cáo tỷ lệ phế phẩm sau khi áp dụng biện pháp kỹ thuật mới là 2% thì có chấp nhận được không với mưc ý nghĩa 5%?