# Chương 1: Sự kiện ngẫu nhiên và phép tính xác suất

Lê Xuân Lý (1)

Hà Nội, tháng 8 năm 2018

(1) Email: lexuanly@gmail.com





## Quy tắc cộng

**Ví dụ 1**

*Có 2 loại phương tiện để sinh viên đi học: phương tiện cá nhân hoặc phương tiện công cộng*
*Phương tiện cá nhận: xe đạp, xe máy, xe hơi,*
*Phương tiện công cộng: bus, taxi, xe ôm, xích lô,*
*Có bao nhiêu cách sinh viên có thể đi học? (sv chỉ chọn một trong các loại trên, không đi bộ hoặc bồ chở).*

Có 3 cách đi bằng phương tiện cá nhân và 4 cách đi bằng phương tiện công cộng.
Có $3 + 4 = 7$ cách.

# Quy tắc cộng

## Ví dụ 2

*Có 3 loại lựa chọn mua bàn ăn: bàn gỗ, bàn sắt hoặc bàn inox.*
*Bàn gỗ: có 3 kiểu,*
*Bàn sắt có 6 kiểu,*
*Bàn inox có 4 kiểu,*
*Có bao nhiêu cách mua 1 bàn ăn.*

Có $3 + 6 + 4 = 13$ cách.







# Quy tắc cộng

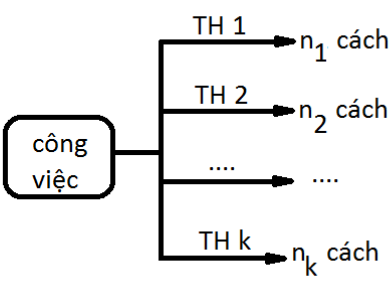


## Chú ý 1.1

Một công việc có thể chia làm $k$ trường hợp:

- trường hợp thứ nhất có $n_1$ cách giải quyết,
- trường hợp thứ 2 có $n_2$ cách giải quyết,
- ...
- trường hợp thứ $k$ có $n_k$ cách giải quyết.

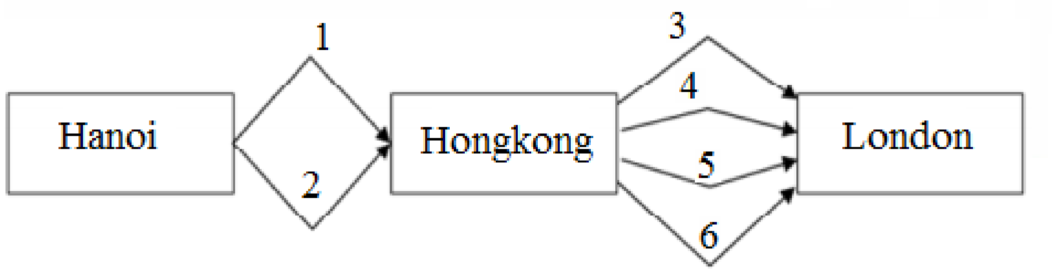
Khi đó có $n_1 + n_2 + \ldots + n_k$ cách giải quyết công việc trên.



# Quy tắc nhân

**Ví dụ 3**

*Để bay từ Hà Nội tới London phải qua trạm dừng chân tại Hong Kong. Có 2 hãng hàng không phục vụ bay từ Hà Nội tới Hong Kong (Vietnam airline, Pacific Airline) và có 4 hãng hàng không phục vụ bay từ Hong Kong tới London (Air Hong Kong Limited, Cathay Pacific Airways, CR Airways, Hong Kong Airlines).*
*Hỏi có bao nhiêu cách bay từ Hà Nội đến London qua trạm dừng chân Hong Kong?*




Để đi theo cách này ta chia làm 2 bước thực hiện:
Bước 1: HN $\Rightarrow$ HK: có 2 cách chọn,
Bước 2: HK $\Rightarrow$ LĐ: có 4 cách chọn,
Số cách đi là: $2.4 = 8$





# Quy tắc nhân

**Ví dụ 4**

*Một người có 5 cái áo,4 cái quần và 2 đôi giày. Hỏi người đó có bao nhiêu cách mặc đồ (gồm 1 áo, 1 quần và 1 đôi giày)*

Công việc chia làm 3 bước:
**Bước 1**: chọn 1 áo: có 5 cách,
**Bước 2**: chọn 1 quần: có 4 cách,
**Bước 3**: chọn 1 đôi giày: có 2 cách,
Số cách mặc đồ: $5.4.2 = 40$ cách.

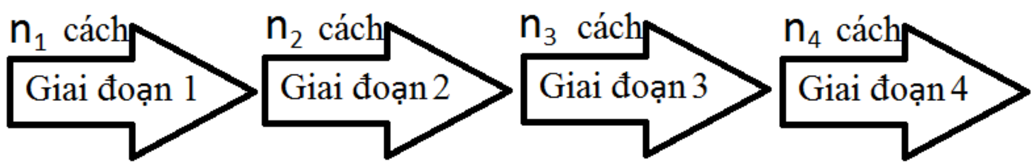
# Quy tắc nhân

**Chú ý 1.2**

Một công việc được chia làm $k$ giai đoạn:

- giai đoạn thứ nhất có $n_1$ cách giải quyết,
- giai đoạn thứ 2 có $n_2$ cách giải quyết,
- ...
- giai đoạn thứ $k$ có $n_k$ cách giải quyết.

Khi đó có $n_1 \times n_2 \ldots \times n_k$ cách giải quyết công việc trên.

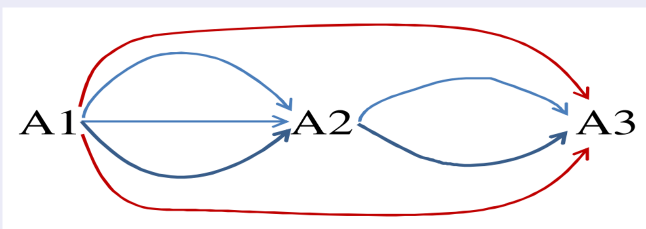



Số cách thực hiện công việc có 4 giai đoạn: $n_1 .n_2 .n_3 .n_4$





# Ví dụ




Có bao nhiêu cách đi từ A1 đến A3

Đi từ A1 đến A3 có 2 trường hợp:

- Đi trực tiếp từ A1 đến A3: có 2 cách
- Đi gián tiếp từ A1 đến A3 thông qua A2: có $3.2 = 6$

Tổng số cách đi từ A1 đến A3: $2 + 6 = 8$.

# Câu hỏi trắc nghiệm

**Có 4 cửa hàng cạnh nhau. Có 4 khách đến, mỗi khách chọn ngẫu nhiên 1 cửa hàng để vào.**

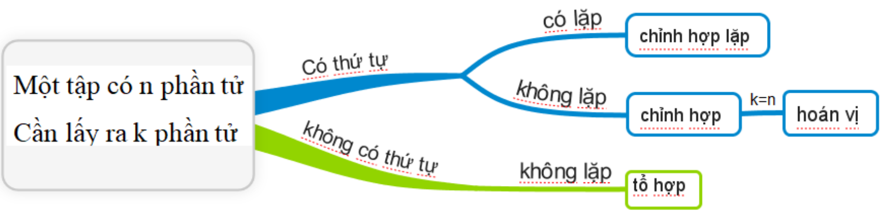
1. số trường hợp chọn cửa hàng là: A. 1 B. 4 C. 24 D. 256
   Đáp án: 1D
2. Số trường hợp chọn cửa hàng sao cho mỗi cửa hàng có đúng 1 khách vào
   A. 1 B. 4 C. 24 D. 256
   Đáp án: 2C





# Giải tích kết hợp

Ta có một tập hợp gồm $n$ phần tử, từ $n$ phần tử này ta sẽ chọn ra $k$ phần tử. Tuỳ vào điều kiện chọn các phần tử như thế nào (có thứ tự, có lặp) thì số cách chọn $k$ phần tử cũng có sự khác nhau.




**Chỉnh hợp lặp**

Chỉnh hợp lặp chập $k$ của $n$ phần tử là số cách chọn $k$ phần tử sao cho:

- Có thứ tự
- Có thể lặp lại

Ký hiệu: $\tilde{A}_n^k$.
**Công thức tính:**

$$\boxed{\tilde{A}_n^k = n^k.} \tag{1.1}$$

# Giải tích kết hợp

## Chỉnh hợp lặp

### Ví dụ 5

*Từ 5 chữ số $1, 2, 3, 4, 5$ lập được bao nhiêu số có 3 chữ số?*

*Giải*:
Chọn 3 chữ số từ 5 chữ số **có thứ tự** và **có thể lặp lại**.
Số các số có 3 chữ số được lập nên là: $\tilde{A}_5^3 = 5^3 = 125$.

### Ví dụ 6

*Xếp 5 cuốn sách khác nhau cho vào 3 ngăn. Hỏi có bao nhiêu cách phân phối sách trong 3 ngăn? (mỗi ngăn có bao nhiêu sách, loại sách gì)*

*Giải*:
Mỗi quyển sách có 3 cách cho vào ngăn.
Có 5 quyển sách.
Vậy số cách xếp là: $\tilde{A}_3^5 = 3^5 = 243$.





# Giải tích kết hợp

## Chỉnh hợp

### Chỉnh hợp

Chỉnh hợp chập $k$ của $n$ phần tử $(k \leq n)$ là số cách chọn $k$ phần tử sao cho:

- Có thứ tự
- Không thể lặp lại

Ký hiệu: $A_n^k$.

$$\boxed{A_n^k = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}.} \tag{1.2}$$

### Ví dụ 7

*Một buổi họp gồm 10 người tham dự, hỏi có mấy cách chọn 1 chủ toạ và 1 thư ký?*

*Giải*:
Chọn 2 người trong 10 người **có thứ tự** và **không lặp lại**.
Số cách chọn là $A_{10}^2 = 10.9 = 90$ (cách).

# Giải tích kết hợp

## Hoán vị

### Hoán vị

- Hoán vị của n phần tử là số cách sắp xếp $n$ phần tử đã cho theo một thứ tự nhất định.
- Ký hiệu: $P_n$.
- Hoán vị là một trường hợp đặc biệt của chỉnh hợp khi $k = n$ $(P_n = A_n^n)$.
- Công thức tính

$$\boxed{P_n = n!}. \tag{1.3}$$





# Giải tích kết hợp

## Hoán vị

### Ví dụ 8

Có 6 người khách cần xếp vào 6 ghế trên một bàn tròn 6 chỗ.
a. Nếu có quan tâm đến khung cảnh xung quanh, hỏi có bao nhiêu cách sắp xếp?
b. Nếu chỉ quan tâm đến người ngồi xung quanh là ai thì sẽ có bao nhiêu cách?

*Giải*:
a. $P_6 = 6! = 720$ (cách).
b. $P_5 = 5! = 120$ (cách).

# Giải tích kết hợp

## Tổ hợp

### Tổ hợp

Tổ hợp chập $k$ của $n$ phần tử ($k \leq n$) là số cách chọn $k$ phần tử sao cho:

- Không có thứ tự
- Không thể lặp lại

Ký hiệu: $C_n^k$.

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!}. \tag{1.4}$$

### Chú ý 1.3

- *Qui ước* $0! = 1$;
- $C_n^k = C_n^{n-k}$;
- $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$.





# Giải tích kết hợp

## Tổ hợp

### Ví dụ 9

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

*Giải*:

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

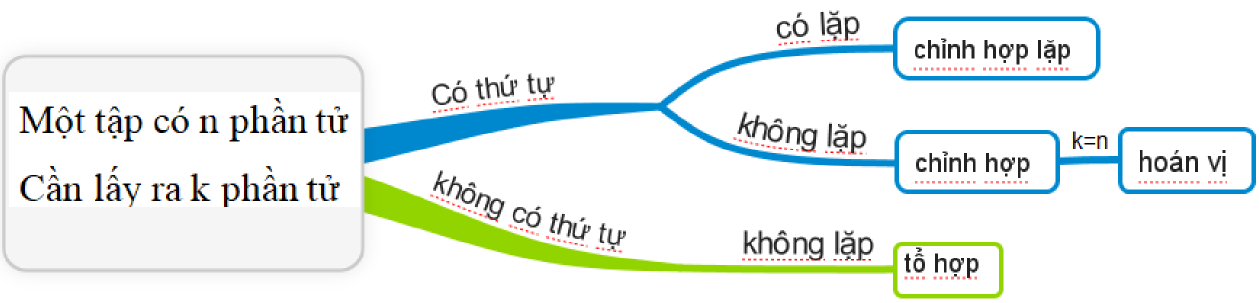
### Ví dụ 10

*Khai triển nhị thức Newton*

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

*trong đó* $a, b \in \mathbb{R}$ *và* $n \in \mathbb{N}^*$.

# Giải tích kết hợp - TỔNG KẾT








# Câu hỏi trắc nghiệm

**III. Một nhóm học sinh gồm 7 nam và 3 nữ. GV cần chọn 5 em.**

1. Số cách chọn 5 em tùy ý
   A. 2520 B. 252 C. 60 D. 30240
   Đáp án: 1B
2. Số cách chọn 5 em có ít nhất 1 nữ và 3 nam
   A. 105 B. 11025 C. 630 D. 210
   Đáp án: 2D

**IV. Một bàn dài có 10 ghế và có 10 học sinh(có bạn An và Bình).**

1. Số cách xếp 10 học sinh tùy ý vào bàn đó là:
   A. 14400 B. 3628800 C. 100 D. 125470
   Đáp án: 1B
2. Số cách xếp 10 học sinh ngồi vào bàn đó để An và Bình ngồi cạnh nhau là:
   A. 362880 B. 80640 C. 725760 D. 40320
   Đáp án: 2C

# Phép thử và sự kiện

### Định nghĩa 2.1

Việc thực hiện một nhóm các điều kiện cơ bản để quan sát một hiện tượng nào đó được gọi là một *phép thử*.
Kết cục: là một kết quả mà ta không chia nhỏ hơn được. Không gian mẫu: tập gồm tất cả các kết cục có thể xảy ra. Ký hiệu: $\Omega$
Sự kiện: là một tập con của không gian mẫu.
Đơn giản hơn: kết quả mà ta quan tâm là *sự kiện*.
Sự kiện được ký hiệu bằng chữ in: A, B, C, ...

### Ví dụ 11

*Phép thử*

- *Khảo sát thời điểm ngủ dậy buổi sáng. Ngày hôm nay mình có ngủ dậy muộn không?*
- *Sáng nay bước ra khỏi nhà. Xét xem bước chân trái hay chân phải ra trước.*
- *Quan sát thời tiết ngày hôm nay. Ngày hôm nay có mưa hay không?*
- *Mua xổ số Vietlott. Hôm nay có trúng xổ số Vietlott không?*





# Phép thử và sự kiện

Như vậy sự kiện chỉ có thể xảy ra nếu ta thực hiện phép thử.

**Sự kiện sơ cấp** : Là sự kiện không thể phân tích được nữa

**Sự kiện chắc chắn** : Là sự kiện luôn xảy ra trong phép thử, ký hiệu là $\Omega$

**Sự kiện không thể** : Là sự kiện không bao giờ xảy ra khi thực hiện phép thử. Ký hiệu là $\emptyset$.

**Sự kiện ngẫu nhiên** : Là sự kiện có thể xảy ra cũng có thể không xảy ra khi thực hiện phép thử.

**Phép thử ngẫu nhiên** : Phép thử mà các kết quả của nó là các sự kiện ngẫu nhiên.

Để thuận tiện, các sự kiện thường được ký hiệu bằng chữ in: $A, B, C, \ldots$

### Ví dụ 12

*Gieo một con xúc xắc, khi đó*

- $\Omega$= *"Gieo được mặt có số chấm* $\leq 6$ *và* $\geq 1$ *" là sự kiện chắc chắn;*
- $\emptyset$= *"Gieo được mặt* $7$ *chấm" là sự kiện không thể;*
- $A$ = *"Gieo được mặt chẵn" là sự kiện ngẫu nhiên.*

# Phép thử và sự kiện

### Ví dụ 13

*Xét một gia đình có 2 con. Gọi:*

- *A: "gia đình có 1 trai và 1 gái"*
- *B: "gia đình có 2 con"*
- *C: "gia đình có 3 con"*

*Sự kiện nào là sự kiện chắc chắn, sk không xảy ra, sự kiện ngẫu nhiên?*

### Ví dụ 14

*Hộp có 8 viên bi trong đó có 6 bi xanh và 2 bi đỏ. Lấy ngẫu nhiên ra 3 bi xem màu. Gọi:*

- *A: "lấy được 3 bi xanh"*
- *B: "lấy được 3 bi màu đỏ"*
- *C: "lấy được 3 bi"*

*Sự kiện nào là sự kiện chắc chắn, sk không xảy ra, sự kiện ngẫu nhiên?*



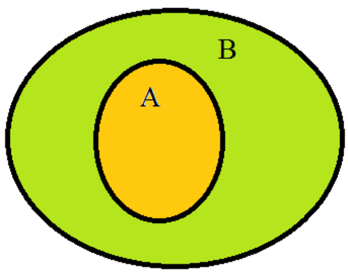


# Quan hệ của các sự kiện

Giả sử $A$ và $B$ là hai sự kiện trong cùng một phép thử.

### Quan hệ kéo theo

Sự kiện $A$ được gọi là kéo theo sự kiện $B$, ký hiệu $A \subset B$ (hoặc $A \Rightarrow B$), nếu $A$ xảy ra thì $B$ xảy ra.




### Quan hệ tương đương

Sự kiện $A$ được gọi là tương đương với sự kiện $B$, ký hiệu $A \Leftrightarrow B$ (hoặc $A = B$), nếu $A \Rightarrow B$ và $B \Rightarrow A$.

### Ví dụ 15

- *Sinh viên mua một tờ vé số. Gọi:*
  *A: "sv có vé số trúng giải đặc biệt"*
  *B: "sv có vé số trúng giải"*
- $A \Rightarrow B$ *hay* $B \Rightarrow A$
- *dùng biểu đồ Ven để minh họa*

### Ví dụ 16

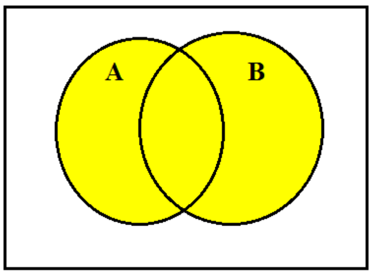
- *Tung một con xúc xắc 1 lần. Gọi:*
  *A: "xúc xắc ra mặt có số chấm chẵn"*
  *B: "xúc xắc ra mặt có số chấm 2 hoặc 4"*
  *C: "xúc xắc ra mặt có số chấm 2, 4, 6"*
  *D: "xúc xắc ra mặt có số chấm nhỏ hơn 4"*
- $A \Rightarrow B$ *hay* $B \Rightarrow A$
- $A \Rightarrow C$ *hay* $C \Rightarrow A$
- $A \Rightarrow D$ *hay* $D \Rightarrow A$





### Sự kiện tổng

Sự kiện $C$ được gọi là tổng của 2 sự kiện $A$ và $B$, ký hiệu là $C = A + B$, nếu $C$ xảy ra khi và chỉ khi ít nhất một trong 2 sự kiện $A$ và $B$ xảy ra.




**A + B**

### Ví dụ 17

*Hai người thợ săn cùng bắn một con thú. Nếu gọi $A$ là sự kiện người thứ nhất bắn trúng con thú và $B$ là sự kiện người thứ 2 bắn trúng con thú, khi đó $C = A + B$ là sự kiện con thú bị bắn trúng.*

# Quan hệ và phép toán của các sự kiện

### Chú ý 2.1

- $A_1 + A_2 + \cdots + A_n$ là sự kiện xảy ra khi có ít nhất một trong $n$ sự kiện đó xảy ra
- Mọi sự kiện ngẫu nhiên đều có thể biểu diễn dưới dạng tổng của một số sự kiện sơ cấp nào đó.
- Sự kiện chắc chắn $\Omega$ là tổng của mọi sự kiện sơ cấp có thể. Do đó $\Omega$ còn được gọi là không gian các sự kiện sơ cấp.

### Ví dụ 18

*Tung một con xúc xắc. Ta có* $6$ *sự kiện sơ cấp* $A_i$ $(i = \overline{1, 6})$, *trong đó* $A_i$ *là sự kiện xuất hiện mặt* $i$ *chấm* $i = 1, 2, \ldots, 6$.

- $A$= *"Xuất hiện mặt có số chấm chẵn", ta suy ra* $A = A_2 + A_4 + A_6$
- $B$ = *"Xuất hiện mặt có số chấm không vượt quá* $3$*", ta suy ra* $B = A_1 + A_2 + A_3$.

*Khi đó* $C = A + B = A_1 + A_2 + A_3 + A_4 + A_6$.

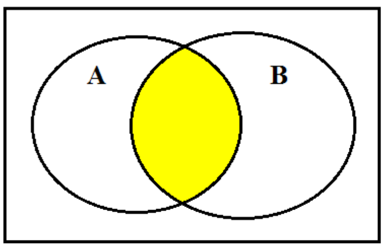




# Quan hệ và phép toán của các sự kiện

### Sự kiện tích

- Sự kiện $C$ được gọi là tích của 2 sự kiện $A$ và $B$, ký hiệu $C = A.B$ (hoặc $AB$), nếu $C$ xảy ra khi và chỉ khi cả $A$ và $B$ cùng xảy ra.
- Tích của $n$ sự kiện $A_1.A_2 \ldots A_n$ là sự kiện xảy ra khi cả $n$ sự kiện cùng xảy ra.




A.B

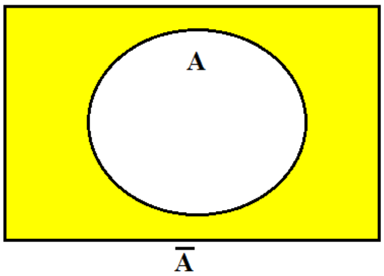
### Ví dụ 19

*Hai người thợ săn cùng bắn một con thú. Nếu gọi* $A$ *là sự kiện người thứ nhất bắn trượt con thú và* $B$ *là sự kiện người thứ* $2$ *bắn trượt con thú, khi đó* $C = A.B$ *là sự kiện con thú không bị bắn trúng.*

# Quan hệ và phép toán của các sự kiện

### Sự kiện đối lập

Sự kiện đối lập với sự kiện A, ký hiệu là $\overline{A}$, là sự kiện xảy ra khi $A$ không xảy ra. Ta có




### Ví dụ 20

*Gieo một con xúc xắc một lần, khi đó*

- $A$ = *"Gieo được mặt chẵn" suy ra* $\overline{A}$= *"Gieo được mặt lẻ"*
- $A$ = *"Gieo được mặt 1 chấm" suy ra* $\overline{A}$= *"Gieo không được mặt 1 chấm"*





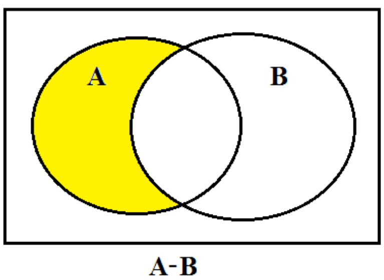
# Quan hệ và phép toán của các sự kiện

### Sự kiện hiệu

Hiệu của 2 sự kiện $A$ và $B$, ký hiệu là $A - B$, là sự kiện xảy ra khi và chỉ khi $A$ xảy ra nhưng $B$ không xảy ra.
Trường hợp hay sử dụng sự kiện hiệu:

$$\bar{A} = \Omega - A$$
$$A = \Omega - \bar{A}$$

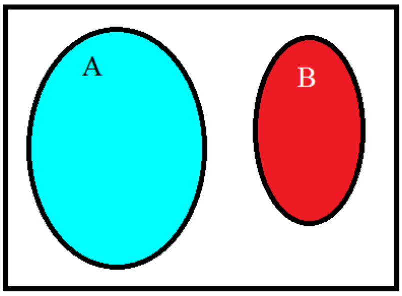
Trường hợp tổng quát: ta biến đổi thành sự kiện tích như sau: $A - B = A.\overline{B}$.



# Quan hệ và phép toán của các sự kiện

## Hai sự kiện xung khắc

Hai sự kiện $A$ và $B$ được gọi là xung khắc với nhau nếu chúng không đồng thời xảy ra trong một phép thử. $A$ và $B$ xung khắc $\Leftrightarrow A.B = \emptyset$.




A.B = Ø

## Ví dụ 21

*Một xạ thủ bắn 1 viên đạn vào bia. Gọi $A$ là sự kiện xạ thủ đó bắn trúng vòng $8$ và $B$ là sự kiện xạ thủ đó bắn trúng vòng $10$. Khi đó ta thấy ngay $AB = \emptyset$ tức là $A, B$ là 2 sự kiện xung khắc với nhau.*





# Quan hệ và phép toán của các sự kiện

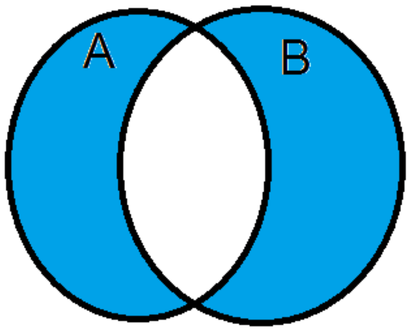
## Các tính chất

- Giao hoán
$$A + B = B + A \qquad A.B = B.A$$
- Kết hợp
$$A + B + C = (A + B) + C = A + (B + C)$$
$$A.B.C = (A.B).C = A.(B.C)$$
- Phân phối của phép cộng và phép nhân
$$A.(B + C) = A.B + A.C$$
- Đặc biệt
$A + A = A \qquad A.A = A$
$A + \Omega = \Omega \qquad A.\Omega = A$
$A + \emptyset = A \qquad A.\emptyset = \emptyset$

# Trắc nghiệm

**I. Miền được tô màu ở hình dưới được biểu diễn bởi:**

A. $(A.\bar{B}).(\bar{A}.B)$
B. $(A+\bar{B})(\bar{A}+B)$
C. $A.\bar{B}+\bar{A}.B$
D. cả 3 kết quả trên đều sai








# Trắc nghiệm

**II. Gieo một con xúc xắc lý tưởng.**
**A: "số chấm xuất hiện là lẻ"**
**B: "số chấm xuất hiện là lớn hơn hoặc bằng 4"**
**C: "số chấm xuất hiện nhiều nhất là 2"**

1. Sự kiện $\bar{A}$ là:
   A. { } B. { 1;3;5} C. { 1;3} D. { 2;4;6}
2. Sự kiện $A.B$ là:
   A. { 5;7} B. { 5;6} C. { 5} D. {1;3;5;6}
3. Sự kiện $B+C$ là:
   A. { $\Phi$ } B. { 1;4;5;6} C. { 1;5;6} D. {1;2;5;6}

# Trắc nghiệm

**III. Có 3 sv A, B, C cùng thi môn XSTK.**
**Gọi $A_i$: "có i sv thi qua môn XSTK", $i = 0, 1, 2, 3$.**

1. Gọi $B$: "sinh viên B thi qua môn XSTK". Sự kiện $A_2.\bar{B}$ là:
   A. sv B thi hỏng B. chỉ có sv B thi đỗ
   C. có 2 sv thi đỗ D. chỉ có sv B thi hỏng
2. Sự kiện $\overline{A_0}.\bar{B}$ là:
   A. sv B thi hỏng B. sv A hoặc C thi đỗ
   C. có 2 sv thi đỗ D. sv A và C thi đỗ
3. Chọn đáp án đúng:
   A. $\overline{A_0}.\bar{B} \subset \overline{A_1}.\bar{B}$ B. $\overline{A_1}.\bar{B} \subset \overline{A_2}$
   C. $\overline{A_0}.\bar{B} = A_1.\bar{B}$ D. $\overline{A_3}.\bar{B} \subset \overline{A_3}$
4. Gọi $H$: "có một sinh viên thi hỏng". Kết quả nào ĐÚNG
   A. $A_1.A_2.\overline{A_3} = H$ B. $\overline{A_1} = H$
   C. $\overline{A_1}.A_2.A_3 \subset H$ D. $\overline{A_2}.A_3 \subset H$





# Xác suất của một sự kiện

**Định nghĩa 3.1**

Xác suất của một sự kiện là một số nằm giữa $0$ và $1$, số này đo lường khả năng xuất hiện của sự kiện đó khi phép thử được thực hiện. Ký hiệu xác suất của sự kiện $A$ là $P(A)$.

**Một số tính chất cơ bản**

- $0 \leq P(A) \leq 1$;
- $P(\Omega) = 1; \quad P(\emptyset) = 0$;
- $P(A) + P\left(\overline{A}\right) = 1$.

# Định nghĩa xác suất theo cổ điển

## Định nghĩa 3.2

Xét một phép thử có hữu hạn kết cục có thể xảy ra (có $n_\Omega$ kết cục), đồng thời các kết cục này là đồng khả năng xuất hiện; trong đó có $n_A$ kết quả thuận lợi cho sự kiện $A$. Khi đó:

$$\boxed{P(A) = \frac{n_A}{n_\Omega} = \frac{\text{Số kết cục thuận lợi cho } A}{\text{Số kết cục có thể xảy ra}}.} \tag{3.1}$$

## Ví dụ 22

Một người gọi điện thoại nhưng lại quên $2$ chữ số cuối của số điện thoại cần gọi mà chỉ nhớ là $2$ chữ số đó khác nhau. Tìm xác suất để người đó chọn ngẫu nhiên $1$ số để gọi thì trúng số cần gọi.

*Giải*:

- Gọi $A$: "Người đó chọn ngẫu nhiên 1 số gọi thì trúng số cần gọi".
- $P(A) = \dfrac{n_A}{n_\Omega} = \dfrac{1}{90}$.



Các định nghĩa xác suất | Định nghĩa xác suất theo cổ điển

# Định nghĩa xác suất theo cổ điển

## Ví dụ 23

Từ bộ bài túlơkhơ 52 cây đã trộn kỹ rút ngẫu nhiên ra 2 cây. Tính xác suất xảy ra các sự kiện sau:

1. $A$: "2 cây rút ra đều là Át";
2. $B$: "2 cây rút ra có 1 cây Át, 1 cây K";
3. $C$: "2 cây rút ra có ít nhất 1 cây Át"

*Giải*:

Số kết cục lấy $2$ cây bài: $n_\Omega = C_{52}^2 = 1326$.

1. $P(A) = \dfrac{n_A}{n_\Omega} = \dfrac{C_4^2}{n_\Omega} = \dfrac{1}{221}$.
2. $P(B) = \dfrac{n_B}{n_\Omega} = \dfrac{C_4^1.C_4^1}{n_\Omega} = \dfrac{8}{663}$.
3. Ta có $\overline{C} =$ "2 cây đều không phải là Át".
   $P(C) = 1 - p(\overline{C}) = 1 - \dfrac{C_{48}^2}{C_{52}^2} = 1 - \dfrac{188}{221} = \dfrac{33}{221}$

# Trắc nghiệm

1. Tung 2 lần liên tiếp một đồng xu (khả năng ra sấp và ngửa như nhau). Xác suất cả 2 lần đều xuất hiện mặt sấp là:
   A. 0 B. 1/4 C. 1/2 D. 1
2. Trong hộp có 10 viên bi cùng kích cỡ (6 trắng 4 đen). Lấy ngẫu nhiên 2 viên bi. Xác suất cả 2 bi màu trắng là:
   A. 1/5 B. 1/3 C. 1/2 D. 1
3. Trong hộp có 10 viên bi cùng kích cỡ (6 trắng 4 đen). Lấy ngẫu nhiên 2 viên bi. Xác suất có 1 bi trắng và 1 bi đen là:
   A. 1/45 B. 10/45 C. 24/45 D. 1





# Định nghĩa xác suất theo quan điểm hình học

**Định nghĩa 3.3**

Giả sử tập hợp vô hạn các kết cục đồng khả năng của một phép thử có thể biểu thị bởi một miền hình học $\Omega$ có độ đo (độ dài, diện tích, thể tích, ...) hữu hạn khác $0$, còn tập các kết cục thuận lợi cho sự kiện $A$ là một miền $A$. Khi đó xác suất của sự kiện $A$ được xác định bởi:

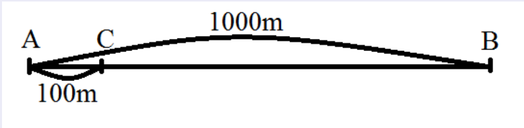
$$\boxed{P(A) = \frac{\text{Độ đo của miền } A}{\text{Độ đo của miền } \Omega} = \frac{|A|}{|\Omega|}} \tag{3.2}$$

Khái niệm đồng khả năng trên $\Omega$ có nghĩa là điểm gieo có thể rơi vào bất kỳ điểm nào của $\Omega$ và xác suất để nó rơi vào một miền con nào đó của $\Omega$ tỉ lệ với độ đo của miền ấy.

# Định nghĩa xác suất theo quan điểm hình học

**Ví dụ 24**

Đường dây điện thoại ngầm nối một tổng đài với một trạm dài 1km. Tính xác suất để dây đứt cách tổng đài không quá 100m.




**Giải**

Rõ ràng nếu dây đồng chất thì khả năng bị đứt tại một điểm bất kỳ trên dây là như nhau, nên tập hợp các kết quả có thể xảy ra có thể biểu thị bằng đoạn thẳng nối tổng đài với trạm dài 1km. Còn sự kiện $A :=$ "Dây bị đứt cách tổng đài không quá 100m" được biểu thị bằng độ dài 100m. Khi đó ta có

$$P(A) = \frac{100}{1000} = 0.1.$$





# Định nghĩa xác suất theo tần suất (theo thống kê)

Do tính đồng khả năng là rất khó có được trong thực tế, nên cần có một cách khác để xác định xác suất của một sự kiện.

**Định nghĩa 3.4**

Giả sử một phép thử có thể thực hiện lặp lại nhiều lần trong những điều kiện giống nhau. Nếu trong $n$ lần thực hiện phép thử trên có $m$ lần xuất hiện sự kiện $A$, khi đó tỉ lệ $f_n(A) = \dfrac{m}{n}$ được gọi là tần suất xuất hiện của sự kiện $A$ trong $n$ phép thử.

Cho số phép thử tăng lên vô hạn:

$$P(A) = \lim_{n\to\infty} f_n(A) = \lim_{n\to\infty} \frac{m}{n}.$$

Thực tế $\boxed{P(A) \approx \dfrac{m}{n}}$ với $n$ đủ lớn.

# Định nghĩa xác suất theo tần suất (theo thống kê)

**Ví dụ 25**

Để xác định xác suất của một người đàn ông $25$ tuổi sẽ bị chết trong vòng $1$ năm sắp tới, người ta theo dõi $100000$ nam thanh niên $25$ tuổi và thấy rằng có $138$ người chết. Vậy xác suất cần tìm xấp xỉ bằng:

$$\frac{138}{100000} = 0.00138.$$

**Chú ý 3.1**

Định nghĩa này chỉ dùng được cho các phép thử ngẫu nhiên có thể lặp lại nhiều lần một cách độc lập trong các điều kiện giống nhau. Ngoài ra để xác định một cách tương đối chính xác giá trị của xác suất ta phải tiến hành một số đủ lớn các phép thử, mà việc này đôi khi không thể thực hiện được do hạn chế về thời gian và kinh phí.





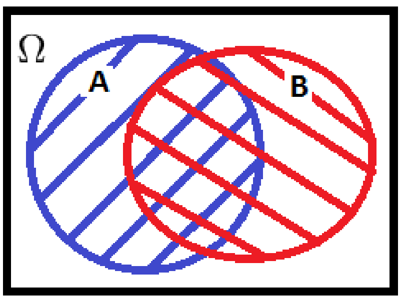
# Công thức cộng xác suất

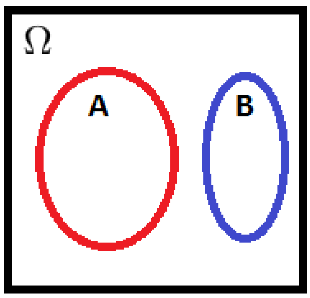
- **Công thức cộng xác suất**: Nếu $A$ và $B$ là hai sự kiện bất kỳ thì ta có

$$\boxed{P(A+B) = P(A) + P(B) - P(AB).} \tag{4.3}$$

- Nếu $A$ và $B$ là hai sự kiện xung khắc thì

$$\boxed{P(A+B) = P(A) + P(B)}. \tag{4.4}$$



# Công thức cộng xác suất

- **Công thức cộng xác suất tổng quát**: Cho $n$ sự kiện bất kỳ $\{A_i\}$, $i = \overline{1, n}$. Khi đó ta có

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \cdots + (-1)^{n-1} P\left(\prod_{i} A_i\right). \quad (4.5)$$

- **Trường hợp đặc biệt**: Khi các sự kiện $A_i$, $i = \overline{1, n}$ xung khắc từng đôi, tức là $A_iA_j = \emptyset\ \forall i \neq j$ thì ta có

$$\boxed{P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).} \quad (4.6)$$





# Công thức cộng xác suất

**Ví dụ 26**

Một lô hàng gồm 10 sản phẩm, trong đó có 2 phế phẩm. Lấy ngẫu nhiên đồng thời từ lô hàng ra 6 sản phẩm. Tìm xác suất để có không quá 1 phế phẩm trong 6 sản phẩm được lấy ra.

# Công thức cộng xác suất

**Bài làm**

Gọi

- $A$: "không có phế phẩm trong sản phẩm"
- $B$: "có đúng 1 phế phẩm trong sản phẩm"
- $C$: "có không quá 1 phế phẩm trong sản phẩm"

Dễ dàng thấy $A$ và $B$ là 2 sự kiện xung khắc và $C = A + B$. Ngoài ra

$$P(A) = \frac{C_8^6}{C_{10}^6} = \frac{2}{15}; \quad P(B) = \frac{C_2^1 C_8^5}{C_{10}^6} = \frac{8}{15}.$$

Do đó $P(C) = P(A+B) = P(A) + P(B) = \dfrac{2}{15} + \dfrac{8}{15} = \dfrac{2}{3}.$





# Công thức cộng xác suất

**Ví dụ 27**

Một lớp có $100$ sinh viên, trong đó có:
$40$ sinh viên giỏi ngoại ngữ, $30$ sinh viên giỏi tin học,
$20$ sinh viên giỏi cả ngoại ngữ lẫn tin học.
Chọn ngẫu nhiên một sinh viên trong lớp. Tìm xác suất để sinh viên đó giỏi ít nhất 1 trong 2 môn trên.

# Công thức cộng xác suất

**Bài làm**

Gọi

- $A$ : "sinh viên đó giỏi ít nhất 1 trong 2 môn ngoại ngữ, tin học"
- $N$ : "sinh viên đó giỏi ngoại ngữ"
- $T$ : "sinh viên đó giỏi tin học"

Dễ thấy $A = T + N$, do đó

$$P(A) = P(T+N) = P(T) + P(N) - P(TN) = \frac{30}{100} + \frac{40}{100} - \frac{20}{100} = \frac{50}{100} = 0.5.$$





# Xác suất có điều kiện

**Định nghĩa 4.1**

Xác suất xảy ra sự kiện $A$ với điều kiện sự kiện $B$ đã xảy ra được gọi là xác suất có điều kiện $B$ của sự kiện $A$. Ký hiệu là $P(A|B)$.

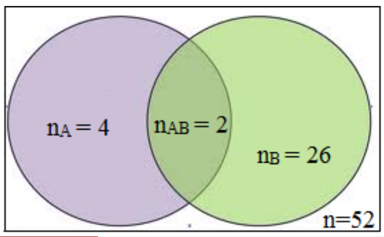
**Ví dụ 28**

Từ một bộ bài tú lơkhơ 52 cây đã trộn kỹ rút ngẫu nhiên ra một cây bài. Biết đó là cây đen, tính xác suất đó là cây át.

**Bài làm**

Gọi $A$ "rút được cây át" và $B$ "rút được cây đen". Xác suất cần tính là $P(A|B)$.

$$P(A|B) = \frac{2}{26} = \frac{n_{AB}}{n_B} = \frac{n_{AB}/n}{n_B/n} = \frac{P(AB)}{P(B)}.$$




# Xác suất có điều kiện

### Công thức tính

$$\boxed{P(A|B) = \frac{P(AB)}{P(B)}.} \tag{4.7}$$





# Công thức nhân xác suất

### Công thức nhân xác suất

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

### Định nghĩa 4.2

Hai sự kiện $A$ và $B$ được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra sự kiện này không làm ảnh hưởng tới việc xảy ra hay không xảy ra sự kiện kia. Ta có:

$$\begin{cases} P(A) & = P(A|B) = P(A|\overline{B}) \\ P(B) & = P(B|A) = P(B|\overline{A}). \end{cases}$$

Hai sự kiện $A$ và $B$ độc lập với nhau khi và chỉ khi

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

### Chú ý 4.1

*Nếu $A$ và $B$ độc lập thì các cặp sau cũng độc lập: $A$ và $\overline{B}$ ; $\overline{A}$ và $B$ ; $\overline{A}$ và $\overline{B}$*

# Công thức nhân xác suất

**Tổng quát**

Cho $n$ sự kiện $A_1, A_2, \ldots, A_n$. Khi đó xác suất tích được tính như sau:

$$\boxed{P(A_1A_2\ldots A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1}).}$$

**Định nghĩa 4.3**

Các sự kiện $A_1, A_2, \ldots, A_n$ được gọi là độc lập (hay độc lập trong tổng thể) nếu việc xảy ra hay không xảy ra của một nhóm bất kỳ $k$ sự kiện $(1 \leq k \leq n)$ không làm ảnh hưởng tới việc xảy ra hay không xảy ra của các sự kiện còn lại.
Khi đó ta có: $P(A_1.A_2\ldots A_n) = P(A_1).P(A_2)\ldots P(A_n)$





# Công thức nhân xác suất

**Ví dụ 29**

Ba xạ thủ độc lập với nhau, mỗi người bắn một viên đạn vào bia với xác suất bắn trúng của từng người tương ứng là $0.7; 0.8$ và $0.9$. Tính xác suất:

1. Có đúng 2 người bắn trúng;
2. Có ít nhất 1 người bắn trúng.

# Công thức nhân xác suất

**Giải**

Gọi $A_i$: "người thứ $i$ bắn trúng bia" với $i = 1, 2, 3$. Theo bài ra ta có $A_1, A_2, A_3$ xung khắc với nhau (từng đôi) và $P(A_1) = 0.7;\quad P(A_2) = 0.8;\quad P(A_3) = 0.9$.

1. Gọi $A$: "Có đúng hai người bắn trúng", khi đó
   $A = A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3$.
   Dùng tính xung khắc của ba số hạng trong tổng và tính độc lập của các sự kiện $A_1, A_2, A_3$ ta có:
   $$\begin{aligned} P(A) &= P\left(A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3\right) \\ &= P\left(A_1A_2\overline{A}_3\right) + P\left(A_1\overline{A}_2A_3\right) + P\left(\overline{A}_1A_2A_3\right) \\ &= P(A_1)P(A_2)P\left(\overline{A}_3\right) + P(A_1)P\left(\overline{A}_2\right)P(A_3) + P\left(\overline{A}_1\right)P(A_2)P(A_3) \\ &= 0.7\times0.8\times(1-0.9) + 0.7\times(1-0.8)\times0.9 + (1-0.7)\times0.8\times0.9 = 0.398. \end{aligned}$$
2. Gọi $B$: "Có ít nhất 1 người bắn trúng bia" suy ra $\overline{B}$: "Không có ai bắn trúng". Ta có $\overline{B} = \overline{A}_1\overline{A}_2\overline{A}_3$, suy ra
   $P(B) = 1 - P(\overline{B}) = 1 - P\left(\overline{A}_1\overline{A}_2\overline{A}_3\right) = 1 - P\left(\overline{A}_1\right)P\left(\overline{A}_2\right)P\left(\overline{A}_3\right) =$
   $1 - 0.3\times0.2\times0.1 = 0.994$.





# Trắc nghiệm

1. Cho $P(A) = 1/3, P(B) = 1/4, P(C) = 1/12$. $A$ và $B$ là 2 sự kiện:
   A. độc lập
   B. xung khắc
   C. không độc lập và không xung khắc
2. Cho $P(A) = 1/3, P(B) = 1/4, P(C) = 6/12$. $A$ và $B$ là 2 sự kiện:
   A. độc lập
   B. xung khắc
   C. không độc lập và không xung khắc
3. Cho $P(A) = 1/3, P(B) = 1/4, P(C) = 7/12$. $A$ và $B$ là 2 sự kiện:
   A. độc lập
   B. xung khắc
   C. không độc lập và không xung khắc

# Công thức nhân xác suất

## Ví dụ 30

*Một người thỏa thuận với vợ sắp cưới như sau: anh ta chỉ cần có con trai. Nếu vợ anh sinh cho anh một đứa con trai thì lập tức dừng lại liền, không sinh nữa. Giả sử một người phụ nữ sinh tối đa $n$ lần, và xác suất sinh con trai ở mỗi lần là 1/2 (khả năng sinh con trai ở mỗi lần sinh không ảnh hưởng tới nhau).*
*a. Hỏi khả năng anh này có con trai là bao nhiêu?*
*b. Hỏi n phải là bao nhiêu thì khả năng anh này có con trai lớn hơn hoặc bằng 90%.*

## Giải

a. Gọi $T_i$ : "sinh con trai ở lần sinh thứ $i$", $i = 0, 1, 2, ..., n$
T: "anh này có con trai ".
$P(T) = 1 - P(\overline{T}) = 1 - P(\overline{T_1}.\overline{T_2}...\overline{T_n})$
$= 1 - 0,5^n$.
b. $P(T) \geq 0,99 \Leftrightarrow 1 - 0,5^n \geq 0,90 \Leftrightarrow 0,5^n \leq 0,01$
$\Leftrightarrow n \geq \dfrac{ln0,1}{ln0,5} \Leftrightarrow n \geq 3,322$
Vậy $n \geq 4$. :(





# Công thức nhân xác suất

## Ví dụ 31

Có 4 que thăm, trong đó có 3 que thăm dài bằng nhau và 1 que thăm ngắn hơn. Bốn người lần lượt lên rút ngẫu nhiên một que thăm. Tính xác suất người thứ $i$ rút được thăm ngắn $(i = 1, 2, 3, 4)$.

## Giải

Gọi $A_i$: "Người thứ $i$ rút được thăm ngắn" với $i = 1, 2, 3, 4$. Ta có

$$\begin{aligned}
P(A_1) &= \frac{1}{4};\\
P(A_2) &= P(\overline{A}_1).P(A_2|\overline{A}_1) = \frac{3}{4}.\frac{1}{3} = \frac{1}{4};\\
P(A_3) &= P(\overline{A}_1\overline{A}_2A_3) = P(\overline{A}_1)\,P(\overline{A}_2|\overline{A}_1)\,P(A_3|\overline{A}_1\overline{A}_2) = \frac{3}{4}.\frac{2}{3}.\frac{1}{2} = \frac{1}{4};\\
P(A_4) &= \frac{1}{4}.
\end{aligned}$$

Vậy khả năng rút được thăm ngắn của 4 người là như nhau và bằng $\dfrac{1}{4}$.

# Công thức Bernoulli

### Định nghĩa 4.4

*(Dãy phép thử Bernoulli)* Tiến hành $n$ phép thử độc lập. Giả sử trong mỗi phép thử chỉ có thể xảy ra một trong hai trường hợp: hoặc sự kiện $A$ xảy ra hoặc sự kiện $A$ không xảy ra. Xác suất xảy ra sự kiện $A$ trong mỗi phép thử luôn bằng $p$. Đó chính là dãy phép thử Bernoulli.

### Công thức Bernoulli

Xác suất để sự kiện $A$ xuất hiện đúng $k$ lần trong $n$ phép thử của dãy phép thử Bernoulli là:

$$\boxed{p_n(k) = C_n^k p^k q^{n-k}, \quad q = 1 - p; k = 0, 1, \dots, n.} \tag{4.8}$$

### Ví dụ 32

- *Gieo một đồng tiền* $10$ *lần. Ta quan tâm ra mặt sấp*
- $5$ *xạ thủ, mỗi người bắn* $1$ *viên vào mục tiêu. Ta quan tâm đến số người bắn trúng*
- *Gieo một con xúc xắc* $100$ *lần, ta quan tâm đến sự kiện ra mặt lục*



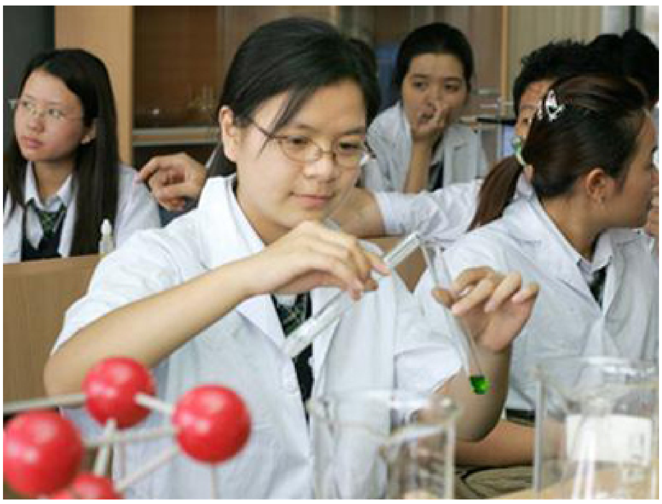


# Công thức Bernoulli

### Ví dụ 33

Xác suất thành công của một thí nghiệm sinh hóa là $40\%$. Một nhóm gồm $9$ sinh viên tiến hành cùng thí nghiệm trên độc lập với nhau. Tìm xác suất để:

1. Có đúng $6$ thí nghiệm thành công
2. Có ít nhất $1$ thí nhiệm thành công



# Công thức Bernoulli

**Giải**

Phép thử là tiến hành thí nghiệm. $A$ là sự kiện thí nghiệm thành công. Ta có

$$p = P(A) = 0.4; \quad q = 1 - p = 0.6; \quad n = 9.$$

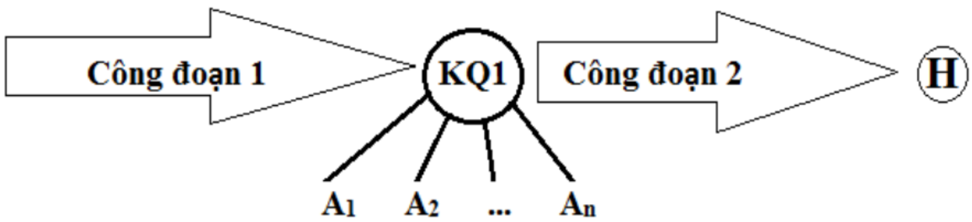
1. Xác suất cần tính: $p_9(6) = C_9^6 p^6 q^3 = C_9^6 (0.4)^6 (0.6)^3 = 0.0743$.
2. Gọi $B$ là sự kiện "có ít nhất 1 thí nghiệm thành công".
   Ta có $\overline{B}$: "không có thí nghiệm nào thành công". Khi đó

$$P(B) = 1 - P\left(\overline{B}\right) = 1 - (0.6)^9 = 0.9899.$$





# Công thức xác suất đầy đủ và công thức Bayes




**Mục tiêu**: Tính xác suất xảy ra kết quả H sau công đoạn 2.
**Khó khăn**: Kết quả công đoạn 2 phụ thuộc vào kết quả công đoạn 1.
Các kết quả của công đoạn 1 được chia làm n tập $A_i$, mỗi một tập sẽ gồm một số kết quả có ảnh hưởng giống nhau đến khả năng xảy ra H.

# Khái niệm nhóm đầy đủ

### Định nghĩa 5.1

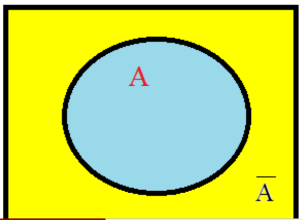
Nhóm các sự kiện $A_1, A_2, \dots, A_n$ $(n \geq 2)$ của một phép thử được gọi là một nhóm đầy đủ nếu thỏa mãn 2 điều kiện:

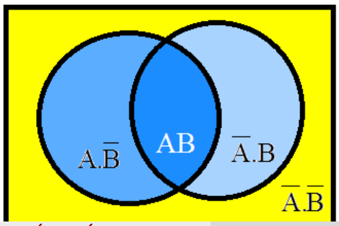
- $A_i A_j = \emptyset \ \forall i \neq j$;
- $A_1 + A_2 + \cdots A_n = \Omega$.

Tính chất: $P(A_1) + P(A_2) + ... + P(A_n) = 1$

### Chú ý 5.1

- Đối với một sự kiện $A$ thì ta có nhóm đầy đủ $\{A, \overline{A}\}$
- Đối với 2 sự kiện A và B,một nhóm đầy đủ: $\{AB, A\overline{B}, \overline{A}B, \overline{A}.\overline{B}\}$.








# Khái niệm nhóm đầy đủ

### Ví dụ 34

Xét phép thử gieo một con xúc xắc 1 lần.

- Gọi $A_i$: "Gieo được mặt $i$ chấm" với $i = 1, 2, \dots, 6$. Ta có nhóm đầy đủ $A_1, A_2, \dots, A_6$.
- Gọi
    - $A$: "Gieo được mặt chẵn"
    - $B$: "Gieo được mặt 1 chấm hoặc 3 chấm"
    - $C$: "Gieo được mặt 5 chấm"

  Khi đó $A, B, C$ là một nhóm đầy đủ.

# Công thức xác suất đầy đủ

## Công thức xác suất đầy đủ

Giả sử $A_1, A_2, \ldots, A_n$ là một nhóm đầy đủ các sự kiện. Xét sự kiện $H$ sao cho $H$ chỉ xảy ra khi một trong các sự kiện $A_1, A_2, \ldots, A_n$ xảy ra. Nói cách khác $H$ xảy ra thì một sự kiện $A_i$ nào đó xảy ra. Khi đó ta có công thức xác suất đầy đủ

$$\boxed{P(H) = \sum_{i=1}^{n} P(A_i).P(H|A_i).} \tag{5.9}$$

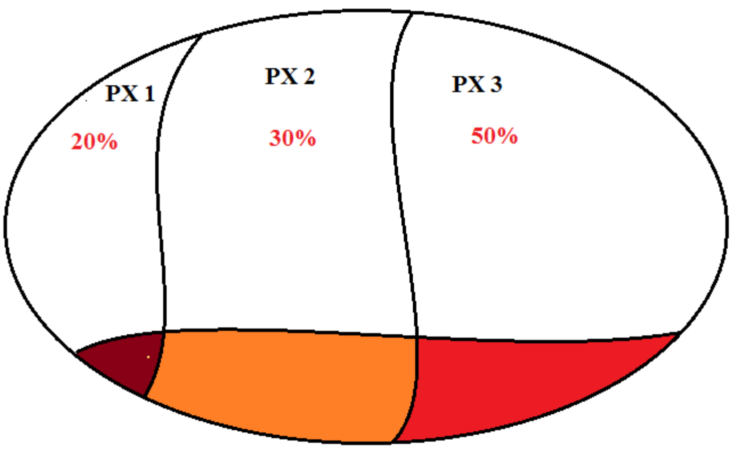




# Công thức xác suất đầy đủ

## Ví dụ 35

Xét một lô sản phẩm có số lượng rất lớn trong đó số sản phẩm do phân xưởng I sản xuất chiếm 20%, phân xưởng II sản xuất chiếm 30%, phân xưởng III sản xuất chiếm 50%. Xác suất phế phẩm của phân xưởng I là $0.001$; phân xưởng II là $0.005$; phân xưởng III là $0.006$. Lấy ngẫu nhiên 1 sản phẩm của lô hàng. Tìm xác suất để sản phẩm đó là phế phẩm.



# Công thức xác suất đầy đủ

**Giải**

Gọi $H$: "Sản phẩm lấy ra là phế phẩm"; $A_i$: "Sản phẩm đó do phân xưởng $i$ sản xuất" $i = 1, 2, 3$. Ta có $\{A_1, A_2, A_3\}$ là một nhóm đầy đủ và

$$\begin{aligned}&P(A_1) = 0.2; \quad P(A_2) = 0.3; \quad P(A_3) = 0.5\\ &P(H|A_1) = 0.001; \quad P(H|A_2) = 0.005; \quad P(H|A_3) = 0.006.\end{aligned}$$

Áp dụng công thức xác suất đầy đủ ta có

$$\begin{aligned}P(H) &= P(A_1).P(H|A_1) + P(A_2).P(H|A_2) + P(A_3).P(H|A_3)\\ &= 0.2 \times 0.001 + 0.3 \times 0.005 + 0.5 \times 0.006 = 0.0047.\end{aligned}$$





# Công thức xác suất đầy đủ

**Ví dụ 36**

Có hai chuồng thỏ. Chuồng thỏ thứ nhất có $3$ thỏ trắng và $3$ thỏ nâu. Chuồng thỏ thứ hai có $6$ thỏ trắng và $4$ thỏ nâu. Bắt ngẫu nhiên $2$ con thỏ từ chuồng thứ nhất bỏ vào chuồng thứ hai rồi sau đó bắt ngẫu nhiên $1$ con thỏ từ chuồng thứ hai ra. Tính xác suất bắt được thỏ nâu từ chuồng thứ hai.

**Giải**

Gọi $A_i$: "Trong $2$ con thỏ bắt từ chuồng một có $i$ con thỏ nâu" , $i = 0, 1, 2$. Ta có $A_0, A_1, A_2$ lập thành một nhóm đầy đủ. Gọi $H$: "Bắt được thỏ nâu từ chuồng hai". Ta có

$$\begin{aligned}&P(A_0) = \frac{C_3^2}{C_6^2} = \frac{1}{5}; \quad P(A_1) = \frac{C_3^1 C_3^1}{C_6^2} = \frac{3}{5}; \quad P(A_2) = \frac{C_3^2}{C_6^2} = \frac{1}{5}\\ &P(H|A_0) = \frac{4}{12} = \frac{1}{3}; \quad P(H|A_1) = \frac{5}{12}; \quad P(H|A_2) = \frac{6}{12} = \frac{1}{2}.\end{aligned}$$

Áp dụng công thức xác suất đầy đủ:

$$P(H) = \sum^{2} P(A_i)\,P(H|A_i) = \frac{1}{5}.\frac{1}{3} + \frac{3}{5}.\frac{5}{12} + \frac{1}{5}.\frac{1}{2} = \frac{5}{12}.$$

# Công thức Bayes

- Trong công thức xác suất đầy đủ, $H$ là sự kiện kết quả, còn các sự kiện $A_i\ i = \overline{1,n}$ là các sự kiện nguyên nhân. Nếu biết nguyên nhân nào xảy ra thì ta xác định được xác suất xảy ra $H$.
- Bây giờ ngược lại, người ta đã biết được kết quả xảy ra $H$, muốn tính xác suất để nguyên nhân thứ $i$ xảy ra là bao nhiêu, tức là đi tính $P\left(A_i|H\right)$. $P(A_i)$ được gọi là xác suất tiên nghiệm, còn $P\left(A_i|H\right)$ được gọi là xác suất hậu nghiệm.

Ta có công thức Bayes:

$$\boxed{P\left(A_i|H\right) = \frac{P(A_i)P(H|A_i)}{\sum_{j=1}^{n} P(A_j).P(H|A_j)}, \quad i = 1, 2, \ldots, n.} \tag{5.10}$$





# Công thức Bayes

**Chứng minh.**

Theo công thức xác suất có điều kiện ta có:

$$P\left(A_i|H\right) = \frac{P(A_iH)}{P(H)} = \frac{P(A_i).P(H|A_i)}{P(H)}.$$

Mặt khác theo công thức xác suất đầy đủ: $P(H) = \sum\limits_{j=1}^{n} P(A_j).P(H|A_j)$. Thay vào công thức trên ta có đpcm. □

# Công thức Bayes

**Ví dụ 37**

Một nhà máy sản xuất bóng đèn có tỷ lệ bóng đèn tốt là $90\%$. Trước khi xuất ra thị trường, mỗi bóng đèn đều được qua kiểm tra chất lượng. Vì sự kiểm tra không tuyệt đối hoàn toàn nên một bóng đèn tốt có xác suất $0.9$ được công nhận là tốt, còn một bóng đèn hỏng có xác suất $0.95$ bị loại bỏ.

1. Tính tỷ lệ bóng qua được kiểm tra chất lượng.
2. Tính tỷ lệ bóng hỏng qua được kiểm tra chất lượng.

**Giải.**

Gọi $A$: "Bóng đèn thuộc loại tốt"; $B$: "Bóng đèn thuộc loại hỏng". Ta có $A, B$ là một nhóm đầy đủ và $P(A) = 0.9; \quad P(B) = 0.1$. Gọi $H$: "Bóng qua được kiểm tra chất lượng", ta có $P(H|A) = 0.9; \quad P(H|B) = 0.05$.

1. Theo công thức xác suất đầy đủ ta có

$$P(H) = P(A).P(H|A) + P(B).P(H|B) = 0.9 \times 0.9 + 0.1 \times 0.05 = 0.815.$$

2. Ta có $P(B|H) = \dfrac{P(B).P(H|B)}{P(H)} = \dfrac{0.1 \times 0.05}{0.815} = 0.0061.$

# Chương 2: Biến ngẫu nhiên và luật phân phối xác suất

Lê Xuân Lý [1]

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

Hà Nội, tháng 2 năm 2018

[1] Email: lexuanly@gmail.com

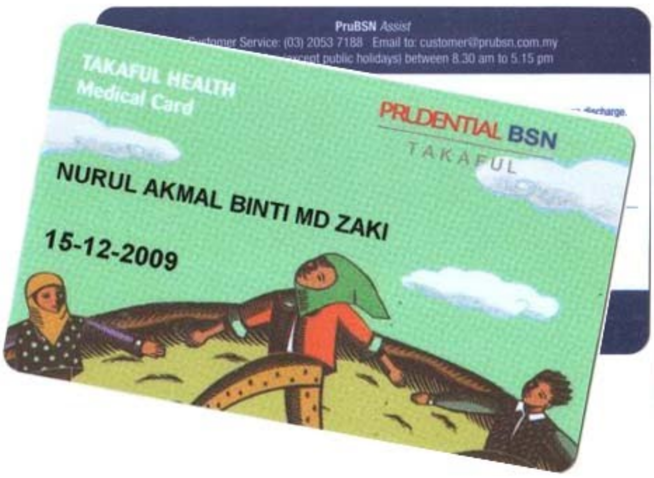




## Ví dụ

Một công ty bảo hiểm bán thẻ bảo hiểm với giá 100 ngàn đồng/1 người/1 năm. Nếu người bảo hiểm gặp rủi ro trong năm đó thì nhận được số tiền bồi thường là 1 triệu đồng. Theo thống kê biết rằng tỷ lệ người tham gia bảo hiểm bị rủi ro trong năm là 0.05.

- Hãy tính tiền lãi trung bình khi bán mỗi thẻ bảo hiểm
- Nếu bán bảo hiểm được cho 10000 khách hàng thì số tiền lãi trung bình thu về được là bao nhiêu?



### Định nghĩa 1.1

*Biến ngẫu nhiên (đại lượng ngẫu nhiên) là một đại lượng mà giá trị của nó là ngẫu nhiên, phụ thuộc vào kết quả phép thử. Ta thường dùng các chữ in hoa để kí hiệu biến ngẫu nhiên:* $X, Y, Z, X_1, X_2, \ldots$*. Còn các giá trị mà biến ngẫu nhiên nhận thường được kí hiệu là chữ thường:* $a, b, c, \ldots, x, y, z, x_1, x_2, \ldots$.

### Ví dụ 1





## Biến ngẫu nhiên

- Gieo một con xúc xắc. Ta quan tâm đến số chấm xuất hiện. Gọi $X$ là số chấm xuất hiện trên mặt con xúc xắc, ta có $X$ là một biến ngẫu nhiên và tập giá trị có thể nhận là $\{1, 2, 3, 4, 5, 6\}$.
- Chọn ngẫu nhiên $3$ đứa trẻ từ một nhóm gồm $6$ bé trai và $4$ bé gái. Ta quan tâm có bao nhiêu bé gái. Gọi $X$ là số bé gái trong nhóm. Khi đó X là một biến ngẫu nhiên và tập giá trị có thể nhận là $\{0, 1, 2, 3\}$.
- Khoảng thời gian giữa $2$ ca cấp cứu ở một bệnh viện nào đó là một biến ngẫu nhiên. Nó có thể nhận giá trị bất kỳ trong khoảng $[0; +\infty)$.
- Nhiệt độ của Hà Nội lúc 6h sáng hàng ngày
- Số iphone phải đi bảo hành
- ...

# Phân loại biến ngẫu nhiên

- **Biến ngẫu nhiên được gọi là rời rạc, nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.**
  + Nói một cách khác đối với biến ngẫu nhiên rời rạc ta có thể liệt kê tất cả các giá trị nó có thể nhận bằng một dãy hữu hạn hoặc vô hạn.
  + **Ví dụ**: số điểm thi của học sinh, số cuộc gọi điện thoại của một tổng đài trong một đơn vị thời gian, số tai nạn giao thông trong một ngày, ...
- **Biến ngẫu nhiên được gọi là liên tục, nếu tập giá trị của nó lấp kín một miền hoặc một số miền của trục số hoặc cũng có thể là cả trục số.**
  + Một miền có dạng $(a; b), [a; b), (a; b], [a; b]$
  + **Ví dụ**: huyết áp của một bệnh nhân, độ dài của một chi tiết máy, tuổi thọ của một loại bóng đèn điện tử,...

Biến ngẫu nhiên
rời rạc

Biến ngẫu nhiên
liên tục





# Hàm phân phối xác suất

**Định nghĩa 1.2**

*Hàm phân phối xác suất của biến ngẫu nhiên $X$, kí hiệu là $F(x)$ và được xác định như sau:*

$$\boxed{F(x) = P(X < x), \quad x \in \mathbb{R}.} \tag{1.1}$$

Hàm phân phối xác suất $F(x)$ phản ánh độ tập trung xác suất ở bên trái của điểm $x$.

**Các tính chất**

- $0 \le F(x) \le 1$
- $\lim\limits_{x \to -\infty} F(x) = 0$ , $\lim\limits_{x \to +\infty} F(x) = 1$
- $F(x)$ là hàm không giảm: $\forall a < b, F(a) \le F(b)$
- $P(a \le X < b) = F(b) - F(a)$

# Bảng phân phối xác suất

**Định nghĩa 2.1**

*Phân bố xác suất của một biến ngẫu nhiên rời rạc $X$ là một bảng trên đó ta ghi cả giá trị mà $X$ có thể nhận kèm theo xác suất để nó nhận các giá trị đó*

| $X = x$ | $x_1$ | $x_2$ | $\ldots$ | $x_n$ | $\ldots$ |
|---|---|---|---|---|---|
| $P(X = x)$ | $p_1$ | $p_2$ | $\ldots$ | $p_n$ | $\ldots$ |

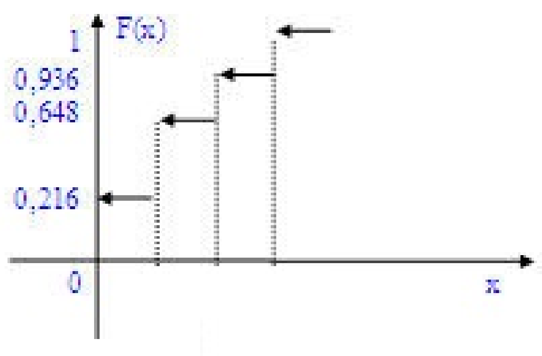
*Trong đó tập các giá trị của $X$ là $\{x_1, x_2, \ldots, x_n\}$ được sắp xếp theo thứ tự tăng dần. Các xác suất $p_i$ thỏa mãn*

- $p_i = P(X = x_i) > 0 \quad \forall i = 1, 2, \ldots;$
- $\sum\limits_i p_i = 1.$
- *Hàm phân phối xác suất của biến ngẫu nhiên rời rạc $X$ :*
  $F(x) = P(X < x) = \sum\limits_{i: x_i < x} P(X = x_i) = \sum\limits_{i: x_i < x} p_i$






# Bảng phân phối xác suất


Câu hỏi: Để lập được bảng phân phối xác suất ta cần làm gì? Trả lời:

- Xác định các giá trị $x_i$ mà $X$ có thể nhận
- Tìm các xác suất $p_i$ tương ứng với các giá trị $x_i$

# Bảng phân phối xác suất

### Ví dụ 1

*Tung một đồng tiền cân đối và đồng chất. Gọi $X$ là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:*

| $X = x$ | $0$ | $1$ |
|---|---|---|
| $P(X = x)$ | $1/2$ | $1/2$ |



Biến ngẫu nhiên rời rạc | Bảng phân phối xác suất

# Bảng phân phối xác suất

### Ví dụ 1

*Tung đồng xu cân đối và đồng chất 2 lần. Gọi $X$ là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:*

| $X = x$ | $0$ | $1$ | $2$ |
|---|---|---|---|
| $P(X = x)$ | $1/4$ | $1/2$ | $1/4$ |

# Bảng phân phối xác suất

**Ví dụ 2**

*Một người đem 10 nghìn đồng đi đánh một số đề. Nếu trúng thì thu được 700 nghìn đồng, nếu trượt thì không được gì. Gọi $X$ (nghìn đồng) là số tiền thu được. Ta có bảng phân phối xác suất của $X$*

| $X = x$ | $0$ | $700$ |
|---|---|---|
| $P(X = x)$ | $^{99}/_{100}$ | $^{1}/_{100}$ |






# Các tham số đặc trưng

## Kỳ vọng

**Kỳ vọng**

- *Kỳ vọng*: là đại lượng đặc trưng cho giá trị trung bình.
  (Đôi khi người ta có thể gọi nó là giá trị trung bình bởi công thức tính của nó chính là tính giá trị trung bình cho trường hợp thu được vô hạn số liệu)
- *Ký hiệu*: $E(X)$ hoặc $EX$
- *Công thức tính*: với $X$ rời rạc ta có: $EX = \sum_i x_i.p_i$

# Các tham số đặc trưng

## Kỳ vọng

### Ví dụ 1

*Tung một đồng tiền cân đối và đồng chất. Gọi $X$ là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:*

| $X = x$ | 0 | 1 |
|---|---|---|
| $P(X = x)$ | $1/2$ | $1/2$ |

*Kỳ vọng của $X$* : $EX = 0.1/2 + 1.1/2 = 1/2$





# Các tham số đặc trưng

## Kỳ vọng

### Ví dụ 2

*Tung đồng xu cân đối và đồng chất 2 lần. Gọi $X$ là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:*

| $X = x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X = x)$ | $1/4$ | $1/2$ | $1/4$ |

*Kỳ vọng của $X$* : $EX = 0.1/4 + 1.1/2 + 2.1/4 = 1$

*Như vậy trong 2 lần tung đồng xu thì trung bình có một lần ra mặt sấp.*

# Các tham số đặc trưng

## Kỳ vọng

### Ví dụ 3

*Một người đem 10 nghìn đồng đi đánh một số đề. Nếu trúng thì thu được 700 nghìn đồng, nếu trượt thì không được gì. Gọi $X$ (nghìn đồng) là số tiền thu được. Ta có bảng phân phối xác suất của $X$*

| $X = x$ | $0$ | $700$ |
|---|---|---|
| $P(X = x)$ | $^{99}/_{100}$ | $^{1}/_{100}$ |

*Kỳ vọng của $X$ : $EX = 0.{}^{99}/_{100} + 700.{}^{1}/_{100} = 7$*
*Như vậy bỏ ra 10 nghìn đồng, trung bình thu được 7 nghìn đồng, người chơi về lâu dài sẽ lỗ 30% tổng số tiền chơi.*





# Các tham số đặc trưng

## Kỳ vọng

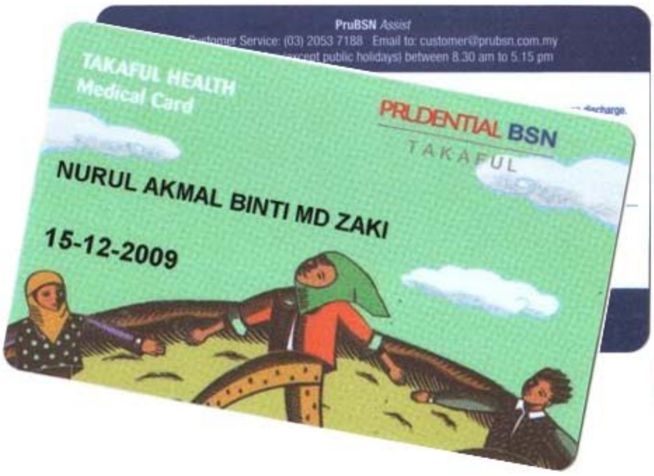
### Các tính chất của kỳ vọng

- $Ec = c$ với $c$ là hằng số
- $E(aX) = a.EX$
- $E(X + b) = EX + b$
  Ta suy ra kết quả: $E(aX + b) = aEX + b$
- Tổng quát với X là biến ngẫu nhiên rời rạc: $Eg(X) = \sum\limits_i g(x_i).p_i$
  Ví dụ: $E(X^2) = \sum\limits_i x_i^2.p_i$
- $E(X + Y) = EX + EY$

### Ví dụ 4

*Một công ty bảo hiểm bán thẻ bảo hiểm với giá 100 ngàn đồng/1 người/1 năm. Nếu người bảo hiểm gặp rủi ro trong năm đó thì nhận được số tiền bồi thường là 1 triệu đồng. Theo thống kê biết rằng tỷ lệ người tham gia bảo hiểm bị rủi ro trong năm là 0.05.*

- *Hãy tính tiền lãi trung bình khi bán mỗi thẻ bảo hiểm*
- *Nếu bán bảo hiểm được cho 10000 khách hàng thì số tiền lãi trung bình thu về được là bao nhiêu?*








# Các tham số đặc trưng

## Phương sai

### Phương sai

- *Phương sai*: trung bình của bình **phương sai** số.
- *Ký hiệu*: $V(X)$ hoặc $VX$
- *Công thức tính*: $VX = E(X - EX)^2$
  Với $(X - EX)$ là sai số, hoặc là độ lệch khỏi giá trị trung bình
  Người ta biến đổi để đưa công thức tính phương sai về dễ tính hơn:

$$\boxed{VX = E(X - EX)^2 = E(X^2) - (EX)^2}$$

Với $X$ là biến ngẫu nhiên rời rạc:

$EX = \sum_{i=1}^{n} x_i.p_i$

$E(X^2) = \sum_{i=1}^{n} x_i^2.p_i$

### Ý nghĩa của phương sai

- Phương sai thể hiện mức độ phân tán dữ liệu xung quanh giá trị trung bình $EX$, phương sai càng lớn thì độ phân tán dữ liệu càng cao và ngược lại.
- Trong công nghiệp, $X$ thường là kích cỡ của các sản phẩm. $VX$ lúc này biểu thị độ chính xác của các sản phẩm.
- Trong chăn nuôi, $X$ thường là chiều cao hay cân nặng của gia súc gia cầm. $VX$ lúc này biểu thị độ tăng trưởng đồng đều của các gia súc gia cầm.
- Trong trồng trọt, $X$ thường là năng suất của giống cây trồng. $VX$ lúc này biểu thị mức độ ổn định của năng suất giống cây trồng.
- Trong kinh tế, $X$ thường là lãi suất thu được của khoản đầu tư. $VX$ lúc này sẽ biểu thị cho mức độ rủi ro của đầu tư.








## Các tham số đặc trưng

### Phương sai

### Ví dụ 1

*Tung một đồng tiền cân đối và đồng chất. Gọi $X$ là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:*

| $X = x$ | $0$ | $1$ |
|---|---|---|
| $P(X = x)$ | $1/2$ | $1/2$ |

$EX = 0.1/2 + 1.1/2 = 1/2$

$E(X^2) = 0^2.1/2 + 1^2.1/2 = 1/2$

*Phương sai* $VX = E(X^2) - (EX)^2 = 1/2 - 1/4 = 1/4$

# Các tham số đặc trưng

## Phương sai

### Ví dụ 2

*Tung đồng xu cân đối và đồng chất 2 lần. Gọi $X$ là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp. Ta có bảng phân phối xác suất sau:*

| $X = x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X = x)$ | $1/4$ | $1/2$ | $1/4$ |

$EX = 0.1/4 + 1.1/2 + 2.1/4 = 1$
$E(X^2) = 0^2.1/4 + 1^2.1/2 + 2^2.1/4 = 3/2$
*Phương sai* $VX = E(X^2) - (EX)^2 = 3/2 - 1^2 = 1/2$
*Nhận xét: Phương sai của VD2 lớn hơn phương sai của VD1 cho ta kết luận rằng biên độ dao động của X xung quanh giá trị trung bình ở VD2 lớn hơn VD1.*





# Các tham số đặc trưng

## Phương sai

### Các tính chất của phương sai

- $Vc = 0$ với $c$ là hằng số
- $V(aX) = a^2.VX$
- $V(X + b) = VX$
  Ta suy ra kết quả: $V(aX + b) = a^2VX$

# Các tham số đặc trưng

## Độ lệch chuẩn

Đơn vị đo của phương sai bằng bình phương đơn vị đo của biến ngẫu nhiên. Để dễ đánh giá mức độ phân tán hơn, người ta đưa ra khái niệm độ lệch chuẩn.

### Độ lệch chuẩn

- *Ý nghĩa*: dùng để đo độ phân tán dữ liệu xung quanh giá trị trung bình $EX$.
- *Ký hiệu*: $\sigma(X)$ hoặc $\sigma$
- *Công thức tính*: $\sigma = \sqrt{VX}$

Ví dụ: Phân tích kỹ thuật giá chứng khoán: SMA(n) và Bollinger Band(n).






### Ví dụ 3

# Các tham số đặc trưng

## Mode

### Mode

- *Khái niệm*: Mode của biến ngẫu nhiên $X$, kí hiệu là $mod(X)$, là giá trị của biến ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận nào đó của nó. Đối với biến ngẫu nhiên rời rạc, mod(X) là giá trị của X ứng với xác suất lớn nhất. Như vậy một biến ngẫu nhiên có thể có một mode hoặc nhiều mode.
- *Ký hiệu*: $mod(X)$





# Các tham số đặc trưng

## Phân vị mức $p$

- *Khái niệm*: Phân vị mức $p$ của biến ngẫu nhiên $X$ là giá trị $z_p$ sao cho.

$$F(z_p) = P(X < z_p) = p$$

  Một số phân vị đặc biệt:
  + Phân vị mức 25% được gọi là tứ phân vị thứ nhất
  + Phân vị mức 50% được gọi là tứ phân vị thứ hai hay trung vị.
  + Phân vị mức 75% được gọi là tứ phân vị thứ ba
- *Trung vị*: Trung vị của biến ngẫu nhiên $X$ là giá trị của $X$ chia phân phối xác suất thành hai phần có xác suất bằng nhau. Kí hiệu là $med(X)$:

$$P(X < med(X) = P(X \geq med(X)) = 0,5$$

  Ta có thể tìm trung vị bằng cách giải phương trình: $F(x) = 0,5$.
  Trong ứng dụng, trung vị là đặc trưng vị trí tốt nhất, nhiều khi tốt hơn cả kỳ vọng, nhất là trong những trường hợp số liệu có nhiều sai sót hoặc sai sót thái quá.

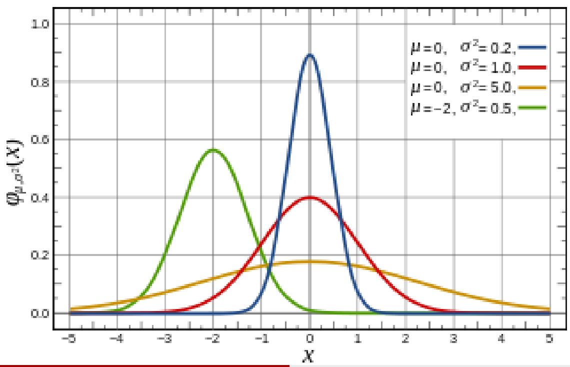
# Hàm mật độ xác suất

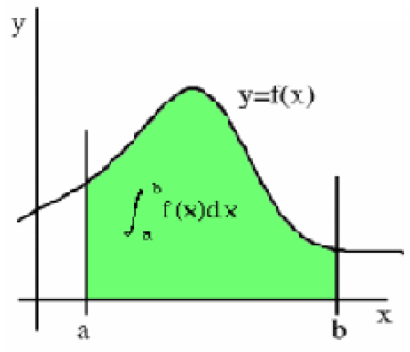
Đối với biến ngẫu nhiên liên tục, không thể dùng bảng phân phối xác suất do xác suất nó nhận tại mỗi điểm luôn bằng "0". Do đó người ta thay thế bằng hàm mật độ xác suất.

**Định nghĩa 3.1**

*Hàm mật độ xác suất của biến ngẫu nhiên liên tục $X$ là hàm $f(x)$ xác định trên $\mathbb{R}$ thỏa mãn:*

- $f(x) \geq 0 \ \forall x \in \mathbb{R}$;
- $P(X \in B) = \int\limits_B f(x)dx \ \ \forall B \subset \mathbb{R}$.








# Hàm mật độ xác suất

**Chú ý 3.1**

*Hàm mật độ xác suất $f(x)$ của biến ngẫu nhiên liên tục $X$ thể hiện mức độ tập trung xác suất của $X$ xung quanh điểm $x$. Tức là với $\Delta_x$ đủ nhỏ cho trước ta có thể tính xấp xỉ:*

$$P\left(x \leq X \leq x + \Delta_x\right) \approx f(x).\Delta_x.$$

*Do đó ta thấy xác suất để $X$ nhận giá trị thuộc lân cận khá bé $(x, x + \Delta_x)$ gần như tỉ lệ thuận với $f(x)$.*

# Hàm mật độ xác suất

## Tính chất

- $\int_{-\infty}^{+\infty} f(x)dx = 1;$
- $P(a \leq X < b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x)dx$
- Hàm phân phối xác suất: $F(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt$

  Từ đó suy ra $f(x) = F'(x)$





# Hàm mật độ xác suất

## Ví dụ 4

*Cho hàm số $f(x) = a.\sin 2x$. Tìm $a$ để hàm này trở thành hàm mật độ xác suất của một biến ngẫu nhiên nhận giá trị trong $[0, \pi/2]$.*

## Lời giải

Để hàm này trở thành hàm mật độ xác suất của một biến ngẫu nhiên nhận giá trị trong $[0, \pi/2]$ thì:

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

Do $\sin 2x \geq 0$ với mọi $x \in [0, \pi/2]$ nên $a \geq 0$. Ta có:

$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_0^{\pi/2} a\sin 2x dx = a$. Vậy $a = 1$.

# Hàm mật độ xác suất

**Ví dụ 5**

*Tuổi thọ của một loài côn trùng là biến ngẫu nhiên $X$(tháng tuổi) có hàm mật độ xác suất* $f(x) = \begin{cases} ax^2(4-x^2), & x \in [0,2] \\ 0, & x \notin [0,2] \end{cases}$.

*a. Xác định* $a$

*b. Tính* $P(0 \leq X \leq 1), P(X > 1)$

*c. Xác định hàm phân phối xác suất* $F(x)$

**Lời giải**

a. Do $ax^2(4-x^2) \geq 0$ với $\forall x \in [0,2]$ nên $a \geq 0$

Ta có $1 = \int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{0}^{2} ax^2(4-x^2)dx = a.\frac{64}{15} \Rightarrow a = \frac{15}{64}$

b. $P(0 \leq X \leq 1) = \int\limits_{0}^{1} f(x)dx = \int\limits_{0}^{1} ax^2(4-x^2)dx = a.\frac{17}{15} = \frac{17}{64} = 0,266$





# Hàm mật độ xác suất

**Lời giải**

b. $P(X > 1) = \int\limits_{1}^{+\infty} f(x)dx = \int\limits_{1}^{2} ax^2(4-x^2)dx = \frac{47}{64} = 0,734$

c. Hàm phân phối $F(x) = \int\limits_{-\infty}^{x} f(t)dt$

$x < 0$ suy ra $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{x} 0dt = 0$

$0 \leq x \leq 2$ suy ra $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{0}^{x} at^2(4-t^2)dt = \frac{15}{64}(\frac{4x^3}{3} - \frac{x^5}{5})$

$x > 2$ suy ra $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{0}^{2} at^2(4-t^2)dt = 1$

# Hàm mật độ xác suất

**Nhận xét**

Qua tính toán trên ta thấy 26.6% côn trùng sống không quá một tháng tuổi, và 73,4% côn trùng sống hơn một tháng tuổi. Do đó ta có thể nhận xét rằng tuổi thọ trung bình của loài này sẽ lớn hơn một tháng tuổi. Tuy nhiên tuổi thọ trung bình của loài côn trùng này chính xác là bao nhiêu?





# Các tham số đặc trưng

## Kỳ vọng

**Kỳ vọng của biến ngẫu nhiên liên tục $X$**

- *Ý nghĩa*: nó đặc trưng cho giá trị trung bình của $X$
- *Ký hiệu*: $E(X)$ hoặc $EX$
- *Công thức tính*: $EX = \int\limits_{-\infty}^{+\infty} x.f(x)dx$
- *Tính chất*:
  $+\ E(aX+b) = a.EX + b$
  $+\ Eg(X) = \int\limits_{-\infty}^{+\infty} g(x).f(x)dx$
  Ví dụ: $g(X) = X^2$ ta có $E(X^2) = \int\limits_{-\infty}^{+\infty} x^2.f(x)dx$

# Các tham số đặc trưng

## Phương sai

### Phương sai của biến ngẫu nhiên liên tục $X$

- *Ý nghĩa*: nó đặc trưng cho độ phân tán dữ liệu xung quanh $EX$
- *Ký hiệu*: $V(X)$ hoặc $VX$
- *Công thức tính*: $\boxed{VX = E(X-EX)^2 = E(X^2) - (EX)^2}$

  với: $EX = \int\limits_{-\infty}^{+\infty} x.f(x)dx$ và $E(X^2) = \int\limits_{-\infty}^{+\infty} x^2.f(x)dx$
- *Tính chất*: $V(aX+b) = a^2VX$





# Các tham số đặc trưng

## Độ lệch chuẩn

### Độ lệch chuẩn

- *Ý nghĩa*: dùng để đo độ phân tán dữ liệu xung quanh giá trị trung bình $EX$.
- *Ký hiệu*: $\sigma(X)$ hoặc $\sigma$
- *Công thức tính*: $\sigma = \sqrt{VX} = \sqrt{E(X^2) - (EX)^2}$

  với $X$ liên tục: $EX = \int\limits_{-\infty}^{+\infty} xf(x)dx$

  $E(X^2) = \int\limits_{-\infty}^{+\infty} x^2f(x)dx$

# Các tham số đặc trưng

## Mode - phân vị mức $p$

### Mode

- *Khái niệm*: Mode của biến ngẫu nhiên $X$, kí hiệu là $mod(X)$, là giá trị của biến ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận nào đó của nó. Đối với biến ngẫu nhiên liên tục, mod(X) là giá trị của X ứng với $f(x)$ đạt cực đại địa phương.
- *Ký hiệu*: $mod(X)$

### Phân vị mức $p$

- *Khái niệm*: Phân vị mức $p$ của biến ngẫu nhiên $X$ là giá trị $z_p$ sao cho.

$$F(z_p) = P(X < z_p) = p$$

- *Trung vị*: Trung vị của biến ngẫu nhiên $X$ là giá trị của $X$ chia phân phối xác suất thành hai phần có xác suất bằng nhau. Kí hiệu là $med(X)$: $P(X < med(X) = P(X \geq med(X)) = 0,5$





# Một số phân phối xác suất thông dụng

Các quy luật thông dụng sẽ học:

### Biến ngẫu nhiên rời rạc

- Luật phân phối nhị thức
- Luật phân phối Poisson

### Biến ngẫu nhiên liên tục

- Phân phối đều liên tục
- Phân phối chuẩn
- Phân phối mũ
- Phân phối Khi bình phương
- Phân phối Student

SAMI 1956

# Phân phối nhị thức (Binomial Distribution)

## Định nghĩa 4.1

*Biến ngẫu nhiên $X$ nhận giá trị trong tập $\{0; 1; 2; ...; n\}$ với xác suất được tính theo công thức Bernoulli:*
$P(X = k) = C_n^k.p^k.(1-p)^{n-k}$ *với* $k = 0, 1, \dots, n; 0 \leq p \leq 1$
*gọi là tuân theo phân phối nhị thức với các tham số $n$ và $p$.*
*Ký hiệu:* $X \sim B(n; p)$

## Các tham số đặc trưng

Với $X \sim B(n; p)$ ta có:

- $EX = np$
- $VX = np(1-p) = npq$ với $q = 1 - p$
- $(n+1)p - 1 \leq mod(X) \leq (n+1)p$





# Phân phối nhị thức

## Ứng dụng

Ta thực hiện $n$ phép thử độc lập cùng điều kiện. Trong mỗi phép thử xác suất xảy ra sự kiện $A$ luôn là $p$. Gọi $X$ là số phép thử xảy ra $A$. Ta có kết quả: $X \sim B(n; p)$

## Ví dụ 1

*Gieo một con xúc xắc 3 lần. Gọi $X$ là số lần ra mặt lục trong 3 lần gieo. Lập bảng phân phối xác suất của $X$, biết rằng khả năng ra mặt lục ở mỗi lần gieo là $1/6$.*
**Gợi ý**:
$X \sim B(n; p)$ *với* $n = 3; p = 1/6$ , $P(X = k) = C_n^k.p^k.(1-p)^{n-k}$

| $X = x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P(X = x)$ | $125/216$ | $75/216$ | $15/216$ | $1/216$ |

# Phân phối nhị thức

**Ví dụ 2**

*Một người chơi đề trong 10 ngày, mỗi ngày người đó chơi 5 số. Tính xác suất trong 10 ngày chơi:*
*+) Người đó trúng được đúng 2 ngày.*
*+) Người đó trúng được ít nhất 2 ngày*
*+) Xác định số ngày trúng có khả năng xảy ra cao nhất?*





# Phân phối nhị thức

Biến nào sau đây là tuân theo phân phối nhị thức:

- Tung một đồng xu 3 lần. Gọi $X$ là số lần được mặt ngửa.
- Hộp có 4 bi trắng và 3 bi xanh. Lấy ngẫu nhiên 3 bi. Gọi $X$ là số bi xanh lấy được theo 2 cách:
  +) Lấy lần lượt 3 bi
  +) Lấy có hoàn lại 3 bi
- Một máy sản xuất ra sản phẩm có tỷ lệ phế phẩm là 2%. Cho máy sản xuất ra 10 sản phẩm. Gọi $X$ là số phế phẩm có được.
- Một xạ thủ bắn 3 phát đạn vào bia. Ở lần bắn sau do rút được kinh nghiệm các lần bắn trước nên xác suất bắn trúng của 3 phát lần lượt là $0,7; 0,8; 0,9$. Gọi X là số phát bắn trúng bia.

# Phân phối Poisson

**Định nghĩa 4.2**

*Biến ngẫu nhiên $X$ nhận giá trị trong tập $\{0; 1; 2; \ldots; n; \ldots\}$ với xác suất :*

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

*gọi là tuân theo phân phối Poisson với tham số $\lambda$*
*Ký hiệu: $X \sim P(\lambda)$*

**Các tham số đặc trưng**

Với $X \sim P(\lambda)$ ta có:

- $EX = \lambda$
- $VX = \lambda$
- $\lambda - 1 \leq mod(X) \leq \lambda$





# Phân phối Poisson

- Quá trình Poisson còn có thể gọi là quá trình đếm.
- Trong tình huống nào ta gặp phân phối Poisson?
  Xét một sự kiện E xuất hiện ở những thời điểm ngẫu nhiên. Giả sử số lần xuất hiện E trong một khoảng thời gian không ảnh hưởng tới xác suất xuất hiện của E trong các khoảng thời gian kế tiếp. Hơn nữa cường độ xuất hiện của E là không thay đổi, nghĩa là số lần trung bình xuất hiện E trong khoảng thời gian tỉ lệ với độ dài khoảng thời gian đó.
  Gọi $X$ là số lần xuất hiện E trong khoảng thời gian $(t_1, t_2)$. Ta có $X \sim P(\lambda)$ với $\lambda = c(t_2 - t_1)$, trong đó c là hằng số được gọi là cường độ xuất hiện của E.
- Phân phối này có nhiều ứng dụng đối với nhiều quá trình có liên quan đến số quan sát đối với một đơn vị thời gian hoặc không gian. Ví dụ: Số cuộc điện thoại nhận được ở một trạm điện thoại trong một phút, số khách hàng đến nhà băng đối với mỗi một chu kỳ 30 phút, số lỗi in sai trong một trang, .... Nói chung dòng vào của một hệ phục vụ (quán bia, hiệu cắt tóc, hiệu sửa xe, trạm điện thoại, một cửa hàng nào đó, ...) là các biến ngẫu nhiên tuân theo luật phân phối Poisson.

### Ví dụ 3

*Ở một tổng đài bưu điện, các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau với tốc độ trung bình 2 cuộc gọi trong một phút. Tìm xác suất để:*
*a) Có đúng 5 cuộc điện thoại trong vòng 2 phút*
*b) Không có cuộc điện thoại nào trong khoảng thời gian 30 giây*
*c) Có ít nhất 1 cuộc điện thoại trong khoảng thời gian 10 giây.*

### Lời giải

a. Gọi $X$ là số cuộc điện thoại xuất hiện trong vòng 2 phút. $X \sim P(\lambda)$
$\lambda$ chính là số cuộc điện thoại trung bình đến trong vòng 2 phút. $\lambda = 4$
$P(X = 5) = e^{-\lambda}\frac{\lambda^5}{5!} = e^{-4}\frac{4^5}{5!} = 0,156$
b. Gọi $X$ là số cuộc điện thoại xuất hiện trong vòng 30 giây. $X \sim P(\lambda)$ với $\lambda = 1$. Ta có
$P(X = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-1} = 0,3679$
c. Gọi $X$ là số cuộc điện thoại xuất hiện trong vòng 10 giây. $X \sim P(\lambda)$ với $\lambda = 1/3$. Ta có $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/3} = 0,2835$

SAMI 1956





### Chú ý 4.1

*Khi $n$ lớn và $p$ nhỏ ($n > 50; p < 0,1$) thì $X \sim B(n;p)$ có thể chuyển thành $X \sim P(\lambda)$ với $\lambda = np$*

### Ví dụ 4

*Trong một lô thuốc, tỷ lệ ống thuốc hỏng là $p = 0,003$. Kiểm nghiệm 1000 ống. Tính xác suất để gặp 3 ống bị hỏng.*
**Lời giải:**
*Gọi $X$ là số ống thuốc hỏng trong 1000 ống. Ta có $X \sim B(n;p)$ với $n = 1000; p - 0,003$*
*Do $n$ lớn và $p$ bé nên ta xấp xỉ $X \sim P(\lambda)$ với $\lambda = np = 3$*

$$P(X = 3) = e^{-\lambda}\frac{\lambda^3}{3!} = e^{-3}\frac{3^3}{3!} = 0,224$$

SAMI 1956

# Phân phối đều rời rạc

**Định nghĩa 4.3**

*Biến ngẫu nhiên $X$ tuân theo phân phối đều rời rạc với tham số $n$ nếu $X$ có bảng phân phối xác suất như sau:*

| $X = x$ | 1 | 1 | ... | n |
|---|---|---|---|---|
| $P(X = x)$ | 1/n | 1/n | ... | 1/n |

*Ký hiệu:* $X \sim U(n)$

**Các tham số đặc trưng**

$$EX = \frac{n+1}{2}$$
$$VX = \frac{n^2-1}{12}$$





# Phân phối đều liên tục

**Định nghĩa 4.4**

*Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều liên tục trên [a;b] nếu X có hàm mật độ:*

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

*Ký hiệu:* $X \sim U([a,b])$

**Các tham số đặc trưng**

$$EX = \frac{a+b}{2}$$
$$VX = \frac{(b-a)^2}{12}$$

# Phân phối đều liên tục - Ví dụ

### Ví dụ 5

*Lịch chạy của xe bus tại một trạm xe bus như sau: chiếc xe bus đầu tiên trong ngày sẽ khởi hành từ trạm này lúc 7 giờ, cứ sau 15 phút sẽ có một xe khác đến trạm. Giả sử một hành khách đến trạm ngẫu nhiên trong khoảng thời gian từ 7 giờ đến 7 giờ 30. Tìm xác suất để hành khách này chờ:*

*a) Ít hơn 5 phút*

*b) Ít nhất 12 phút.*

**Lời giải**:

*Gọi X là số phút sau 7 giờ hành khách đến trạm, ta có* $X \sim U([0, 30])$

*a) Hành khách chờ ít hơn 5 phút nếu đến trạm giữa 7 giờ 10 và 7 giờ 15 hoặc giữa 7 giờ 25 và 7 giờ 30. Do đó xác suất cần tìm là:*

$P(10 < X < 15) + P(25 < X < 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$

*b) Hành khách chờ ít nhất 12 phút nếu đến trạm giữa 7 giờ và 7 giờ 03 hoặc giữa 7 giờ 15 và 7 giờ 18. Xác suất cần tìm là:*

$P(0 < X < 3) + P(15 < X < 18) = \frac{3}{30} + \frac{3}{30} = 0,2$





# Phân phối chuẩn

### Định nghĩa 4.5

*Biến ngẫu nhiên* $X$ *được gọi là tuân theo phân phối chuẩn với hai tham số* $\mu$ *và* $\sigma^2$ *(với* $\sigma > 0$*) nếu hàm mật độ của* $X$ *có dạng:*

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

*Ký hiệu:* $X \sim N(\mu, \sigma^2)$

### Các tham số đặc trưng

$EX = \mu$

$VX = \sigma^2$

$mod(X) = med(X) = \mu$

Mục tiêu là ta tính xác suất dạng $P(a < X < b)$

### Phân phối chuẩn tắc

Đặc biệt: $X \sim N(0;1)$ với $(\mu = 0, \sigma = 1)$, $X$ được gọi là tuân theo phân phối chuẩn tắc (hay chuẩn hoá).
Hàm mật độ xác suất hay còn gọi là hàm mật độ Gauss:

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

Để tính xác suất ta dùng hàm Laplace: $\phi(x) = \int\limits_0^x \varphi(t)dt$

**Tính chất**:

- $\phi(x)$ là hàm lẻ, tăng thực sự.
- $\phi(+\infty) = 0,5$
- $X \sim N(0;1)$ ta có: $P(a < X < b) = \phi(b) - \phi(a)$
- Giá trị của hàm Laplace được tính sẵn thành bảng số liệu.

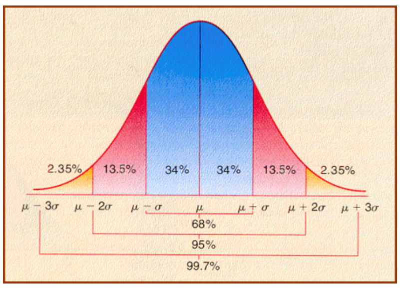


Một số luật phân phối xác suất thông dụng | Phân phối chuẩn

### Phân phối chuẩn tổng quát

**Kết quả**: Nếu $X \sim N(\mu;\sigma^2)$ ta có $Z = \frac{X-\mu}{\sigma} \sim N(0;1)$
Từ đó ta xây dựng được công thức tính:

- $P(X < a) = 0,5 + \phi(\dfrac{a-\mu}{\sigma})$
- $P(X > a) = 0,5 - \phi(\dfrac{a-\mu}{\sigma})$
- $P(a \le X < b) = \phi(\dfrac{b-\mu}{\sigma}) - \phi(\dfrac{a-\mu}{\sigma})$
- $P(|X-\mu| < \varepsilon) = 2\phi(\dfrac{\varepsilon}{\sigma})$



# Phân phối chuẩn - Ví dụ

**Ví dụ 6**

*Độ dài một chi tiết máy giả sử tuân theo luật phân phối chuẩn với giá trị trung bình là 20 cm và độ lệch chuẩn là 0,5 cm. Tính xác suất khi chọn ngẫu nhiên ra một chi tiết thì độ dài của nó:*
*a) lớn hơn 20 cm*
*b) bé hơn 19,5 cm*
*c) nằm trong khoảng 19 cm – 21 cm*
**Lời giải**:
*Gọi $X$ (cm) là độ dài chi tiết máy đã chọn.* $X \sim N(\mu, \sigma^2)$, $\mu = 20$, $\sigma = 0,5$.

- $P(X > 20) = 0,5 - \phi(\frac{20-\mu}{\sigma}) = 0,5 - \phi(0) = 0,5$
- $P(X < 19,5) = 0,5 + \phi(\frac{19,5-\mu}{\sigma}) = 0,5 + \phi(-1) = 0,5 - \phi(1) = 0,5 - 0,3413 = 0,1587$
- $P(19 < X < 21) = \phi(\frac{21-\mu}{\sigma}) - \phi(\frac{19-\mu}{\sigma}) = \phi(2) - \phi(-2) = 2\phi(2) = 2.0,4772 = 0,9544$





# Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Với $X \sim B(n;p)$ thoả mãn $np(1-p) > 20$.
Khi đó ta xấp xỉ $X \sim N(\mu, \sigma^2)$ với $\mu = np, \sigma^2 = np(1-p)$
Tuy nhiên vì chúng ta xấp xỉ một phân phối rời rạc bằng một phân phối liên tục, nên cần một sự hiệu chỉnh để giảm sai số. Cụ thể với $k, k_1, k_2$ là số tự nhiên ta có:

$$P(X = k) = \phi(\frac{k+0,5-\mu}{\sigma}) - \phi(\frac{k-0,5-\mu}{\sigma})$$
$$P(k_1 \leq X \leq k_2) = \phi(\frac{k_2+0,5-\mu}{\sigma}) - \phi(\frac{k_1-0,5-\mu}{\sigma})$$

# Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

**Ví dụ 7**

*Kiểm tra chất lượng 1000 sản phẩm với tỷ lệ chính phẩm 0,95. Tìm xác suất để số chính phẩm trong lô kiểm tra từ 940 đến 960.*

**Lời giải**

: Gọi $X$ là biến ngẫu nhiên chỉ số chính phẩm trong lô sản phẩm kiểm tra, ta có $X \sim B(1000; 0,95)$
Với $n = 1000, p = 0,95$, ta có $np = 950$ và $npq = 47,5$ đủ lớn nên ta xấp xỉ $X \sim N(950; 47,5)$:

$$P(940 \leq X \leq 960) = \phi(\frac{960 + 0,5 - 950}{\sqrt{47,5}}) - \phi(\frac{940 + 0,5 - 950}{\sqrt{47,5}})$$
$$= \phi(1,52) - \phi(-1,52) = 2\phi(1,52) = 0,8716$$





# Phân phối chuẩn - Ý nghĩa

Phân phối chuẩn được Gauss phát minh năm 1809 nên cũng có khi nó được mang tên là phân phối Gauss.
Ta thấy biến ngẫu nhiên tuân theo phân phối chuẩn nhận giá trị trên cả trục số, tuy nhiên có thể xấp xỉ một số biến ngẫu nhiên không nhận tất cả các giá trị trên $\mathbb{R}$ theo phân phối chuẩn, đó là do qui tắc $3 - \sigma$, tức là nếu ta có xác suất $X$ rơi vào miền có xác suất bằng 0,9974 rất gần 1, nên hầu hết người ta chỉ cần quan tâm đến các giá trị trong lân cận $3 - \sigma$ của kỳ vọng.
Phân phối chuẩn chiếm vị trí quan trọng trong lý thuyết xác suất, là vị trí trung tâm trong các kết luận thống kê sau này. Trong thực tế, ví dụ trong lĩnh vực kinh tế, khoa học xã hội, ... nhiều phân phối không giống phân phối chuẩn, nhưng phân phối của trung bình cộng đối với mỗi trường hợp lại có thể xem là phân phối chuẩn miễn là cỡ mẫu n đủ lớn.

# Phân phối mũ

## Định nghĩa 4.6

*Biến ngẫu nhiên $X$ được gọi là tuân theo phân phối mũ với tham số $\lambda > 0$ nếu nó có hàm mật độ xác suất có dạng:*

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

*Ký hiệu:* $X \sim \mathcal{E}(\lambda)$

## Các tham số đặc trưng

$EX = \dfrac{1}{\lambda}$

$VX = \dfrac{1}{\lambda^2}$





# Phân phối mũ

Ta có $P(X > x) = e^{\lambda x}$
Phân phối mũ có tính chất không nhớ:

$$P(X > t + s | X > t) = P(X > s)$$

**Ý nghĩa**: Phân phối mũ có nhiều nhiều ứng dụng trong thực tiễn. Nói chung với một giả thiết nào đó, khoảng thời gian giữa hai lần xuất hiện của một sự kiện E nào đó sẽ có phân phối mũ. Vì lý do này phân phối mũ còn có tên gọi là phân phối của thời gian chờ đợi ("Waiting time distribution"). Ví dụ khoảng thời gian giữa 2 ca cấp cứu ở một bệnh viện, khoảng thời gian giữa 2 lần hỏng hóc của một chiếc máy, khoảng thời gian giữa 2 trận lụt hay động đất, . . .

# Phân phối mũ

**Ví dụ 8**

*Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân phối mũ với kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.*

**Lời giải**

Gọi $X$ là tuổi thọ của mạch. $X$ tuân theo phân phối mũ với tham số $\lambda = \dfrac{1}{EX} = \dfrac{1}{6,25}$

$P(X \leq 5) = 1 - e^{-5\lambda} = 1 - e^{-0,8} = 0,5506$

Vậy có khoảng 55% mạch điện tử bán ra phải thay thế trong thời gian bảo hành





# Phân phối Khi bình phương

**Định nghĩa 4.7**

*Giả sử $X_i, (i = 1, 2, \ldots, n)$ là các biến ngẫu nhiên độc lập cùng phân phối chuẩn tắc. Biến ngẫu nhiên $Y = \sum\limits_{i=1}^{n} X_i^2$ được gọi là tuân theo phân phối Khi bình phương với $n$ bậc tự do.*
*Ký hiệu: $Y \sim \chi^2(n)$*

**Các tham số đặc trưng**

- $EY = n$
- $VY = 2n$

# Phân phối Student

### Định nghĩa 4.8

*Giả sử* $X \sim N(0;1)$ *và* $Y \sim \chi^2(n)$ *là hai biến ngẫu nhiên độc lập. Khi đó:*

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

*được gọi là tuân theo phân phối Student với* $n$ *bậc tự do.*
*Ký hiệu:* $T \sim T(n)$

### Các tham số đặc trưng

- $ET = 0$
- $VT = \dfrac{n}{n-2}$





### Chú ý

- Phân phối Student có cùng dạng và tính đối xứng như phân phối chuẩn nhưng nó phản ánh tính biến đổi của phân phối sâu sắc hơn. Phân phối chuẩn không thể dùng để xấp xỉ phân phối khi mẫu có kích thước nhỏ. Trong trường hợp này ta dùng phân phối Student.
- Khi bậc tự do n tăng lên (n > 30) thì phân phối Student tiến nhanh về phân phối chuẩn. Do đó khi n > 30 ta có thể dùng phân phối chuẩn thay thế cho phân phối Student.

# Chương 3: Biến ngẫu nhiên nhiều chiều

3rd July 2017





## Các khái niệm cơ sở

- Ở chương trước chúng ta quan tâm đến xác suất của biến ngẫu nhiên riêng rẽ. Nhưng trong thực tế nhiều khi ta phải xét đồng thời nhiều biến khác nhau có quan hệ tương hỗ (ví dụ khi nghiên cứu về sinh viên một trường đại học thì cần quan tâm đến chiều cao, cân nặng, tuổi, ...). Do đó dẫn đến khái niệm biến ngẫu nhiên nhiều chiều hay véctơ ngẫu nhiên.
- Để cho đơn giản, ta nghiên cứu biến ngẫu nhiên hai chiều $(X, Y)$, trong đó $X, Y$ là các biến ngẫu nhiên một chiều. Hầu hết các kết quả thu được đều có thể mở rộng khá dễ dàng cho trường hợp biến ngẫu nhiên $n$ chiều.
- Biến ngẫu nhiên hai chiều được gọi là rời rạc (liên tục) nếu các thành phần của nó là các biến ngẫu nhiên rời rạc (liên tục).

# Các khái niệm cơ sở

**Định nghĩa 3.1**

Hàm phân phối xác suất của biến ngẫu nhiên hai chiều $(X, Y)$ được xác định như sau

$$\boxed{F(x, y) = P(X < x,\ Y < y), \quad x, y \in \mathbb{R}.} \tag{3.1}$$

Nhiều tài liệu gọi hàm trên là hàm phân phối xác suất đồng thời của hai biến $X$ và $Y$.

**Tính chất**

- $0 \le F(x, y) \le 1,\ \forall x, y \in \mathbb{R}$;
- $F(x, y)$ là hàm không giảm theo từng đối số;
- $F(-\infty, y) = F(x, -\infty) = 0,\ \forall x, y \in \mathbb{R}$ và $F(+\infty, +\infty) = 1$;
- Với $x_1 < x_2,\ y_1 < y_2$ ta luôn có

  $$P(x_1 \le X \le x_2,\ y_1 \le y \le y_2) = F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1).$$





# Các khái niệm cơ sở

**Tính chất (tiếp)**

- Các hàm

  $$\begin{aligned} F(x, +\infty) &= P(X < x,\ Y < +\infty) = P(X < x) =: F_X(x) \\ F(+\infty, y) &= P(X < +\infty,\ Y < y) = P(Y < y) =: F_Y(x) \end{aligned}$$

  là các hàm phân phối riêng của các biến ngẫu nhiên $X$ và $Y$ và còn được gọi là các *phân phối biên* của biến ngẫu nhiên hai chiều $(X, Y)$.

**Định nghĩa 3.2**

Hai biến ngẫu nhiên $X, Y$ được gọi là *độc lập* nếu

$$F(x, y) = F_X(x).F_Y(y), \quad \forall x, y \in \mathbb{R}.$$

# Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

## Định nghĩa 3.3

Bảng phân phối xác suất của biến ngẫu nhiên hai chiều $(X, Y)$ rời rạc được xác định như sau

| X \ Y | $y_1$ | $y_2$ | $\dots$ | $y_j$ | $\dots$ | $y_n$ | $\sum_j$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | $p_{11}$ | $p_{12}$ | $\dots$ | $p_{1j}$ | $\dots$ | $p_{1n}$ | $P(X = x_1)$ |
| $x_2$ | $p_{21}$ | $p_{22}$ | $\dots$ | $p_{2j}$ | $\dots$ | $p_{2n}$ | $P(X = x_2)$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_i$ | $p_{i1}$ | $p_{i2}$ | $\dots$ | $p_{ij}$ | $\dots$ | $p_{in}$ | $P(X = x_i)$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_m$ | $p_{m1}$ | $p_{m2}$ | $\dots$ | $p_{mj}$ | $\dots$ | $p_{mn}$ | $P(X = x_m)$ |
| $\sum_i$ | $P(Y = y_1)$ | $P(Y = y_2)$ | $\dots$ | $P(Y = y_j)$ | $\dots$ | $P(Y = y_n)$ | 1 |



Luật phân phối xác suất của biến ngẫu nhiên nhiều chiều | Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

# Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

Trong đó

$$\boxed{p_{ij} = P\left(X = x_i,\ Y = y_j\right) \quad \forall i = \overline{1, m},\ j = \overline{1, n}.}$$

Kích thước bảng này có thể chạy ra vô hạn khi $m, n$ chạy ra vô hạn.

## Tính chất

- $p_{ij} \geq 0 \quad \forall i, j;$
- $\sum\limits_{i,j} p_{ij} = 1;$
- Hàm phân phối xác suất được xác định theo công thức $F(x, y) = \sum\limits_{i,j:\ x_i < x,\ y_j < y} p_{ij};$
- Các phân phối biên được xác định như sau:

$$\begin{aligned} P(X = x_i) &= \sum_j P(X = x_i,\ Y = y_j) = \sum_j p_{ij} \\ P(Y = y_j) &= \sum_i P(X = x_i,\ Y = y_j) = \sum_i p_{ij}. \end{aligned}$$

# Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

**Ví dụ 1**

Cho bảng phân phối xác suất đồng thời của $(X, Y)$ như sau:

| $X \backslash Y$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.10 | 0.25 | 0.10 |
| 2 | 0.15 | 0.05 | 0.35 |

Tìm bảng phân phối xác suất của $X$ và $Y$, sau đó tính $F(2; 3)$.





# Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

**Giải**

Lấy tổng của hàng, cột tương ứng ta thu được

| $X$ | 1 | 2 |
|---|---|---|
| $P(X = x)$ | 0.45 | 0.55 |

| $Y$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(Y = x)$ | 0.25 | 0.30 | 0.45 |

Ta có

$$F(2,3) = \sum_{x_i<2} \sum_{y_j<3} p_{ij} = p_{11} + p_{12} = 0.35.$$

# Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

### Chú ý 3.1

- Hai biến ngẫu nhiên $X, Y$ được gọi là độc lập với nhau nếu ta có

$$\boxed{P(X = x_i,\ Y = y_j) = P(X = x_i).P(Y = y_j),\quad \forall i = \overline{1, m},\ j = \overline{1, n}}$$

- Các xác suất có điều kiện vẫn được tính như thông thường, tức là

$$P\left(X = x_i | Y = y_j\right) = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)} \quad \text{hoặc}$$
$$P\left(X = x_i | Y \in D\right) = \frac{P(X = x_i,\ Y \in D)}{P(Y \in D)}$$

Công thức cũng tương tự với $P\left(Y = y_j | X = x_i\right),\ P\left(Y = y_j | X \in D\right)$.





# Phân phối xác suất của biến ngẫu nhiên hai chiều liên tục

### Định nghĩa 3.4

Hàm hai biến không âm, liên tục $f(x, y)$ được gọi là *hàm mật độ xác suất đồng thời* của biến ngẫu nhiên hai chiều liên tục $(X < Y)$ nếu nó thỏa mãn

$$\boxed{P\left((X, Y) \in \mathcal{D}\right) = \iint\limits_{\mathcal{D}} f(x, y)dxdy \quad \forall \mathcal{D} \subset \mathbb{R}^2.} \tag{3.2}$$

### Tính chất

- $F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u, v)dudv;$
- $\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f(x, y)dxdy.$

# Phân phối xác suất của biến ngẫu nhiên hai chiều liên tục

### Tính chất (tiếp)

- $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$;
- Các hàm mật độ biên
  - theo $x$ : $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)dy$;
  - theo $y$ : $f_Y(y) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)dx$.
- Hai biến ngẫu nhiên $X$ và $Y$ được gọi là độc lập nếu $f(x,y) = f_X(x).f_Y(y) \ \ \forall x,y$.
- Hàm mật độ có điều kiện của $X$ khi đã biết $Y = y$:

$$\varphi(x|y) = \frac{f(x,y)}{f_Y(y)}.$$





# Kỳ vọng và phương sai của các thành phần

### Trường hợp $(X,Y)$ rời rạc

$$EX = \sum_i P(X = x_i) = \sum_i \sum_j x_i p_{ij}; \quad EY = \sum_j y_j P(Y = y_j) = \sum_i \sum_j y_j p_{ij}$$

$$VX = \sum_i \sum_j x_i^2 p_{ij} - (EX)^2; \quad VY = \sum_i \sum_j y_j^2 p_{ij} - (EY)^2.$$

### Trường hợp $(X,Y)$ liên tục

$$EX = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x.f(x,y)dxdy; \quad EY = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y.f(x,y)dxdy$$

$$VX = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2.f(x,y)dxdy - (EX)^2; \quad VY = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2.f(x,y)dxdy - (EY)^2.$$

# Kỳ vọng và phương sai của các thành phần

## Chú ý 4.1

Đối với biến ngẫu nhiên $Z = g(X,Y)$ ta có

$$EZ = E\left[g(X,Y)\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y).f(x,y)dxdy$$





# Hiệp phương sai và hệ số tương quan

## Định nghĩa 4.1

Cho biến ngẫu nhiên hai chiều $(X,Y)$, hiệp phương sai của hai thành phần $X$ và $Y$, kí hiệu là $\mu_{XY}$ , được xác định bởi

$$\boxed{\mu_{XY} = E\left[(X-EX)(Y-EY)\right] = E(XY) - EX.EY,} \tag{4.3}$$

trong đó $E(XY)$ được xác định theo công thức

$$E(XY) = \begin{cases} \sum\limits_i \sum\limits_j x_i y_j p_{ij}, & \text{đối với biến ngẫu nhiên rời rạc} \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xy.f(x,y), & \text{đối với biến ngẫu nhiên liên tục} \end{cases}$$

# Hiệp phương sai và hệ số tương quan

**Định nghĩa 4.2**

Ta nói rằng $X$ và $Y$ *không tương quan* nếu $\mu_{XY} = 0$.

**Nhận xét**

- $\mu_{XY} = \mu_{YX}$;
- Phương sai chính là trường hợp riêng của hiệp phương sai ($VX = \mu_{XX}$, $VY = \mu_{YY}$);
- Nếu $X, Y$ độc lập thì ta có $E(XY) = EX.EY$. Khi đó $\mu_{XY} = 0$, tức là $X$ và $Y$ không tương quan. Vậy ta có, nếu hai biến ngẫu nhiên độc lập thì không tương quan. Điều ngược lại chưa chắc đã đúng.



Các tham số đặc trưng của biến ngẫu nhiên hai chiều Hiệp phương sai và hệ số tương quan

# Hiệp phương sai và hệ số tương quan

**Định nghĩa 4.3**

*Ma trận hiệp phương sai* của biến ngẫu nhiên hai chiều $(X, Y)$ được xác định bởi

$$\Gamma = \begin{bmatrix} \mu_{XX} & \mu_{XY} \\ \mu_{YX} & \mu_{YY} \end{bmatrix}$$

**Định nghĩa 4.4**

*Hệ số tương quan* của hai biến ngẫu nhiên $X$ và $Y$, ký hiệu là $\rho_{XY}$ và được xác định theo công thức

$$\boxed{\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y}.} \tag{4.4}$$

**Chú ý 4.2**

- Có thể chứng minh được $|\rho_{XY}| \leq 1$. Nếu $\rho_{XY} = \pm 1$ ta nói hai biến ngẫu nhiên $X$ và $Y$ có tương quan tuyến tính;
- Nếu $\rho_{XY} = 0$ ta nói hai biến ngẫu nhiên $X$ và $Y$ là không tương quan.

# Hàm của một biến ngẫu nhiên

Nếu ta xác định là một hàm của biến ngẫu nhiên $X$ thì $Z$ trở thành một biến ngẫu nhiên mới. Ta sẽ tìm hàm phân phối xác suất cho $Z$ trong một số trường hợp đơn giản.

**Định nghĩa 5.1**

Cho biến ngẫu nhiên $X$ có hàm phân phối xác suất. Khi đó hàm phân phối xác suất của $Z$ được xác định theo cách sau:

$$F_Z(z) = P(Z < z) = P\left(g(X) < z\right) = P(X \in D), \tag{5.5}$$

trong đó $D = \{x | g(x) < z\}$.
Tuy nhiên tùy vào từng bài có thể có các cách giải ngắn hơn.





# Hàm của một biến ngẫu nhiên

**Ví dụ 2**

Cho biến ngẫu nhiên $X$ có bảng phân phối xác suất

| $X$ | $-1$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|---|
| $P(X = x)$ | $0.1$ | $0.2$ | $0.3$ | $0.2$ | $0.2$ |

Xác định luật phân phối xác suất của $Z = X^2$ và tìm kỳ vọng của $Z$.

**Giải**

Ta có $X \in \{-1, 0, 1, 2, 3\}$, suy ra $Z \in \{0, 1, 4, 9\}$ với các xác suất tương ứng:

$$P(Z = 0) = P(X = 0) = 0.2; \qquad P(Z = 1) = P(X = 1) + P(X = -1) = 0.4;$$
$$P(X = 4) = P(X = 2) = 0.2; \qquad P(Z = 9) = P(X = 3) = 0.2.$$

| $Z$ | $0$ | $1$ | $4$ | $9$ |
|---|---|---|---|---|
| $P(Z = z)$ | $0.2$ | $0.4$ | $0.2$ | $0.2$ |

Kỳ vọng $EZ = \sum_i z_i p_i = 3$.

# Hàm của một biến ngẫu nhiên

**Ví dụ 3**

Thanh $AB$ dài 10cm bỗng nhiên bị gãy ở một điểm $C$ bất kỳ. Hai đoạn $AC$ và $BC$ được dùng làm hai cạnh của một hình chữ nhật. Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ diện tích hình chữ nhật đó.

**Giải**

Gọi $X$ là biến ngẫu nhiên chỉ độ dài đoạn $AC$, ta có $X \sim U(0;10)$. Gọi $Y$ là biến ngẫu nhiên chỉ diện tích hình chữ nhật, ta có $Y = X(10-X)$. Do
$X \in (0;10) \Rightarrow Y = X(10-X) \in (0;25)$. Vậy ta có hàm phân phối xác suất của $Y$ là
$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1, & y > 25 \end{cases}$.
Với $0 < y \leq 25$ ta có

$$\begin{aligned} F_Y(y) &= P(Y<y) = P(X(10-X)<y) = P\left(X^2 - 10X + y > 0\right) \\ &= P\left(X < 5 - \sqrt{25-y}\right) + P\left(X > 5 + \sqrt{25-y}\right) \\ &= P\left(0 < X < 5 - \sqrt{25-y}\right) + P\left(10 > X > 5 + \sqrt{25-y}\right) = \frac{5-\sqrt{25-y}}{5}. \end{aligned}$$





# Hàm của hai biến ngẫu nhiên

Xét biến ngẫu nhiên $Z = g(X,Y)$, trong đó $(X,Y)$ là biến ngẫu nhiên hai chiều đã biết luật phân phối. Ta sẽ xét luật phân phối xác suất của $Z$ trong một số trường hợp đơn giản theo cách sau:

$$F_Z(z) = P(Z<z) = P\left(g(X,Y)<z\right) = P\left((X,Y) \in D\right),$$

trong đó $D\{(x,y)|g(x,y)<z\}$.
Đối với biến ngẫu nhiên hai chiều liên tục $(X,Y)$ với hàm mật độ đồng thời $f(x,y)$ ta có

$$P\left((X,Y)\in D\right) = \iint\limits_D f(x,y)dxdx,$$

đồng thời kỳ vọng

$$EZ = E\left(g(X,Y)\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y).f(x,y)dxdy.$$

# Hàm của hai biến ngẫu nhiên

### Ví dụ 4

Hai người bạn hẹn gặp nhau ở công viên trong khoảng thời gian từ 17h đến 18h. Họ hẹn nhau nếu người nào đến trước thì sẽ đợi người kia trong vòng 10 phút. Sau 10 phút đợi nếu không gặp sẽ về. Thời điểm đến của hai người là ngẫu nhiên và độc lập với nhau trong khoảng thời gian trên. Tính xác suất hai người gặp được nhau.

### Giải

Quy gốc thời gian về lúc 17h. Gọi $X, Y$ là biến ngẫu nhiên chỉ thời điểm người $A, B$ đến, ta có $X, Y \sim U(0; 60)$. Do $X, Y$ độc lập nên chúng có hàm mật độ đồng thời $f(x,y) = \begin{cases} \dfrac{1}{3600}, & (x,y) \in [0;60]^2 \\ 0, & \text{ngược lại} \end{cases}$. Gọi $Z$ là biến ngẫu nhiên chỉ khoảng thời gian giữa thời điểm hai người đến. Ta có $Z = |X - Y|$. Khi đó, xác suất hai người gặp nhau là

$$P(Z < 10) = P(|X - Y| < 10) = P((X, Y) \in D),$$

trong đó $D$ là giao miền $|X - Y| < 10$ và hình vuông $[0; 60]^2$. Vậy

$$P(Z < 10) = \frac{S_D}{\ldots} = \frac{1100}{\ldots} = \frac{11}{\ldots}$$





# Luật số lớn

### Bất đẳng thức Trebyshev

**Định lý 1**: Cho $Y$ là biến ngẫu nhiên không âm. Khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$P(Y \geq \epsilon) < \frac{E(Y^2)}{\epsilon^2}$$

### Chứng minh

Ta chứng minh cho trường hợp $Y$ là biến ngẫu nhiên liên tục.

$$P(Y \geq \epsilon) = \int_{\epsilon}^{+\infty} f(y)dy = \frac{1}{\epsilon^2}\int_{\epsilon}^{+\infty} \epsilon^2.f(y)dy \leq \frac{1}{\epsilon^2}\int_{\epsilon}^{+\infty} y^2.f(y)dy$$

$$\leq \frac{1}{\epsilon^2}\int_{0}^{+\infty} y^2.f(y)dy = \frac{E(Y^2)}{\epsilon^2}$$

Tuy nhiên dấu bằng không thể đồng thời xảy ra ở cả 2 dấu $\leq$ nên ta có ĐPCM.

# Luật số lớn

## Bất đẳng thức Trebyshev

**Định lý 2**: Cho $X$ là biến ngẫu nhiên có $EX = \mu, VX = \sigma^2$ hữu hạn. Khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$P(|X - \mu| \geq \epsilon) < \frac{\sigma^2}{\epsilon^2}$$

hay tương đương

$$P(|X - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

## Chứng minh

Ta chứng minh cho trường hợp $X$ là biến ngẫu nhiên liên tục.
Ta chỉ cần đặt $Y = |X - \mu|$, lập tức áp dụng định lý 1 ta có ĐPCM.





# Luật số lớn

Áp dụng định lý 2 với $X = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ ta có luật số lớn Trebyshev

## Luật số lớn Trebyshev

Nếu dãy các biến ngẫu nhiên $X_1, X_2, ...X_n, ...$ độc lập, có kỳ vọng hữu hạn và phương sai bị chặn ($VX_i \leq C$ với C là hằng số), khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i| < \epsilon) = 1$$

## Hệ quả

Nếu dãy các biến ngẫu nhiên $X_1, X_2, ...X_n, ...$ độc lập, có cùng kỳ vọng ($EX_i = \mu$) và phương sai bị chặn ($VX_i \leq C$ với C là hằng số), khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \epsilon) = 1$$

# Luật số lớn Bernoulli

Áp dụng luật số lớn Trebyshev với trường hợp $X_i \sim B(1,p)$ chính là số lần xảy ra $A$ trong phép thử thứ $i$ ta có luật số lớn Bernoulli.

**Luật số lớn Bernoulli**

Xét $n$ phép thử độc lập, cùng điều kiện.
Trong mỗi phép thử, xác suất xảy ra $A$ luôn là $p$.
$m$ là số lần xảy ra $A$ trong $n$ phép thử.
khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$\lim_{n\to+\infty} P(|\frac{m}{n} - p| < \epsilon) = 1$$

Với luật số lớn Bernoulli ta đã chứng minh được điều thừa nhận trong phần ĐỊNH NGHĨA XÁC SUẤT THEO THỐNG KÊ, đó là với $n \to +\infty$ thì $\frac{m}{n} \to p$





# Định lý giới hạn trung tâm

**Định lý giới hạn trung tâm**

Giả sử $\{X_n\}$ là dãy biến ngẫu nhiên độc lập cùng phân phối với $EX_i = \mu, VX_i = \sigma^2$.
Đặt $\overline{X_n} = \sum\limits_{i=1}^{n} X_i$. Khi đó với $n$ đủ lớn ta có:

$$\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$$

hay là

$$\frac{\overline{X_n} - \mu}{\sigma}\sqrt{n} \sim N(0;1)$$

# Chương 4: Thống kê - Ước lượng tham số

Lê Xuân Lý [1]

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

Hà Nội, tháng 8 năm 2014

[1] Email: `lexuanly@gmail.com`





## Tổng thể

Khi nghiên cứu về một vấn đề người ta thường khảo sát trên một dấu hiệu nào đó, các dấu hiệu này được thể hiện trên nhiều phần tử.

**Định nghĩa 1.1**

*Tập hợp các phần tử mang dấu hiệu ta quan tâm được gọi là tổng thể hay đám đông (population).*



# Một số lý do không thể khảo sát toàn bộ tổng thể

- **Giới hạn về thời gian, tài chính**: Ví dụ muốn khảo sát xem chiều cao của thanh niên VN hiện nay có tăng lên hay không ta phải khảo sát toàn bộ thanh niên VN (giả sử là 40 triệu người). Để khảo sát hết sẽ tốn nhiều thời gian và kinh phí. Ta có thể khảo sát một triệu thanh niên VN, từ chiều cao trung bình thu được ta suy ra chiều cao trung bình của người VN.
- **Phá vỡ tổng thể nghiên cứu**: Ví dụ ta cất vào kho $N = 10000$ hộp sản phẩmvà muốn biết tỷ lệ hộp hư sau 1 năm bảo quản. Ta phải kiểm tra từng hộp để xác định số hộp hư $M = 300$, tỷ lệ hộp hư trong kho là $M/N$. Một hộp sản phẩm sau khi kiểm tra thì mất phẩm chất, và vì vậy sau khi kiểm tra cả kho thì cũng "tiêu" luôn kho. Ta có thể lấy ngẫu nhiên $n = 100$ hộp ra kiểm tra, giả sử có $m = 9$ hộp bị hư. Tỷ lệ hộp hư 9% ta suy ra tỷ lệ hộp hư của cả kho.
- **Không xác định được chính xác tổng thể**: Ví dụ muốn khảo sát tỷ lệ người bị nhiễm HIV qua đường tiêm chích là bao nhiêu. Tổng thể lúc này là toàn bộ người bị nhiễm HIV, nhưng ta không thể xác định chính xác là bao nhiêu người (những người xét nghiệm thì bệnh viện biết, những người không xét nghiệm thì ...). Do đó ta chỉ biết một phần tổng thể. Ngoài ra số người bị nhiễm HIV mới và bị chết do HIV thay đổi liên tục nên tổng thể thay đổi liên tục.





# Tập mẫu

Do đó người ta nghĩ ra cách thay vì khảo sát tổng thể, người ta chỉ cần chọn ra một tập nhỏ để khảo sát và đưa ra quyết định.

**Định nghĩa 1.2**

- *Tập mẫu là tập con của tổng thể và có tính chất tương tự như tổng thể.*
- *Số phần tử của tập mẫu được gọi là kích thước mẫu.*

**Câu hỏi**: Làm sao chọn được tập mẫu có tính chất tương tự như tổng thể để các kết luận của tập mẫu có thể dùng cho tổng thể ?

# Một số cách chọn mẫu cơ bản

## Một số cách chọn mẫu

- Chọn mẫu ngẫu nhiên có hoàn lại: Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó. Sau đó trả phần tử đó lại tổng thể trước khi lấy 1 phần tử khác. Tiếp tục như thế n lần ta thu được một mẫu có hoàn lại gồm n phần tử.
- Chọn mẫu ngẫu nhiên không hoàn lại: Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó rồi để qua một bên, không trả lại tổng thể. Sau đó lấy ngẫu nhiên 1 phần tử khác, tiếp tục như thế n lần ta thu được một mẫu không hoàn lại gồm n phần tử.
- Chọn mẫu phân nhóm: Đầu tiên ta chia tập nền thành các nhóm tương đối thuần nhất, từ mỗi nhóm đó chọn ra một mẫu ngẫu nhiên. Tập hợp tất cả mẫu đó cho ta một mẫu phân nhóm. Phương pháp này dùng khi trong tập nền có những sai khác lớn. Hạn chế là phụ thuộc vào việc chia nhóm.
- Chọn mẫu có suy luận: dựa trên ý kiến của chuyên gia về đối tượng nghiên cứu để chọn mẫu.





# Biểu diễn dữ liệu

Từ tổng thể ta trích ra tập mẫu có n phần tử. Ta có $n$ số liệu.

## Dạng liệt kê

Các số liệu thu được ta ghi lại thành dãy số liệu:

$$x_1, x_2, \ldots, x_n$$

## Dạng rút gọn

Số liệu thu được có sự lặp đi lặp lại một sô giá trị thì ta có dạng rút gọn sau:

- Dạng tần số: $(n_1 + n_2 + \ldots + n_k = n)$

| Giá trị | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| Tần số | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

- Dạng tần suất: $(p_k = n_k/n)$

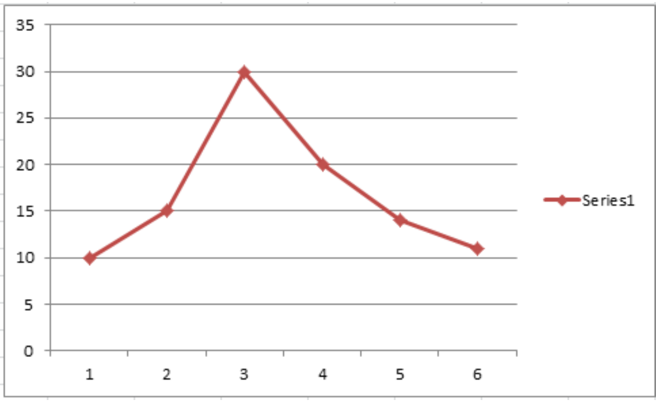
| Giá trị | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| Tần suất | $p_1$ | $p_2$ | $\ldots$ | $p_k$ |

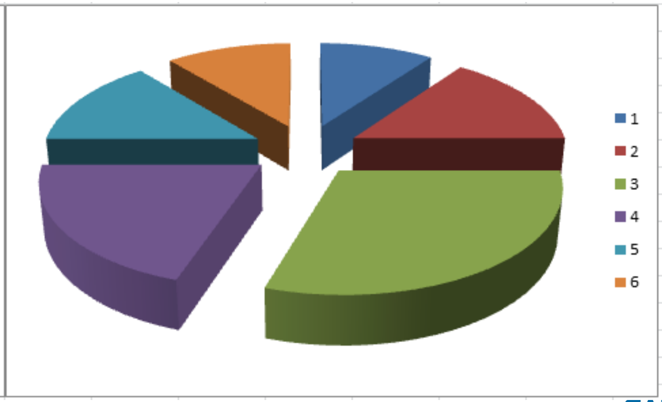
# Biểu diễn dữ liệu

## Ví dụ dạng rút gọn

Ta có bảng số liệu như sau:

| Giá trị | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Tần số | 10 | 15 | 30 | 20 | 14 | 11 |
| Tần suất | 0.10 | 0.15 | 0.30 | 0.20 | 0.14 | 0.11 |








# Biểu diễn dữ liệu

## Dạng khoảng

Dữ liệu thu được nhận giá trị trong $(a, b)$. Ta chia $(a, b)$ thành $k$ miền con bởi các điểm chia: $a_0 = a < a_1 < a_2 < ... < a_{k-1} < a_k = b$.

- Dạng tần số: $(n_1 + n_2 + \ldots + n_k = n)$

| Giá trị | $(a_0 - a_1]$ | $(a_1 - a_2]$ | $\ldots$ | $(a_{k-1} - a_k]$ |
|---|---|---|---|---|
| Tần số | $n_1$ | $n_2$ | $\ldots$ | $n_k$ |

- Dạng tần suất: $(p_k = n_k/n)$

| Giá trị | $(a_0 - a_1]$ | $(a_1 - a_2]$ | $\ldots$ | $(a_{k-1} - a_k]$ |
|---|---|---|---|---|
| Tần suất | $p_1$ | $p_2$ | $\ldots$ | $p_k$ |

- Một số vấn đề chú ý:
  - $k = 5 \to 15$.
  - Độ dài các khoảng thường chia bằng nhau.

# Biểu diễn dữ liệu

### Dạng khoảng

- Nếu độ dài các khoảng bằng nhau ta có thể chuyển về dạng rút gọn.

| Giá trị | $x_1$ | $x_2$ | $\ldots$ | $x_k$ |
|---|---|---|---|---|
| Tần suất | $p_1$ | $p_2$ | $\ldots$ | $p_k$ |

Trong đó $x_i$ là điểm đại diện cho $(a_{i-1}, a_i]$ thường được xác định là trung điểm của miền: $x_i = \frac{1}{2}(a_{i-1} + a_i)$

- Dạng rút gọn thường được thể hiện bằng đồ thị dạng đường hoặc dạng hình tròn.
- Dạng khoảng thường được thể hiện bằng đồ thị dạng hình cột.





# Mẫu ngẫu nhiên

Tổng thể được đặc trưng bởi dấu hiệu nghiên cứu $X$ là một biến ngẫu nhiên. Do đó khi nói về $X$ là nói về tổng thể.
Từ tổng thể trích ra $n$ phần tử làm một tập mẫu. Ta có 2 loại tập mẫu: *mẫu ngẫu nhiên* và *mẫu cụ thể*
Gọi $X_i$ là biến ngẫu nhiên chỉ giá trị thu được của phần tử thứ $i, i = 1, 2, \ldots, n$. Ta có $X_1, X_2, \ldots, X_n$ là $n$ biến ngẫu nhiên độc lập và có cùng phân phối với biến ngẫu nhiên $X$.

### Định nghĩa 2.1

- *Mẫu ngẫu nhiên: là véctơ $W_X = (X_1, X_2, \ldots, X_n)$, trong đó mỗi thành phần $X_i$ là một biến ngẫu nhiên. Các biến ngẫu nhiên này độc lập và có cùng phân phối xác suất với $X$.*
- *Mẫu cụ thể: là véctơ $W_x = (x_1, x_2, \ldots, x_n)$, trong đó mỗi thành phần $x_i$ là một giá trị cụ thể.*
- *Với một mẫu ngẫu nhiên thì có nhiều mẫu cụ thể ứng với các lần lấy mẫu khác nhau.*

# Mẫu ngẫu nhiên

## Ví dụ 1

*Một kệ chứa các đĩa nhạc với giá như sau:*

| *Giá (ngàn đồng)* | 20 | 25 | 30 | 34 | 40 |
|---|---|---|---|---|---|
| *Số đĩa* | 35 | 10 | 25 | 17 | 13 |

*Ta cần lấy 4 đĩa có hoàn lại để khảo sát.*
*Ta xét trong 2 trường hợp:*

- *Xét về mặt định lượng: giá của từng đĩa là bao nhiêu?*
- *Xét về mặt định tính: đĩa đó có phải đĩa lậu không?*
  *(Đĩa lậu là đĩa có giá dưới 25 ngàn đồng)*





# Mẫu ngẫu nhiên

## Xét tổng thể về mặt định lượng

Lấy ngẫu nhiên một đĩa nhạc trong kệ. Gọi $X$ là giá của đĩa nhạc này. Ta có bảng phân phối xác suất của $X$.

| $X$ | 20 | 25 | 30 | 34 | 40 |
|---|---|---|---|---|---|
| $P$ | $0,35$ | $0,10$ | $0,25$ | $0,17$ | $0,13$ |

- Lấy ngẫu nhiên có hoàn lại 4 đĩa nhạc từ kệ.
  Gọi $X_i$ là giá của đĩa nhạc thứ $i$ lấy được, $i = 1, 2, 3, 4$.
- Ta thấy các biến $X_i$ độc lập và có cùng phân phối xác suất với $X$.
- Ta có $W_X = (X_1, X_2, X_3, X_4)$ là một mẫu ngẫu nhiên.
- Bây giờ ta khảo sát giá cụ thể của 4 đĩa lấy ra, ta thấy:
  - Đĩa 1: giá 20 ngàn đồng
  - Đĩa 2: giá 30 ngàn đồng
  - Đĩa 3: giá 20 ngàn đồng
  - Đĩa 4: giá 40 ngàn đồng

  Lập $W_x = (x_1, x_2, x_3, x_4) = (20, 30, 20, 40)$, đây là mẫu cụ thể.

# Mẫu ngẫu nhiên

### Xét tổng thể về mặt định tính

Đĩa có giá dưới 25 ngàn đồng là đĩa "lậu". Lấy ngẫu nhiên một đĩa từ kệ.
Gọi $X$ là số đĩa lậu lấy được.

| $X$ | $0$ | $1$ |
|---|---|---|
| $P$ | $0,65$ | $0,35$ |

- Lấy ngẫu nhiên có hoàn lại 4 đĩa nhạc từ kệ.
  Gọi $X_i$ là só đĩa lậu lấy được khi lấy một đĩa lần thứ $i$, $i = 1, 2, 3, 4$.
- Ta thấy các biến $X_i$ độc lập và có cùng phân phối xác suất với $X$.
- Ta có $W_X = (X_1, X_2, X_3, X_4)$ là một mẫu ngẫu nhiên.
- Bây giờ ta khảo sát giá cụ thể của 4 đĩa lấy ra, ta thấy:
  • Đĩa 1: giá 20 ngàn đồng $\rightarrow x_1 = 1$
  • Đĩa 2: giá 30 ngàn đồng $\rightarrow x_2 = 0$
  • Đĩa 3: giá 20 ngàn đồng $\rightarrow x_3 = 1$
  • Đĩa 4: giá 40 ngàn đồng $\rightarrow x_4 = 0$
  Lập $W_x = (x_1, x_2, x_3, x_4) = (1, 0, 1, 0)$, đây là mẫu cụ thể.





# Các đặc trưng mẫu

### Thống kê

Cho $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên.
Biến ngẫu nhiên $Y = g(X_1, X_2, ..., X_n)$ (với $g$ là một hàm nào đó) được gọi là một thống kê

### Các tham số đặc trưng

- *Xét tổng thể về mặt định lượng*: tổng thể được đặc trưng bởi dấu hiệu nghiên cứu $X$, ($X$ là biến ngẫu nhiên). Ta có:
  • Trung bình tổng thể: $EX = \mu$
  • Phương sai tổng thể: $VX = \sigma^2$
  • Độ lệch chuẩn của tổng thể: $\sigma$.
- *Xét tổng thể về mặt định tính*: tổng thể có kích thước$N$, trong đó có $M$ phần tử có tính chất $A$. Khi đó $p = M/N$ gọi là tỷ lệ xảy ra $A$ của tổng thể.

# Các đặc trưng mẫu

## Trung bình mẫu

Cho $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên.

- Thống kê - **Trung bình mẫu ngẫu nhiên**:

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

- Mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì $\overline{X}$ nhận giá trị:

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

  $\overline{x}$ được gọi là **trung bình mẫu**.

  Nếu mẫu dạng rút gọn thì: $\overline{x} = \frac{1}{k}\sum\limits_{i=1}^{n} x_i n_i$





# Các đặc trưng mẫu

## Phương sai mẫu(chưa hiệu chỉnh)

Cho $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên.

- Thống kê - **Phương sai mẫu ngẫu nhiên**:

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

- Mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì $S^2$ nhận giá trị:

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

  $S^2$ được gọi là **Phương sai mẫu (chưa hiệu chỉnh)**.

- Vấn đề: $E(S^2) = \dfrac{n-1}{n}\sigma^2$

# Các đặc trưng mẫu

### Phương sai mẫu hiệu chỉnh

Ta phải hiệu chỉnh đi để thu được giá trị thay thế $\sigma^2$ tốt hơn.

- Thống kê - **Phương sai mẫu ngẫu nhiên hiệu chỉnh**:

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

- Mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ có mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì $s^2$ nhận giá trị:

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

  $s^2$ được gọi là **Phương sai mẫu hiệu chỉnh**.
- $s$ được gọi là **độ lệch chuẩn mẫu hiệu chỉnh**.





# Ước lượng điểm

### Vấn đề

Cho biến ngẫu nhiên gốc $X$ có phân phối xác suất đã biết nhưng chưa biết tham số $\theta$ nào đó.
Mẫu số liệu thu thập được của $X$ là: $(x_1, x_2, ..., x_n)$.
Khi đó $\overline{\theta} = g(x_1, x_2, ..., x_n)$ được gọi là một ước lượng điểm của $\theta$
Muốn biết ước lượng này tốt hay xấu ta phải so sánh với $\theta$.

### Ước lượng không chệch

Thống kê $\overline{\theta}$ được gọi là ước lượng không chệch của $\theta$ nếu thoả mãn: $E\overline{\theta} = \theta$

### Kết quả

Cho biến ngẫu nhiên $X$ có $EX = \mu, VX = \sigma^2$. Mẫu số liệu quan sát $(x_1, x_2, ..., x_n)$.
Khi đó ta có kết quả:

- Ước lượng không chệch cho $\mu$ là: $\overline{x}$
- Ước lượng không chệch cho $\sigma^2$ là: $s^2$

# Xác định ước lượng điểm

**Ví dụ 2**

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

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

*a. Tính trung bình mẫu, độ lệch chuẩn mẫu hiệu chỉnh.*
*b. Những thửa ruộng có năng suất từ 48 tạ/ha trở lên là những thửa ruộng có năng suất cao. Tính trung bình mẫu, độ lệch chuẩn mẫu hiệu chỉnh của những thửa ruộng có năng suất cao.*

$\overline{x} = 46; s = 3,30$





# Xác định ước lượng điểm

**Ví dụ 3**

*Quan sát tuổi thọ của một số người ta có bảng số liệu sau:*

| Tuổi(năm) | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|
| Số người | 5 | 14 | 25 | 6 |

*Tính trung bình mẫu, độ lệch chuẩn mẫu hiệu chỉnh của biến ngẫu nhiên $X$ chỉ tuổi thọ của con người.*

$\overline{x} = 41,4; s = 8,271$

# Ước lượng khoảng

- Cho biến ngẫu nhiên $X$ có $EX = \mu, VX = \sigma^2$.
  Mẫu cụ thể của $X$ là $(x_1, x_2, ..., x_n)$
  **Chú ý**: nếu cỡ mẫu $n \leq 30$ thì ta phải thêm điều kiện $X \sim N(\mu, \sigma^2)$.
- Bài toán đặt ra là tìm khoảng ước lượng cho $\mu$ với xác suất xảy ra bằng $(1-\alpha)$ cho trước. Điều đó tương đương với việc tim khoảng $(a, b)$ sao cho:

$$P(a < \mu < b) = 1 - \alpha$$

  - $(a, b)$ được gọi là khoảng tin cậy (hoặc khoảng ước lượng) của $\mu$.
  - $(1-\alpha)$ được gọi là độ tin cậy.





# Ước lượng khoảng cho kỳ vọng

**Trường hợp 1: $\sigma$ đã biết**

- Chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0; 1)$
- Xét cặp số không âm $\alpha_1, \alpha_2$ thoả mãn: $\alpha_1 + \alpha_2 = \alpha$ và các phân vị chuẩn tắc $u_{\alpha_1}, u_{1-\alpha_2}$:
  - $P(Z < u_{\alpha_1}) = \alpha_1$. Do tính chất của phân phối chuẩn tắc: $u_{\alpha_1} = -u_{1-\alpha_1}$
  - $P(Z < u_{1-\alpha_2}) = 1 - \alpha_2$

  Suy ra $P(-u_{1-\alpha_1} < Z < u_{1-\alpha_2}) = P(u_{\alpha_1} < Z < u_{1-\alpha_2})$
  $= P(Z < u_{1-\alpha_2}) - P(Z < u_{\alpha_1}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$
- $1 - \alpha = P(-u_{1-\alpha_1} < Z < u_{1-\alpha_2}) = P(-u_{1-\alpha_1} < \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} < u_{1-\alpha_2})$
  $= P(\overline{X} - u_{1-\alpha_2}\dfrac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{1-\alpha_1}\dfrac{\sigma}{\sqrt{n}})$
- Từ mẫu cụ thể $(x_1, x_2, .., x_n)$, ta có khoảng ước lượng cho $\mu$ với độ tin cậy $1 - \alpha$ là:

$$(\overline{x} - u_{1-\alpha_2}\frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\alpha_1}\frac{\sigma}{\sqrt{n}})$$

  Như vậy có vô số khoảng ước lượng cho $\mu$.

# Ước lượng khoảng cho kỳ vọng

### Trường hợp 1: $\sigma$ đã biết

- Khoảng ước lượng đối xứng ($\alpha_1 = \alpha_2 = \alpha/2$):

$$\boxed{(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})}, \text{hàm laplace: } \phi(u_{1-\frac{\alpha}{2}}) = 0,5 - \frac{\alpha}{2}$$

  trong đó $\epsilon = u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ gọi là *độ chính xác của ước lượng*.

  **Chú ý**: Khoảng đối xứng là khoảng ước lượng có độ dài ngắn nhất.

- Khoảng ước lượng một phía ($\alpha_1 = \alpha; \alpha_2 = 0$):

$$\boxed{(-\infty; \overline{x} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}})}, \text{hàm laplace: } \phi(u_{1-\alpha}) = 0,5 - \alpha$$

- Khoảng ước lượng một phía ($\alpha_1 = 0; \alpha_2 = \alpha$):

$$\boxed{(\overline{x} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}; +\infty)}, \text{hàm laplace: } \phi(u_{1-\alpha}) = 0,5 - \alpha$$





# Ước lượng khoảng cho kỳ vọng

### Ví dụ

Doanh thu của một cửa hàng là biến ngẫu nhiên $X$(triệu/tháng) có độ lệch chuẩn 2 triệu/tháng. Điều tra ngẫu nhiên doanh thu của 500 cửa hàng có qui mô tương tự nhau ta tính được doanh thu trung bình là 10 triệu/tháng. Với độ tin cậy 95% hãy ước lượng khoảng cho doanh thu trung bình của cửa hàng thuộc qui mô đó.

### Bài làm

- $X$ là doanh thu của cửa hàng loại đang xét, $EX = \mu$ , $VX = \sigma^2$ với $\sigma = 2$
  Chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0; 1)$
- Khoảng ước lượng đối xứng cho doanh thu trung bình $\mu$ là:
  $(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$
- Với $\overline{x} = 10, \sigma = 2, n = 500$
  $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$
- Thay các số liệu vào khoảng trên ta có kết quả: (9,825 ; 10,175)

# Ước lượng khoảng cho kỳ vọng

### Trường hợp 2: $\sigma$ chưa biết

Do $\sigma$ chưa biết nên ta thay thế bằng $s$.

- Chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{s}\sqrt{n} \sim t(n-1)$
- Làm tương tự như trường hợp 1, ta chỉ thay phân vị chuẩn bằng phân vị Student.
- Mẫu cụ thể $(x_1, x_2, .., x_n)$, ta có khoảng ước lượng cho $\mu$ với độ tin cậy $1-\alpha$ là:

$$(\overline{x} - t(n-1, 1-\alpha_2)\frac{s}{\sqrt{n}}; \overline{x} + t(n-1, 1-\alpha_1)\frac{s}{\sqrt{n}})$$

- **Chú ý**:
  $n > 30$ thì phân phối chuẩn và phân phối student bậc tự do $(n-1)$ có thể coi là một.
  Do đó nếu $n > 30$ ta có thể chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{s}\sqrt{n} \sim N(0;1)$
  Khoảng ước lượng cho $\mu$ với độ tin cậy $1-\alpha$ là:

$$(\overline{x} - u_{1-\alpha_2}\frac{s}{\sqrt{n}}; \overline{x} + u_{1-\alpha_1}\frac{s}{\sqrt{n}})$$





# Ước lượng khoảng cho kỳ vọng

### Trường hợp 2: $\sigma$ chưa biết

- Khoảng ước lượng đối xứng ($\alpha_1 = \alpha_2 = \alpha/2$):

$$\boxed{(\overline{x} - t(n-1, 1-\frac{\alpha}{2})\frac{s}{\sqrt{n}}; \overline{x} + t(n-1, 1-\frac{\alpha}{2})\frac{s}{\sqrt{n}})}$$

- Khoảng ước lượng một phía ($\alpha_1 = \alpha; \alpha_2 = 0$):

$$\boxed{(-\infty; \overline{x} - t(n-1, 1-\alpha)\frac{s}{\sqrt{n}})}$$

- Khoảng ước lượng một phía ($\alpha_1 = 0; \alpha_2 = \alpha$):

$$\boxed{(\overline{x} - t(n-1, 1-\alpha)\frac{s}{\sqrt{n}}; +\infty)}$$

# Ước lượng khoảng cho kỳ vọng

### Ví dụ

*Ví dụ trước sẽ hợp với thực tế hơn nếu ta sửa lại như sau:*
Doanh thu của một cửa hàng là biến ngẫu nhiên $X$(triệu/tháng). Điều tra ngẫu nhiên doanh thu của 500 cửa hàng có qui mô tương tự nhau ta tính được doanh thu trung bình là 10 triệu/tháng và độ lệch chuẩn mẫu hiệu chỉnh là 2 triệu/tháng. Với độ tin cậy 95% hãy ước lượng khoảng cho doanh thu trung bình của cửa hàng thuộc qui mô đó.

### Bài làm

- $X$(triệu/tháng) là doanh thu của cửa hàng loại đang xét, $EX = \mu$ , $VX = \sigma^2$
  Chọn thống kê: $Z = \dfrac{\overline{X} - \mu}{s}\sqrt{n} \sim t(n-1)$
- Khoảng ước lượng đối xứng cho doanh thu trung bình $\mu$ là:
  $(\overline{x} - t(n-1, 1-\frac{\alpha}{2})\dfrac{s}{\sqrt{n}}; \overline{x} + t(n-1, 1-\frac{\alpha}{2})\dfrac{s}{\sqrt{n}})$
- Với $\overline{x} = 10, s = 2, n = 500$
  $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow t(n-1, 1-\frac{\alpha}{2}) = t(499; 0,975) = 1,96$
- Thay các số liệu vào khoảng trên ta có kết quả: (9,825 ; 10,175)





# Ước lượng khoảng cho tỷ lệ

### Bài toán

Xác suất xảy ra sự kiện $A$ là $p$.
Do không biết $p$ nên người ta thực hiện $n$ phép thử độc lập, cùng điều kiện.
Trong đó có $m$ phép thử xảy ra $A$.
$f = {}^{m}/_{n}$ là ước lượng điểm không chệch cho $p$.
**Câu hỏi**: Với độ tin cậy $(1-\alpha)$ hãy ước lượng khoảng cho $p$.

### Cách giải quyết: tương tự cách làm cho kỳ vọng

- Chọn thống kê: $Z = \dfrac{f-p}{\sqrt{p(1-p)}}\sqrt{n} \sim N(0;1)$
- Tuy nhiên do khó giải quyết nên người ta thay $p$ dưới mẫu bởi $f$ cho dễ tính.
  Thống kê trở thành: $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n} \sim N(0;1)$
- Mẫu cụ thể $(x_1, x_2, .., x_n)$, ta có khoảng ước lượng cho $p$ với độ tin cậy $1-\alpha$ là:

$$(f - u_{1-\alpha_2}\sqrt{\frac{f(1-f)}{n}}, f + u_{1-\alpha_1}\sqrt{\frac{f(1-f)}{n}})$$

# Ước lượng khoảng cho tỷ lệ

## Các trường hợp ước lượng hay dùng

- Khoảng ước lượng đối xứng ($\alpha_1 = \alpha_2 = \alpha/2$):

$$\boxed{(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}, f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}})}$$

- Khoảng ước lượng một phía ($\alpha_1 = \alpha; \alpha_2 = 0$):

$$\boxed{(-\infty; f + u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}})}$$

- Khoảng ước lượng một phía ($\alpha_1 = 0; \alpha_2 = \alpha$):

$$\boxed{(f - u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}}; +\infty)}$$

- **Chú ý**: Do tỷ lệ chỉ nhận giá trị từ 0 đến 1 nên ta có thể thay giá trị $-\infty$ bằng 0 và $+\infty$ bằng 1 trong khoảng ước lượng một phía.





# Ước lượng khoảng cho tỷ lệ

## Ví dụ

Tại một bến xe, kiểm tra ngẫu nhiên 100 xe thấy có 30 xe xuất phát đúng giờ. Với độ tin cậy 95% hãy ước lượng khoảng cho tỷ lệ xe xuất phát đúng giờ.

## Bài làm

- Gọi $p$ là tỷ lệ xe xuất phát đúng giờ.
  Chọn thống kê: $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n} \sim N(0;1)$
- Khoảng ước lượng đối xứng cho tỷ lệ xe xuất phát đúng giờ là:
  $(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}, f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}})$
- Với $n = 100, m = 30 \Rightarrow f = {}^{m}/_{n} = 0,3$
  $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$
- Thay các số liệu vào khoảng trên ta có kết quả: (0,21 ; 0,39)

# Chương 5: Kiểm định giả thuyết

Lê Xuân Lý [(1)]

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

Hà Nội, tháng 8 năm 2014

(1) Email: lexuanly@gmail.com





## Kiểm định giả thuyết cho kỳ vọng

- Giả thuyết thống kê: Trong nhiều lĩnh vực của đời sống kinh tế xã hội, chúng ta thường nêu ra các nhận xét khác nhau về đối tượng quan tâm. Những nhận xét như vậy có thể đúng hoặc sai. Vấn đề kiểm tra tính đúng sai của nhận xét sẽ được gọi là kiểm định.
- Kiểm định giả thuyết là bài toán đi xác định có nên chấp nhận hay bác bỏ một khẳng định về giá trị của một tham số của tổng thể.

### Bài toán

- Cho biến ngẫu nhiên $X$ có $EX = \mu, VX = \sigma^2$.
  Mẫu cụ thể của $X$ là $(x_1, x_2, ..., x_n)$
  **Chú ý**: nếu cỡ mẫu $n \leq 30$ thì ta phải thêm điều kiện $X \sim N(\mu, \sigma^2)$.
- Bài toán đặt ra là ta cần so sánh giá trị kỳ vọng $\mu$ với một số $\mu_0$ cho trước.

| | | | |
|---|---|---|---|
| Giả thuyết $H_0$ | $\mu = \mu_0$ | $\mu \leq \mu_0$ | $\mu \geq \mu_0$ |
| Đối thuyết $H_1$ | $\mu \neq \mu_0$ | $\mu > \mu_0$ | $\mu < \mu_0$ |

- Tuy nhiên do giả thuyết luôn có dấu "=" nên người ta chỉ cần viết giả thuyết $H_0 : \mu = \mu_0$

# Kiểm định giả thuyết một mẫu

## Cách giải quyết

- Từ bộ số liệu đã cho $x_1, x_2, ..., x_n$ ta tính được giá trị quan sát $k$.
- Ta chia được trục số thành 2 phần, trong đó một phần là $W_\alpha$
  +) Nếu $X \in W_\alpha$ thì bác bỏ $H_0$ và chấp nhận $H_1$
  +) Nếu $X \notin W_\alpha$ thì ta không có cơ sở bác bỏ $H_0$

## Sai lầm mắc phải

Có 2 loại sai lầm c ó thể mắc phải

- Sai lầm loại 1: Bác bỏ $H_0$ trong khi $H_0$ đúng.
  Xác suất xảy ra sai lầm loại 1: $\alpha = P(k \in W_\alpha | H_0$ đúng$)$
  $\alpha$ được gọi là **mức ý nghĩa**
- Sai lầm loại 2: Chấp nhận $H_0$ trong khi $H_0$ sai.
  Xác suất xảy ra sai lầm loại 2: $\beta = P(k \notin W_\alpha | H_0$ sai$)$
- Mục tiêu là cực tiểu cả 2 sai lầm, tuy nhiên điều đó là rất khó khăn. Người ta chọn cách cố định sai lầm loại 1 và cực tiểu sai lầm loại 2.





# Kiểm định giả thuyết một mẫu

| | Bản chất tổng thể | |
|---|---|---|
| Kết luận | $H_0$ đúng | $H_0$ Sai |
| Chấp nhận $H_0$ | Quyết định đúng | Sai lầm loại II |
| Bác bỏ $H_0$ | Sai lầm loại I | Quyết định đúng |

Quan hệ của thực tế và quyết định toán học

# Kiểm định giả thuyết một mẫu

### Các bước làm một bài kiểm định

- **Bước 1**: Gọi biến ngẫu nhiên, xây dựng cặp giả thuyết - đối thuyết
- **Bước 2**: Chọn tiêu chuẩn kiểm định
  Tính giá trị quan sát $k$
- **Bước 3**: Xác định miền bác bỏ $H_0 : W_\alpha$
- **Bước 4**: Kiểm tra xem giá trị quan sát $k \in W_\alpha$ hay không và ra quyết định.





# Kiểm định cho kỳ vọng - $\sigma^2$ đã biết

### Trường hợp 1: $\sigma^2$ đã biết

- Chọn tiêu chuẩn kiểm định: $\boxed{Z = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \sim N(0;1)}$ nếu giả thuyết $H_0$ đúng.
- Từ mẫu cụ thể $(x_1, x_2, .., x_n)$, ta tính được giá trị quan sát: $k = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | $(-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$ |
| $\mu = \mu_0$ | $\mu > \mu_0$ | $(u_{1-\alpha}; +\infty)$ |
| $\mu = \mu_0$ | $\mu < \mu_0$ | $(-\infty; -u_{1-\alpha})$ |

# Kiểm định cho kỳ vọng - $\sigma^2$ đã biết

### Ví dụ

Doanh thu của một cửa hàng là biến ngẫu nhiên $X$(triệu/tháng) có độ lệch chuẩn 2 triệu/tháng. Điều tra ngẫu nhiên doanh thu của 500 cửa hàng có qui mô tương tự nhau ta tính được doanh thu trung bình là 10 triệu/tháng. Có người cho rằng thu nhập trung bình của cửa hàng loại đó phải trên 9 triệu/tháng. Với mức ý nghĩa 5% có thể kết luận gì về nhận xét trên.

### Bài làm

- $X$ là doanh thu của cửa hàng loại đang xét, $EX = \mu$ , $VX = \sigma^2$ với $\sigma = 2$
  Cặp giả thuyết: $H_0 : \mu = \mu_0$ và $H_1 : \mu > \mu_0$ (với $\mu_0 = 9$)
- Chọn tiêu chuẩn kiểm định: $Z = \dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \sim N(0;1)$ nếu $H_0$ đúng
  Giá trị quan sát $k = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \dfrac{10-9}{2}\sqrt{500} = 11,18$
- Với $\alpha = 0,05$, miền bác bỏ $H_0$:
  $W_\alpha = (u_{1-\alpha}; +\infty) = (u_{0,95}; +\infty) = (1,645; +\infty)$
- Do $k \in W_\alpha$ nên ta bác bỏ $H_0$ và chấp nhận $H_1$. Nghĩa là nhận xét đó là đúng





# Kiểm định cho kỳ vọng - $\sigma^2$ chưa biết

### Trường hợp 2: $\sigma^2$ chưa biết

Do $\sigma$ chưa biết nên ta thay thế bằng $s$.

- Chọn tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{\overline{X} - \mu_0}{s}\sqrt{n} \sim t(n-1)}$ nếu giả thuyết $H_0$ đúng.
- Từ mẫu cụ thể $(x_1, x_2, .., x_n)$, ta tính được giá trị quan sát: $k = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | $(-\infty; -t(n-1; 1-\frac{\alpha}{2})) \cup (t(n-1; 1-\frac{\alpha}{2}); +\infty)$ |
| $\mu = \mu_0$ | $\mu > \mu_0$ | $(t(n-1; 1-\alpha); +\infty)$ |
| $\mu = \mu_0$ | $\mu < \mu_0$ | $(-\infty; -t(n-1; 1-\alpha))$ |

# Kiểm định cho kỳ vọng - $\sigma^2$ chưa biết

### Chú ý

Nếu $n > 30$ thì ta có thể chuyển từ tiêu chuẩn kiểm định theo phân phối Student sang phân phối chuẩn, nghĩa là ta có thể dùng :

- Chọn tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{\overline{X} - \mu_0}{s}\sqrt{n} \sim N(0;1)}$ nếu giả thuyết $H_0$ đúng.
- Từ mẫu cụ thể $(x_1, x_2, .., x_n)$, ta tính được giá trị quan sát: $k = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0$ : $W_\alpha$ |
|---|---|---|
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | $(-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$ |
| $\mu = \mu_0$ | $\mu > \mu_0$ | $(u_{1-\alpha}; +\infty)$ |
| $\mu = \mu_0$ | $\mu < \mu_0$ | $(-\infty; -u_{1-\alpha})$ |





# Kiểm định cho kỳ vọng - $\sigma^2$ chưa biết

### Ví dụ: Ví dụ trước sẽ được sửa hợp với thực tế hơn

Doanh thu của một cửa hàng là biến ngẫu nhiên $X$(triệu/tháng). Điều tra ngẫu nhiên doanh thu của 500 cửa hàng có qui mô tương tự nhau ta tính được doanh thu trung bình là 10 triệu/tháng và độ lệch chuẩn mẫu hiệu chỉnh là 2 triệu/tháng. Có người cho rằng thu nhập trung bình của cửa hàng loại đó phải trên 9 triệu/tháng. Với mức ý nghĩa 5% có thể kết luận gì về nhận xét trên.

### Bài làm

- $X$ là doanh thu của cửa hàng loại đang xét, $EX = \mu$ , $VX = \sigma^2$
  Cặp giả thuyết: $H_0 : \mu = \mu_0$ và $H_1 : \mu > \mu_0$ (với $\mu_0 = 9$)
- Chọn tiêu chuẩn kiểm định: $Z = \dfrac{\overline{X} - \mu_0}{s}\sqrt{n} \sim t(n-1)$ nếu $H_0$ đúng
  Giá trị quan sát $k = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n} = \dfrac{10 - 9}{2}\sqrt{500} = 11,18$
- Với $\alpha = 0,05$, miền bác bỏ $H_0$:
  $W_\alpha = (t(n-1; 1-\alpha); +\infty) = (t(499; 0,95); +\infty) = (1,645; +\infty)$
- Do $k \in W_\alpha$ nên ta bác bỏ $H_0$ và chấp nhận $H_1$. Nghĩa là nhận xét đó là đúng

# Kiểm định cho kỳ vọng - $\sigma^2$ chưa biết

### Chú ý

Do $n > 30$ nên ta hoàn toàn có thể chuyển phân phối Student thành phân phối chuẩn. Bài giải có thể làm như sau:

### Bài làm

- $X$ là doanh thu của cửa hàng loại đang xét, $EX = \mu$ , $VX = \sigma^2$
  Cặp giả thuyết: $H_0 : \mu = \mu_0$ và $H_1 : \mu > \mu_0$ (với $\mu_0 = 9$)
- Chọn tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{\overline{X} - \mu_0}{s}\sqrt{n} \sim N(0;1)}$ nếu $H_0$ đúng
  Giá trị quan sát $k = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n} = \dfrac{10 - 9}{2}\sqrt{500} = 11,18$
- Với $\alpha = 0,05$, miền bác bỏ $H_0$:
  $\boxed{W_\alpha = (u_{1-\alpha}; +\infty) = (u_{0,95}; +\infty) = (1,645; +\infty)}$
- Do $k \in W_\alpha$ nên ta bác bỏ $H_0$ và chấp nhận $H_1$. Nghĩa là nhận xét đó là đúng





# Kiểm định cho tỷ lệ

### Bài toán

Xác suất xảy ra sự kiện $A$ là $p$.
Do không biết $p$ nên người ta thực hiện $n$ phép thử độc lập, cùng điều kiện.
Trong đó có $m$ phép thử xảy ra $A$.
$f = {}^{m}/_{n}$ là ước lượng điểm không chệch cho $p$.
**Câu hỏi**: Hãy so sánh $p$ với giá trị $p_0$ cho trước.

### Cách giải quyết: tương tự cách làm cho kỳ vọng

- Bài toán đặt ra là ta cần so sánh $p$ với giá trị $p_0$ cho trước.

| Giả thuyết $H_0$ | $p = p_0$ | $p \leq p_0$ | $p \geq p_0$ |
|---|---|---|---|
| Đối thuyết $H_1$ | $p \neq p_0$ | $p > p_0$ | $p < p_0$ |

- Tuy nhiên do giả thuyết luôn có dấu "=" nên người ta chỉ cần viết giả thuyết $H_0 : p = p_0$

# Kiểm định giả thuyết cho tỷ lệ

### Cách giải quyết

- Cách xử lý tương tự như với kỳ vọng
- Chọn tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \sim N(0;1)}$ nếu giả thuyết $H_0$ đúng.
- Từ mẫu thu thập, ta tính được giá trị quan sát: $k = Z = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ với $f = \dfrac{m}{n}$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $p = p_0$ | $p \neq p_0$ | $(-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$ |
| $p = p_0$ | $p > p_0$ | $(u_{1-\alpha}; +\infty)$ |
| $p = p_0$ | $p < p_0$ | $(-\infty; -u_{1-\alpha})$ |





# kiểm định cho tỷ lệ

### Ví dụ

Tại một bến xe, kiểm tra ngẫu nhiên 100 xe thấy có 35 xe xuất phát đúng giờ. Với mức ý nghĩa 5% có thể khẳng định được rằng tỷ lệ xe xuất phát đúng giờ thấp hơn 40% hay không?

### Bài làm

- Gọi $p$ là tỷ lệ xe xuất phát đúng giờ.
  Cặp giả thuyết: $H_0 : p = p_0$ và $H_1 : p < p_0$ (với $p_0 = 0,4$)
- Tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \sim N(0;1)}$ nếu giả thuyết $H_0$ đúng.

  Giá trị quan sát $k = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \dfrac{35/100 - 0,4}{\sqrt{0,4.0,6}}\sqrt{100} = -1,02$
- Với $\alpha = 0,05$, miền bác bỏ $H_0$:
  $\boxed{W_\alpha = (-\infty; -u_{1-\alpha}) = (-\infty; -u_{0,95}) = (-\infty; -1,645)}$
- Do $k \notin W_\alpha$ nên ta không có cơ sở bác bỏ $H_0$. Nghĩa là không thể khẳng định.

# Kiểm định cho phương sai

## Bài toán

Cho biến ngẫu nhiên $X$ có $EX = \mu, VX = \sigma^2$.
Mẫu cụ thể của $X$ là $(x_1, x_2, ..., x_n)$
**Chú ý**: nếu cỡ mẫu $n \leq 30$ thì ta phải thêm điều kiện $X \sim N(\mu, \sigma^2)$.
**Câu hỏi**: Hãy so sánh $\sigma^2$ với giá trị $\sigma_0^2$ cho trước.

## Cách giải quyết

- Bài toán đặt ra là ta cần so sánh $\sigma^2$ với giá trị $\sigma_0^2$ cho trước.

| Giả thuyết $H_0$ | $\sigma^2 = \sigma_0^2$ | $\sigma^2 \leq \sigma_0^2$ | $\sigma^2 \geq \sigma_0^2$ |
|---|---|---|---|
| Đối thuyết $H_1$ | $\sigma^2 \neq \sigma_0^2$ | $\sigma^2 > \sigma_0^2$ | $\sigma^2 < \sigma_0^2$ |

- Tuy nhiên do giả thuyết luôn có dấu "=" nên người ta chỉ cần viết giả thuyết $H_0 : \sigma^2 = \sigma_0^2$





# Kiểm định cho phương sai

## Cách làm

- Tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1)}$ nếu giả thuyết $H_0$ đúng.
- Từ mẫu cụ thể $(x_1, x_2, .., x_n)$, ta tính được giá trị quan sát: $k = \dfrac{(n-1)s^2}{\sigma_0^2}$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 \neq \sigma_0^2$ | $(0; \chi^2_{n-1;\frac{\alpha}{2}}) \cup (\chi^2_{n-1;1-\frac{\alpha}{2}}; +\infty)$ |
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 > \sigma_0^2$ | $(\chi^2_{n-1;1-\alpha}; +\infty)$ |
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 < \sigma_0^2$ | $(-\infty; \chi^2_{n-1;\alpha})$ |

# kiểm định cho phương sai

**Ví dụ**

Đo đường kính 12 sản phẩm của một dây chuyền sản xuất, người kỹ sư kiểm tra chất lượng 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 kỹ sư có kết luận gì?

**Bài làm**:

- $X$ là đường kính sản phẩm, $EX = \mu$ , $VX = \sigma^2$
  Cặp giả thuyết: $H_0 : \sigma^2 = \sigma_0^2$ và $H_1 : \sigma^2 > \sigma_0^2$ (với $\sigma_0 = 0,2$)
- Chọn tiêu chuẩn kiểm định: $\boxed{Z = \dfrac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1)}$ nếu $H_0$ đúng
  Giá trị quan sát $k = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{11.0,3^2}{0,2^2} = 24,75$
- Với $\alpha = 0,05$, miền bác bỏ $H_0$:
  $\boxed{W_\alpha = (\chi^2_{n-1;1-\alpha}; +\infty) = (\chi^2_{11;0,95}; +\infty) = (19,6752; +\infty)}$
- Do $k \in W_\alpha$ nên ta bác bỏ $H_0$ và chấp nhận $H_1$. Nghĩa là dây chuyền cần điều chỉnh vì độ biến động lớn hơn mức cho phép.

# Kiểm định giả thuyết cho kỳ vọng

## Bài toán

- Cho hai biến ngẫu nhiên $X$ có $EX = \mu_1, VX = \sigma_1^2$ và $Y$ có $EY = \mu_2, VY = \sigma_2^2$. Mẫu cụ thể của $X$ là $(x_1, x_2, ..., x_{n_1})$, của $Y$ là $(y_1, y_2, ..., y_{n_2})$.
  **Chú ý**: Nếu cỡ mẫu nhỏ thì ta phải thêm giả thuyết biến ngẫu nhiên gốc tuân theo phân phối CHUẨN.
- Bài toán đặt ra là ta cần so sánh giá trị kỳ vọng $\mu_1$ với $\mu_2$.

| Giả thuyết $H_0$ | $\mu_1 = \mu_2$ | $\mu_1 \leq \mu_2$ | $\mu_1 \geq \mu_2$ |
|---|---|---|---|
| Đối thuyết $H_1$ | $\mu_1 \neq \mu_2$ | $\mu_1 > \mu_2$ | $\mu_1 < \mu_2$ |

- Tuy nhiên do giả thuyết luôn có dấu "=" nên người ta chỉ cần viết giả thuyết $H_0 : \mu_1 = \mu_2$





# Kiểm định cho kỳ vọng - $\sigma_1^2, \sigma_2^2$ đã biết

## Trường hợp 1: $\sigma_1^2, \sigma_2^2$ đã biết

- Chọn tiêu chuẩn kiểm định:
  $\boxed{Z = \dfrac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0;1)}$ nếu giả thuyết $H_0$ đúng thì $\mu_1 - \mu_2 = 0$.
- Từ mẫu cụ thể $(x_1, x_2, .., x_{n_1}), (y_1, y_2, ..., y_{n_2})$, ta tính được giá trị quan sát:
  $$k = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\mu_1 = \mu_2$ | $\mu_1 \neq \mu_2$ | $(-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$ |
| $\mu_1 = \mu_2$ | $\mu_1 > \mu_2$ | $(u_{1-\alpha}; +\infty)$ |
| $\mu_1 = \mu_2$ | $\mu_1 < \mu_2$ | $(-\infty; -u_{1-\alpha})$ |

# Kiểm định cho kỳ vọng - $\sigma_1^2, \sigma_2^2$ chưa biết

**Trường hợp 2: $\sigma_1^2, \sigma_2^2$ chưa biết**

- Chọn tiêu chuẩn kiểm định:
$$\boxed{Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1+n_2-2)}$$
nếu giả thuyết $H_0$ đúng thì $\mu_1 - \mu_2 = 0$.
- Từ mẫu cụ thể $(x_1, x_2, .., x_{n_1}), (y_1, y_2, ..., y_{n_2})$, ta tính được giá trị quan sát:
$$k = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\mu_1 = \mu_2$ | $\mu_1 \neq \mu_2$ | $(-\infty; -t(n_1+n_2-2; 1-\frac{\alpha}{2})) \cup (t(n_1+n_2-2; 1-\frac{\alpha}{2}); +\infty)$ |
| $\mu_1 = \mu_2$ | $\mu_1 > \mu_2$ | $(t(n_1+n_2-2; 1-\alpha); +\infty)$ |
| [illegible] | [illegible] | [illegible] |





# Kiểm định cho kỳ vọng - $\sigma_1^2, \sigma_2^2$ chưa biết

**Chú ý: $\sigma_1^2, \sigma_2^2$ chưa biết, $n_1, n_2$ lớn**

- Chọn tiêu chuẩn kiểm định:
$$\boxed{Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0; 1)}$$
nếu giả thuyết $H_0$ đúng thì $\mu_1 - \mu_2 = 0$.
- Từ mẫu cụ thể $(x_1, x_2, .., x_{n_1}), (y_1, y_2, ..., y_{n_2})$, ta tính được giá trị quan sát:
$$k = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\mu_1 = \mu_2$ | $\mu_1 \neq \mu_2$ | $(-\infty; -u_{1-\frac{\alpha}{2}}) \cup u_{1-\frac{\alpha}{2}}; +\infty)$ |
| $\mu_1 = \mu_2$ | $\mu_1 > \mu_2$ | $(u_{1-\alpha}; +\infty)$ |
| $\mu_1 = \mu_2$ | $\mu_1 < \mu_2$ | $(-\infty; -u_{1-\alpha})$ |

# Kiểm định 2 mẫu cho kỳ vọng

**Ví dụ**

Khảo sát điểm thi môn Xác suất thống kê của sinh viên 2 lớp A, B ta có kết quả:
•Trường A: $n = 64, \overline{x} = 7,32, s_1 = 1,09$
•Trường B: $n = 68, \overline{x} = 7,66, s_1 = 1,12$
Với mức ý nghĩa 1% có thể kết luận rằng kết quả thi của lớp B cao hơn của lớp A hay không?





# Kiểm định 2 mẫu cho kỳ vọng

**Bài làm**

- Gọi $X, Y$ là điểm thi môn XSTK của lớp A, B tương ứng.
  $EX = \mu_1, VX = \sigma_1^2$ và $EY = \mu_2, VX = \sigma_2^2$
  Cặp giả thuyết: $H_0 : \mu_1 = \mu_2$ và $H_1 : \mu_1 < \mu_2$
- Chọn tiêu chuẩn kiểm định: $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim N(0;1)$ nếu $H_0$ đúng.
  Giá trị quan sát $k = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{7,32 - 7,66}{\sqrt{\dfrac{1,09^2}{64} + \dfrac{1,12^2}{68}}} = -31,43$
- Với $\alpha = 0,01$, miền bác bỏ $H_0$:
  $\boxed{W_\alpha = (-\infty; -u_{1-\alpha}) = (-\infty; -u_{0,99}) = (-\infty; -2,33)}$
- Do $k \in W_\alpha$ nên ta bác bỏ $H_0$ và chấp nhận $H_1$. Nghĩa là kết luận là đúng

# Kiểm định 2 mẫu cho tỷ lệ

## Bài toán

Giả sử $p_1, p_2$ tương ứng là tỷ lệ các phần tử mang dấu hiệu A nào đó của tổng thể thứ nhất và tổng thể thứ hai.
Mẫu của tổng thể thứ nhất: Thực hiện $n_1$ phép thử độc lập cùng điều kiện, có $m_1$ phép thử xảy ra sự kiện A.
Mẫu của tổng thể thứ hai: Thực hiện $n_2$ phép thử độc lập cùng điều kiện, có $m_2$ phép thử xảy ra sự kiện A.
**Câu hỏi**: Hãy so sánh $p_1$ với $p_2$.

## Cách giải quyết

- Bài toán đặt ra là ta cần so sánh $p_1$ và $p_2$.

| Giả thuyết $H_0$ | $p_1 = p_2$ | $p_1 \leq p_2$ | $p_1 \geq p_2$ |
|---|---|---|---|
| Đối thuyết $H_1$ | $p_1 \neq p_2$ | $p_1 > p_2$ | $p_1 < p_2$ |

- Tuy nhiên do giả thuyết luôn có dấu "=" nên người ta chỉ cần viết giả thuyết $H_0 : p_1 = p_2$





# Kiểm định giả thuyết 2 mẫu cho tỷ lệ

## Cách giải quyết

- Chọn tiêu chuẩn kiểm định:
$$\boxed{Z = \frac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})(\frac{1}{n_1}+\frac{1}{n_2})}} \sim N(0;1)}$$ nếu giả thuyết $H_0$ đúng.
- Từ mẫu thu thập, ta tính được giá trị quan sát: $k = \dfrac{f_1 - f_2}{\sqrt{\overline{f}(1-\overline{f})(\frac{1}{n_1}+\frac{1}{n_2})}}$

với $f_1 = \dfrac{m_1}{n_1}, f_2 = \dfrac{m_2}{n_2}, \overline{f} = \dfrac{m_1+m_2}{n_1+n_2} = \dfrac{n_1.f_1 + n_2.f_2}{n_1+n_2}$

- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $p_1 = p_2$ | $p_1 \neq p_2$ | $(-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$ |
| $p_1 = p_2$ | $p_1 > p_2$ | $(u_{1-\alpha}; +\infty)$ |
| $p_1 = p_2$ | $p_1 < p_2$ | $(-\infty; -u_{1-\alpha})$ |

# Kiểm định giả thuyết 2 mẫu cho tỷ lệ

## Ví dụ

Kiểm tra các sản phẩm được chọn ngẫu nhiên của 2 nhà máy sản xuất ta được số liệu sau:
• Nhà máy thứ nhất: kiểm tra 100 sản phẩm có 20 phế phẩm.
• Nhà máy thứ hai : kiểm tra 120 sản phẩm có 36 phế phẩm.
Với mức ý nghĩa $\alpha = 0,05$ có thể coi tỷ lệ phế phẩm của nhà máy thứ ai cao hơn của nhà máy thứ nhất hay không?

**Bài làm**:
Gọi $p_1, p_2$ lần lượt là tỷ lệ phế phẩm của nhà máy thứ nhất và thứ hai.
$n_1 = 100, m_1 = 20$ và $n_2 = 120, m_2 = 36$.
• Cặp giả thuyết: $H_0 : p_1 = p_2$ , $H_1 : p_1 < p_2$
• Với $f_1 = \dfrac{m_1}{n_1} = 0,2; f_2 = \dfrac{m_2}{n_2} = 0,3; \overline{f} = \dfrac{m_1 + m_2}{n_1 + n_2} = 0,227$ Giá trị quan sát

$$k = \frac{f_1 - f_2}{\sqrt{\overline{f}(1 - \overline{f})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0,2 - 0,3}{\sqrt{0,227(1 - 0,227)(\frac{1}{100} + \frac{1}{120})}} = 1,763$$

• Với $\alpha = 0,05$ ta có miền bác bỏ $H_0$ :
$W_\alpha = (-\infty; -u_{1-\alpha}) = (-\infty; -u_{0,95}) = (-\infty; -1,645)$
• Do $k \in W_\alpha$ nên ta bác bỏ $H_0$, chấp nhận $H_1$.





# Kiểm định 2 mẫu cho phương sai

## Bài toán

- Cho hai biến ngẫu nhiên $X$ có $EX = \mu_1, VX = \sigma_1^2$ và $Y$ có $EY = \mu_2, VY = \sigma_2^2$. Mẫu cụ thể của $X$ là $(x_1, x_2, ..., x_{n_1})$, của $Y$ là $(y_1, y_2, ..., y_{n_2})$.
- Bài toán đặt ra là ta cần so sánh giá trị kỳ vọng $\sigma_1^2$ với $\sigma_2^2$.

| | | | |
|---|---|---|---|
| Giả thuyết $H_0$ | $\sigma_1^2 = \sigma_2^2$ | $\sigma_1^2 \le \sigma_2^2$ | $\sigma_1^2 \ge \sigma_2^2$ |
| Đối thuyết $H_1$ | $\sigma_1^2 \ne \sigma_2^2$ | $\sigma_1^2 > \sigma_2^2$ | $\sigma_1^2 < \sigma_2^2$ |

- Tuy nhiên do giả thuyết luôn có dấu "=" nên người ta chỉ cần viết giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$

# Kiểm định 2 mẫu cho phương sai

### Cách làm

- Tiêu chuẩn kiểm định: $\boxed{K = \dfrac{s_1^2.\sigma_2^2}{s_2^2.\sigma_1^2}}$

  nếu giả thuyết $H_0$ đúng ta có $K \sim F(n_1 - 1, n_2 - 1)$.
- Từ mẫu cụ thể $(x_1, x_2, .., x_{n_1}), (y_1, y_2, ..., y_{n_2})$, suy ra giá trị quan sát: $k = \dfrac{s_1^2}{s_2^2}$
- Miền bác bỏ $H_0$ được xác định cho 3 trường hợp như sau:

| $H_0$ | $H_1$ | Miền bác bỏ $H_0 : W_\alpha$ |
|---|---|---|
| $\sigma_1^2 = \sigma_2^2$ | $\sigma_1^2 \neq \sigma_2^2$ | $(0; F(n_1-1; n_2-1; \frac{\alpha}{2})) \cup (F(n_1-1; n_2-1; 1-\frac{\alpha}{2}); +\infty)$ |
| $\sigma_1^2 = \sigma_2^2$ | $\sigma_1^2 > \sigma_2^2$ | $(F(n_1-1; n_2-1; 1-\alpha); +\infty)$ |
| $\sigma_1^2 = \sigma_2^2$ | $\sigma_1^2 < \sigma_2^2$ | $(0; F(n_1-1; n_2-1; \alpha))$ |

**Chú ý**: $F(n_1 - 1; n_2 - 1; p) = \dfrac{1}{F(n_1 - 1; n_2 - 1; 1 - p)}$





# kiểm định cho phương sai

### Ví dụ

Hai máy A, B cùng gia công một loại chi tiết máy. Người ta muốn kiểm tra xem hai máy có độ chính xác như nhau hay không. Để làm điều đó người ta tiến hành lấy mẫu và thu được kết quả sau:
Máy A: 135 138 136 140 138 135 139
Máy B: 140 135 140 138 135 138 140
Với mức ý nghĩa 5% hãy kiểm tra xem 2 máy có độ chính xác như nhau hay không? Biết rằng kích thước của chi tiết do máy làm ra tuân theo phân phối chuẩn.

# kiểm định cho phương sai

**Ví dụ**

- Gọi $X, Y$ là đường kính chi tiết do máy A và B làm ra
  $X \sim N(\mu_1; \sigma_1^2)$ và $Y \sim N(\mu_2; \sigma_2^2)$
  Cặp giả thuyết: $H_0 : \sigma_1^2 = \sigma_2^2$ và $H_1 : \sigma_1^2 \neq \sigma_2^2$
- Chọn tiêu chuẩn kiểm định: $K = \dfrac{s_1^2}{s_2^2} \sim F(n_1 - 1; n_2 - 1)$ nếu $H_0$ đúng
  Với mẫu số liệu ta có $s_1^2 = 3,905; s_2^2 = 5$
  Giá trị quan sát $k = \dfrac{s_1^2}{s_2^2} = \dfrac{3,905}{5} = 0,781$
- Với $\alpha = 0,05$, miền bác bỏ $H_0$:
  $W_\alpha = (-\infty; F(n_1 - 1; n_2 - 1; \frac{\alpha}{2})) \cup (F(n_1 - 1; n_2 - 1; 1 - \frac{\alpha}{2}); +\infty)$
  Với mức ý nghĩa $\alpha = 0,05$ , $n_1 = n_2 = 7$ ta có $F(6; 6; 0,025) = 0,17$ và $F(6; 6; 0,975) = 5,82$
  $W_\alpha = (0; 0,17) \cup (5,82; +\infty)$
- Do $k \notin W_\alpha$ nên ta chấp nhận $H_0$. Nghĩa là độ chính xác của 2 máy là như nhau.