BÀI GIẢNG
XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC

Ths Nguyễn Văn Du

CHƯƠNG MỞ ĐẦU

GIẢI TÍCH TỔ HỢP

§1 – CÁC KHÁI NIỆM CƠ BẢN

- **1.1 – BÀI TOÁN CỦA GIẢI TÍCH TỔ HỢP**
- Từ tập hợp $A = \{a_1, a_2, \ldots, a_n\}$ ta lấy ngẫu nhiên k phần tử kèm theo một điều kiện ràng buộc nào đó.
- Vấn đề đặt ra là: Hãy tính số cách chọn ra k phần tử đó
- Đây là bài toán cơ bản của giải tích tổ hợp

1.2 - NGUYEÂN LYÙ CỘNG

Neáu moät coâng vieäc ñöôïc chia thaønh k tröôøng hôïp thöïc hieän:

- Tröôøng hôïp 1: coù n_1 caùch thöïc hieän
- Tröôøng hôïp 2: coù n_2 caùch thöïc hieän
- ...
- Tröôøng hôïp k: coù n_k caùch thöïc hieän

Thì công vieäc đó có $n_1 + n_2 + \ldots + n_k$ cách thöïc hieän

1.3 – NGUYÊN LÝ NHÂN

- Neáu moät coâng vieäc ñöôïc chia laøm k giai ñoaïn ñeå thöïc hieän:
- Giai ñoaïn 1: coù n_1 caùch thöïc hieän
- Giai ñoaïn 2: coù n_2 caùch thöïc hieän
- ...
- Giai ñoaïn k: coù n_k caùch thöïc hieän
- Thì công vieäc đó có n_1 n_2 …n_k cách thöïc hieän

VÍ DỤ ÁP DỤNG

- Cho taäp hôïp: A = {0,1,2,3,4,5}
- Ngöôøi ta laäp moät soá töï nhieân coù 4 chöõ soá khaùc nhau ñoâi moät_
- a) Hoûi coù bao nhieâu soá ñöôïc laäp ?
- b) Trong caùc soá ñöôïc laäp coù bao nhieâu soá chaün, bao

GIAÛI

a) Giaû söû soá phaûi laäp coù daïng x = $a_1a_2a_3a_4$

ÔÛ vò trí a_1 ta coù 5 caùch choïn, coøn 5 chöõ soá

ÔÛ vò trí a_2 ta coù 5 caùch choïn, coøn 4 chöõ soá

ÔÛ vò trí a_3 ta coù 4 caùch choïn, coøn 3 chöõ soá

ÔÛ vò trí a_4 ta coù 3 caùch choïn

b) Giaû söû soá chaün phaûi laäp coù daïng

x = $a_1a_2a_3a_4$

Tröôøng hôïp 1:

Soá chaün coù taän cuøng laø soá 0: x = $a_1a_2a_3$0

- ÔÛ vò trí a_1 ta coù 5 caùch choïn, coøn 4 chöõ soá
- ÔÛ vò trí a_2 ta coù 4 caùch choïn, coøn 3 chöõ soá

Tröôøng hôïp 2:

Soá chaün coù taän cuøng laø soá khaùc 0:

x = $a_1a_2a_3a_4$

- ÔÛ vò trí a_4 ta coù 2 caùch choïn, coøn 5 chöõ soá
- ÔÛ vò trí a_1 ta coù 4 caùch choïn, coøn 4 chöõ soá
- ÔÛ vò trí a_2 ta coù 4 caùch choïn, coøn 3 chöõ soá

- Theo nguyeân lyù coäng ta coù 60 + 96 = 156 soá chaün ñöôïc laäp thoûa maõn ñeà baøi
- Do ñoù coù: 300 – 156 = 144 soá leû thoûa maõn ñeà baøi

§2 – CHỈNH HỢP VAØ HOAÙN VÒ

2.1 - ÑÒNH NGHÓA

Cho A laø taäp hôïp coù n phaàn töû.

1) Moãi caùch sắp xếp k phaàn töû cuûa A theo một trình tự nhất định ñöôïc goïi laø moät chỉnh hôïp chaäp k cuûa n phaàn töû ñoù.

2) Moãi caùch sắp xếp n phaàn töû cuûa A theo một trình tự nhất định ñöôïc goïi laø moät hoaùn vò cuûa n

2.2 - COAÂNG THÖÙC

1) Neáu ta goïi A_n^k laø soá caùc chỉnh hôïp chaäp k cuûa n phaàn töû thì ta coù coâng thöùc: $$A_n^k = \frac{n!}{(n-k)!}$$

2) Neáu ta goïi P_n laø soá caùc hoaùn vò cuûa n phaàn töû thì ta coù coâng thöùc:

$$P_n = n!$$

3 - Ví duï

Ví duï 1

Moät lôùp hoïc coù 30 sinh vieân. Ngöôøi ta thaønh laäp moät ban caùn söï coù 3 ngöôøi, trong ñoù moät ngöôøi laøm lôùp tröôûng, moät ngöôøi laø lôùp phoù, moät ngöôøi laøm thuû quyõ maø khoâng cho ai kieâm nhieäm.

Hoûi coù bao nhieâu caùch

Giaûi:

Moãi caùch thaønh laäp Ban caùn söï thoûa maõn ñeà baøi laø moät chænh hôïp chaäp 3 cuûa 30, do ñoù ta coù A_{30}^3 caùch thaønh laäp.

Cuï theå laø:

$$A_{30}^{3} = \frac{30!}{(30-3)!} = \frac{30.29.28.27!}{27!}$$

$$= 30.29.28 = 24360$$

Ví duï 2

Trong moät buoåi daï hoäi, coù 5 chaøng trai vaø 5 coâ gaùi muoán gheùp ñoâi moät caùch ngaãu nhieân ñeå thaønh laäp nhöõng caëp khieâu vuõ. Hoûi coù bao nhieâu caùch thaønh laäp caùc caëp khieâu vuõ nhö vaäy?

GIAÛI

Moãi caùch thaønh laäp nhöõng caëp khieâu vuõ chính laø moät hoaùn vò cuûa 5 phaàn töû. Do ñoù ta coù:

5! = 5.4.3.2.1 = 120 caùch

§3 - TOÅ HÔÏP

3.1 - Ñònh nghóa

Cho A laø taäp hôïp coù n phaàn töû. Moãi caùch thaønh laäp moät taäp hôïp coù k phaàn töû cuûa A ñöôïc goïi laø moät toå hôïp chaäp k cuûa n phaàn töû ñoù.

3.2 - Coâng thöùc

Neáu ta goïi C_n^k laø soá caùc toå hôïp chaäp k cuûa n phaàn töû thì ta coù coâng thöùc:

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

3 – Tính chất cơ bản

$$C_n^k = C_n^{n-k} \qquad \left(k = \overline{0, n-1}\right)$$

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^{n-k} \qquad \left(k = \overline{0, n-1}\right)$$

$$C_n^0 = C_n^n = 1$$

$$C_n^1 = n$$

- Nhị thức Newton

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

- Suy ra

$$C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$$

3.3 - Ví duï

Moät lôùp hoïc coù 30 sinh vieân. Ngöôøi ta thaønh laäp moät ban caùn söï coù 3 ngöôøi

Hoûi coù bao nhieâu caùch thaønh laäp?

Giaûi

Moãi caùch thaønh laäp ban caùn söï nhö vaäy laø moät toå hôïp chaäp 3 cuûa 30. Do ñoù ta coù:

$$C_{30}^{3}=\frac{30!}{(30-3)!3!}=\frac{30.29.28.27!}{27!.3.2.1}=\frac{30.29.28}{6}=4060 \text{ caùch}$$

Chương 1

CÁC KHÁI NIỆM CƠ BẢN CỦA LÝ THUYẾT XÁC SUẤT

PHẦN A

BIẾN CỐ NGẪU NHIÊN

§1 - BIẾN CỐ NGẪU NHIÊN

- **1.1 – Phép thử và biến cố**
- **1 – Định nghĩa**
- Một thí nghiệm dùng để nghiên cứu một đại lượng hay một hiện tượng nào đó được gọi là phép thử. Ký hiệu một phép thử là T
- Mỗi phép thử đều cho ta một kết cục. Kết cục đó được gọi là một biến cố ngẫu nhiên. Ký hiệu biến cố ngẫu nhiên là A, B, C …
- **2 – Ví dụ**
- Tung một đồng tiền đồng chất cân đối là một phép thử. Kết cục xảy ra là: Đồng tiền xuất hiện

- Mặt sấp (S) hoặc xuất hiện mặt ngửa (N).Ta có: S và N là những biến cố
- Gieo một con xúc sắc đồng chất cân đối là một phép thử. Kết cục có thể xảy ra là: Con xúc sắc xuất hiện mặt một chấm A_1, hai chấm A_2, ba chấm A_3, bốn chấm A_4, năm chấm A_5, sáu chấm A_6. Ta có: A_1, A_2, A_3, A_4, A_5, A_6 là những biến cố

1.2 – Các loại biến cố

- **1 – Biến cố sơ cấp:** Là những biến cố loại trừ nhau trong cùng một phép thử
- Tập hợp các biến cố sơ cấp của một phép thử còn gọi là không gian các biến cố sơ cấp và ký hiệu là Ω
- **Ví dụ:**
- Tung một đồng tiền đồng chất cân đối ta thấy không gian các BCSC của phép thử này là:
- $$\Omega = \{N,S\}$$
- Tung một con xúc sắc đồng chất cân đối ta thấy không gian các BCSC của phép thử này là:
- $$\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$$

- **2 – Biến cố chắc chắn:** Là biến cố nhất định phải xảy ra khi phép thử được thực hiện. Ký hiệu là Ω
- **3 – Biến cố không thể có:** Là biến cố không thể xảy ra khi phép thử được thực hiện. Ký hiệu là $\emptyset$
- **4 – Biến cố đồng khả năng:** Là những biến cố có khả năng xuất hiên ngang nhau khi thực hiện một phép thử

§2 – CÁC PHÉP TOÁN VỀ BIẾN CỐ

- **2.1 – Tổng của các biến cố**
- **1- Định nghĩa**
- Tổng của hai biến cố A và B trong cùng một phép thử là một biến cố C ký hiệu là C = A + B. Biến cố này xảy ra khi và chỉ khi có ít nhất một biến cố A hoặc B xảy ra khi phép thử đươc thực hiện
- Tổng của n biến cố A_1, A_2, … , A_n trong cùng một phép thử là một biến cố C ký hiệu là
- $C = A_1 + A_2 + \ldots + A_n$. Biến cố này xảy ra khi và chỉ khi có ít nhất một biến cố A_i nào đó xảy ra khi phép thử được thực hiện

2 – Tính chất

- Cho A, B, C là những biến cố trong cùng phép thử ta có:
- $A + B = B + A$
- $(A + B) + C = A + (B + C)$
- $A + A = A$
- $A + \varnothing = A$

3 – Ghi chú

- Một biến cố sơ cấp không thể biểu diễn được dưới dạng tổng của các biến cố khác

Ví dụ

- Tung một con xúc sắc đồng chất, cân đối ta thấy:
- A_1, A_2, A_3 , A_4, A_5 , A_6 là những biến cố sơ cấp. Ta thấy các biến cố này không biểu diễn được thành tổng của các biến cố khác
- Gọi C, L tương ứng là các biến cố con xúc sắc xuất hiện mặt chẵn hay lẻ chấm. Ta có:
- $C = A_2 + A_4 + A_6$; $L = A_1 + A_3 + A_5$
- Như vậy C và L không phải là các biến cố sơ cấp

2.2 – Tích của các biến cố

- **1 – Định nghĩa**
- Tích của hai biến cố A và B trong cùng một phép thử là một biến cố C ký hiệu là C = A B. Biến cố này xảy ra khi và chỉ khi cả hai biến cố A và B xảy ra khi phép thử được thực hiện
- Tích của n biến cố $A_1, A_2, \dots, A_n$ trong cùng một phép thử là một biến cố C ký hiệu là
- $$C = A_1 A_2 \dots A_n$$
- Biến cố này xảy ra khi và chỉ khi tất cả các biến cố A_i đều xảy ra khi phép thử được thực hiện

2 – Tính chất

- Cho A, B, C là những biến cố trong cùng phép thử ta có:
- A B = B A
- (A B)C = A(B C)
- A A = A; A Ø = Ø; A Ω = A
- A(B+C) = AB + AC
- A+BC = (A+B)(A+C)

Ví dụ

- Lớp học có 30 sinh viên dự thi môn XSTK; Gọi A_i là biến cố sinh viên i thi đậu; A là biến cố có ít nhất một sinh viên đậu, B là biến cố tất cả sinh viên đều thi đậu.
- Ta có:
- $A = A_1 + A_2 + \ldots + A_{30}$
- $B = A_1 A_2 \ldots A_{30}$

§3 – MỐI QUAN HỆ GIỮA CÁC BIẾN CỐ

- **3.1 – Quan hệ kéo theo và quan hệ bằng nhau**
- Nếu biến cố A xảy ra luôn luôn làm cho biến cố B xảy ra thì ta nói biến cố A kéo theo biến cố B và ký hiệu là A0B
- Nếu biến cố A kéo theo biến cố B và ngược lại biến cố B kéo theo biến cố A thì ta nói biến cố A bằng biến cố B và ký hiệu là A = B
- **3.2 – Biến cố xung khắc**
- Hai biến cố A và B trong cùng phép thử được gọi là xung khắc với nhau nếu A và B không đồng thời xảy ra khi phép thử được thực hiện. Ký hiệu AB = Ø

- Các biến cố $A_1, A_2, \ldots, A_n$ trong cùng phép thử được gọi là xung khắc nhau từng đôi một nếu bất kỳ hai biến cố nào trong hệ cũng xung khắc với nhau
- **3.3 – Quan hệ đối lập**
- Hai biến cố A và B trong cùng phép thử được gọi là đối lập với nhau nếu chúng là những biến cố xung khắc và khi thực hiện phép thử chỉ xuất hiện biến cố A hoặc biến cố B. Ký hiệu là

$$A = \overline{B} \ hay \ B = \overline{A}$$

- Vậy:

$$A = \overline{B} \Leftrightarrow \begin{cases} AB = \varnothing \\ A + B = \Omega \end{cases}$$

3.4 – Hệ đầy đủ các biến cố

- Hệ các biến cố $A_1, A_2, \ldots, A_n$ trong cùng một phép thử được gọi là một hệ đầy đủ các biến cố nếu chúng là một hệ xung khắc với nhau từng đôi một và khi phép thử được thực hiện chỉ xuất hiện một trong các biến cố A_i
- Vậy:

$A_1, A_2, ..., A_n$ laø moät heä ñaày ñuû

$$\Leftrightarrow \begin{cases} \forall i \neq j : A_i A_j = \varnothing & \left(\forall i, j = \overline{1, n}\right) \\ A_1 + A_2 + \ldots + A_n = \Omega \end{cases}$$

3.5 – Tính chất (Luật đối ngẫu Đơmoocgăng)

$$\overline{A+B} = \overline{A}.\overline{B}$$

$$\overline{A_1 + A_2 + \ldots + A_n} = \overline{A_1}.\overline{A_2}.\ldots.\overline{A_n}$$

$$\overline{A.B} = \overline{A} + \overline{B}$$

$$\overline{A_1.A_2\ldots A_n} = \overline{A_1} + \overline{A_2} + \ldots + \overline{A_n}$$

PHẦN B

XÁC SUẤT

§1 – CÁC ĐỊNH NGHĨA XÁC SUẤT

- **1.1 – Định nghĩa xác suất theo lối cổ điển**
- Giả sử một phép thử T có n biến cố sơ cấp đồng khả năng; A là biến cố trong cùng phép thử và có m biến cố sơ cấp có lợi cho A (nghĩa là số khả năng xảy ra biến cố A)
- Ta gọi tỉ số m/n là xác suất của biến cố A và ký hiệu là p(A)

$$p\left(A\right)=\frac{\text{Số BCSC có lợi cho A}}{\text{Số BCSC đồng khả năng}}$$

1.2 – Định nghĩa xác suất bằng thống kê

- Một phép thử T được lặp lại nhiều lần trong những điều kiên giống nhau. Nếu trong n lần thực hiện phép thử có k lần xuất hiện biến cố A thì tỉ số $f_n(A) = k/n$ được gọi là tần suất xuất hiện biến cố A trong n lần thử
- Khi số phép thử n tăng lên vô hạn thì $f_n(A)$ dao động xung quanh một giá trị ổn định. Ta gọi giá trị đó là xác suất của biến cố A và ký hiệu là p(A)

1.3 – Định nghĩa xác suất bằng hình học

- Ta coi một hình chữ nhật là biến cố chắc chắn Ω; mỗi điểm trong hình chữ nhật được coi là biến cố sơ cấp; mỗi miền con A của hình chữ nhật được coi là biên cố ngẫu nhiên; tập hợp Ø được coi là biến cố không thể có.
- Ta định nghĩa xác suất của biến cố A là tỉ số giữa diện tích của miền con A và diện tích của miền Ω

1.4 – Các tính chất cơ bản của xác suất

- Với mọi biến cố A ta có:

$$0 \le p(A) \le 1$$

$$p(\Omega) = 1$$

$$p(\phi) = 0$$

$$p(\overline{A}) = 1 - p(A)$$

1.5 – Ví dụ minh họa

Ví dụ 1

- Một bình đựng 12 bi trong đó có 8 bi trắng và 4 bi đỏ. Lấy ngẫu nhiên từ trong bình ra 4 bi
- a) Tính xác suất để 4 bi được chọn có đúng ba bi trắng
- b) Tính xác suất để 4 bi được chọn có ít nhất ba bi trắng

Giải

- a) Gọi A là biến cố 4 bi được chọn có đúng 3 bi trắng ta có: p(A) = m/n
- n là số BCSC đồng khả năng. Đó chính là số trường hợp chọn 4 bi từ 12 bi mà không phân biệt mầu. Ta có: $n = C_{12}^{4} = 495$
- m là số BCSC có lợi cho A. Đó chính là số trường hợp chọn 4 bi từ 12 bi trong đó có đúng ba bi mầu trắng. Ta có: $m = C_{8}^{3} C_{4}^{1} = 224$
- Suy ra:
- p(A) = m/n = 224/495 = 0,453 (45,3%)

b) Gọi B là biến cố có ít nhất ba bi trắng.Ta có:

$$p(B) = m/n$$

- n là số BCSC đồng khả năng. Đó chính là số trường hợp chọn 4 bi từ 12 bi mà không phân biệt mầu. Ta có: $n = C_{12}^{4} = 495$
- m là số BCSC có lợi cho B. Đó chính là số trường hợp chọn 4 bi từ 12 bi trong đó có ít nhất ba bi mầu trắng. Có 2 trường hợp xảy ra:
- TH1: 3 trắng – 1 đỏ: Trường hợp này có $C_{8}^{3} C_{4}^{1}$ cách chọn
- TH2: 4 trắng – 0 đỏ: Trường hợp này có $C_{8}^{4} C_{4}^{0}$ cách chọn

Suy ra:

- $m = C_8^3 C_4^1 + C_8^4 C_4^0 = 224 + 70 = 294$
- Vậy p(B) = m/n = 294/495 = 59,4%
- **Ví dụ 2:**
- Một tập hợp có N phần tử, trong đó có N_A phần tử loại A. Người ta chọn ngẫu nhiên n phần tử.
- Tính xác suất sao cho trong n phần tử chọn ra có đúng k phần tử loại A

Giải

- Gọi F là biến cố: "*Trong n phần tử chọn ra có đúng k phần tử loại A*". Ta có:

$$p(F) = \frac{\mathcal{M}}{\mathcal{A}}$$

$$\begin{cases} \mathcal{M} = C_{N_A}^{k} C_{N-N_A}^{n-k} \\ \mathcal{A} = C_N^n \end{cases} \Rightarrow p(F) = \frac{C_{N_A}^{k} C_{N-N_A}^{n-k}}{C_N^n}$$

- Người ta thường gọi đây là công thức xác suất lựa chọn

§2 – CÁC CÔNG THỨC XÁC SUẤT

- **2.1 – Công thức cộng xác suất**
- **2.1.1 – Công thức**
- **1 – Các biến cố xung khắc**
- Nếu A và B là hai biến cố xung khắc với nhau ta có:
- $$p(A+B) = p(A) + p(B)$$
- Nếu A_1, A_2, … , A_n là những biến cố xung khắc nhau từng đôi một ta có:
- $p(A_1 + A_2 + \ldots + A_n) = p(A_1) + p(A_2) + \ldots + p(A_n)$

- **2 – Các biến cố bất kỳ**
- Cho A và B là những biến cố bất kỳ ta có:
- $$p(A+B) = p(A) + p(B) - p(AB)$$
- Cho A, B và C là những biến cố bất kỳ ta có:
- $$p(A+B+C) = p(A) + p(B) + p(C) - p(AB) - p(AC) - p(BC) + p(ABC)$$

3 - Ví dụ

- **Ví duï 1**
- *Moät loâ haøng coù 100 saûn phaåm trong ñoù coù 5 pheá phaåm. Loâ haøng ñöôïc chaáp nhaän neáu choïn ngaãu nhieân ra 50 saûn phaåm ñeå kieåm tra thì soá pheá phaåm khoâng quaù 1.*
- *Tìm xaùc suaát ñeå loâ haøng ñöôïc chaáp nhaän.*
- ***Giải***

- A_0 laø *bieán coá trong 50 saûn phaåm khoâng coù pheá phaåm naøo*; A_1 laø *bieán coá trong 50 saûn phaåm coù 1 pheá phaåm.*
- Khi ñoù A_0, A_1 laø hai bieán coá xung khaéc vaø ta coù: $A = A_0 + A_1$. Suy ra:

$$P(A) = P(A_0 + A_1) = P(A_0) + P(A_1)$$

$$= \frac{C_5^0 \times C_{95}^{50}}{C_{100}^{50}} + \frac{C_5^1 \times C_{95}^{49}}{C_{100}^{50}} = 0,18(18\%)$$

Ví dụ 2

- *Thaêm doø 100 ngöôøi trong moät Caâu Laïc Boä thaáy coù 80 ngöôøi thích nhaïc Vaên Cao, 70 ngöôøi thích nhaïc Trònh Coâng Sôn, 60 ngöôøi thích nhaïc cuûa caû hai oâng. Choïn ngaãu nhieân 1 ngöôøi trong soá ñöôïc thaêm thaêm doø.*
- *Tính xaùc suaát ñeå ngöôøi naøy thích nhaïc cuûa ít nhaát 1 trong 2 nhaïc só.*

Giải

- Goïi: A_1 laø *Bieán coá ngöôøi ñöôïc choïn thích nhaïc Vaên Cao*; A_2 laø *Bieán coá ngöôøi ñöôïc choïn thích nhaïc Trònh Coâng Sôn*; F laø *Bieán coá ngöôøi ñöôïc choïn thích nhaïc cuûa ít nhaát 1 trong 2 nhaïc só.*
- Ta thaáy A_1, A_2 khoâng phaûi laø hai bieán coá xung khaéc vaø: $F = A_1 + A_2$

$$\Rightarrow P(F) = P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$$

$$= \frac{80}{100} + \frac{70}{100} - \frac{60}{100} = \frac{90}{100} = 0,9$$

2.2 – Công thức nhân xác suất

- **2.2.1 – Xác suất có điều kiện**
- **1 – Định nghĩa**
- Cho A và B là những biến cố trong cùng phép thử. Xác suất của biến cố A khi biến cố B xảy ra được gọi là xác suất có điều kiện của A theo B. Ký hiệu là p(A/B)
- **2 - Ví dụ**
- Năm người bắt thăm mua 3 món hàng cùng loại. Hỏi người bắt thăm trước hay sau có lợi thế hơn?

Giải

- Gọi A_i (i =1,..,5) là biến cố người thứ i bắt trúng.
  Ta thấy xác suất bắt trúng của từng người như sau:

$$p(A_1) = \frac{3}{5} \qquad p(A_2 / A_1) = \frac{2}{4} \qquad p(A_2 / \overline{A_1}) = \frac{3}{4}$$

$$p(A_3 / A_1 A_2) = \frac{1}{3} \qquad p\left(A_3 / \overline{A_1} A_2\right) = p\left(A_3 / A_1 \overline{A_2}\right) = \frac{2}{3}$$

$$p(A_3 / \overline{A_1}\,\overline{A_2}) = \frac{3}{3}$$

2.2.2 – Công thức nhân xác suất thứ nhất

- Cho A và B là hai biến cố bất kỳ trong cùng phép thử ta có: $p(AB) = p(A)p(B/A) = p(B)p(A/B)$
- Cho $A_1, A_2, \ldots, A_n$ là những biến cố bất kỳ trong cùng phép thử ta có:
- $p(A_1 A_2 \ldots A_n) = p(A_1)p(A_2/A_1)p(A_3/A_1A_2) \ldots$
- $p(A_n / A_1 A_2 \ldots A_{n-1})$

2.2.3 – Công thức nhân xác suất thứ hai

- **1 – Sự độc lập của các biến cố**
- 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 việc xảy ra hay không xảy ra của biến cố kia
- Các biến cố A_1, A_2, … , A_n được gọi là độc lập từng đôi nếu một cặp hai biến cố bất kỳ trong hệ đều độc lập với nhau

- Các biến cố $A_1, A_2, \dots, A_n$ được gọi là độc lập toàn thể nếu việc xảy ra hay không xảy ra của một nhóm k biến cố bất kỳ trong hệ không ảnh hưởng đến việc xảy ra hay không của các biến cố khác trong hệ đó
- **2 – Công thức nhân xác suất thứ hai**
- Cho A và B là hai biến cố độc lập. Ta có:
- $$p(AB) = p(A)p(B)$$
- Cho $A_1, A_2, \dots, A_n$ là hệ độc lập toàn thể ta có:
- $$p(A_1A_2 \dots A_n) = p(A_1)p(A_2) \dots p(A_n)$$

VÍ DỤ ÁP DỤNG

Ví dụ 1

Moät chaøng trai vieát thö cho 3 coâ baïn gaùi, do

ñaõng trí neân anh ta boû thư vaøo phong bì moät

caùch ngaãu nhieân.

Tính xaùc suaát ñeå khoâng coâ naøo nhaän ñöôïc ñuùng

thö cuûa mình.

Giải

- Goïi A_i laø *bieán coá coâ thöù i nhaän ñuùng thö cuûa*
- *mình* (i = 1, 2, 3)
- Goïi F laø *bieán coá khoâng coâ naøo nhaän ñöôïc ñuùng thö cuûa mình*. Ta coù:

$$F = \overline{A}_1.\overline{A}_2.\overline{A}_3 \quad \Rightarrow p\left(F\right) = p\left(\overline{A}_1.\overline{A}_2.\overline{A}_3\right)$$

$$= p\left(\overline{A}_1\right).p\left(\overline{A}_2 / \overline{A}_1\right).p\left(\overline{A}_3 / \overline{A}_1.\overline{A}_2\right) = \frac{2}{3}.\frac{1}{2}.\frac{1}{1} = \frac{1}{3}$$

Ví dụ 2

- Một thủ kho có một chùm chìa khóa gồm 9 chiếc bề ngoài giống hệt nhau, trong đó chỉ có 2 chiếc mở được. Anh ta thử ngẫu nhiên từng chìa, chiếc nào thử không trúng thì bỏ ra.
- Tính xác suất để người thủ kho mở được ở lần mở thứ 3

Giải

- Gọi A_i là biến cố người thủ kho mở được ở lần mở thứ i (i = 1…9); F là biến cố người thủ kho mở được cửa ở lần thứ 3.
- Khi đó ta có biểu diễn:

$$F = \overline{A_1}.\overline{A_2}.A_3 \Rightarrow p\left(F\right) = p\left(\overline{A_1}.\overline{A_2}.A_3\right)$$

$$= p\left(\overline{A_1}\right) p\left(\overline{A_2} / \overline{A_1}\right) p\left(A_3 / \overline{A_1}.\overline{A_2}\right)$$

$$= \frac{7}{9} \times \frac{6}{8} \times \frac{2}{7} = \frac{1}{6}$$

- **Ví dụ 3**
- Lô hàng chứa 20 sản phẩm trong đó có 2 sản phẩm xấu. Chọn lần lượt không hòan lại mỗi lần một sản phẩm cho đến khi thấy đủ 2 sản phẩm xấu thì dừng.
- Tính các suất để việc chọn dừng lại ở bước chọn thứ 4

Giải

- Gọi F là biến cố việc chọn dừng lại ở bước thứ 4
- Gọi X_i là biến cố sản phẩm xấu chọn ở bước thứ i (i = 1…20) ta có:

$$F = X_1\overline{X}_2\overline{X}_3X_4 + \overline{X}_1X_2\overline{X}_3X_4 + \overline{X}_1\overline{X}_2X_3X_4$$

$$\Rightarrow P(F) = P\left(X_1\overline{X}_2\overline{X}_3X_4\right) + P\left(\overline{X}_1X_2\overline{X}_3X_4\right) + P\left(\overline{X}_1\overline{X}_2X_3X_4\right)$$

- Tính các xác suất:

$$P\left(X_1\overline{X}_2\overline{X}_3X_4\right) = P(X_1)\,P\left(\overline{X}_2/X_1\right)P\left(\overline{X_3}/X_1\overline{X}_2\right)P\left(X_4/X_1\overline{X}_2\overline{X}_3\right)$$

$$= \frac{2}{20}\times\frac{18}{19}\times\frac{17}{18}\times\frac{1}{17} = \frac{1}{190}$$

$$P\left(\overline{X}_1 X_2 \overline{X}_3 X_4\right) = P\left(\overline{X}_1\right) P\left(X_2 / \overline{X}_1\right) P\left(\overline{X}_3 / \overline{X}_1 X_2\right) P\left(X_4 / \overline{X}_1 X_2 \overline{X}_3\right)$$

$$= \frac{18}{20} \times \frac{2}{19} \times \frac{17}{18} \times \frac{1}{17} = \frac{1}{190}$$

$$P\left(\overline{X}_1 \overline{X}_2 X_3 X_4\right) = P\left(\overline{X}_1\right) P\left(\overline{X}_2 / \overline{X}_1\right) P\left(X_3 / \overline{X}_1 \overline{X}_2\right) P\left(X_4 / \overline{X}_1 \overline{X}_2 X_3\right)$$

$$= \frac{18}{20} \times \frac{17}{19} \times \frac{2}{18} \times \frac{1}{17} = \frac{1}{190}$$

$$\Rightarrow P\left(F\right) = \frac{1}{190} + \frac{1}{190} + \frac{1}{190} = \frac{3}{190} = 0,016 \quad \left(1,6\%\right)$$

Ví dụ 4

- Một xưởng có ba máy làm việc. Trong một ca máy thứ nhất cần sửa chữa với xác suất là 0,15; máy thứ hai là 0,1; máy thứ ba là 0,12.
- Tính xác suất sao cho trong một ca làm việc có ít nhất một máy cần sửa chữa
- **Giải**
- Gọi A_i là biến cố máy thứ i cần phải sửa chữa
- (i = 1,2,3). Ta thấy A_1, A_2, A_3 độc lập
- Gọi F là *biến cố có ít nhất một máy cần sửa chữa*
- Khi đó ta có:

$$F = A_1 + A_2 + A_3 \Rightarrow \overline{F} = \overline{A_1 + A_2 + A_3} = \overline{A}_1 \overline{A}_2 \overline{A}_3$$

$$\Rightarrow P\left(\overline{F}\right) = P\left(\overline{A}_1\right) P\left(\overline{A}_2\right) P\left(\overline{A}_3\right)$$

$$= \left[1 - P\left(A_1\right)\right]\left[1 - P\left(A_2\right)\right]\left[1 - P\left(A_3\right)\right]$$

$$= \left(1 - 0{,}15\right) \times \left(1 - 0{,}1\right) \times \left(1 - 0{,}12\right)$$

$$= 0{,}85 \times 0{,}9 \times 0{,}88 = 0{,}6732$$

$$\Rightarrow P\left(F\right) = 1 - P\left(\overline{F}\right) = 1 - 0{,}6732 = 0{,}3268 \ \left(32{,}68\%\right)$$

Ví dụ 5

- Một người có ba viên đạn bắn độc lập vào một mục tiêu. Xác suất bắn trúng mục tiêu của mỗi viên đều bằng nhau và bằng 0,7.
- Tính xác suất để ba viên đạn bắn ra:
- A) Không có viên nào trúng mục tiêu
- B) Có một viên trúng mục tiêu
- C) Có hai viên trúng mục tiêu
- D) Có ba viên trúng mục tiêu

Giải

- Gọi A_i là biến cố viên thứ I trúng mục tiêu (i=1,2,3). Ta thấy A_1, A_2, A_3 độc lập
- Gọi A, B, C, D lần lượt là biến cố có 0, 1, 2, 3 viên trúng mục tiêu. Ta có:

$$A = \overline{A}_1 \overline{A}_2 \overline{A}_3$$

$$\Rightarrow P(A) = P\left(\overline{A}_1\right) P\left(\overline{A}_2\right) P\left(\overline{A}_3\right)$$

$$= 0,3 \times 0,3 \times 0,3 = 0,027$$

$$B = \overline{A}_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 \overline{A}_3 + \overline{A}_1 \overline{A}_2 A_3$$

$$\Rightarrow P(B) = P\left(\overline{A}_1\right) P(A_2) P\left(\overline{A}_3\right)$$

$$+ P(A_1) P\left(\overline{A}_2\right) P\left(\overline{A}_3\right) + P\left(\overline{A}_1\right) P\left(\overline{A}_2\right) P(A_3)$$

$$= 0,3 \times 0,7 \times 0,3 + 0,7 \times 0,3 \times 0,3$$

$$+ 0,3 \times 0,3 \times 0,7 = 0,189$$

$$C = A_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 A_3 + \overline{A}_1 A_2 A_3$$

$$\Rightarrow p(C) = p(A_1)\, p(A_2)\, p(\overline{A}_3)$$

$$+ p(A_1)\, p(\overline{A}_2)\, p(A_3) + p(\overline{A}_1)\, p(A_2)\, p(A_3)$$

$$= 0,7 \times 0,7 \times 0,3 + 0,7 \times 0,3 \times 0,7$$

$$+ 0,3 \times 0,7 \times 0,7 = 0,441$$

Tương tự

$$D = A_1 A_2 A_3 \quad \Rightarrow p\left(D\right) = p\left(A_1\right) p\left(A_2\right) p\left(A_3\right)$$

$$= 0,7 \times 0,7 \times 0,7 = 0,343$$

Ví dụ 6

- Một người có 3 viên đạn bắn độc lập vào một mục tiêu; xác suất bắn trúng mục tiêu đều bằng nhau và bằng 0,6. Người đó bắn theo nguyên tắc: Nếu bắn trúng mục tiêu hay bắn hết đạn thì dừng lại. Tính xác suất để người đó bắn ra:
- 1 viên
- 2 viên
- 3 viên

Giải

- Gọi A_i là biến cố viên thứ I trúng mục tiêu (i=1,2,3). Ta thấy A_1, A_2, A_3 độc lập
- Gọi A, B, C lần lượt là biến cố xạ thủ bắn ra 1, 2, 3. Ta có:

$$A = A_1 \Rightarrow P(A) = P(A_1) = 0,6$$

$$B = \overline{A_1} A_2 \Rightarrow P(B) = P\left(\overline{A_1}\right) P(A_2) = 0,4 \times 0,6 = 0,24$$

$$C = \overline{A_1}\,\overline{A_2} A_3 + \overline{A_1}\,\overline{A_2}\,\overline{A_3} = \overline{A_1}\,\overline{A_2}\left(A_3 + \overline{A_3}\right) = \overline{A_1}\,\overline{A_2}$$

$$\Rightarrow P(C) = P\left(\overline{A_1}\,\overline{A_2}\right) = P\left(\overline{A_1}\right) P\left(\overline{A_2}\right) = 0,4 \times 0,4 = 0,16$$

- **2.3 – Công thức xác suất Becnuli**
- **2.3.1 – Dãy phép thử Becnuli**
- Một dãy n phép thử được gọi là dãy phép thử Becnuli nếu thỏa mãn các điều kiện sau:
- Dãy các phép thử được tiến hành độc lập với nhau
- Trong qúa trình thực hiện phép thử chỉ xuất hiện biến cố A hoặc biến cố đối lập với biến cố A
- Xác suất xuất hiện biến cố A (xác suất thành công) trong mỗi lần thử đều bằng một hằng số p không đổi

2.3.2 – Công thức xác suất Becnuli

- Cho một dãy n phép thử Becnuli với xác suất thành công là p.
- Khi đó xác suất để biến cố A trong phép thử đó xuất hiện đúng k lần được tính theo công thức:

$$P_n(k) = C_n^k p^k (1-p)^{n-k} \quad (0 \le k \le n)$$

Bài toán về công thức xác suất Becnuli

- **Ví dụ**
- Một cuộc thăm dò cho thấy tỉ lệ những hộ dân có sử dụng loại sản phẩm X trong thành phố là 65%. Chọn ngẫu nhiên 12 hộ dân trong thành phố. Tính xác suất để 12 hộ dân này:
- A) Có 5 hộ sử dụng loại sản phẩm X
- B) Có ít nhất hai gia đình sử dụng loại sản phẩm X

Giải

- Mỗi lần kiểm tra một hộ dân ta coi đó là một phép thử, ta có số phép thử là n =12. Xác suất để một hộ có sử dụng loại sản phẩm X trong mỗi lần thử là p = 0,65. Vậy việc chọn kiểm tra hộ dân như trên là dãy phép thử Becnuli với số lần thử là n = 12 và xác suất thành công là 0,65
- A)Gọi A là biến cố có đúng 5 hộ dân sử dụng sản phẩm X. Theo công thức xác suất Becnuli ta có

$$p(A) = p_{12}(5) = C_{12}^{5}(0,65)^{5}(1-0,65)^{12-5} = 0,059(5,9\%)$$

- B) Gọi B là biến cố có ít nhất hai hộ sử dụng sản phẩm X. Khi đó biến cố đối lập của B là biến cố có không quá 1 hộ sử dụng và ta có:

$$p\left(\overline{B}\right) = p_{12}(0) + p_{12}(1)$$

$$= C_{12}^{0}\left(0,65\right)^{0}\left(1-0,65\right)^{12-0} + C_{12}^{1}\left(0,65\right)^{1}\left(1-0,65\right)^{12-1}$$

$$= 0,001$$

$$\Rightarrow p(B) = 1 - p\left(\overline{B}\right) = 1 - 0,001 = 0,999\left(99,9\%\right)$$

Ví dụ 2

- Một xạ thủ bắn độc lập 100 phát đạn vào một mục tiêu; xác suất bắn trúng mục tiêu của mỗi viên đều bằng nhau và bằng 0,8.
- Tính xác suất để 100 viên đạn mà người đó bắn ra có 80 viên trúng mục tiêu.
- **Giải**
- Mỗi lần bắn ra một viên đạn ta coi như là một phép thử, ta có số phép thử là n = 100

- Xác suất để một viên trúng mục tiêu trong mỗi lần thử là p = 0,8. Vậy việc bắn súng như trên là dãy phép thử Becnuli với số lần thử là n = 100 và xác suất thành công là p = 0,8.
- Gọi A là biến cố “có 80 viên trúng mục tiêu”.
- Theo công thức Becnuli ta có:

$$p(A) = p_{100}(80)$$

$$= C_{100}^{80}(0,8)^{80}(1-0,8)^{100-80} = 0,099(9,9\%)$$

2.4 – Công thức xác suất đầy đủ - Công thức Bayes

- **1- Công thức xác suất đầy đủ**
- Cho $A_1, A_2, \ldots, A_n$ là một hệ đầy đủ các biến cố; F là một biến cố bất kỳ trong cùng phép thử. Khi đó ta có công thức:

$$P(F) = P(A_1)P(F/A_1) + P(A_2)P(F/A_2) + \ldots$$
$$+ P(A_n)P(F/A_n)$$

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

- **2 – Công thức Bayes**
- Cho $A_1, A_2, \ldots, A_n$ là một hệ đầy đủ các biến cố; F là một biến cố bất kỳ trong cùng phép thử. Khi đó ta có công thức:

$$P(A_k / F) = \frac{P(A_k)P(F / A_k)}{P(F)} \quad \left(1 \le k \le n\right)$$

Chứng minh

- 1) Vì $A_1, A_2, \ldots, A_n$ là một hệ đầy đủ nên:

$$A_1 + A_2 + \ldots + A_n = \Omega$$

$$\Rightarrow F\Omega = F\left(A_1 + A_2 + \ldots + A_n\right)$$

$$\Leftrightarrow F = FA_1 + FA_2 + \ldots + FA_n$$

$$\Rightarrow P\left(F\right) = P\left(FA_1\right) + P\left(FA_2\right) + \ldots + P\left(FA_n\right)$$

$$= P\left(A_1\right)P\left(F/A_1\right) + P\left(A_2\right)P\left(F/A_2\right)$$

$$+ \ldots + P\left(A_n\right)P\left(F/A_n\right)$$

- 2) Ta có:

$$p(FA_k) = p(F)\,p(A_k / F)$$

$$\Rightarrow p(A_k / F) = \frac{p(FA_k)}{p(F)} = \frac{p(A_k)\,p(F / A_k)}{p(F)}$$

Các bài toán về công thức xác suất đầy đủ

- **Ví dụ 1**
- Sản phẩm X trên thị trường là do ba cơ sở sản xuất cung cấp. Cơ sở I chiếm 30% lượng hàng với tỉ lệ phế phẩm là 1%; Cơ sở II chiếm 50% lượng hàng với tỉ lệ phế phẩm là 3%; Cơ sở III chiếm 20% lượng hàng với tỉ lệ phế phẩm là 5%
- A) Một người mua một sản phẩm X trên thị trường. Tính xác suất để người đó mua phải sản phẩm phế phẩm
- B) Một người mua một sản phẩm X trên thị trường và mua phải sản phẩm phế phẩm. Theo anh chị thì sản phẩm phế phẩm này có khả năng là của cơ sở nào nhiều nhất?

Giải

- A) Gọi F là biến cố khách hàng mua phải sản phẩm phế phẩm
- Gọi A_i là biến cố sản phẩm X là của cơ sở i sản xuất (i=1,2,3). Ta thấy A_1, A_2, A_3 lập thành một hệ đầy đủ các biến cố
- Suy ra:

$$P(F) = P(A_1)P(F/A_1) + P(A_2)P(F/A_2) + P(A_3)P(F/A_3)$$

$$= 0,3 \times 0,01 + 0,5 \times 0,03 + 0,20 \times 0,05$$

$$= 0,003 + 0,015 + 0,010 = 0,028 (2,8\%)$$

- Vậy xác suất để khách hàng mua phải sản phẩm phế phẩm là 2,8%

B) Theo công thức Bayes ta có:

$$P(A_1/F)=\frac{P(A_1)P(F/A_1)}{P(F)}=\frac{0,3\times 0,01}{0,028}=\frac{3}{28}$$

$$P(A_2/F)=\frac{P(A_2)P(F/A_2)}{P(F)}=\frac{0,5\times 0,03}{0,028}=\frac{15}{28}$$

$$P(A_3/F)=\frac{P(A_3)P(F/A_3)}{P(F)}=\frac{0,2\times 0,05}{0,028}=\frac{10}{28}$$

- Ta thấy xác suất 15/28 lớn nhất. Do đó sản phẩm phế phẩm có khả năng là của cơ sở II là nhiều nhất

Ví dụ 2

- Có hai bình đựng bi: Bình thứ nhất chứa 12 bi trong đó có 8 bi mầu trắng và 4 bi mầu đỏ; Bình thứ hai chứa 12 bi trong đó có 9 bi mầu trắng và 3 bi mầu đỏ. Người ta lấy ngẫu nhiên một bi từ bình I bỏ vào bình II, sau đó lấy ngẫu nhiên bốn bi từ bình II ra ngoài.
- Tính xác suất để 4 bi lấy ra từ bình II có đúng 2 bi mầu trắng

Giải

- Gọi A_1 là biến cố bi lấy ra từ bình I có mầu trắng;
- Gọi A_2 là biến cố bi lấy ra từ bình I có mầu đỏ.
- Ta thấy $\{A_1, A_2\}$ lập thành một hệ đầy đủ
- Gọi F là biến cố có đúng 2 bi mầu trắng. Theo công thức xác suất đầy đủ ta có:

$$P(F) = P(A_1)P(F/A_1) + P(A_2)P(F/A_2)$$

$$= \frac{8}{12}\frac{C_{10}^2 C_3^2}{C_{13}^4} + \frac{4}{12}\frac{C_9^2 C_4^2}{C_{13}^4} = \frac{8}{12}\frac{45\times 3}{715} + \frac{4}{12}\frac{36\times 6}{715} = \frac{1944}{8580}$$

$$= 0,2265(22,65\%)$$

Ví dụ 3

- Trong một thúng cam có 42% cam Trung Quốc, 24% cam Thái Lan, 26% cam Campuchia và 8% cam Việt Nam.Trong số đó có một số cam bị hư gồm 20% cam TQ, 10% cam TL, 12% cam CPC và 2% cam VN.
- 1) Tính xác suất để một người mua phải một trái cam TQ bị hư
- 2) Tính xác suất để một người mua phải một trái cam bị hư

- 3) Biết một người mua phải một trái cam hư. Tính xác suất để trái cam ấy là cam của CPC
- 4) Biết một người mua phải một trái cam hư. Tính xác suất để trái cam ấy không phải là cam của VN
- **Giải**
- Gọi A_1, A_2, A_3, A_4 lần lượt là biến cố cam của TQ, TL, CPC, VN .Ta thấy A_1, A_2, A_3, A_4 lập thành một hệ đầy đủ các biến cố

- 1) Gọi A là biến cố trái cam hư là của TQ. F là biến cố trái cam bị hư. Ta có:
- $A = A_1F \Rightarrow p(A) = p(A_1F) = p(A_1)p(F/A_1)$
- $= 0,42x0,2 = 0,084$ (8,4%)
- 2) $p(F) = p(A_1)p(F/A_1) + p(A_2)p(F/A_2)$
- $+ p(A_3)p(F/A_3) + p(A_4)p(F/A_4)$
- $= 0,42x0,2 + 0,24x0,1 + 0,26x0,12$
- $+ 0,08x0,02$
- $= 0,1408$ (14,08%)

$$3)\ p\left(A_3/F\right)=\frac{p\left(A_3\right)p\left(F/A_3\right)}{p\left(F\right)}=\frac{0,26\times 0,12}{0,1408}=0,222\left(22,2\%\right)$$

$$4)\ p\left(\overline{A_4}/F\right)=\frac{p\left(\overline{A_4}F\right)}{p\left(F\right)}$$

$$\overline{A_4}F=\left(A_1+A_2+A_3\right)F=A_1F+A_2F+A_3F$$

$$\Rightarrow p\left(\overline{A_4}F\right)=p\left(A_1F+A_2F+A_3F\right)=p\left(A_1F\right)+p\left(A_2F\right)+p\left(A_3F\right)$$

$$=p\left(A_1\right)p\left(F/A_1\right)+p\left(A_2\right)p\left(F/A_2\right)+p\left(A_3\right)p\left(F/A_3\right)$$

$$=0,42\times 0,2+0,24\times 0,1+0,26\times 0,12=0,1392$$

$$\Rightarrow p\left(\overline{A_4}/F\right)=\frac{0,1392}{0,1408}=0,9886\left(98,86\%\right)$$

Cách khác

$$p\left(\overline{A_4}/F\right)=1-p\left(A_4/F\right)$$

$$p\left(A_4/F\right)=\frac{p\left(A_4\right)p\left(F/A_4\right)}{p\left(F\right)}=\frac{0,08\times 0,02}{0,1408}=0,0114$$

$$\Rightarrow p\left(\overline{A_4}/F\right)=1-0,0114=0,9886\left(98,86\%\right)$$

Chương 2

ĐẠI LƯỢNG NGẪU NHIÊN

§1 – ĐẠI LƯỢNG NGẪU NHIÊN

- **1.1 – Khái niệm ĐLNN**
- **1 – Định nghĩa**
- ĐLNN là một đại lượng biến thiên, phụ thuộc vào nhiều yếu tố ngẫu nhiên mà trong kết quả của phép thử nó chỉ mang một giá trị có thể được
- Như vậy ĐLNN là một quy tắc cho tương ứng mỗi biến cố sơ cấp trong không gian các biến cố sơ cấp của một phép thử với một số thực duy nhất
- Ký hiệu ĐLNN: X, Y, Z, …
- Giá trị của ĐLNN: x, y, z, …

2 – Phân loại

- Khi tập các giá trị của ĐLNN X là hữu hạn hoặc vô hạn đếm được thì X được gọi là ĐLNN rời rạc và ký hiệu: (X = a), (X = b), …
- Khi ĐLNN X lấy một giá trị tùy ý trong đoạn [a,b] thì ta nói X là ĐLNN liên tục và ký hiệu: (a#X#b); (a <X<b); (X#b),

3 – Ghi chú

- Khi có ĐLNN (X = a) thì ta có thể coi đó là một biến cố xác suất ghi nhận hiện tượng: "Đại lượng ngẫu nhiên X nhận giá trị a"
- Nếu X và Y là hai ĐLNN ứng với hai phép thử khác nhau thì ta nói X và Y là hai ĐLNN độc lập và ta có:
- $$p\left(X = a\right)\left(Y = b\right) = p\left(X = a\right) p\left(Y = b\right)$$

§2 – QUY LUẬT PHÂN PHỐI XÁC SUẤT

- **2.1 – Quy luật PPXS của ĐLNN rời rạc**
- **2.1.1 – Bảng PPXS**
- Cho X là một ĐLNN rời rạc với các giá trị có thể được là $x_1, x_2, \ldots, x_n$. Đặt $p_i = p(X=x_i)$ $(i=1\ldots n)$
- Ta lập một bảng liệt kê tất cả các giá trị của X cùng với các xác suất của nó và gọi bảng này là bảng PPXS của ĐLNN X:

X	x_1	x_2	…	x_n
p_X	p_1	p_2	…	p_n

Ghi chú

- Vì $x_1, x_2, \ldots, x_n$ là các giá trị có thể của X nên chúng là những số thực khác nhau đôi một do đó $(X = x_1)$; $(X = x_2)$;…; $(X = x_n)$ lập thành một hệ đầy đủ các biến cố. Suy ra:
- $(X = x_1) + (X = x_2) + \ldots + (X = x_n) = \Omega$
- 0$p(X = x_1)+p(X = x_2) + \ldots + p(X = x_n) =p(\Omega)$
- 0$p_1 + p_2 + \ldots + p_n = 1$
- Điều này thể hiện một quy luật cho tất cả ĐLNN.
- Người ta gọi đó là *Quy luật phân phối xác suất*

2.1.2 – Hàm PPXS

- **1 – Định nghĩa**
- Hàm số F(x) xác định trên R cho bởi công thức:
- F(x) = p(X<x) được gọi là hàm PPXS của đại lượng ngẫu nhiên X
- **2 – Định lý**
- Cho X là một ĐLNN rời rạc có các giá trị:
- $x_1 < x_2 < \ldots < x_n$ với bảng PPXS là:

X	x_1	x_2	…	x_n
p_X	p_1	p_2	…	p_n

- Khi đó hàm PPXS của X có dạng:

$$F(x)=\begin{cases} 0 & khi\ x \le x_1 \\ p_1 & \text{khi } x_1 < x \le x_2 \\ p_1 + p_2 & \text{khi } x_2 < x \le x_3 \\ \ldots & \\ p_1 + p_2 + \ldots + p_{n-1} & \text{khi } x_{n-1} < x \le x_n \\ 1 & \text{khi } x > x_n \end{cases}$$

Bài tập lập bảng phân phối xác suất

- **Bài 1**
- Một bình chứa 10 viên bi trong đó có 6 bi trắng và 4 bi đỏ. Người ta lấy ngẫu nhiên từ trong bình ra 3 bi. Gọi X là số bi trắng có trong ba bi được chọn.
- Hãy tìm quy luật PPXS của X
- Tìm hàm PPXS của X
- **Giải**
- Gọi X là số bi trắng có trong ba bi được chọn thì X là ĐLNN có các giá trị là 0, 1, 2, 3

Ta tính các xác suất

$$p_0 = p\left(X = 0\right) = \frac{C_6^0 C_4^3}{C_{10}^3} = \frac{4}{120}$$

$$p_1 = p\left(X = 1\right) = \frac{C_6^1 C_4^2}{C_{10}^3} = \frac{36}{120}$$

$$p_2 = p\left(X = 2\right) = \frac{C_6^2 C_4^1}{C_{10}^3} = \frac{60}{120}$$

$$p_3 = p\left(X = 3\right) = \frac{C_6^3 C_4^0}{C_{10}^3} = \frac{20}{120}$$

Kiểm tra ta được:

- $p_0 + p_1 + p_2 + p_3 = 1$
- Suy ra bảng PPXS của X là:

X	0	1	2	3
p_X	4/120	36/120	60/120	20/120

Hàm PPXS của X là

$$F(x)=\begin{cases}0 & khi\ x\le 0\\ \dfrac{4}{120} & \text{khi } 0<x\le 1\\ \dfrac{36+4}{120} & \text{khi } 1<x\le 2\\ \dfrac{36+4+60}{120} & \text{khi } 2<x\le 3\\ 1 & \text{khi x > 3}\end{cases} \Leftrightarrow F(x)=\begin{cases}0 & khi\ x\le 0\\ \dfrac{4}{120} & \text{khi } 0<x\le 1\\ \dfrac{40}{120} & \text{khi } 1<x\le 2\\ \dfrac{100}{120} & \text{khi } 2<x\le 3\\ 1 & \text{khi x > 3}\end{cases}$$

Baøi 2

- Moät xaï thuû baén ñoäc laäp 3 vieân ñaïn vaøo moät bia, xaùc suaát truùng cuûa moãi vieân ñaïn baèng nhau vaø baèng 0,7. Goïi X laø soá vieân ñaïn truùng bia (trong 3 vieân ñaïn baén).
- Haõy laäp baûng PPXS cuûa X.
- **Giaûi**
- Goïi X laø ñaïi löôïng ghi nhaän soá vieân ñaïn baén truùng muïc tieâu.

Ta tính caùc xaùc suaát

- Goïi Ai laø vieân ñaïn thöù i baén truùng bia (i = 1,2,3).
- Khi ñoù {A_1, A_2, A_3} ñoäc laäp vaø ta coù:

$$(X=0)=\overline{A_1}\,\overline{A_2}\,\overline{A_3}$$

$$\Rightarrow P(X=0)=P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})=(1-0,7)(1-0,7)(1-0,7)=0,027$$

$$(X=1)=\overline{A_1}A_2\overline{A_3}+A_1\overline{A_2}\,\overline{A_3}+\overline{A_1}\,\overline{A_2}A_3$$

$$\Rightarrow P\left(X=1\right)=P(\overline{A_1})P(A_2)P(\overline{A_3})+$$

$$+P(A_1)P(\overline{A_2})P(\overline{A_3})+P(\overline{A_1})P(\overline{A_2})P(A_3)$$

$$=3\times 0,3\times 0,7\times 0,3=0,189$$

$$\left(X=2\right)=A_1A_2\overline{A}_3+A_1\overline{A}_2A_3+\overline{A}_1A_2A_3$$

$$\Rightarrow P\left(X=2\right)=P(A_1)P(A_2)P(\overline{A}_3)+$$

$$+P(A_1)P(\overline{A}_2)P(A_3)+P(\overline{A}_1)P(A_2)P(A_3)$$

$$=3\times 0,7\times 0,7\times 0,3=0,441$$

$$\left(X=3\right)=A_1A_2A_3\Rightarrow P\left(X=3\right)=P(A_1)P(A_2)P(A_3)$$

$$=0,7\times 0,7\times 0,7=0,343$$

Kieåm tra $p_0 + p_1 + p_2 + p_3 = 1$

Suy ra baûng PPXS cuûa X laø

X	0	1	2	3
p_X	0,027	0,189	0,441	0,343

2.2 – Quy luật PPXS của ĐLNN liên tục

- **2.2.1 – Hàm mật độ xác suất**
- **1 – Định nghĩa**
- Cho X là một ĐLNN liên tục. Hàm f(x) xác định trên R được gọi là hàm mật độ xác suất của X nếu thỏa mãn các điều kiện như sau:

$$(1)\quad f(x) > 0\,(\forall x \in R)$$

$$(2)\quad \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$(3)\quad p(a \le X \le b) = \int_a^b f(x)dx$$

2 – Ghi chú:

- Cho X là một ĐLNN liên tục có hàm mật độ xác suất là f(x). Khi đó xác suất p(a<X<b) chính là diện tích hình thang cong được giới hạn bởi các đường:
- x = a, x = b, y = f(x) và y = 0
- **2.2.2 – Hàm PPXS của ĐLNN liên tục**
- **1 – Định nghĩa**
- Hàm số F(x) xác định trên R cho bởi công thức: F(x) = p(X<x) được gọi là hàm PPXS của ĐLNN X

2 – Ghi chú

- Khi X là một ĐLNN liên tục có hàm mật độ xác suất là f(x) thì hàm PPXS của X có dạng là:

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

3 – Các tính chất cơ bản của hàm PPXS

- (1) $\forall x \in R: 0 \le F(x) \le 1$
- (2) F(x) là hàm không giảm. Nghĩa là nếu
  $x_1 < x_2$ thì $F(x_1) \le F(x_2)$
- (3) Khi $x \to +\infty$ thì $F(x) \to 1$
- Khi $x \to -\infty$ thì $F(x) \to 0$
- (4) $[F(x)]' = f(x)$

Bài tập về ĐLNN liên tục

- **Ví dụ 1**
- Cho X là một ĐLNN liên tục có hàm mật độ xác suất là

$$f(x) = \begin{cases} a\cos x khi -\frac{\pi}{2} \le x \le \frac{\pi}{2} \\ 0 khi |x| > \frac{\pi}{2} \end{cases}$$

A) Tính hệ số a

B) Tìm hàm mật độ F(x)

C) Tính

$$p\left(-\frac{\pi}{4} \le X \le \frac{\pi}{4}\right)$$

Giải

A) Theo tính chất của hàm mật độ xác suất ta có

$$\int_{-\infty}^{+\infty} f\left(x\right)dx = 1 \Leftrightarrow \underbrace{\int_{-\infty}^{-\frac{\pi}{2}} f\left(x\right)dx}_{0} + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f\left(x\right)dx + \underbrace{\int_{\frac{\pi}{2}}^{+\infty} f\left(x\right)dx}_{0} = 1$$

$$\Leftrightarrow \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a\cos x dx = 1 \Leftrightarrow a\sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 1 \Leftrightarrow 2a = 1 \Rightarrow a = \frac{1}{2}$$

b) Theo tính chất của hàm PPXS ta có

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

$$Khi \quad x \le -\frac{\pi}{2} : F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-\frac{\pi}{2}} 0dx = 0$$

$$Khi \quad -\frac{\pi}{2} < x \le \frac{\pi}{2} : F(x) = \int_{-\infty}^{-\frac{\pi}{2}} f(t)\,dt + \int_{-\frac{\pi}{2}}^{x} f(t)\,dt$$

$$= \int_{-\infty}^{-\frac{\pi}{2}} 0dt + \int_{-\frac{\pi}{2}}^{x} \frac{1}{2}\cos t dt = \frac{1}{2}\sin t \Big|_{-\frac{\pi}{2}}^{x} = \frac{1}{2}(\sin x + 1)$$

$$Khi \quad x > \frac{\pi}{2}: F(x) = \int_{-\infty}^{-\frac{\pi}{2}} f(t)\,dt + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(t)\,dt + \int_{\frac{\pi}{2}}^{x} f(t)\,dt$$

$$= \underbrace{\int_{-\infty}^{-\frac{\pi}{2}} 0dt}_{0} + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}\cos t dt + \underbrace{\int_{\frac{\pi}{2}}^{x} 0dt}_{0} = \frac{1}{2}\sin t\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{1}{2}\left(1-(-1)\right) = 1$$

Suy ra hàm PPXS của X là:

$$F(x) = \begin{cases} 0 & khi \quad x \le -\dfrac{\pi}{2} \\ \dfrac{1}{2}(\sin x + 1) & khi \quad -\dfrac{\pi}{2} < x \le \dfrac{\pi}{2} \\ 1 & khi \quad x > \dfrac{\pi}{2} \end{cases}$$

- C) Theo tính chất của hàm phân phối xác suất ta có:

$$P\left(-\frac{\pi}{4} < X < \frac{\pi}{4}\right) = F\left(\frac{\pi}{4}\right) - F\left(-\frac{\pi}{4}\right)$$

$$= \frac{1}{2}\left(\sin\frac{\pi}{4} + 1\right) - \frac{1}{2}\left[\sin\left(-\frac{\pi}{4}\right) + 1\right]$$

$$= \frac{1}{2} \times \frac{\sqrt{2}}{2} - \frac{1}{2} \times \left(-\frac{\sqrt{2}}{2}\right) = \frac{\sqrt{2}}{2} = 0,7 \quad (70\%)$$

Ví dụ 2

Cho X là ĐLNN liên tục có hàm PPXS:

$$F(x) = \begin{cases} 0 & khi \;\; x \le 1 \\ a(x-1)^2 & khi \;\; 1 < x \le 4 \\ 1 & khi \;\; x > 4 \end{cases}$$

- A) Tìm hệ số a
- B) Tính p(2# X # 3)
- **Giải**

A) Vì F(x) là hàm liên tục nên nó liên tục tại x = 4. Khi đó:

$$\lim_{x\to 4^-} F(x) = \lim_{x\to 4^+} F(x) = F(4) \Leftrightarrow a(4-1)^2 = 1 \Leftrightarrow a = \frac{1}{9}$$

B) Theo tính chất của hàm PPXS ta có:

$$p(2 \le X \le 3) = F(3) - F(2) = \frac{1}{9}(3-1)^2 - \frac{1}{9}(2-1)^2 = \frac{1}{3} = 0,33(33\%)$$

§3 – CÁC ĐẶC TRƯNG CỦA MỘT ĐLNN

- **3.1 – Kỳ vọng toán học**
- **1 – Định nghĩa**
- (1) Cho X là ĐLNN rời rạc có bảng PPXS là:

X	x_1	x_2	…	x_n
p_X	p_1	p_2	…	p_n

Kỳ vọng toán học của ĐLNN X là một số thực được cho bởi công thức:

$$E(X) = p_1x_1 + p_2x_2 + \dots + p_n x_n$$

- (2) Cho X là một ĐLNN liên tục có hàm mật độ xác suất là f(x). Kỳ vọng toán học của ĐLNN X là một số thực được cho bởi công thức:

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

Ví dụ

- Theo dõi thu nhập của nhân viên trong một công ty có 100 người người ta thu được một bảng số liệu như sau:

Mức thu nhập (triệu đồng/tháng)	3	4	5	6	7
Số người	15	35	10	30	10

- Nếu gọi X là mức thu nhập trung bình của một
- người trong công ty thì X là một ĐLNN có các giá trị 2,4,5,6,7 và bảng PPXS của X là:

x_i	3	4	5	6	7
p_i	0,15	0,35	0,10	0,30	0,10

$$\Rightarrow EX = 0,15\times 3 + 0,35\times 4 + 0,1\times 5 + 0,3\times 6 + 0,1\times 7$$

$$= \frac{15\times 3 + 35\times 4 + 10\times 5 + 30\times 6 + 10\times 7}{100} = 4,85$$

- Như vậy:
- Kỳ vọng của một ĐLNN chính là giá trị trung bình theo xác suất của các giá trị của ĐLNN đó

2 – Tính chất cơ bản của kỳ vọng

- (1) $E(C) = C$ (C là hằng số)
- (2) $E(CX) = CE(X)$ (C là hằng số)
- (3) $E(X+Y) = E(X) + E(Y)$
- (4) Nếu X và Y độc lập thì
- $$E(XY) = E(X)E(Y)$$

3.2 – Phương sai và độ lệch tiêu chuẩn

- **1 – Định nghĩa**
- (1) Phương sai của ĐLNN X là một số thực được định nghĩa bằng công thức:
- $$D(X) = E[X - E(X)]^2$$
- (2) Độ lệch tiêu chuẩn của ĐLNN X là một số thực được định nghĩa theo công thức:

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

2 – Ghi chú

- (1) Nếu X là ĐLNN rời rạc với bảng PPXS là:

X	x_1	x_2	…	x_n
p_X	p_1	p_2	…	p_n

- Thì phương sai của X được tính theo công thức:

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

- (2) Nếu X là một ĐLNN liên tục có hàm mật độ xác suất là f(x) thì phương sai của X được tính theo công thức:

$$D(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\,dx$$

3 – Tính chất cơ bản của phương sai

- (1) Biểu thức khác của phương sai:
- $$D(X) = E(X^2) - [E(X)]^2$$
- (2) $D(C) = 0$ (là hằng số)
- (3) $D(CX) = C^2 D(X)$ (là hằng số)
- (4) Nếu X và Y độc lập thì
- $$D(X+Y) = D(X) + D(Y)$$

Ví dụ

- Cho X là một ĐLNN có bảng PPXS như sau:

x_i	3	4	5	6	7
p_i	0,15	0,35	0,10	0,30	0,10

- Ta có:

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

$$E(X^2) = 0,15 \times 3^2 + 0,35 \times 4^2 + 0,1 \times 5^2 + 0,3 \times 6^2 + 0,1 \times 7^2$$

$$= 26,75$$

Suy ra:

- Phương sai của X là:

$$D(X) = 26,75 - (4,85)^2 = 3,2275$$

- Độ lệch tiêu chuẩn của X là:

$$\sigma = \sqrt{D(X)} = \sqrt{3,2275} = 1,7965$$

3.3 – Mốt

(1) Cho X là ĐLNN rời rạc có bảng PPXS là:

X	x_1	x_2	…	x_n
p_X	p_1	p_2	…	p_n

Nếu xác suất $p(X = x_0)$ lớn nhất thì ta nói x_0 là mốt của X và ký hiệu là $mod(X) = x_0$

(2) Cho X là ĐLNN liên tục với hàm mật độ xác suất là f(x). Nếu f(x) đạt cực đại tại x_0 thì ta nói x_0 là mốt của X và cũng ký hiệu là $mod(X) = x_0$

§4 – MỘT SỐ QUY LUẬT PHÂN PHỐI THÔNG DỤNG

- **4.1 – Quy luật phân phối siêu bội**
- **1 – Định nghĩa**
- Cho X là một ĐLNN rời rạc với các giá trị nguyên 0, 1, … , n. Ta nói X là ĐLNN có quy luật phân phối siêu bội với các tham số N, N_A, n nếu công thức xác suất của nó được thực hiện theo công thức xác suất lựa chọn:

$$P(X=k)=\frac{C_{N_A}^{k}C_{N-N_A}^{n-k}}{C_N^n}\left(0\le k\le n\right)$$

- Ký hiệu X H(N, N_A, n)

2 – Các đặc trưng của phân phối siêu bội

- Cho X là ĐLNN có quy luật phân phối siêu bội
- X H(N, N_A, n). Khi đó X có các đặc trưng như sau:
- (1) Kỳ vọng:
- $E(X) = np$ với $p = N_A/N$
- (2) Phương sai:

$$D(X) = npq\frac{N-n}{N-1} \quad \left(p = \frac{N_A}{N};\ q = 1-p\right)$$

4.2 – Quy luật phân phối nhị thức

- **1 – Định nghĩa**
- Cho X là một ĐLNN rời rạc với các giá trị nguyên 0, 1, … , n. Ta nói X là ĐLNN có quy luật phân phối nhị thức với các tham số là n và p nếu công thức xác suất của nó được thực hiện theo công thức xác suất Becnuli:
- $$p\left(X = k\right) = C_n^k p^k \left(1 - p\right)^{n-k} \left(0 \le k \le n\right)$$
- Ký hiệu X B(n,p)
- **2 - Ghi chú**
- Quy luật phân phối nhị thức gắn liền với một dãy phép thử Becnuli với số lần thử là n và xác suất thành công là p

3 – Các đặc trưng của phân phối nhị thức

- Cho X là một ĐLNN có quy luật phân phối nhị thức X B(n,p). Khi đó ta có:
- (1) modX = k trong đó k là một số nguyên sao cho
- np – q # k # np – q + 1 với q = 1 – p
- (2) Kỳ vọng E(X) = np
- (3) Phương sai D(X) = npq với q = 1 – p

4 – Ví dụ

- Xác suất để một máy sản xuất ra phế phẩm là 0,02.
- A) Tính xác suất để trong 10 sản phẩm do máy sản xuất có không quá một phế phẩm
- B) Một ngày máy sản suất được 250 sản phẩm. Tìm số sản phẩm trung bình và số phế phẩm tin chắc nhất của máy đó trong một ngày

Giải

- A) Mỗi lần máy sản suất ra một phế phẩm ta coi là một phép thử, do đó số phép thử là n =10. Xác suất để máy sản xuất ra phế phẩm trong mỗi lần thử là p = 0,02, do đó nếu gọi X là số phế phẩm trong 10 sản phẩm sản xuất ra thì X có phân phối nhị thức $X \sim B(10; 0,02)$
- Gọi A là biến cố có không quá 1 phế phẩm.
- Ta có: $A = (0 \le X \le 1) = (X = 0) + (X = 1)$

Suy ra:

- p(A) = p(X=0) + p(X=1)
- $= C^0_{10}(0{,}02)^0\,(1 - 0{,}02)^{10-0}$
- $+ C^1_{10}(0{,}02)^1\,(1 - 0{,}02)^{10-1} = 0{,}983$
- B) Gọi X là số sản phẩm tin chắc trong ngày ta có $X \sim B(250;\ 0{,}02)$. Khi đó:
- E(X) = np = 250 x 0,02 = 5
- Số sản phẩm tin chắc trong ngày chính là modX:

- modX = k sao cho $np - q \leq k \leq np - q + 1$
- $np - q = 250 \times 0{,}02 - 0{,}98 = 4{,}02$
- $np - q + 1 = 5{,}02$
- $\Rightarrow 4{,}02 \leq k \leq 5{,}02 \Rightarrow k = 5$
- Vậy số phế phẩm tin chắc nhất trong ngày là 5

4.3 – Quy luật phân phối Poisson

- **1 – Định lý**
- Cho biểu thức $f(k) = C^k_n\, p^k\, (1 - p)^{n-k}$ ta có:
- Nếu np = a với a cố định và với mọi k cố định ta có:

$$\lim_{n\to\infty} f(k) = \lim_{n\to\infty}\left[C_n^k p^k (1-p)^{n-k}\right] = \frac{e^{-a}a^k}{k!}$$

- **2 – Định nghĩa**
- Cho X là một ĐLNN rời rạc có vô số giá trị nguyên 0, 1, 2, …Ta nói rằng X là ĐLNN có quy luật phân phối Poisson với tham số a > 0 nếu các công thức tính xác suất của nó có dạng:

$$P(X=k)=\frac{e^{-a}a^{k}}{k!}\quad (k=0,1,...)$$

Ký hiệu X ~ P(a)

3 – Các đặc trưng của phân phối Poisson

- Cho X là ĐLNN có quy luật phân phối Poisson $X \sim P(a)$. Ta có:
- (1) Kỳ vọng $E(X) = a$
- (2) Phương sai $D(X) = a$

4 - Ứng dụng

- Do định lý trên , phân phối Poisson được áp dụng trong các trường hợp sau:
- (1) Khi p khá nhỏ và n khá lớn thì xấp xỉ với phân phối nhị thức B(n,p)
- (2) Tính xác suất p (X = k) với X là ĐLNN biểu thị số lần xuất hiện biến cố A nào đó trong một khoảng thời gian nhất định

Ví dụ:

- Một cửa hàng trung bình trong một giờ bán được 4 sản phẩm. Tính xác suất để trong một giờ của hàng bán được nhiều hơn 5 sản phẩm.
- **Giải**
- Gọi X là số sản phẩm mà cửa hàng bán được trong một giờ, ta có: $X \sim P(4)$
- Gọi A là biến cố trong một giờ của hàng bán được nhiều hơn 5 sản phẩm. Ta có:

$$A=\left(X>5\right)\Rightarrow \overline{A}=\overline{\left(X>5\right)}=\left(X\le 5\right)$$

$$=\left(X=1\right)+\left(X=2\right)+\ldots+\left(X=5\right)$$

$$\Rightarrow p\left(\overline{A}\right)=p\left(X=1\right)+p\left(X=2\right)+\ldots+p\left(X=5\right)$$

$$=\frac{e^{-4}4^{0}}{0!}+\frac{e^{-4}4^{1}}{1!}+\ldots+\frac{e^{-4}4^{5}}{5!}=0,7851$$

$$\Rightarrow p\left(A\right)=1-0,7851=0,2149\left(21,49\%\right)$$

4.4 – Quy luật phân phối chuẩn

- **1 – Định nghĩa**
- Cho X là một ĐLNN liên tục có hàm mật độ xác suất là f(x). Ta nói rằng X là ĐLNN có quy luật phân phối chuẩn với các tham số là μ và σ^2 nếu hàm mật độ xác suất của nó có dạng:

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

- Ký hiệu: X N(μ; σ^2)

- Khi $\mu = 0$ và $\sigma^2 = 1$ ta có phân phối chuẩn tắc
- X N(0; 1) với hàm mật độ xác suất:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

2 – Các đặc trưng của phân phối chuẩn

- Cho X là ĐLNN có phân phối chuẩn $X \sim N(\mu; \sigma^2)$
- Ta có:
- (1) Kỳ vọng $E(X) = \mu$
- (2) Phương sai $D(X) = \sigma^2$
- (3) $mod(X) = \mu$

3 – Công thức tính xác suất của ĐLNN có quy luật phân phối chuẩn

- **Định lý**
- Cho X là ĐLNN có phân phối chuẩn
- X N(μ; σ²)
- Ta có công thức:

$$p\left(a < X < b\right) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

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

Chứng minh

$$X \in N\left(\mu, \sigma^2\right) \Rightarrow f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\Rightarrow p\left(a < X < b\right) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$t = \frac{x-\mu}{\sigma} \Rightarrow x = t\sigma + \mu \Rightarrow dx = \sigma dt$$

$$Khi: x = a \Rightarrow t = \frac{a-\mu}{\sigma}; Khi: x = b \Rightarrow t = \frac{b-\mu}{\sigma}$$

$$\Rightarrow p\left(a \le X \le b\right) =$$

$$= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}} \sigma dt = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= \int_{\frac{a-\mu}{\sigma}}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + \int_{0}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= \int_0^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt - \int_0^{\frac{a-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)\left(\text{vôùi } \phi(x) = \int_0^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\right)$$

$$\Rightarrow p(a < X < b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

Khi a=-∞ thì : $\phi(X < b) = \phi\left(\frac{b-\mu}{\sigma}\right) + \frac{1}{2}$

Ghi chú

- Hàm (x) trên đây được gọi là hàm tích phân Laplace.
- Hàm này là một hàm lẻ và được tính sẵn trong bảng F từ x = 0 đến x = 4,09.
- Ta có:

$$\lim_{x\to+\infty}\int_0^x \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt = \int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt = \frac{1}{2}$$

- Mà (4,09) = 0,4999. Vì thế khi x > 5 thì ta xấp xỉ (x) ≈ 0,5

- **Ví dụ:** $\Phi(1{,}96) = 0{,}475$; $\Phi(2{,}34) = 0{,}490$
- $\Phi(5{,}96) = 0{,}5$ …
- **5 – Phân vị chuẩn**
- Cho X là ĐLNN có quy luật phân phối chuẩn
- $X \sim N(\mu; \sigma^2)$. Số thực x_p là phân vị bậc p của X nếu thỏa mãn 2 điều kiện:

$$\begin{cases} x_{1-p} = -x_p & \left(0 < p < 1\right) \\ p\left(X < x_p\right) = \Phi\left(x_p\right) + \dfrac{1}{2} \end{cases}$$

Ví dụ

- Cho X có phân phối chuẩn $N(2100; 200^2)$
- Tính p(1700 < X < 2200)
- **Giải:**

$$p\left(1700 < X < 2200\right) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{2200-2100}{200}\right) - \phi\left(\frac{1700-2100}{200}\right)$$

$$= \phi\left(0,5\right) - \phi\left(-2\right) = \phi\left(0,5\right) + \phi\left(2\right)$$

$$= 0,1915 + 0,4772 = 0,6687\left(66,87\%\right)$$

4.5 – Quy luật phân phối chi bình phương

- **1 – Định lý**
- Cho X_1 , X_2 , … , X_n là những ĐLNN độc lập có phân phối chuẩn với cùng trung bình μ và cùng phương sai σ^2.
- Đặt:

$$\overline{X} = \frac{1}{n}\left[\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \left(\frac{X_2 - \mu}{\sigma}\right)^2 + \ldots + \left(\frac{X_n - \mu}{\sigma}\right)^2\right]$$

- Khi đó $\overline{X}$ là một ĐLNN có hàm PPXS là:

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

Trong ñoù $\Gamma(t)=\int\limits_{0}^{+\infty}x^{t-1}e^{-x}\,dx$ (Haøm Gamma)

2 – Định nghĩa

- Cho X là một ĐLNN liên tục. Ta nói rằng X là ĐLNN có phân phối chi bình phương với n bậc tự do (n nguyên dương) và ký hiệu là X $\sim \chi^2_n$ nếu hàm mật độ xác suất của nó là:

$$f\left(x\right)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\times\Gamma\left(\dfrac{n}{2}\right)}\times x^{\frac{n}{2}-1}\times e^{-\frac{x}{2}} & khi\ \ x>0\\ 0 & khi\ \ x\leq 0\end{cases}$$

3 – Phân vị chi bình phương

- Cho X là ĐLNN liên tục có phân phối chi bình phương $X \sim \chi^2_n$.
- Số thực x_p là phân vị bậc p của X nếu thỏa mãn 2 điều kiện:

$$\begin{cases} x_p > 0 \\ p\left(X < x_p \right) = p \end{cases}$$

- Các giá trị của x_p được trình bày ở bảng I

4.5 – Quy luật phân phối student

- **1 – Định lý**
- Cho X và Y là các ĐLNN liên tục, độc lập với nhau, trong đó X có phân phối chuẩn tắc X N(0,1); Y có phân phối chi bình phương Y χ^2_n . Đặt:

$$T = \frac{X\sqrt{n}}{\sqrt{Y}}$$

- Khi đó T là một ĐLNN liên tục có hàm mật độ xác suất là

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\times\Gamma\left(\frac{n}{2}\right)}\times\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{n}}\qquad (x\in R)$$

2 – Định nghĩa

- Cho T là một ĐLNN liên tục. Ta nói rằng T có phân phối student nếu T có hàm mật độ xác suất là:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\times\Gamma\left(\frac{n}{2}\right)}\times\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{n}} \qquad (x \in R)$$

- và ký hiệu T T_n

3 – Phân vị Student

- Cho T là một ĐLNN có phân phối Student T □ T_n . Ta gọi số thực t_p là phân vị bậc p của T nếu thỏa mãn hai điều kiện:

$$\begin{cases} t_{1-p} = -t_p & \left(0 < p < 1\right) \\ p\left(T < t_p\right) = p \end{cases}$$

- Các giá trị t_p trình bày trong bảng H

4.5 – MỘT SỐ ĐỊNH LÝ QUAN TRỌNG

Định lý 1

- Cho X là một ĐLNN có quy luật phân phối nhị thức B(n;p). Nếu n P 30 , p quá bé gần 0 hoặc không quá lớn gần 1 thì ta có thể coi X có phân phối Poisson với tham số a = np và ta có công thức xấp xỉ:

$$p\left(X = k\right) = C_n^k p^k \left(1 - p\right)^{n-k} \approx \frac{e^{-a} a^k}{k!} \quad \left(a = np\right)$$

Định lý 2 (định lý giới hạn địa phương)

- Cho X là một ĐLNN có quy luật phân phối nhị thức B(n;p). Nếu n P 30 , p không quá gần 0 và không quá gần 1,thì X được xấp xỉ phân phối chuẩn với trung bình μ = np và phương sai σ² = npq . Và ta có công thức:

$$p\left(X=k\right)=C_n^k p^k\left(1-p\right)^{n-k}\approx\frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right)$$

$$\left(\text{Vôùi}\,\varphi(x)=\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}\right)$$

Ghi chú

- Hàm

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

- Được gọi là hàm Gauss. Hàm này được tính sẵn trong bảng E từ x = 0 đến x = 4,09
- Khi x > 4,09 thì ta xấp xỉ

$$\varphi(x)\approx 0$$

- Ví dụ:

$$\varphi(0{,}15)\approx 0{,}3945;\varphi(2{,}5)\approx 0{,}0175;\varphi(5{,}5)\approx 0$$

Định lý 3 (định lý giới hạn tích phân)

- Cho X là một ĐLNN có quy luật phân phối nhị thức B(n;p). Nếu n P 30 , p không quá gần 0 và không quá gần 1, thì X được xấp xỉ phân phối chuẩn với trung bình μ = np và phương sai
- δ^2 = npq . Nghĩa là ta có công thức xấp xỉ:

$$p\left(a<X<b\right)\approx\phi\left(\frac{b-\mu}{\sigma}\right)-\phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\left(\mu=np;\sigma=\sqrt{npq}\right)$$

$$\left(\text{Vôùi}\ \phi(x)=\int_0^x\frac{1}{\sqrt{2\pi}}.e^{-\frac{t^2}{2}}dt\right)$$

Ví duï

- Nhaø maùy X saûn xuaát moät loaïi saûn phaåm coù soá saûn phaåm loaïi A chieám tæ leä 66,87%
- Choïn ngaãu nhieân 100 saûn phaåm (trong raát nhieàu saûn phaåm). Tính xaùc suaát ñeå:
- a) Trong 100 saûn phaåm coù ñuùng 70 saûn phaåm loaïi A.
- b) Trong 100 saûn phaåm coù khoâng quaù 60 saûn phaåm loaïi A.

Giaûi

- Moãi laàn kieåm tra moät saûn phaåm ta coi laø moät pheùp thöû. Ta coù soá pheùp thöû laø n = 100. Xaùc suaát ñeå moät saûn phaåm thuoäc loaïi A trong moãi laàn thöû laø   p = 0,6687.
- Do ñoù neáu goïi X laø soá saûn phaåm loaïi A trong 100 saûn phaåm thì X laø ÑLNN coù phaân phoái nhò thöùc: X ~ B(100; 0,6687)

- a) Goïi A laø bieán coá coù ñuùng 70 saûn phaåm loaïi A.
- Ta coù bieåu dieãn A = (X = 70)

$$\Rightarrow P\left(A\right) = P\left(X = 70\right)$$

$$= C_{100}^{70}\left(0,6687\right)^{70}\left(1 - 0,6687\right)^{100-70}$$

$$= 0,0681\left(6,81\%\right)$$

- B) Goïi B laø bieán coá coù khoâng quaù 60 saûn phaåm loaïi A .
- Ta coù bieåu dieãn: B = $(0 \le X \le 60)$
- Vì n = 100 > 30, xaùc suaát p = 0,6687 khoâng quaù gaàn 0 vaø khoâng quaù gaàn 1 neân:

$$P(B) = P(0 \le X \le 60) \approx \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\mu = np = 100.0,6687 = 66,87$$

$$\sigma = \sqrt{npq} = \sqrt{100.0,6687.(1-0,6687)}$$

$$= 4,7068$$

$$\Rightarrow P\left(B\right) = \phi\left(\frac{60-66,87}{4,7068}\right) - \phi\left(\frac{0-66,87}{4,7068}\right)$$

$$= \phi\left(-1,46\right) - \phi\left(-14,21\right)$$

$$= \phi\left(14,21\right) - \phi\left(1,46\right)$$

$$= 0,5 - 0,4279$$

$$= 0,0721\left(7,21\%\right)$$

Bài tập áp dụng

- **Ví duï 1**
- Troïng löôïng cuûa moät loaïi saûn phaåm ñöôïc quan saùt laø moät ñaïi löôïng ngaãu nhieân coù phaân phoái chuaån vôùi trung bình 50kg vaø phöông sai 100kg^2
- Nhöõng saûn phaåm coù troïng löôïng töø 45kg ñeán 70kg ñöôïc xeáp vaøo loaïi A. Choïn ngaãu nhieân 100 saûn phaåm (trong raát nhieàu saûn phaåm). Tính xaùc suaát ñeå:
- a) Trong 100 saûn phaåm coù ñuùng 70 saûn phaåm loaïi A.

Giaûi

- Tröôùc heát ta tính *tæ leä saûn phaåm loaïi A:*
- Goïi W laø troïng löôïng cuûa loaïi saûn phaåm ñaõ cho, ta thaáy W laø ÑLNN coù phaân phoái chuaån:
- W N(50; 100)
- Goïi F laø bieán coá saûn phaåm ñöôïc xeáp vaøo loaïi A, ta coù: F = (45 # W # 70). Suy ra:

$$P(F) = P(45 \le W \le 70) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\Rightarrow P\left(F\right) = \phi\left(\frac{70-50}{10}\right) - \phi\left(\frac{45-50}{10}\right) = \phi(2) - \phi(-0,5)$$

$$= \phi(2) + \phi(0,5) = 0,4772 + 0,1915 = 0,6687$$

- Suy ra xaùc suaát ñeå moät saûn phaåm thuoäc loaïi
- A laø p = 0,6687

Ta tính xaùc suaát theo yeâu caàu cuûa baøi toaùn:

- Moãi laàn kieåm tra moät saûn phaåm ta coi laø moät pheùp thöû, ta coù n = 100. Xaùc suaát ñeå moät saûn phaåm thuoäc loaïi A trong moãi laàn thöû laø p = 0,6687.
- Do ñoù neáu goïi X laø soá saûn phaåm loaïi A trong 100 saûn phaåm ñaõ choïn ra thì X laø ÑLNN coù phaân phoái nhò thöùc: X □ B(100; 0,6687)

- a) Goïi A laø bieán coá coù ñuùng 70 saûn phaåm loaïi A. Ta coù: A = (X = 70)

$$\Rightarrow P\left(A\right) = P\left(X = 70\right)$$

$$= C_{100}^{70}\left(0,6687\right)^{70}\left(1-0,6687\right)^{100-70}$$

$$\approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right)$$

$$\mu = np = 100\times 0,6687 = 66,87$$

$$\sigma = \sqrt{npq} = \sqrt{100\times 0,6687\times(1-0,6687)} = 4,7068$$

$$\Rightarrow P(A) \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right)$$

$$= \frac{1}{4,7068}\varphi\left(\frac{70-66,87}{4,7068}\right)$$

$$= \frac{1}{4,7068}\varphi(0,66)$$

$$= \frac{0,3209}{4,7068} = 0,0681(6,81\%)$$

- b) Goïi B laø bieán coá coù khoaûng quaù 60 saûn phaåm loaïi A ta coù: B =(0# X #60)

$$\Rightarrow P\left(B\right) = P\left(0 \le X \le 60\right) \approx \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{60-66,87}{4,7068}\right) - \phi\left(\frac{0-66,87}{4,7068}\right)$$

$$= \phi\left(-1,46\right) - \phi\left(-14,21\right) = \phi\left(14,21\right) - \phi\left(1,46\right)$$

$$= 0,5 - 0,4297 = 0,0721\left(7,21\%\right)$$

Ví duï 2

- *Saûn phaåm trong moät nhaø maùy ñöôïc ñoùng thaønh töøng kieän, moãi kieän goàm 14 saûn phaåm trong ñoù coù 8 saûn phaåm loaïi A vaø 6 saûn phaåm loaïi B. Khaùch haøng choïn caùch kieåm tra nhö sau: töø moãi kieän laáy ra 4 saûn phaåm; neáu thaáy soá saûn phaåm thuoäc loaïi A nhieàu hôn soá saûn phaåm thuoäc loaïi B thì môùi nhaän kieän ñoù; ngöôïc laïi thì loaïi boû. Kieåm tra 100 kieän (trong raát nhieàu kieän). Tính xaùc suaát ñeå*
- *coù 42 kieän ñöôïc nhaän.*
- *coù töø 40 ñeán 45 kieän ñöôïc nhaän.*
- *coù ít nhaát 42 kieän ñöôïc nhaän*

Giaûi

- Tröôùc heát ta tính *tæ leä nhöõng kieän ñöôïc nhaän* nghóa laø tìm xaùc suaát ñeå moät kieän haøng ñöôïc nhaän
- Goïi X laø soá saûn phaåm loaïi A trong kieän haøng
- Goïi F laø bieán coá kieän haøng ñöôïc nhaän ta coù:
- $$F = (X = 3) + (X = 4)$$
- Suy ra:

$$\Longrightarrow P\left(F\right) = P(X = 3) + P(X = 4)$$

$$= \frac{C_8^3 C_6^1}{C_{14}^4} + \frac{C_8^4 C_6^0}{C_{14}^4} = 0,4056\left(40,56\%\right)$$

Tính caùc xaùc suaát theo yeâu caàu

- Moãi laàn kieåm tra moät kieän haøng ta coi ñoù laø moät pheùp thöû khi ñoù soá pheùp thöû laø n = 100. Xaùc suaát ñeå moät kieän haøng ñöôïc nhaän trong moãi laàn thöû laø p = 0,4056.
- Do ñoù neáu goïi X laø soá kieän haøng ñöôïc nhaän trong 100 kieän haøng ñöôïc kieåm tra thì X laø ÑLNN coù phaân phoái nhò thöùc X $\in$ B(n, p) vôùi n = 100, p = 0,4056

- a) Goïi A laø bieán coá coù ñuùng 42 kieän ñöôïc nhaän. Ta coù:

$$A = \left(X = 42 \right)$$

$$\Rightarrow P\left(A \right) = P\left(X = 42 \right)$$

$$= C_{100}^{42}\left(0,4056 \right)^{42}\left(1 - 0,4056 \right)^{100-42}$$

$$= 0,0779 \quad \left(7,79\% \right)$$

- Vaäy xaùc suaát ñeå trong 100 kieän choïn ra coù ñuùng 42 kieän ñöôïc mua laø 7,79%

- b) Goïi B laø bieán coá coù töø 40 ñeán 45 kieän ñöôïc nhaän , ta coù:

$$B = \left(40 \le X \le 45\right) \Rightarrow P\left(B\right) = P\left(40 \le X \le 45\right)$$

$$\approx \phi\left(\frac{45-\mu}{\sigma}\right) - \phi\left(\frac{40-\mu}{\sigma}\right)$$

$$\mu = np = 100 \times 0{,}4056 = 40{,}56$$

$$\sigma = \sqrt{npq} = \sqrt{100 \times 0{,}4056 \times (1-0{,}4056)} = 4{,}9101$$

- Suy ra:

$$P\left(B\right) = \phi\left(\frac{45-40,56}{4,9101}\right) - \phi\left(\frac{40-40,56}{4,9101}\right)$$

$$= \phi(0,90) - \phi(-0,11)$$

$$= \phi(0,90) + \phi(0,11)$$

$$= 0,3159 + 0,0438 = 0,3597 = 35,97\%$$

- Vaäy xaùc suaát ñeå 100 kieän choïn ra coù töø 40 ñeán 45 kieän ñöôïc nhaän laø 35,97%

c) Goïi C laø bieán coá coù ít nhaát 42 kieän ñöôïc nhaän, ta coù:

$$C = \left(42 \le X \le 100\right)$$

$$\Rightarrow P\left(C\right) = P\left(42 \le X \le 100\right) = \phi\left(\frac{100-\mu}{\sigma}\right) - \phi\left(\frac{42-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{100-40,56}{4,9101}\right) - \phi\left(\frac{42-40,56}{4,9101}\right) = \phi(12) - \phi(0,29)$$

$$= 0,50 - 0,1141 = 0,3859 = 38,59\%.$$

- Vaäy xaùc suaát ñeå 100 kieän choïn ra coù ít nhaát 42 kieän ñöôïc nhaän laø 38,59%

Chương 3

ĐẠI CƯƠNG VỀ THỐNG KÊ TOÁN HỌC

§1 – LÝ THUYẾT MẪU

- **1.1 – ĐÁM ĐÔNG**
- **1- Khái niệm đám đông**
- Đám đông là một tập hợp các phần tử chứa đựng những thông tin hay dấu hiệu mà ta cần nghiên cứu.
- Ký hiệu một đám đông là Ω
- Dấu hiệu của một đám đông ký hiệu là X^*

2 – Ghi chú

- Khi nghiên cứu một đám đông ta quan tâm đến dấu hiệu X* của đám đông đó. Dấu hiệu X^* có thể thay đổi từ cá thể này sang cá thể khác .
- Vì vậy:
- Ta có thể coi việc nghiên cứu một đám
- đông là nghiên cứu một đại lượng ngẫu
- nhiên X^*

3- Các khái niệm liên quan

- Kích thước của đám đông: là số phần tử N của đám đông
- Giá trị của dấu hiệu X^*: $x_1, x_2, \ldots, x_k$
- Tần số $N_1, N_2, \ldots, N_k$: số phần tử mang dấu hiệu tương ứng $x_1, x_2, \ldots, x_k$
- Tần suất của các phần tử mang dấu hiệu
- x_i là tỉ số $p_i = N_i/N$

Ghi chú:

- Ta luôn có: $\sum N_i = N$ và $\sum p_i = 1$
- Bảng cơ cấu đám đông: Do ta coi X^* như là một ĐLNN nên ta gọi bảng PPXS của nó là bảng cơ cấu đám đông:

X^*	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

4- Các đặc trưng của đám đông

1) Trung bình đám đông:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \ hay \ \mu = \sum_{i=1}^{k} x_i p_i = \frac{1}{N}\sum_{i=1}^{k} x_i N_i$$

2) Phương sai đám đông:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$hay : \sigma^2 = \sum_{i=1}^{k}(x_i - \mu)^2 p_i = \frac{1}{N}\sum_{i=1}^{k}(x_i - \mu)^2 N_i$$

3) Độ lệch chuẩn: $$\sigma = \sqrt{\sigma^2}$$

4) Tỉ lệ đám đông: p = M/N (M là số phần tử có tính chất mà ta quan tâm)

5 – Ghi chú

Việc nghiên cứu một đám đông thường gặp một số khó khăn như sau:

- Do kích thước của đám đông quá lớn nên tốn kém vật chất và thời gian

- Khi làm việc với quy mô lớn người ta khó kiểm soát được quá trình nghiên cứu

- Nếu đám đông biến động, các phần tử thay đổi thường xuyên thì việc nghiên cứu trên cả đám đông khó thực hiện được…
- Để khắc phục được tình trạng này người ta tìm cách chọn mẫu và nghiên cứu trên mẫu và từ kết quả thu được trên mẫu ta có kết luận cho cả đám đông

1.2 – CÁC VẤN ĐỀ VỀ MẪU

1- Mẫu

- Từ đám đông ta chọn ra một tập hợp có n phần tử và nghiên cứu dấu hiệu X^* trên chúng. Tập hợp này được gọi là một mẫu

Ghi chú:

Khi nghiên cứu một mẫu ta quan tâm đến dấu hiệu X^* của đám đông trên mẫu nên ta coi một mẫu như là một ĐLNN

2 – Các khái niệm liên quan đến mẫu

- Kích thước mẫu: là số phần tử của tập hợp mẫu
- Giá trị của dấu hiệu X^*: $x_1, x_2, \ldots, x_k$
- Tần số n_i của x_i: số phần tử mang giá trị x_i (i=1..k)
- Tần suất của x_i là tỉ số: n_i/n (i=1..k)

3 – Các đặc trưng của mẫu

Trung bình mẫu:

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

Trung bình bình phương:

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

Phương sai mẫu:

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

- Phương sai mẫu hiệu chỉnh:

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

Ta có:

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

- Tỉ lệ mẫu:

$$f_n = f_n\left(A\right) = \frac{m}{n}$$

4 – Các phương pháp chọn mẫu

- 1) Mẫu cơ học: Chia đám đông thành n tập nhỏ sau đó chọn từ mỗi tập một phần tử làm đại diện
- 2) Mẫu điển hình: Chia đám đông thành n tập nhỏ có tính điển hình. Sau đó chọn từ mỗi tập một phần tử làm đại diện
- 3) Mẫu dãy: Chia đám đông thành n dãy sau đó chọn từ mỗi dãy một phần tử làm đại diện

5 – Các phương pháp sắp xếp số liệu

- 1) Sắp xếp theo giá trị quan sát:

X^*	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Ta gọi bảng này là bảng tần số của mẫu

- 2) Sắp xếp số liệu dạng khoảng:

X^*	$x_1' – x_1''$	$x_2' – x_2''$	…	$x_k' – x_k''$
n_i	n_1	n_2	…	n_k

Khi đó để lập bảng tần số ta chọn:

$$x_i = (x_i' + x_i'')/2$$

6 - Phương pháp tính các đặc trưng mẫu

- **Ví dụ:**
- Cho bảng số liệu:

x_i	4	5	7	9
n_i	10	15	13	12

Tính trung bình mẫu và phương sai mẫu hiệu chỉnh

Giải

- Lập bảng tính:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
4	10	40	160
5	15	75	375
7	13	91	637
9	12	108	972
Tổng	50	314	2144

Suy ra:

$$\overline{x} = \frac{1}{n}\left(\sum_{i=1}^{k} n_i x_i\right) = \frac{1}{50}\times 314 = 6,28$$

$$\overline{x^2} = \frac{1}{n}\left(\sum_{i=1}^{k} n_i x_i^2\right) = \frac{1}{50}\times 2144 = 42,88$$

$$\hat{S}^2 = \overline{x^2} - \left(\overline{x}\right)^2 = 42,88 - (6,28)^2 = 3,4416$$

$$\Rightarrow \hat{S} = \sqrt{3,4416} = 1,8552$$

$$S^2 = \frac{n}{n-1}\hat{S}^2 = \frac{50}{49}\times 3,4416 = 3,5118$$

$$\Rightarrow S = \sqrt{3,5118} = 1,8740$$

Phương pháp sử dụng máy tính bỏ túi F_x570 MS

- **Bước 1:** Vào chương trình thống kê (SD)
- Mod – SD – xóa dữ liệu có trong máy – vào lại SD
- **Bước 2:** Nhập số liệu
- x_i ; n_i – M^+ (i = 1…n)
- **Bước 3:** Đọc kết quả
- Gọi n; $\sum x_i$; $\sum x_i^2$: Bấm shift – sum
- Gọi

$$\overline{x};\ \overset{0}{S}\ (x\sigma n);\ S\ (x\sigma n - 1)$$ Bấm shift - var

Phương pháp sử dụng máy tính bỏ túi F_x570ES

- **Bước 1:**
- Vào chương trình
- Bấm sift – mod – (mũi tên xuống) – chọn 4 – chọn 1 – mod – chọn 3 – chọn 1
- Máy xuất hiện 2 cột để nhập số liệu:

	X	Freq
1		
2		
3		

- **Bước 2:**
- Nhập số liệu theo cột – nhập xong bấm AC
- **Bước 3:**
- Đọc kết quả
- Gọi $\sum x_i$; $\sum x_i^2$: Bấm Sift – 1 – chọn 4
- Gọi n; xσn; xσn-1: Bấm Sift – 1 – chọn 5

§2 – LÝ THUYẾT ƯỚC LƯỢNG

- **2.1 – Bài toán ước lượng**
- **1 - Định nghĩa**
- Cho một đám đông có tham số θ chưa biết (θ có thể là μ; σ hay p).
- Người ta chọn ra một mẫu có n quan sát cho dấu hiệu
- X^* là $x_1, x_2, \ldots, x_n$
- Vấn đề đặt ra là: Căn cứ vào mẫu quan sát hãy tìm
- một giá trị gần đúng θ' cho θ
- Bài toán này được gọi là Bài toán ước lượng tham số
- đám đông

2 - Phân loại bài toán ước lượng

1) Ước lượng điểm: là loại ước lượng chỉ phụ thuộc vào mẫu quan sát
2) Ước lượng khoảng: là loại ước lượng phụ thuộc vào mẫu quan sát và một xác suất cho trước (còn gọi là độ tin cậy)

2.2 – Phương pháp ước lượng điểm

1) Để ước lượng điểm trung bình đám đông ta dùng trung bình mẫu để ước lượng: $\mu \approx \overline{x}$
2) Để ước lượng điểm phương sai đám đông ta dùng phương sai mẫu hiệu chỉnh để ước lượng: $\sigma^2 \approx S^2$

3) Để ước lượng điểm tỉ lệ đám đông ta dùng tỉ lệ mẫu để ước lượng: $p \approx f_n$

2.3 – Ước lượng khoảng cho trung bình

2.3.1 - Bài toán:

- Cho một đám đông có trung bình μ chưa biết.

  Người ta lấy ra một mẫu có kích thước là n, tính được trung bình mẫu và phương sai mẫu hiệu chỉnh.

- Vấn đề đặt ra là: Căn cứ vào kết quả tính được trên mẫu, với độ tin cậy là $1 - \alpha$ cho trước hãy tìm khoảng ước lượng cho trung bình của đám đông

2.3.2 – Phương pháp giải quyết

Người ta chọn một khoảng (μ_1, μ_2) sao cho:

$$p\,(\mu_1 \# \mu \# \mu_2) = 1 - \alpha$$

2.3.3 – Công thức ước lượng

Trường hợp 1: Phương sai của đám đông σ^2 đã biết. Ta có công thức:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha \frac{\sigma}{\sqrt{n}}$$

Trong đó t_α được tính như sau: Từ hệ thức □(t_α) = (1 – α)/2 tra bảng F ta tính được t_α

Trường hợp 2:

- Phương sai của đám đông chưa biết và kích thước mẫu lớn (n P 30) ta có công thức ước lượng:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha \frac{S}{\sqrt{n}}$$

- **Trường hợp 3:**
- Phương sai của đám đông chưa biết và kích thước mẫu nhỏ (n < 30) ta có công thức ước lượng:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha^{n-1} \frac{S}{\sqrt{n}}$$

- Trong đó t_α^{n-1} được tra ở bảng H

Ghi chú:

Ñaët $\varepsilon = t_\alpha \dfrac{\sigma}{\sqrt{n}}; \varepsilon = t_\alpha \dfrac{S}{\sqrt{n}} \; hay \; \varepsilon = t_\alpha^{n-1} \dfrac{S}{\sqrt{n}}$

Ta coù: $\mu_1, \mu_2 = \overline{x} \mp \varepsilon$

- Và gọi ε là độ chính xác của ước lượng.

CÁC VÍ DỤ

- **Ví dụ 1**
- Khảo sát chiều cao của 36 sinh viên người
- ta tính được:

$$\overline{x} = 66\ inches\left(1inches = 2,54cm\right)$$

- Biết độ lệch tiêu chuẩn của một người
- trưởng thành là σ = 3inches, hãy ước
- lượng chiều cao trung bình của một sinh
- viên với độ tin cậy là 95%

Giải

- Gọi μ là ước lượng chiều cao trung bình của một sinh viên với độ tin cậy 95%
- Ta có công thức ước lượng:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha \frac{\sigma}{\sqrt{n}}$$

- Nhưng

$$\phi\left(t_\alpha\right) = \frac{1-\alpha}{2} = \frac{0,95}{2} = 0,475 \Rightarrow t_\alpha = 1,96$$

Suy ra:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha \frac{\sigma}{\sqrt{n}} = 66 \mp 1,96 \frac{3}{\sqrt{36}}$$

$$\varepsilon = 1,96 \frac{3}{\sqrt{36}} = 0,98$$

$$\Rightarrow \begin{cases} \mu_1 = 66 - 0,98 = 65,02 \\ \mu_2 = 66 + 0,98 = 66,98 \end{cases} \Rightarrow 65,02 < \mu < 66,98 \left(95\% \right)$$

Vậy chiều cao trung bình của một sinh viên nằm trong khoảng 65,02 inches đến 66,98 inches với độ tin cậy 95%

Ví dụ 2

- Lấy ngẫu nhiên 100 hộp sữa trong một kho sữa, cân trọng lượng và tính được:

$$\overline{x} = 396g; S^2 = 25g^2$$

- Hãy ước lượng trọng lượng trung bình của một hộp sữa trong kho với độ tin cậy là 96%

Giải

- Gọi μ là ước lượng trọng lượng trung bình của một hộp sữa trong kho với độ tin cậy là 96%. Do n =100 > 30 nên ta có công thức ước lượng:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha . \frac{S}{\sqrt{n}}$$

- Nhưng:

$$\phi\left(t_\alpha\right) = \frac{1-\alpha}{2} = \frac{0,96}{2} = 0,480 \Rightarrow t_\alpha = 2,05$$

Suy ra: $\mu_1, \mu_2 = 396 \mp 2,05 \times \frac{5}{\sqrt{100}}$

$$\varepsilon = 2,05 \times \frac{5}{\sqrt{100}} = 1,025$$

$$\Rightarrow \begin{cases} \mu_1 = 396 - 1,025 = 394,975 \\ \mu_2 = 396 + 1,025 = 397,025 \end{cases}$$

$$\Rightarrow 394,975 \le \mu \le 397,025 \left(96\% \right)$$

- Vậy trọng lượng trung bình của một hộp sữa trong kho vào khoảng từ 394,975g đến 397,025g với độ tin cậy 96%

Ví dụ 3

- Lấy ngẫu nhiên 25 hộp sữa trong một kho sữa, cân trọng lượng và tính được:

$$\overline{x} = 395g; S^2 = 24g^2$$

- Ước lượng trọng lượng trung bình của một hộp sữa trong kho với độ tin cậy là 95%

Giải

- Gọi μ là ước lượng trọng lượng trung bình của một hộp sữa trong kho với độ tin cậy là 95%. Do n = 25 < 30 nên ta có công thức ước lượng:

$$\mu_1, \mu_2 = \overline{x} \mp t_\alpha^{n-1} \frac{S}{\sqrt{n}}$$

- Nhưng

$$t_\alpha^{n-1} = t_{0,05}^{24} = 2,0639$$

Suy ra: $\mu_1, \mu_2 = 395 \mp 2,0639 \times \frac{4,9}{\sqrt{25}}$

$$\varepsilon = 2,0639 \times \frac{4,9}{\sqrt{25}} = 2,02$$

$$\Rightarrow \begin{cases} \mu_1 = 395 - 2,02 = 392,98 \\ \mu_2 = 395 + 2,02 = 397,02 \end{cases}$$

$$\Rightarrow 392,98 \le \mu \le 397,02 \left(95\% \right)$$

Vậy trong lượng trung bình của một hộp sữa trong kho vào khoảng từ 392,98g đến 397,02g với độ tin cậy 95%

Ví dụ 4

- Lấy ngẫu nhiên 100 hộp sữa trong một kho sữa, cân trọng lượng và tính được: $\bar{x} = 396g; S^2 = 25g^2$

a) Muốn ước lượng trong lượng trung bình của một hộp sữa trong kho đạt được độ chính xác là 1gam và độ tin cậy 99% thì cần điều tra trên bao nhiêu hộp?

b) Muốn ước lượng trong lượng trung bình của một hộp sữa trong kho đạt được độ chính xác là 1gam thì độ tin cậy của ước lượng sẽ là bao nhiêu?

Giải

a) Theo công thức ước lượng trung bình đám đông ta có:

$$\varepsilon = t_{\alpha} . \frac{S}{\sqrt{n}} \Rightarrow \varepsilon^2 = \left(t_{\alpha}\right)^2 \frac{S^2}{n}$$

$$\Rightarrow n = \frac{\left(t_{\alpha}\right)^2 S^2}{\varepsilon^2} \approx \frac{\left(2,58\right)^2 \times 25}{1^2} \approx 166,41 \Rightarrow n = 166$$

Vậy ta phải điều tra ít nhất là 166 hộp

b) Theo công thức ước lượng trung bình đám đông ta có:

$$\varepsilon = t_\alpha . \frac{S}{\sqrt{n}} \Rightarrow \varepsilon^2 = \left(t_\alpha\right)^2 \frac{S^2}{n} \Rightarrow \left(t_\alpha\right)^2 = \frac{n\varepsilon^2}{S^2} = \frac{100\times 1^2}{25} = 4$$

$$\Rightarrow t_\alpha = 2 \Rightarrow \varphi\left(t_\alpha\right) = \varphi\left(2\right) = 0,4772$$

$$\Rightarrow 1-\alpha = 2\varphi\left(t_\alpha\right) = 0,9544\left(95,44\%\right)$$

Vậy độ tin cậy của ước lượng là 95,44%

Ghi chú:

Bài toán trong ví dụ 4 còn gọi là bài toán xác định các chỉ tiêu của bài tóan ước lượng trung bình đám đông

2.4 – Ước lượng khoảng cho tỉ lệ

1 - Bài toán: Cho một đám đông có tỉ lệ p chưa biết. Người ta chọn ra một mẫu có kích thước là n, tính được tỉ lệ mẫu là f_n.

- Vấn đề đặt ra là: Căn cứ vào kết quả trên mẫu, hãy tìm khoảng ước lượng cho tỉ lệ đám đông với độ tin cậy 1 – α cho trước

2 – Phương pháp giải quyết: Người ta tìm một khoảng (p_1, p_2) sao cho:

$$p\,(p_1 \# \; p \; \# p_2) = 1 - \alpha$$

3 – Công thức ước lượng

$$p_1, p_2 = f_n \mp t_\alpha \sqrt{\frac{f_n\left(1 - f_n\right)}{n}}$$

Ghi chú: Nếu ta đặt

$$\varepsilon = t_\alpha \sqrt{\frac{f_n(1 - f_n)}{n}}$$

Thì công thức ước lượng có dạng:

$$p_1, p_2 = f_n \mp \varepsilon$$

Ta gọi ε là độ chính xác của ước lượng

Ví dụ 1

- Lô trái cây của một chủ hàng được đóng thành từng sọt mỗi sọt chứa 100 trái. Kiểm tra 50 sọt thấy có 450 trái không đạt tiêu chuẩn
- a) Ước lượng tỉ lệ trái cây không đạt tiêu chuẩn của lô hàng với độ tin cậy 95%.
- b) Muốn lượng tỉ lệ trái cây không đạt tiêu chuẩn với độ chính xác là 0,005 thì độ tin cậy là bao nhiêu
- c) Muốn lượng tỉ lệ trái cây không đạt tiêu chuẩn với độ chính xác là 0,01 và độ tin cậy là 99% thì cần kiểm tra bao nhiêu sọt

Giải

- a) Gọi p là ước lượng tỉ lệ trái cây không đạt tiêu chuẩn của lô hàng với độ tin cậy 95% ta có công thức ước lượng:

$$p_1, p_2 = f_n \mp t_\alpha \sqrt{\frac{f_n\left(1-f_n\right)}{n}}$$

$$Do: f_n = \frac{450}{50\times 100} = 0,09; 1-\alpha = 95\% \Rightarrow t_\alpha = 1,96$$

$$\Rightarrow p_1, p_2 = 0,09 \mp 1,96\sqrt{\frac{0,09\left(1-0,09\right)}{5000}}$$

Khi đó

$$\varepsilon = 1{,}96\sqrt{\frac{0,09(1-0,09)}{5000}} = 0,008$$

$$\Rightarrow \begin{cases} p_1 = 0,09 - 0,008 = 0,082 \\ p_2 = 0,09 + 0,008 = 0,098 \end{cases}$$

$$\Rightarrow 0,082 \le p \le 0,098 (95\%)$$

Vậy tỉ lệ trái cây không đạt chuẩn nằm trong khoảng từ 8,2% đến 9,8% với độ tin cậy là 95%

b) Theo công thức ước lượng tỉ lệ đám đông ta có:

$$\varepsilon = t_\alpha \sqrt{\frac{f_n(1-f_n)}{n}} \Rightarrow \varepsilon^2 = (t_\alpha)^2 \frac{f_n(1-f_n)}{n}$$

$$\Rightarrow (t_\alpha)^2 = \frac{n\varepsilon^2}{f_n(1-f_n)} = \frac{5000\times(0,005)^2}{0,09(1-0,09)} = 1,5376 \Rightarrow t_\alpha = 1,24$$

$$\Rightarrow 1-\alpha = 2\varphi(t_\alpha) = 2\varphi(1,24) = 2\times 0,3925 = 0,785 \quad (78,5\%)$$

- Vậy độ tin cậy đạt được là 78,5%

c) Theo công thức ước lượng tỉ lệ đám đông ta có:

$$\varepsilon = t_\alpha \sqrt{\frac{f_n\left(1-f_n\right)}{n}} \Rightarrow \varepsilon^2 = \left(t_\alpha\right)^2 \frac{f_n\left(1-f_n\right)}{n}$$

$$\Rightarrow n = \frac{f_n\left(1-f_n\right)\left(t_\alpha\right)^2}{\varepsilon^2} \approx \frac{0,09\left(1-0,09\right)\left(2,58\right)^2}{\left(0,01\right)^2} = 5451,59$$

$\Rightarrow n = 5452$ trái hay 55 sọt

- Vậy ta phải kiểm tra ít nhất là 55 sọt

Ví dụ 2 (Bài toán đếm cá)

- *Muoán bieát ñöôïc soá caù trong moät hoà lôùn ngöôøi ta baét leân 2000 con, ñaùnh daáu xong laïi thaû xuoáng hoà. Sau ñoù ngöôøi ta baét leân 400 con vaø thaáy coù 80 con ñöôïc ñaùnh daáu.*
- *Vôùi ñoä tin caäy 95% haõy öôùc löôïng soá caù trong hoà.*
- **Giaûi**
- Ta goïi N laø soá caù trong hoà, p laø tæ leä caù coù ñaùnh daáu trong hoà ta coù: $p = \frac{2000}{N}$

- Tröôùc heát ta öôùc löôïng tæ leä p soá con caù ñaõ ñaùnh daáu trong hoà:
- Theo coâng thöùc öôùc löôïng tæ leä ñaùm ñoâng ta coù:

$$p_1, p_2 = f_n \mp t_\alpha \sqrt{\frac{f_n(1-f_n)}{n}}$$

$$f_n = \frac{80}{400} = 0,2;\ 1\text{-}\ \alpha = 95\% \Rightarrow t_\alpha = 1,96$$

$$p_1, p_2 = 0,2 \mp 1,96 \times \sqrt{\frac{0,2 \times \left(1-0,2\right)}{400}}$$

$$\varepsilon = 1,96 \times \sqrt{\frac{0,2 \times \left(1-0,2\right)}{400}} = 0,0392$$

Suy ra:

$$\Rightarrow \begin{cases} p_1 = 0,2 - 0,0392 = 0,1608 \\ p_2 = 0,2 + 0,0392 = 0,2392 \end{cases} \Rightarrow 0,1608 \le p \le 0,2392 (95\%)$$

$$Do: p = \frac{2000}{N} \Rightarrow N = \frac{2000}{p}$$

$$0,1608 \le p \le 0,2392 \Rightarrow \frac{1}{0,2392} \le \frac{1}{p} \le \frac{1}{0,1608}$$

$$\Rightarrow \frac{2000}{0,2392} \le \frac{2000}{p} \le \frac{2000}{0,1608} \Rightarrow 8361\left(con\right) \le N \le 12438\left(con\right)$$

Vaäy öôùc löôïng soá caù trong hoà coù khoaûng töø 8361 con ñeán 12438 con vôùi ñoä tin caäy laø 98%

§3 – LÝ THUYẾT KIỂM ĐỊNH

- **3.1 – KIỂM ĐỊNH GIẢ THIẾT THAM SỐ**
- **3.1.1 – Kiểm định 2 phía**
- **A – Bài toán kiểm định trung bình**

Cho một đám đông có trung bình μ chưa biết.

Đặt một giả thiết H_0 về μ : “$\mu = \mu_0$”

Vấn đề đặt ra là với mức ý nghĩa α cho trước hãy kiểm tra giả thiết H_0 là đúng hay sai

Bài toán này được gọi là bài toán kiểm định giả thiết trung bình đám đông

Giải quyết

- **Trường hợp 1: Biết phương sai σ^2**
- **Bước 1:**
- Đặt giả thiết không H_0 : “$\mu = \mu_0$ “
- Đặt đối thiết H_1 : “μ s μ_0 “
- **Bước 2:**
- Tính thống kê:

$$t = \frac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{\sigma}$$

- Tính t_α

Bước 3: Kết luận

1) Nếu | t |# t_α thì ta chấp nhận giả thiết H_0

Và bác bỏ H_1

2) Nếu | t | > t_α thì ta bác bỏ giả thiết H_0 và

Chấp nhận H_1

Ghi chú: Khi giả thiết H_0 bị bác bỏ ta có kết luận:

$$Khi : \overline{x} > \mu_0 \Rightarrow \mu > \mu_0$$

$$Khi : \overline{x} < \mu_0 \Rightarrow \mu < \mu_0$$

- **Trường hợp 2: Biết phương sai σ^2 và n 30**
- **Bước 1:**
- Đặt giả thiết không H_0 : "$\mu = \mu_0$ "
- Đặt đối thiết H_1 : "μ s μ_0 "
- **Bước 2:**
- Tính thống kê: $$t = \frac{\left(\bar{x} - \mu_0\right)\sqrt{n}}{S}$$
- Tính t_α

Bước 3: Kết luận

1) Nếu | t |# t_α thì ta chấp nhận giả thiết H_0

Và bác bỏ H_1

2) Nếu | t | > t_α thì ta bác bỏ giả thiết H_0 và

Chấp nhận H_1

Ghi chú: Khi giả thiết H_0 bị bác bỏ ta có kết luận:

$$Khi : \overline{x} > \mu_0 \Rightarrow \mu > \mu_0$$

$$Khi : \overline{x} < \mu_0 \Rightarrow \mu < \mu_0$$

- **Trường hợp 3: Chưa biết phương sai σ^2**
- **n < 30**
- **Bước 1:**
- Đặt giả thiết không H_0 : "$\mu = \mu_0$ "
- Đặt đối thiết H_1 : "μ s μ_0 "
- **Bước 2:**
- Tính thống kê:

$$t = \frac{\left(\bar{x} - \mu_0\right)\sqrt{n}}{S}$$

- Tính t^{n-1}_{α}

Bước 3: Kết luận

1) Nếu | t |# t^{n-1}_{α} thì ta chấp nhận giả thiết H_0 và bác bỏ H_1

2) Nếu | t | > t^{n-1}_{α} thì ta bác bỏ giả thiết H_0 và Chấp nhận H_1

Ghi chú: Khi giả thiết H_0 bị bác bỏ ta có kết luận:

$$Khi : \overline{x} > \mu_0 \Rightarrow \mu > \mu_0$$

$$Khi : \overline{x} < \mu_0 \Rightarrow \mu < \mu_0$$

Ví dụ áp dụng

- **Ví dụ 1**
- Giám đốc một xí nghiệp cho biết lương trung bình của một công nhân là 2,5 triệu đồng / tháng
- Chọn ngẫu nhiên 36 công nhân thì thấy lương trung bình là 2,3 triệu đồng / tháng. Biết độ lệch tiêu chuẩn là 0,5 triệu hãy kiểm định tuyên bố của ông giám đốc với mức ý nghĩa 5%.
- **Giải**
- Gọi μ là mức lương trung bình của một công nhân. Đặt giả thiết H_0:"$\mu = 2,5$". Ta kiểm định H_0 ở mức ý nghĩa 5%

Giải

- Gọi μ là mức lương trung bình của một công nhân.
- Đặt giả thiết không H_0:"μ = 2,5".
- Với đối thiết H_0:"μ s 2,5".
- Ta kiểm định H_0 ở mức ý nghĩa 5%
- Tính:

$$t = \frac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{\sigma} = \frac{\left(2,3 - 2,5\right)\sqrt{36}}{0,5} = 2,4$$

$$\alpha = 0,05 \Rightarrow t_\alpha = 1,96 \Rightarrow \left|t\right| > t_\alpha$$

- Suy ra bác bỏ H_0, chấp nhận H_1
- Lời của giám đốc không tin cậy được.
- Thực tế mức lương trung bình của một
- công nhân thấp hơn 2,5 triệu đồng/tháng

Ví dụ 2

- Trước đây trọng lượng thanh niên độ tuổi 20 của vùng A là 45kg. Năm nay người ta chọn ngẫu nhiên 100 thanh niên và xác định được trọng lượng trung bình là 48kg; phương sai mẫu hiệu chỉnh là $100kg^2$
- Với mức ý nghĩa 5% hãy kiểm tra xem trọng lượng của thanh niên trong vùng có thay đổi hay không?

Giải

- Gọi μ là trọng lượng trung bình của thanh niên trong vùng hiện nay
- Đặt giả thiết không H_0:"μ = 45"
- Đối thiết H_1:"μ s 45" và kiểm định H_0 ở mức ý nghĩa 5%.
- Ta có

$$t = \frac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{S} = \frac{\left(48 - 45\right)\sqrt{100}}{10} = 3$$

$$\alpha = 0,05 \Rightarrow t_\alpha = 1,96 \Rightarrow |\, t \,| > t_\alpha$$

- Suy ra bác bỏ H_0, chấp nhận H_1
- Kết luận trọng lượng của thanh niên trong vùng có tăng lên.

Ví dụ 3

- Một công ty sản xuất pin tuyên bố rằng pin của họ có tuổi thọ trung bình là 21,5 giờ. Một cơ quan kiểm tra chất lượng kiểm tra ngẫu nhiên 6 chiếc pin và thu được kết quả về tuổi thọ của 6 chiếc pin này là: 19, 18, 22, 20, 16, 25.
- Kiểm tra lại thông tin trên với mức ý nghĩa là 5%
- **Giải**
- Gọi μ là tuổi thọ trung bình của loại pin do công ty sản xuất. Đặt giả thiết H_0:"μ = 21,5" và kiểm định H_0 ở mức ý nghĩa 5%.

Tính toán ta được

$$\overline{x} = \frac{19+18+22+20+16+25}{6} = 20$$

$$\overline{x^2} = \frac{19^2+18^2+22^2+20^2+16^2+25^2}{6} = 408,3$$

$$\hat{S}^2 = \overline{x^2} - \left(\overline{x}\right)^2 = 8,3 \Rightarrow S^2 = \frac{n}{n-1}\hat{S}^2 = 10 \Rightarrow S = 3,16$$

$$t = \frac{\left(\overline{x}-\mu_0\right)\sqrt{n}}{S} = \frac{\left(20-21,5\right)\sqrt{6}}{3,16} = 1,16$$

$$t_\alpha^{n-1} = t_{0,05}^{5} = 2,571 \Rightarrow |t| < t_\alpha^{n-1}$$

Ta chấp nhận H_0 ở mức ý nghĩa α = 5%

Vậy tuyên bố của công ty là đúng

B – Bài toán kiểm định về tỉ lệ

Cho một đám đông có tỉ lệ p chưa biết. Đặt một giả thiết H_0 về p : "$p = p_0$"

Vấn đề đặt ra là với mức ý nghĩa α cho trước hãy kiểm tra giả thiết H_0 là đúng hay sai

Bài toán này được gọi là bài toán kiểm định giả thiết tỉ lệ đám đông

Giải quyết

- **Bước 1:**
- Đặt giả thiết không H_0 : “$p = p_0$ “
- Đặt đối thiết H_1 : “p s p_0 “
- **Bước 2:**
- Tính thống kê:

$$t = \frac{\left(p - p_0\right)\sqrt{n}}{\sqrt{p_0\left(1 - p_0\right)}}$$

- Tính t_α (tra bảng G)

Bước 3: Kết luận

1) Nếu | t |# t_α thì ta chấp nhận giả thiết H_0 và bác bỏ H_1

2) Nếu | t | > t_α thì ta bác bỏ giả thiết H_0 và Chấp nhận H_1

Ghi chú: Khi giả thiết H_0 bị bác bỏ ta có kết luận:

$$Khi : f_n > p_0 \Rightarrow p > p_0$$

$$Khi : f_n < \mu_0 \Rightarrow p < p_0$$

Ví dụ áp dụng

- **Ví dụ 1**
- Một tài liệu cũ cho biết tỉ lệ hộ dân thích xem chương trình “Phim Việt” là 80%. Kiểm tra ngẫu nhiên 36 hộ dân thì thấy có 25 hộ thích xem chương trình này. Với mức ý nghĩa 5% có chấp nhận tài liệu này hay không?
- **Giải**
- Gọi p là tỉ lệ hộ dân thích xem chương trình “Phim Việt”.
- Đặt giả thiết H_0: ”p = 0,8”, Đối thiết H_1: ”p s 0,8”
- Ta kiểm định H_0 ở mức ý nghĩa 5%

Ta có:

$$t = \frac{\left(f_n - p_0 \right)\sqrt{n}}{\sqrt{p_0(1 - p_o)}}$$

$$f_n = \frac{25}{36} = 0,7 \Rightarrow t = \frac{\left(0,7 - 0,8\right)\sqrt{36}}{\sqrt{0,8(1-0,8)}} = -1,5$$

$$t_\alpha = 1,96 \Rightarrow \left|t\right| < t_\alpha$$

Vậy ta chấp nhận giả thiết H_0; Bác bỏ H_1

Nghĩa là thông tin cho rằng tỉ lệ 80% người thích chương trình “Phim Việt” có thể tin cậy được

Ví dụ 2

- Một bản báo cáo nói rằng 80% gia đình trong thành phố có máy vi tính ở nhà. Để kiểm tra người ta chọn ngẫu nhiên 80 gia đình trong thành phố và thấy có 68 hộ có máy tính cá nhân.
- Kiểm định thông tin trên với mức ý nghĩa $\alpha = 2\%$
- **Giải**
- Gọi p là tỉ lệ hộ dân có máy vi tính ở nhà.
- Đặt giả thiết H_0: ”p = 0,8”, Đối thiết H_1: ”p s 0,8”
- Ta kiểm định H_0 ở mức ý nghĩa 2%

Ta có

$$t=\frac{\left(f_n-p_0\right)\sqrt{n}}{\sqrt{p_0(1-p_o)}}$$

$$f_n=\frac{68}{80}=0,85\Rightarrow t=\frac{\left(0,85-0,8\right)\sqrt{80}}{\sqrt{0,8(1-0,8)}}=2,5$$

$$\alpha=2\%\Rightarrow t_\alpha=2,34\Rightarrow\left|t\right|>t_\alpha$$

Ta bác bỏ giả thiết H_0 ; chấp nhận H_1

Thực tế số gia đình có máy tính cá nhân ở nhà trong thành phố cao hơn

Ví dụ 3

- Tỉ lệ phế phẩm của một nhà máy trước đây là 5%. Năm nay nhà máy áp dụng một biện pháp cải tiến kỹ thuật mới. Để nghiên cứu tác dụng của biện pháp kỹ thuật mới, người ta lấy ngẫu nhiên một mẫu gồm 800 sản phẩm kiếm tra thấy có 24 phế phẩm.
- A) Cho biết kết luận về biện pháp cải tiến kỹ thuật với mức ý nghĩa 1%

- B) Với mức ý nghĩa 5%, nếu nhà máy báo cáo tỉ lệ phế phẩm sau khi áp dụng biện pháp mới là 2% thì có tin được không
- **Giải**
- Gọi p là tỉ lệ phế phẩm sau khi áp dụng biện pháp cải tiến kỹ thuật mới.
- Đặt giả thiết H_0: ”p = 0,05”
- Đối thiết H_1:”p s 0,05”
- Ta kiểm định H_0 ở mức ý nghĩa 1%

- Ta có

$$t = \frac{\left(f_n - p_0 \right)\sqrt{n}}{\sqrt{p_0(1-p_o)}}$$

$$f_n = \frac{24}{800} = 0,03 \Rightarrow t = \frac{\left(0,03-0,05\right)\sqrt{800}}{\sqrt{0,05(1-0,05)}} = -2,59$$

$$\alpha = 1\% \Rightarrow t_\alpha = 2,58 \Rightarrow |t| > t_\alpha$$

- Vậy bác bỏ H_0, chấp nhận H_1
- Suy ra biện pháp kỹ thuật mới có tác dụng tốt do làm giảm tỉ lệ phế phẩm

- B) Gọi p là tỉ lệ phế phẩm sau khi áp dụng biện pháp cải tiến kỹ thuật mới.
- Đặt giả thiết H_0: ”p = 0,02”
- Đối thiết H_1:”p s 0,02”
- Ta kiểm định H_0 ở mức ý nghĩa 5%
- Ta có

$$t = \frac{\left(f_n - p_0\right)\sqrt{n}}{\sqrt{p_0(1-p_o)}} = \frac{\left(0,03-0,02\right)\sqrt{800}}{\sqrt{0,02(1-0,02)}} = 2,02$$

$$\alpha = 5\% \Rightarrow t_\alpha = 1,96 \Rightarrow \left|t\right| > t_\alpha$$

- Suy ra bác bỏ H_0, chấp nhận H_1
- Vậy báo cáo của nhà máy tỉ lệ phế phẩm là 2% không đúng theo thực tế.

B – Bài toán kiểm định về phương sai

Cho một đám đông có phương sai σ chưa biết. Đặt một giả thiết H_0 về σ : “$\sigma = \sigma_0$”

Vấn đề đặt ra là với mức ý nghĩa α cho trước hãy kiểm tra giả thiết H_0 là đúng hay sai

Bài toán này được gọi là bài toán kiểm định giả thiết phương sai đám đông

Giải quyết

- **Bước 1:**
- Đặt giả thiết không H_0 : "$\sigma^2 = \sigma^2_0$ "
- Đặt đối thiết H_1 : "$\sigma\ ^2$s σ^2_0 "
- **Bước 2:**
- Tính thống kê: $$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$
- Tính $$\chi^2_{\frac{\alpha}{2}}(n-1); \chi^2_{1-\frac{\alpha}{2}}(n-1)$$

Bước 3: Kết luận

1) Nếu $$\chi^2_{\frac{\alpha}{2}}(n-1) \le \chi^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)$$

thì ta chấp nhận giả thiết H_0 và bác bỏ H_1

2) Nếu $$\chi^2_{\frac{\alpha}{2}}(n-1) > \chi^2 \vee \chi^2_{1-\frac{\alpha}{2}}(n-1) < \chi^2$$

thì ta bác bỏ giả thiết H_0 và chấp nhận H_1

Ghi chú:

Khi giả thiết H_0 bị bác bỏ ta có kết luận:

$$Khi: S^2 > \sigma_0^2 \Rightarrow \sigma^2 > \sigma_0^2$$

$$Khi: S^2 < \sigma_0^2 \Rightarrow \sigma^2 < \sigma_0^2$$

Ví dụ áp dụng

- Nếu máy móc hoạt động bình thường thì kích thước của một loại sản phẩm là một ĐLNN có quy luật phân phối chuẩn với phương sai $\sigma^2 = 25cm^2$.
- Nghi ngờ máy hoạt động không bình thường, người ta đó thử 20 sản phẩm và tính được $S^2 = 27{,}5\ cm^2$
- Với mức ý nghĩa 2%, hãy kết luận về điều nghi ngờ này

Giải

- Gọi σ^2 là phương sai của kích thước sản phẩm hiện nay.
- Đặt giả thiết không H_0 : "$\sigma^2 = 25$ "
- Đặt đối thiết H_1 : "σ^2 s 25 "
- Tính:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{19 \times 27,5}{25} = 20,9$$

- Tra bảng I ta được:

$$\chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{\frac{0,02}{2}}(20-1) = \chi^2_{0,01}(19) = 7,6327$$

$$\chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{1-\frac{0,02}{2}}(20-1) = \chi^2_{0,99}(19) = 36,1908$$

$$\Rightarrow \chi^2_{\frac{\alpha}{2}}(n-1) \le \chi^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)$$

- Vậy chấp nhận H_0 bác bỏ H_1
- Suy ra máy hoạt động bình thường

3.1.2 – Kiểm định một phía

- **1 – Kiểm định trung bình**
- **A) Phía phải**
- Giả thiết không: H_0: "$\mu = \mu_0$"
- Giả thiết đối: H_1: "$\mu > \mu_0$"
- (1) Nếu n 30, biết σ:
- - Tính

$$t = \frac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{\sigma}$$

- - Tính $t_{2\alpha}$

- - Nếu $|t| \le t_{2\alpha}$: Chấp nhận H_0 bác bỏ H_1
- - Nếu $|t| > t_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1
- Nếu $n \ge 30$, chưa biết σ: Ta thay σ bằng S và lập luận như trên
- (2) Nếu $n < 30$, biết σ:
- Tính

$$t = \frac{\left(\bar{x} - \mu_0\right)\sqrt{n}}{\sigma}; \quad t_{2\alpha}$$

- Nếu $t \le t_{2\alpha}$: Chấp nhận H_0 bác bỏ H_1
- Nếu $t > t_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1

- (2) Nếu n < 30, không biết σ:
- Tính

$$t = \frac{\left(\bar{x} - \mu_0\right)\sqrt{n}}{S}; \quad t_{2\alpha}^{n-1}$$

- Nếu $t \le t^{n-1}_{2\alpha}$: Chấp nhận H_0 bác bỏ H_1
- Nếu $t > t^{n-1}_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1

A) Phía trái

- Giả thiết không: H_0: "$\mu = \mu_0$"
- Giả thiết đối: H_1: "$\mu < \mu_0$"
- Tiến hành kiểm định như đối với phía phải chỉ lưu ý tính t theo công thức

$$t = \frac{\left(\mu_0 - \overline{x}\right)\sqrt{n}}{\sigma} \quad hay \quad t = \frac{\left(\mu_0 - \overline{x}\right)\sqrt{n}}{S}$$

Ví dụ áp dụng

- Một công ty có một hệ thống máy vi tính có thể xử lý 1200 hóa đơn trong một giờ. Công ty vừa cho nhập về một hệ thống máy mới. Cho chạy thử 40 giờ thì thấy số hóa đơn xử lý một giờ trung bình là 1260 với độ lệch tiêu chuẩn là 215
- Với mức ý nghĩa 5% hãy nhận xét hệ thống mới có tốt hơn hệ thống cũ hay không

Giải

- Gọi μ là số thời gian xử lý trung bình trong một giờ của hệ thống máy vi tính mới
- Đặt giả thiết không: H_0: "$\mu = 1200$" (nghĩa là HT mới tốt bằng HT cũ)
- Giả thiết đối: H_1: "$\mu > \mu_0$" (nghĩa là HT mới tốt hơn HT cũ)
- Ta tính:

$$t = \frac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{S} = \frac{\left(1260 - 1200\right)\sqrt{40}}{215} = 1,76$$

- Tra bảng F ta tính: $t_{2\alpha} = t_{2x0,05} = t_{0,1}$

$$\phi\left(t_{2\alpha}\right) = \frac{1 - 2\alpha}{2} = \frac{1 - 0,1}{2} = 0,45 \Rightarrow t_{2\alpha} = 1,64$$

- Suy ra $t > t_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1
- Vậy hệ thống máy mới tốt hơn hệ thống

2 – Kiểm định tỉ lệ

- **A) Phía phải**
- Giả thiết không: H_0: "$p = p_0$"
- Giả thiết đối: H_1: "$p > p_0$"
- Tính

$$t = \frac{\left(f_n - p_0\right)\sqrt{n}}{\sqrt{p_0(1-p_o)}};\ t_{2\alpha}$$

- Nếu $t \le t_{2\alpha}$: Chấp nhận H_0 bác bỏ H_1
- Nếu $t > t_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1

B) Phía trái

- Giả thiết không: H_0: “p = p_0”
- Giả thiết đối: H_1: “p > p_0”
- Tính

$$t = \frac{\left(p_0 - f_n\right)\sqrt{n}}{\sqrt{p_0(1 - p_o)}};\ t_{2\alpha}$$

- Nếu $t \leq t_{2\alpha}$: Chấp nhận H_0 bác bỏ H_1
- Nếu $t > t_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1

C) Ví dụ

- Một báo cáo nói rằng tỷ lệ gia đình ở TP Hồ Chí Minh có máy tính cá nhân là 85%. Một mẫu kiểm tra gồm 80 hộ dân cho thấy có 72 hộ có máy tính ở nhà.
- Với mức ý nghĩa 2% hãy xét xem tỉ lệ hộ dân trong TP có máy tính cá nhân có cao hơn tỷ lệ chung hay không?

Giải

- Gọi p là tỷ lệ hộ dân hiện có máy tính cá nhân ở nhà.
- Đặt giả thiết không: H_0:“p = 0,85”
- Giả thiết đối: H_0: “p > 0,85”
- Tính

$$t = \frac{\left(f_n - p_0\right)\sqrt{n}}{\sqrt{p_0(1-p_o)}};\ f_n = \frac{72}{80} = 0,9$$

$$\Rightarrow t = \frac{\left(0,9-0,80\right)\sqrt{80}}{\sqrt{0,80\left(1-0,80\right)}} = 2,236$$

$$\alpha = 0,02 \Rightarrow t_{2\alpha} = t_{0,04}$$

$$\phi\left(t_{2\alpha}\right) = \phi\left(t_{0,04}\right) = \frac{1-0,04}{2} = 0,48 \Rightarrow t_{2\alpha} = 2,06$$

- Suy ra t > $t_{2\alpha}$: Bác bỏ H_0 chấp nhận H_1
- Vậy tỷ lệ hộ dân có máy tính cá nhân ở nhà cao hơn so với báo cáo

3 – Kiểm đinh phương sai

- **A) Phía phải**
- Đặt giả thiết không H_0 : "$\sigma^2 = \sigma^2_0$ "
- Đặt đối thiết H_1 : "$\sigma^2 > \sigma^2_0$ "
- Tính: $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ $\quad \chi^2_{1-\alpha}(n-1)$
- Nếu $\chi^2 > \chi^2_{1-\alpha}(n-1)$
- Thì bác bỏ H_0 chấp nhận H_1

B) Phía trái

- Đặt giả thiết không H_0 : "$\sigma^2 = \sigma^2_0$ "
- Đặt đối thiết H_1 : "$\sigma^2 < \sigma^2_0$ "
- Tính:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \qquad \chi^2_\alpha(n-1)$$

-
- Nếu $\chi^2 < \chi^2_\alpha(n-1)$
- Thì bác bỏ H_0 chấp nhận H_1

C) Ví dụ

- Đo đường kính 12 sản phẩm của một dây chuyền sản xuất người ta tính được S = 0,3. Biết rằng nếu độ biến động của các sản phẩm lớn hơn 0,2 thì dây chuyền sản xuất phải dừng lại để điều chỉnh.
- Với mức ý nghĩa 5%, người kiểm tra có kết luận như thế nào?

Giải

- Gọi là độ biến động của đường kính sản phẩm.
- Đặt giả thiết không H_0 : "$\sigma^2 = (0,2)^2 = 0,04$"
- Đặt đối thiết H_1 : "$\sigma^2 > 0,04$"
- Tính:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(12-1)(0,3)^2}{0,04} = 24,75$$

$$\chi^2_{1-\alpha}(n-1) = \chi^2_{1-0,05}(12-1) = \chi^2_{0,95}(11) = 19,6752$$

- Suy ra

$$\chi^2 < \chi^2_\alpha(n-1)$$

- Vậy bác bỏ giả thiết H_0 chấp nhận H_1
- Do đó dây chuyền cần phải điều chỉnh vì độ biến động lớn hơn mức cho phép

§4 – LÝ THUYẾT SO SÁNH

4.1 – So sánh hai trung bình

4.1.1 – Bài toán

Cho hai đám đông có trung bình μ_1, μ_2 chưa biết.

Đặt một giả thiết H_0 : "$\mu_1 = \mu_2$"

Vấn đề đặt ra là với mức ý nghĩa α cho trước hãy kiểm tra giả thiết H_0 là đúng hay sai

4.1.2 – Phương pháp

Từ đám đông thứ nhất chọn ra một mẫu có kích thước là n_1; tính trung bình mẫu và phương sai mẫu hiệu chỉnh $\overline{x}_{n1};\ S_{n1}$

Từ đám đông thứ nhất chọn ra một mẫu có kích thước là n_2; tính trung bình mẫu và phương sai mẫu hiệu chỉnh $\overline{x_{n2}}; S_{n2}$

Trường hợp 1: Biết các phương sai σ_1^2 và σ_2^2

Bước 1:

Tính t theo công thức: $$t = \frac{\left|\overline{x_{n1}} - \overline{x_{n2}}\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Tính t_α

Bước 2 : Kết luận

1) Nếu t # t_α thì ta chấp nhận giả thiết H_0

2) Nếu $t > t_\alpha$ thì ta bác bỏ giả thiết H_0

Trường hợp 2: Không biết các phương sai σ_1^2 và σ_2^2

1) Nếu n_1, n_2 P 30 ta <u>tiến hành kiểm định:</u>

- Tính t theo công thức: $$t = \frac{\left|\overline{x_{n1}} - \overline{x_{n2}}\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$
- Tính t_α
- So sánh t với t_α và kết luận

Nếu t # t_α thì ta chấp nhận giả thiết H_0

Nếu $t > t_\alpha$ thì ta bác bỏ giả thiết H_0

2) Nếu n_1 hoặc n_2 < 30 ta tiến hành kiểm định:

- Tính t theo công thức:

$$t = \frac{\left|\overline{x_{n1}} - \overline{x_{n2}}\right|}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} \text{ vôù } S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- Tính $t_\alpha^{n1 + n2 - 2}$
- So sánh t với $t_\alpha^{n1 + n2 - 2}$ và kết luận:

Nếu t # $t_\alpha^{n1 + n2 - 2}$ thì ta chấp nhận giả thiết H_0

Nếu t > $t_\alpha^{n1 + n2 - 2}$ thì ta bác bỏ giả thiết H_0

Ghi chú:

Khi giả thiết H_0 bị bác bỏ thì ta kết luận $\mu_1 \neq \mu_2$ ở mức ý nghĩa α.

Tuy nhiên:

Nếu $x_{n1} > x_{n2}$ thì ta kết luận $\mu_1 > \mu_2$

Nếu $x_{n1} < x_{n2}$ thì ta kết luận $\mu_1 < \mu_2$

Ví dụ áp dụng

- **Ví dụ 1**
- Từ đám đông thứ nhất người ta lấy một mẫu kích thước là 40; tính được trung bình mẫu là 130. Từ đám đông thứ hai người ta lấy một mẫu kích thước là 50; tính được trung bình mẫu là 140. Biết phương sai của đám đông thứ nhất là 80, của đám đông thứ hai là 100.
- So sánh trung bình của hai đám đông với mức ý nghĩa $\alpha = 1\%$

Giải

- Theo giả thiết của bài toán ta có

$$n_1 = 40; \overline{x_{n1}} = 130; \sigma_1^2 = 80$$

$$n_2 = 50; \overline{x_{n2}} = 140; \sigma_2^2 = 100$$

$$\left.\begin{array}{l} t = \dfrac{\left|\overline{x_{n1}} - \overline{x_{n2}}\right|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{|130 - 140|}{\sqrt{\dfrac{80}{40} + \dfrac{100}{50}}} = 5 \\ \alpha = 1\% \Longrightarrow t_\alpha = 2,58 \end{array}\right\} \Longrightarrow t > t_\alpha$$

- Vậy giả thiết H_0 bị bác bỏ ở mức ý nghĩa α = 1%

Nhưng

$$\overline{x_{n1}} < \overline{x_{n2}} \Rightarrow \mu_1 < \mu_2$$

- Vậy trung bình của đám đông thứ nhất nhỏ hơn trung bình của đám đông thứ hai

Ví dụ 2

- Kiểm tra chất lượng của một loại sản phẩm do hai nhà máy sản xuất người ta chọn ngẫu nhiên từ mỗi nhà máy một mẫu gồm 225 sản phẩm và xác định các khuyết tật của mỗi sản phẩm. Ở nhà máy thứ nhất người ta tính số khuyết tật trung bình là 2,3 với phương sai mẫu hiệu chỉnh là 3,24. Ở nhà máy thứ hai người ta tính số khuyết tật trung bình là 2,5 với phương sai mẫu hiệu chỉnh là 4,06.
- So sánh số khuyết tật trung bình của loại sản phẩm do hai nhà máy sản xuất với mức ý nghĩa là $\alpha = 5\%$

Giải

- Gọi μ_1, μ_2 lần lượt là số khuyết tật trung bình của loại sản phẩm do hai nhà máy 1 và 2 sản xuất. Đặt giả thiết H_0:" $\mu_1 = \mu_2$" và kiểm định H_0 ở mức ý nghĩa $\alpha = 5\%$
- Từ giả thiết ta có:

$$n_1 = 225; \overline{x_{n1}} = 2,3; S_{n1}^2 = 3,24$$

$$n_2 = 225; \overline{x_{n2}} = 2,5; S_{n2}^2 = 4,06$$

Do n_1 và n_2 > 30 nên ta tính

$$t = \frac{\left|\overline{x_{n1}} - \overline{x_{n2}}\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{\left|2,3 - 2,5\right|}{\sqrt{\frac{3,24}{225} + \frac{4,06}{225}}} = \frac{0,2 \times 15}{\sqrt{7,3}} = \frac{3}{2,7} = 1,1$$

$$\alpha = 5\% \Rightarrow t_\alpha = 1,96 \Rightarrow t > t_\alpha$$

Vậy ta chấp nhận giả thiết H_0 ở mức ý nghĩa α = 5%

Suy ra số khuyết tật trung bình của loại sản phẩm do hai nhà máy là như nhau

Ví dụ 3

- Hai công ty cùng sản xuất một loại pin viễn thông. Kiểm ta công ty A 10 chiếc thì thấy tuổi thọ trung bình là 4,8 năm, với độ lệch tiêu chuẩn là 1,1 năm; Kiểm ta công ty B 12 chiếc thì thấy tuổi thọ trung bình là 4,3 năm, với độ lệch tiêu chuẩn là 0,9 năm.
- So sánh tuổi thọ trung bình của loại pin viễn thông do hai công ty sản xuất với mức ý nghĩa là $\alpha = 5\%$

Giải

- Gọi μ_1, μ_2 lần lượt là tuổi thọ trung bình của loại pin do hai công ty A và B sản xuất. Đặt giả thiết H_0:" $\mu_1 = \mu_2$" và kiểm định H_0 ở mức ý nghĩa α = 1%
- Từ giả thiết ta có:

$$n_1 = 10; \overline{x_{n1}} = 4,8; S_1 = 1,1$$

$$n_2 = 12; \overline{x_{n2}} = 4,3; S_2 = 0,9$$

$$\Rightarrow S^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2} = \frac{9 \times (1,1)^2 + 11 \times (0,9)^2}{10 + 12 - 2} = 0,99$$

Do n_1 và n_2 < 30 nên ta tính:

$$\left.\begin{array}{l} t = \dfrac{\left|\overline{x_{n1}} - \overline{x_{n2}}\right|}{\sqrt{\dfrac{S}{n_1} + \dfrac{S}{n_2}}} = \dfrac{|4,8 - 4,3|}{\sqrt{\dfrac{0,99}{10} + \dfrac{0,99}{12}}} = 1,174 \\ \alpha = 5\% \Rightarrow t_{\alpha}^{n_1+n_2-2} = t_{0,05}^{20} = 2,845 \end{array}\right\} \Rightarrow t < t_{\alpha}^{n_1+n_2-2}$$

Vậy ta chấp nhận giả thiết H_0 ở mức ý nghĩa α = 5%

Suy ra số tuổi thọ trung bình của loại và B sản xuất là như nhau

4.2 – So sánh hai tỉ lệ

4.2.1 – Bài toán

Cho hai đám đông có các tỉ lệ p_1, p_2 chưa biết.

Đặt một giả thiết H_0 : "$p_1 = p_2$"

Vấn đề đặt ra là: Với mức ý nghĩa α cho trước hãy kiểm tra giả thiết H_0 là đúng hay sai

Bài toán này được gọi là bài toán so sánh hai tỉ lệ của hai đám đông

4.2.2 – Phương pháp kiểm định

Từ đám đông thứ nhất chọn ra một mẫu có kích thước là n_1; tính tỉ lệ mẫu f_{n1}

Từ đám đông thứ hai chọn ra một mẫu có kích thước là n_2; tính tỉ lệ mẫu f_{n2}

<u>Tiến hành kiểm định:</u>

- Tính t theo công thức:

$$t = \frac{\left| f_{n1} - f_{n2} \right|}{\sqrt{\mathring{f}_n \left(1 - \mathring{f}_n\right) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- Với tỉ lệ mẫu chung:

$$\mathring{f}_n = \frac{n_1 f_{n1} + n_2 f_{n2}}{n_1 + n_2}$$

- Tính t_α
- So sánh t với t_α và kết luận:

Nếu t # t_α thì ta chấp nhận giả thiết H_0

Nếu t > t_α thì ta bác bỏ giả thiết H_0

Ghi chú:

Khi giả thiết H_0 bị bác bỏ thì ta kết luận $p_1 \neq p_2$ ở mức ý nghĩa α. Tuy nhiên:

Nếu $f_{n1} > f_{n2}$ thì ta kết luận $p_1 > p_2$

Nếu $f_{n1} < f_{n2}$ thì ta kết luận $p_1 < p_2$

Ví dụ áp dụng

- Công ty nước giải khát Côca – Côla tiến hành cải tiến công thức cho sản phẩm của mình. Với công thức cũ, khi cho 500 người dùng thử thì thấy có 120 người ưa thích nó. Với công thức mới, khi cho 1000 người dùng thử thì thấy có 300 người ưa thích nó.
- So sánh tỷ lệ những người ưa thích Côca – Côla theo hai công thức cũ và mới và đưa ra kết luận về công thức mới ở mức ý nghĩa $\alpha = 2\%$

Giải

- Gọi p_1, p_2 lần lượt là tỉ lệ những người ưa thích Côca – Côla theo công thức mới và cũ. Đặt giả thiết H_0:" $p_1 = p_2$" và kiểm định H_0 ở mức ý nghĩa α = 2%
- Từ giả thiết ta có:

$$n_1 = 1000; m_1 = 300 \Rightarrow f_{n1} = \frac{300}{1000} = 0,3$$

$$n_2 = 500; m_2 = 120 \Rightarrow f_{n2} = \frac{120}{500} = 0,24$$

$$\Rightarrow \mathring{f}_n = \frac{n_1 f_{n1} + n_2 f_{n2}}{n_1 + n_2} = \frac{1000 \times 0,3 + 500 \times 0,24}{1000 + 500} = 0,28$$

Tính các test thống kê

$$t = \frac{\left| f_{n1} - f_{n2} \right|}{\sqrt{\mathring{f}_n\left(1-\mathring{f}_n\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\left|0,3-0,24\right|}{\sqrt{0,28\left(1-0,28\right)\left(\frac{1}{1000}+\frac{1}{500}\right)}} = 2,4$$

$$\alpha = 2\% \Rightarrow t_\alpha = 2,34 \Rightarrow t > t_\alpha$$

Ta bác bỏ H_0 ở mức ý nghĩa α = 2%

Do $f_{n1} > f_{n2}$ nên $p_1 > p_2$ do đó tỉ lệ những người ưa thích Côca – Côla theo công thức mới cao hơn. Vậy công thức mới có hiệu quả tốt

§5 – Ví dụ tổng hợp

- *Theo dõi chất lượng của một loại sản phẩm do một nhà máy sản suất người ta tiến hành đo tạp chất có trong sản phẩm đó. Kiểm tra trên một số sản phẩm ta thu được kết quả như sau:*

Löôïng taïp chaát(%)	0,5	1	1,5	2	2,5	3	3,5
soá saûn phaåm	100	80	135	40	10	15	20

1) *Ước lượng số lượng tạp chất trung bình có trong loại sản phẩm trên với độ tin cậy là 98%*
2) *Những sản phẩm có lượng tạp chất dưới 1% là những sản phẩm loại A. Biết rằng số sản phẩm do nhà máy sản suất ra là 20000 sản phẩm, hãy ước lượng số sản phẩm loại A do nhà máy sản suất với độ tin cậy 96%.*
3) *Khi ước lượng trọng lượng trung bình của loại sản phẩm do nhà máy sản suất có độ chính xác là 0,06 thì độ tin cậy của ước lượng là bao nhiêu?*
4) *Tiến hành khảo sát 600 sản phẩm cùng loại do nhà máy thứ hai sản suất người ta thấy có 240 sản phẩm loại A. Với mức ý nghĩa 2% có kết luận được tỉ lệ sản phẩm loại A của hai nhà máy bằng nhau được hay không?*

Giaûi

- 1) Tính toaùn treân maãu ta thu ñöôïc keát quaû nhö sau

$$n = 400; \quad \sum n_i x_i = 552,25; \quad \sum n_i x_i^2 = 1011,25;$$

$$\overline{x} = \frac{1}{n} \sum n_i x_i = 1,38; \quad \overline{x^2} = \frac{1}{n} \sum n_i x_i^2 = 2,53$$

$$\hat{S}^2 = \overline{x^2} - \left(\overline{x} \right)^2 = 0,620$$

$$\hat{S} = 0,788$$

$$S^2 = \frac{n}{n-1} \hat{S}^2 = 0,623$$

$$S = 0,789$$

$$f_n = \frac{180}{400} = 0,45$$

2) Giaûi quyeát caùc yeâu caàu cuûa baøi toaùn

- a) Gọi μ là ước lượng số lượng tạp chất trung bình có trong loại sản phẩm trên với độ tin cậy là 98% ta thấy: n = 400 > 30 nên ta có công thức ước lượng:

$$\mu_1, \mu_2 = \bar{x} \mp t_\alpha \frac{S}{\sqrt{n}} \Leftrightarrow \mu_1, \mu_2 = 1,38 \mp 2,34 \frac{0,789}{\sqrt{400}}$$

$$\varepsilon = 2,34 \frac{0,789}{\sqrt{400}} = 0,092$$

$$\Rightarrow \begin{cases} \mu_1 = 1,38 - 0,092 = 1,288 \\ \mu_2 = 1,38 + 0,092 = 1,472 \end{cases} \Rightarrow 1,288 \le \mu \le 1,472 \left(98\%\right)$$

b) Gọi p là ước lượng tỉ lệ những sản phẩm loại A ta có công thức ước lượng:

$$p_1, p_2 = f_n \mp t_\alpha \sqrt{\frac{f_n(1-f_n)}{n}} = 0,45 \mp 2,06\sqrt{\frac{0,45(1-0,45)}{400}}$$

$$\varepsilon = 2,06\sqrt{\frac{0,45(1-0,45)}{400}} = 0,05$$

$$\Rightarrow \begin{cases} p_1 = 0,45 - 0,05 = 0,4 \\ p_2 = 0,45 + 0,05 = 0,5 \end{cases} \Rightarrow 0,4 \le p \le 0,5 (96\%)$$

- Gọi P là số sản phẩm loại A ta có ước lượng:

$$0,4 \times 20000 \le P \le 0,5 \times 20000 \Leftrightarrow 8000sp \le P \le 10000sp (96\%)$$

- c) Theo công thức ước lượng trung bình đám đông ta có:

$$\varepsilon = t_{\alpha}\frac{S}{\sqrt{n}} \Rightarrow \varepsilon^{2} = \left(t_{\alpha}\right)^{2}\frac{S^{2}}{n}$$

$$\Rightarrow \left(t_{\alpha}\right)^{2} = \frac{n\varepsilon^{2}}{S^{2}} = \frac{400\times(0,06)^{2}}{0,623} = 2,311$$

$$\Rightarrow t_{\alpha} = 1,52 \Rightarrow \varphi\left(t_{\alpha}\right) = \varphi\left(1,52\right) = 0,4357$$

$$\Rightarrow 1-\alpha = 2\varphi\left(t_{\alpha}\right) = 2\times0,4357 = 0,8714(87,14\%)$$

d) Theo đề ra ta có:

$$n_1 = 400,\ f_{n1} = \frac{180}{400} = 0,45 \text{ vaø } n_2 = 600,\ f_{n2} = \frac{240}{600} = 0,4$$

Gọi p_1 và p_2 lần lượt là tỷ lệ sản phẩm loại A của hai nhà máy.

Đặt giả thiết H_0: “$p_1 = p_2$” ta kiểm định giả thiết trên với mức ý nghĩa 1%

Ta có

$$\overset{\circ}{f}_n = \frac{n_1 f_{n1} + n_2 f_{n2}}{n_1 + n_2} = \frac{400 \times 0,45 + 600 \times 0,4}{400 + 600} = 0,42$$

$$\Rightarrow t = \frac{\left| f_{n1} - f_{n2} \right|}{\sqrt{\overset{\circ}{f}_n \left(1 - \overset{\circ}{f}_n\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\left|0,45 - 0,40\right|}{\sqrt{0,42(1 - 0,42)\left(\frac{1}{400} + \frac{1}{600}\right)}} = 1,569$$

$$\alpha = 1\% \Rightarrow t_\alpha = 2,58 \Rightarrow t < t_\alpha$$

- Vậy ta chấp nhận giả thiết H_0 với mức ý nghĩa 1%, nghĩa là tỷ lệ sản phẩm loại A của hai nhà máy là như nhau