Chương 2

PHÂN PHỐI XÁC SUẤT
VÀ THỐNG KÊ TOÁN

Nội dung

2.1. Biến ngẫu nhiên

2.2. Quy luật phân phối xác suất

2.3. Tham số đặc trưng cho biến ngẫu nhiên

2.3.1.Tham số đặc trưng cho xu hướng trung tâm

2.3.2. Tham số đặc trưng cho độ phân tán

2.3.3. Tham số đặc trưng cho dạng phân phối xác suất

2.4. Tham số đặc trưng cho hệ hai biến ngẫu nhiên

2.5. Các dạng phân phối xác suất thông dụng

2.6. Ước lượng thống kê

2.7. Kiểm định giả thuyết thống kê

2.1. Biến ngẫu nhiên

- Việc thực hiện một nhóm các điều kiện cơ bản để quan sát một hiện tượng nào đó có xảy ra hay không được gọi là thực hiện một **phép thử**, còn hiện tượng có thể xảy ra trong **kết quả của phép thử** đó được gọi là **biến cố**.
- **Ví dụ:** Gieo con súc sắc đồng chất trên mặt phẳng (phép thử). Kết quả số chấm có thể xuất hiện là biến cố (tất yếu, bất khả, **ngẫu nhiên**).
- Xác suất của một biến cố là con số đặc trưng khả năng khách quan xuất hiện biến cố đó khi thực hiện phép thử.
- **"Xác suất xuất hiện biến cố A trong một phép thử là tỷ số giữa số kết cục thuận lợi cho A và tổng số các kết cục đồng khả năng có thể xảy ra khi thực hiện phép thử đó".**
- **"Xác suất của biến cố là giới hạn của tần suất xuất hiện biến cố đó khi số phép thử tăng lên vô hạn".**
- Ký hiệu xác suất xảy ra biến cố A là $P(A) \approx f(A)$ $\mathbf{0 \le P(A) \le 1}$

2.1. Biến ngẫu nhiên

- **"Một biến số được gọi là ngẫu nhiên nếu trong kết quả phép thử nó sẽ nhận một và chỉ một trong các giá trị có thể có của nó tùy thuộc vào sự tác động của các nhân tố ngẫu nhiên".**
- Biến ngẫu nhiên là một đại lượng phụ thuộc vào kết quả của phép thử ngẫu nhiên nào đó.
- **Ví dụ 1:** Gieo con xúc sắc. Gọi biến ngẫu nhiên là **số chấm xuất hiện**. Biến ngẫu nhiên này phụ thuộc kết quả phép thử và có thể nhận 1 giá trị nguyên từ 1-6
- **Ví dụ 2:** Biến ngẫu nhiên **nhiệt độ** của một phản ứng hóa học trong một khoảng thời gian nào đó. Biến ngẫu nhiên này nhận giá trị trong khoảng [t^0min-t^0max].
- Các biến ngẫu nhiên thường được ký hiệu bằng các chữ lớn X, Y, Z,… còn các giá trị của chúng được ký hiệu bằng các chữ nhỏ x, y, z...

2.1. Biến ngẫu nhiên

Phân loại

- **Biến ngẫu nhiên rời rạc** (Discrete Random Variable)
  - X là một biến ngẫu nhiên rời rạc nếu giá trị có thể có của X lập nên một tập hữu hạn hoặc có thể đếm được.
  - Biến ngẫu nhiên rời rạc có thể liệt kê được tất cả các giá trị có thể có của biến.
  - **Ví dụ 1**: Gọi **X** là **Số điểm thu được** khi tung xúc sắc. **X** là **biến ngẫu nhiên rời rạc** vì các giá trị có thể có của nó là một tập hữu hạn X = 1,2,3,4,5,6.
  - **Ví dụ 2**: Một phân xưởng có 5 máy phát. Gọi X là **Số máy hỏng trong một ca**. X là biến ngẫu nhiên rời rạc vì các giá trị có thể có của X = 0,1,2,3,4,5.
  - **Ví dụ 3**: Gọi X là **Số người vào siêu thị trong một ngày**. X là biến ngẫu nhiên rời rạc vì các giá trị có thể có của X lập nên một tập hợp có thể đếm được X = 0,1,2,3...

2.1. Biến ngẫu nhiên

Phân loại

- **Biến ngẫu nhiên liên tục** (Continuous Random Variable)
  - X là một biến ngẫu nhiên liên tục nếu giá trị có thể có của X có thể lấp đầy một khoảng trên trục số.
  - Biến ngẫu nhiên liên tục không thể liệt kê được tất cả các giá trị có thể có của biến đó.
  - **Ví dụ 1:** Phép thử là bắn vào bia. Gọi X là **Khoảng cách từ điểm chạm của viên đạn đến tâm bia**. X là biến ngẫu nhiên liên tục vì các giá trị có thể có của X lấp đầy một khoảng trên trục số và không thể kể ra tất cả các giá trị có thể có của X. Chỉ có thể nói X nằm trong khoảng (a,b) nào đó.
  - **Ví dụ 2:** Gọi X là **Năng suất lúa vụ mùa của tỉnh**. X là biến ngẫu nhiên liên tục.
  - **Ví dụ 3:** Gọi X là **Độ dài chi tiết máy được sản xuất ra**. X là biến ngẫu nhiên liên tục.

2.2. Quy luật phân phối xác suất

Định nghĩa

- **Quy luật phân phối xác suất của biến ngẫu nhiên là sự tương ứng giữa các giá trị có thể có của biến ngẫu nhiên và các xác suất tương ứng với các giá trị đó.**
- Có 3 phương pháp mô tả quy luật phân phối xác suất thông dụng của biến ngẫu nhiên: **Bảng phân phối xác suất, hàm phân phối xác suất và hàm mật độ xác suất.**
- **Bảng phân phối xác suất:**
  - Bảng phân phối xác suất chỉ dùng để mô tả quy luật phân phối xác suất của các biến ngẫu nhiên rời rạc.
  - Giả sử biến ngẫu nhiên rời rạc **X** nhận các giá trị có thể có $\mathbf{x_1, x_2 \ldots x_n}$ với các xác suất tương ứng $\mathbf{p_1, p_2 \ldots p_n}$
  - Bảng phân phối xác suất của biến ngẫu nhiên rời rạc X có dạng

X	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_i}$	$\mathbf{\ldots x_n}$	$0 \le \mathbf{p_i} \le 1$
P	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_i}$	$\mathbf{\ldots p_n}$	$\sum \mathbf{p_i} = 1$

2.2. Quy luật phân phối xác suất

Ví dụ bảng phân phối xác suất

- **Ví dụ 1:** Tung xúc sắc. Gọi X là "Số chấm xuất hiện". Hãy tìm quy luật phân phối xác suất của X?
- **Giải:** Vì X là biến ngẫu nhiên rời rạc với cá giá trị có thể có

  X = 1,2,3,4,5,6 với xác suất tương ứng đều bằng 1/6

  Bảng quy luật phân phối xác suất của X có dạng:

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

- **Ví dụ 2:** Trong hộp có 10 sản phẩm trong đó có 6 chính phẩm. Lấy ngẫu nhiên 2 sản phẩm. Tìm quy luật phân phối xác suất của số chính phẩm được lấy ra?
- **Giải:**

  Gọi X là Số chính phẩm lấy ra trong 2 sản phẩm, X là biến ngẫu nhiên rời rạc có các giá trị có thể X = 0,1,2.

  Cần tìm các xác suất tương ứng với các giá trị X có thể có.

2.2. Quy luật phân phối xác suất

Ví dụ bảng phân phối xác suất

- Số chính phẩm là 6 và số phế phẩm là 4.
- Xác xuất P(X=0) Xác suất không lấy được chính phẩm nào (2 phế phẩm). Xác suất xảy ra:
  - P(X=0) = $\mathbf{C_4^2/C_{10}^2 = 6/45 = 2/15}$
- Xác xuất P(X=1) Xác suất lấy được 1 chính phẩm nào đó (1 chính phẩm và 1 phế phẩm). Xác suất xảy ra:
  - P(X=1) = $\mathbf{C_6^1C_4^1/C_{10}^2 = 24/45 = 8/15}$
- Xác xuất P(X=2) Xác suất lấy được 2 chính phẩm nào đó (0 phế phẩm). Xác suất xảy ra:
  - P(X=2) = $\mathbf{C_6^2/C_{10}^2 = 15/45 = 5/15}$
- Bảng luật phân phối xác suất:

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

2.2. Quy luật phân phối xác suất

Ví dụ bảng phân phối xác suất

- **Ví dụ 3:** Một xạ thủ bắn 3 phát, xác suất bắn trúng mục tiêu của mỗi phát là 0.6. Hãy lập bảng phân phối xác suất của số đạn trúng mục tiêu?
- **Giải:** Gọi X là số đạn bắn trúng mục tiêu, các giá trị có thể có của X= 0,1,2,3. Tìm xác suất tương ứng với các giá trị có thể có của X.

Xác suất $P(X= x) = p(x) = \mathbf{C_n^x p^x q^{n-x}}$ Với **(n=3, p=0.6)**

- **Bảng phân phối xác suất:**

X	0	1	2	3
P	0.064	0.288	0.432	0.216

2.2. Quy luật phân phối xác suất

Hàm phân phối xác suất

- **Hàm phân phối xác suất của biến ngẫu nhiên X, ký hiệu F(x), là xác suất để biến X nhận giá trị nhỏ hơn x, với x là số thực bất kỳ.**
- Hàm phân phối xác suất áp dụng được đối với cả biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

$$\mathbf{F(x) = P(X<x)}$$

- Đối với biến ngẫu nhiên rời rạc, hàm phân phối xác suất: $\mathbf{F(x)=\sum_{x_i<x} P_i}$
- **F(x) phản ánh độ tập trung xác suất ở bên trái một số thực (x)**

- **Ví dụ 1:** Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

X	1	3	4
P	0.1	0.5	0.4

Hãy tìm hàm phân phối xác suất của X?

- **Giải:**

Nếu $x \le 1$ biến cố (X<x) là biến cố không thể có, do đó **F(x) = 0**

Nếu $1<x \le 3$ biến cố (X<x) chỉ xảy ra khi x = 1, do đó **F(x) = 0.1**

Nếu $3<x\le 4$ biến cố (X<x) sẽ xảy ra khi x =1 hoặc khi x =3, do đó **F(x) = 0.1+0.5=0.6**

Nếu x>4 biến cố (X<x) sẽ xảy ra khi x =1 hoặc x = 3 hoặc x = 4, do đó **F(x) = 0.1+0.5+0.4 = 1**

2.2. Quy luật phân phối xác suất

Hàm phân phối xác suất

- $F(x) = \begin{cases} 0 \text{ với } x \le 1 \\ 0.1 \text{ với } 1 < x \le 3 \\ 0.6 \text{ với } 3 < x \le 4 \\ 1 \text{ với } x > 4 \end{cases}$
- **Tính chất của Hàm phân phối xác suất**
  - $0 \le F(x) \le 1$
  - $x_2 > x_1$ thì $F(x_2) > F(x_1)$
  - $F(-\infty) = 0 \qquad F(\infty) = 1$
- **Hệ quả của Hàm phân phối xác suất:**
  - $P(a \le x \le b) = F(b) - F(a)$
  - Nếu X là biến ngẫu nhiên liên tục

    $P(X = x) = 0$

    $P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b)$

2.2. Quy luật phân phối xác suất

Hàm mật độ xác suất

- Hàm phân phối xác suất không thể đặc trưng cho xác suất để biến ngẫu nhiên liên tục nhận một giá trị xác định và khó xác định.
- **Hàm mật độ xác suất của biến ngẫu nhiên X, ký hiệu là f(x), được xác định theo biểu thức: f(x) = F'(x)**
- Hàm mật độ xác suất chỉ áp dụng với biến ngẫu nhiên liên tục.

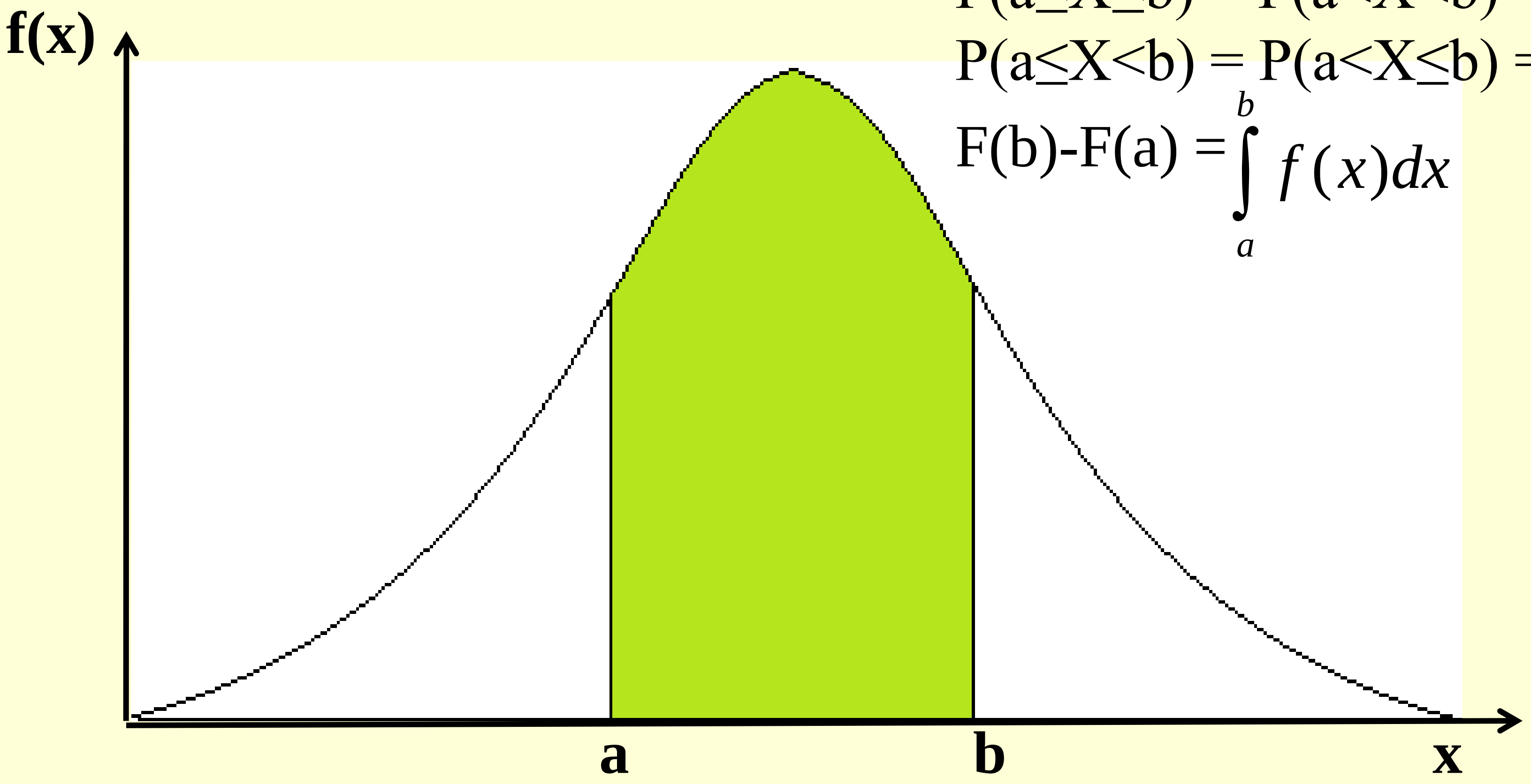
$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b) = \int_{a}^{b} f(x)dx$$

- **Hàm mật độ xác suất của biến ngẫu nhiên (X) tại mỗi điểm (x) cho biết mức độ tập trung xác suất tại điểm đó.**

2.2. Quy luật phân phối xác suất

Hàm mật độ xác suất



$P(a \le X \le b) = P(a < X < b) =$
$P(a \le X < b) = P(a < X \le b) =$
$F(b) - F(a) = \int_a^b f(x)dx$

2.3. Các tham số đặc trưng của biến ngẫu nhiên

- **Các tham số đặc trưng của biến ngẫu nhiên được chia thành 3 loại:**
  - Các tham số đặc trưng cho xu hướng trung tâm
  - Các tham số đặc trưng cho độ phân tán của biến ngẫu nhiên
  - Các tham số đặc trưng cho dạng phân phối xác suất
- **Các tham số đặc trưng cho xu hướng trung tâm**
  - Kỳ vọng toán, Trung vị, Mốt
- **Các tham số đặc trưng cho độ phân tán của biến ngẫu nhiên**
  - Phương sai, Độ lệch chuẩn, Hệ số biến thiên...
- **Các tham số đặc trưng cho dạng phân phối xác suất**
  - Hệ số bất đối xứng
  - Hệ số nhọn

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Định nghĩa:** Cho X là 1 biến ngẫu nhiên, giá trị trung bình hay kỳ vọng toán học (gọi tắt là kỳ vọng) của X được ký hiệu là E(X) và được tính theo công thức:

$$E(X) = \mu = \sum_{i=1}^{n} x_i p_i \quad \textbf{Biến rời rạc} \qquad E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \textbf{Biến liên tục}$$

- ***Chú ý***: Nếu mẫu ngẫu nhiên cho dưới dạng tần suất

X	**X_1**	**X_2**	**X_5**	**…**	**X_k**
n_i	**n_1**	**n_2**	**n_3**	**…**	**n_k**

thì trung bình mẫu được tính:

$$\overline{X} = \frac{n_1 X_1 + n_2 X_2 + n_3 X_3 + \ldots + n_k X_k}{n_1 + n_2 + n_3 + \ldots + n_k} = \frac{\sum_{i=1}^{n} n_i X_i}{\sum_{i=1}^{n} n_i}$$

2.3.1. Tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Các tính chất của kỳ vọng toán:**

1. $E(X + Y) = E(X) + E(Y)$

$E(W + X + Y + Z) = E(W) + E(X) + E(Y) + E(Z)$

2. $E(bX) = bE(X)$ **b: const**

3. $E(b) = b$

4. $E(X.Y) = E(X)*E(Y)$ **X và Y là 2 biến ngẫu nhiên độc lập**

- ***(Hai biến ngẫu nhiên độc lập với nhau nếu quy luật phân phối xác suất của biến ngẫu nhiên này không phụ thuộc gì vào việc biến ngẫu nhiên kia nhận giá trị bao nhiêu).***

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Ví dụ 1**: Cho mẫu quan sát (X_i) với i = 1, 2, ..., 10 của biến ngẫu nhiên X

X	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
n_i	**3**	**7**	**5**	**6**	**5**	**8**	**4**	**2**	**6**	**4**

$$E(X) = \frac{\sum_{i=1}^{n} n_i X_i}{\sum_{i=1}^{n} n_i} = \frac{n_1 X_1 + n_2 X_2 + n_3 X_3 + \dots + n_k X_k}{n_1 + n_2 + n_3 + \dots + n_k}$$

$$E(X) = \frac{1*3 + 2*7 + 3*5 + 4*6 + 5*5 + 6*8 + 7*4 + 8*2 + 9*6 + 10*4}{3+7+5+6+5+8+4+2+6+4} = 5.34$$

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Ví dụ 2:** Một người mua 10nghìn đồng xổ số lôtô 2 số. Anh ta sẽ ta sẽ thắng gấp 70 lần tiền mua nếu trùng với 2 số cuối của giải độc đắc gần nhất sắp tới. Anh ta sẽ không được đồng nào nếu không trùng. Hãy tìm số tiền thắng trung bình của một lần chơi như vậy? Biết thêm rằng xác suất thắng và thua là 1% và 99%. Xác suất trúng tối thiểu là bao nhiêu thì anh ta có cơ hội hòa sau mỗi lần chơi?
- **Giải:** Kỳ vọng số tiền thắng trung bình:

  E(X) = 0đ*99%+700000đ*1% = 7000đ

  Số tiền mất trung bình của một lần chơi:

  10000đ – 7000đ = 3000đ

  E(X) = 0đ*q% + 700000đ*(1-q%) = 10000đ

  q% = 1 – 1/70 = 0.9857 (98.57%)

  p% = 1 – 0.9857 = 0.0143 (1.43%)

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Ví dụ 3:** Một dự án được Viện thiết kế C soạn thảo cho cả 2 bên A và B xét duyệt một cách độc lập. Xác suất để A và B chấp nhận dự án khi xét duyệt là 0.7 và 0.8. Nếu chấp nhận dự án thì A phải trả cho C 4 triệu USD còn ngược lại thì phải trả 1 triệu USD. Với B nếu chấp nhận dự án phải trả cho C là 10 triệu USD, ngược lại phải trả 3 triệu USD. Chi phí cho thiết kế là 10 triệu USD và thuế 10% trên doanh thu. Hỏi C có nên nhận thiết kế hay không?
- **Giải:** Để quyết định xem có nên nhận thiết kế hay không, thì C phải tính số lãi kỳ vọng mà C có thể nhận được từ A và B.
- Gọi X là số lãi mà C có thể nhận được sau khi trừ mọi chi phí và P là xác suất các trường hợp có thể có của X (phụ thuộc quyết định của A và B).

P. án	A_kB_k	AB_k	A_kB	**AB**
X	-6.4	-3.7	-0.1	2.6
P	0.06	0.14	0.24	0.56

- **E(X) = 0.53 > 0   C vẫn có thể chấp nhận thiết kế**

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Ví dụ 4:** Một cửa hàng sách dự định nhập vào một số cuốn niên giám thống kê. Nhu cầu hàng năm về loại sách này cho trong Bảng phân phối xác suất

Nhu cầu j (cuốn)	**20**	**21**	**22**	**23**	**24**	**25**
Xác suất P_j	**0.30**	**0.25**	**0.18**	**0.14**	**0.10**	**0.03**

  Cửa hàng này mua vào với giá 7$/cuốn và bán ra với giá 10$/cuốn, song đến cuối năm thì phải bán hạ giá còn 4$/cuốn trước khi niên giám thống kê năm tới được xuất bản. Cửa hàng muốn xác định số lượng nhập vào sao cho lợi nhuận kỳ vọng là lớn nhất?

- **Giải:** Gọi (i) là số lượng sách cần nhập và (j) là nhu cầu. Lợi nhuận (Pij) sẽ phụ thuộc vào số lượng sách nhập và nhu cầu thực tế về loại sách đó. Có thể xây dựng Bảng liệt kê các kết quả khác có thể có từ những chiến lược nhập hàng khác nhau. **Bảng lợi nhuận có điều kiện**

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- **Lợi nhuận có điều kiện được xác định bằng biểu thức:**
- $P_{ij} = 10.j - 7.i + 4(i-j)$ Với $j \le i$

  $= 10.i - 7.i = 3.i$ Với $j > i$

Nhu cầu	P_j / i	0.3	0.25	0.18	0.14	0.10	0.03
		20	21	22	23	24	25
Lượng Hàng Nhập	20	60	60	60	60	60	60
	21	57	63	63	63	63	63
	22	54	60	66	66	66	66
	23	51	57	63	69	69	69
	24	48	54	60	66	72	72
	25	45	51	57	63	69	75

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Kỳ vọng toán học của biến ngẫu nhiên

- Chiến lược của cửa hàng phải ***chọn số lượng sách cần nhập (i) để cực đại lợi nhuận kỳ vọng***. Với mỗi lượng nhập (i) lợi nhuận kỳ vọng (PE) được tính:

$$\mathbf{PE} = \sum_{j} \mathbf{P_j * P_{ij}}$$

- **Giá trị lợi nhuận kỳ vọng tùy thuộc vào số lượng nhập**

Số lượng nhập (i)	20	**21**	22	23	24	25
LN kỳ vọng PE(i)	60.00	**61.20**	60.90	59.52	57.30	54.48

- **Vậy chiến lược mang lại lợi nhuận kỳ vọng tối đa là nhập 21 cuốn sách**

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

MỐT (Mode) Mo

- **Khái niệm:** Mốt là giá trị của biến ngẫu nhiên tương ứng với
  - Xác suất lớn nhất nếu là biến ngẫu nhiên rời rạc
  - Cực đại của hàm mật độ xác suất nếu là biến ngẫu nhiên liên tục
- Có thể gặp biến ngẫu nhiên không có Mốt hoặc nhiều giá trị Mốt
- Đối với dãy số lượng biến, Mốt là lượng biến có tần số lớn nhất
- **Cách xác định Mốt:**
  - **Không có khoảng cách tổ:** Mốt là lượng biến có tần số lớn nhất
  - **Có khoảng cách tổ:**
    - ***Khoảng cách tổ đều:***
      - *Xác định tổ chứa Mốt: Tổ có tần số lớn nhất*
      - *Xác định giá trị gần đúng của Mốt theo công thức*

$$M_0 = x_{M_0 \min} + h_{M_0} \frac{f_{M_0} - f_{M_{0-1}}}{(f_{M_0} - f_{M_{0-1}}) + (f_{M_0} - f_{M_{0+1}})}$$

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

MỐT (Mode) Mo

- ***Khoảng cách tổ không đều:***
  - *Xác định tổ chứa Mốt: Tổ có mật độ phân phối lớn nhất (Tỷ số giữa tần số và khoảng cách tổ)*
  - *Xác định giá trị gần đúng của Mốt theo công thức*

$$M_0 = x_{M_0 \min} + h_{M_0} \frac{d_{M_0} - d_{M_{0-1}}}{(d_{M_0} - d_{M_{0-1}}) + (d_{M_0} - d_{M_{0+1}})}; \quad d_i = \frac{f_i}{h_i}$$

- **Ví dụ:** Có tài liệu về doanh số bán của 50 trạm xăng dầu thuộc 1 Tỉnh trong tháng 12 như sau. Xác định Mốt về doanh số bán của 50 cửa hàng trên?

Doanh số (triệu đồng)	Số trạm	
200-300	8	
300-400	10	
400-500	**20**	**Tổ chứa Mốt**
500-600	7	
600-700	5	
Tổng	**50**	

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

MỐT (Mode) Mo

$$M_0 = x_{M_0 \min} + h_{M_0} \frac{f_{M_0} - f_{M_{0-1}}}{(f_{M_0} - f_{M_{0-1}}) + (f_{M_0} - f_{M_{0+1}})} = 400 + 100 \frac{20-10}{(20-10)+(20-7)} = 443.48 tr\text{đ}$$

- ***Như vậy đa số các trạm xăng dầu được khảo sát trên có mức doanh số trong tháng 12 khoảng 443.48 triệu đồng***

Ví dụ: Có tài liệu về doanh thu của 79 cửa hàng trong tháng 12 như sau. Hãy xác định Mốt của doanh thu các cửa hàng.

Doanh thu (triệu đồng)	Cửa hàng	Khoảng cách tổ	Mật độ phân phối
200-400	8	200	0.04
400-500	12	100	0.12
500-600	**25**	**100**	**0.25**
600-800	25	200	0.125
800-1000	9	200	0.045
Tổng	**79**		

**** Mật độ phân phối** $d_i = f_i/h_i$*

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

MỐT (Mode) Mo

$$M_0 = x_{M_0 \min} + h_{M_0} \frac{d_{M_0} - d_{M_{0-1}}}{(d_{M_0} - d_{M_{0-1}}) + (d_{M_0} - d_{M_{0+1}})}; \quad d_i = \frac{f_i}{h_i}$$

$$M_0 = 500 + 100 \frac{0.25 - 0.12}{(0.25 - 0.12) + (0.25 - 0.125)} = 550.9tr\text{đ}$$

- ***Như vậy đa số các cửa hàng có mức doanh thu trong tháng 12 khoảng 550.9 triệu đồng.***
- Mốt có ưu điểm không chịu ảnh hưởng của các lượng biến đột biến
- Mốt kém nhạy bén với sự biến thiên của tiêu thức
- Mốt cho biết đa số, khuynh hướng, phong trào.
- Mốt ứng dụng nhiều nhất trong nghiên cứu nhu cầu của thị trường về kích cỡ loại sản phẩm nào đó (quần áo, giày dép...)
- Mốt ứng dụng ít hơn số trung bình và số trung vị

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Số trung vị (Median) Me

- **Số trung vị (Median):** Số trung vị (Me) là giá trị nằm chính giữa tập hợp các giá trị có thể có của biến ngẫu nhiên. Đó là giá trị chia phân phối của biến ngẫu nhiên thành 2 phần bằng nhau.

$\mathbf{F(X_i) \leq 0.5 \leq F(X_{i+1})}$ **X** biến ngẫu nhiên rời rạc

$$\int_{-\infty}^{Me} f(x)dx = 0.5$$

X biến ngẫu nhiên liên tục

- **Xác định Số trung vị khi biến ngẫu nhiên rời rạc cho ở dạng Bảng tần suất**
  - **Không có khoảng cách tổ:** Giá trị của lượng biến ở vị trí (n+1)/2
    - Nếu (n) lẻ thì số trung vị là lượng biến đứng vị trí chính giữa
    - Nếu (n) chẵn thì số trung vị là trung bình hai lượng biến của hai đơn vị đứng giữa
  - **Có khoảng cách tổ:**
    - Xác định tổ chứa trung vị (Tổ đầu tiên có tần số tích lũy tiến lớn hơn hoặc bằng $(\sum f_i+1)/2$
    - Xác định trị số gần đúng của trung vị theo công thức

$$M_e = x_{M_e \min} + h_{M_e} \frac{\sum_{i=1}^{n} f_i / 2 - S_{M_e - 1}}{f_{M_e}}$$

2.3.1 Các tham số đặc trưng cho xu hướng trung tâm

Số trung vị (Median) Me

- **Ví dụ:**

Doanh thu (triệu đồng)	Cửa hàng	Tần số tích lũy
200-400	8	8
400-500	12	20
500-600	25	45
600-800	25	70
800-1000	9	79

Tổ chứa Trung vị là tổ thứ 3 vì Tần số tích lũy > (79+1)/2

$$M_e = x_{M_e \min} + h_{M_e} \frac{\sum_{i=1}^{n} f_i / 2 - S_{M_e - 1}}{f_{M_e}} = 500 + 100 \frac{79/2 - 20}{25} = 578$$

- ***Như vậy là một nửa số cửa hàng có doanh thu dưới 578 triệu đồng và một nửa số cửa hàng có doanh thu trên 578 triệu đồng.***

- Số trung vị biểu hiện mức độ đại biểu của hiện tượng nhưng không san bằng bù trừ chênh lệch gữa các lượng biến. Số trung vị có thể dùng thay thế số trung bình cộng.

2.3.2 Các tham số đặc trưng cho độ phân tán

Phương sai

- **Định nghĩa:** Phương sai của biến ngẫu nhiên X, ký hiệu V(X) được định nghĩa như sau:

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

$$V(X)=\sum_{i=1}^{n}[x_i-E(X)]^2 p_i=\sum_{i=1}^{n}x_i^2 p_i-(\sum_{i=1}^{n}x_i p_i)^2$$ Nếu biến ngẫu nhiên rời rạc

$$V(X)=\int_{-\infty}^{+\infty}[x-E(X)]f(x)dx$$ Nếu biến ngẫu nhiên liên tục

$$\mathbf{V(X) = \sigma^2 = E(X^2) - [E(X)]^2}$$ Công thức hay được dùng

- **Ý nghĩa:** Phương sai đo độ phân tán của các giá trị biến ngẫu nhiên quanh kỳ vọng (giá trị trung bình) của nó.
- **Ứng dụng thực tế:**
  - Trong kỹ thuật phương sai đặc trưng cho **mức độ phân tán** của các kích thước chi tiết gia công, hay sai số của thiết bị
  - Trong quản trị và kinh doanh phương sai đặc trưng cho **mức độ rủi ro** của các quyết định đầu tư.

2.3.2 Các tham số đặc trưng cho độ phân tán

Phương sai

- **Các tính chất của phương sai:**

1. $V(C) = 0$ C: const

2. $V(CX) = C^2V(X)$

3. $V(C+X) = V(X)$

4. $V(X \pm Y) = V(X)+V(Y)$ X, Y là 2 biến ngẫu nhiên độc lập

5. $V(X.Y) = [E(Y)]^2V(X)+[E(X)]^2V(Y)+V(X)V(Y)$

X, Y là 2 biến NN độc lập

6. $V\sum X_i = \sum V(X_i)$ X_i là các biến NN độc lập

7. $V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X,Y)$ X, Y là 2 biến NN phụ thuộc

8. $V(aX \pm bY) = a^2V(X)+b^2V(Y) \pm 2abCov(X,Y)$ X,Y là 2 biến NN phụ thuộc

2.3.2 Các tham số đặc trưng cho độ phân tán

Phương sai

- **Ví dụ:** Một nhà đầu tư đang cân nhắc giữa việc đầu tư vào 2 dự án A và B trong 2 lĩnh vực độc lập với nhau. Khả năng thu hồi vốn sau 2 năm (tính bằng %) của 2 dự án là các biến ngẫu nhiên có Bảng phân phối xác suất như sau. Chọn phương án có tỷ lệ thu hồi vốn đầu tư kỳ vọng cao hơn? Phương án ít rủi ro hơn?

X_A	**65**	**67**	**68**	**69**	**70**	**71**	**73**
P_A	0.04	0.12	0.16	0.28	0.24	0.08	0.08
X_B	**66**	**68**	**69**	**70**	**71**		
P_B	0.12	0.28	0.32	0.20	0.08		

- $E(X_A) = \sum X_A * P_A = 69.16\%$ $\quad V(X_A) = E[(X_A) - E(X_A)]^2 = 3.0944$

- $E(X_B) = \sum X_B * P_B = 68.72\%$ $\quad V(X_B) = E[(X_B) - E(X_B)]^2 = 1.8016$

2.3.2 Các tham số đặc trưng cho độ phân tán

Độ lệch chuẩn, Hệ số biến thiên

- **Độ lệch chuẩn:** Độ lệch chuẩn của biến ngẫu nhiên X, ký hiệu $\boldsymbol{\sigma_X}$ được định nghĩa như sau:

$$\sigma_X = \sqrt{V(X)}$$

- Khi đánh giá mức độ phân tán của biến ngẫu nhiên theo đơn vị đo của biến ngẫu nhiên thường tính độ lệch chuẩn chứ không dùng phương sai (Đơn vị đo của phương sai bằng bình phương đơn vị đo của biến ngẫu nhiên).
- **Hệ số biến thiên:** Hệ số biến thiên, ký hiệu **CV**, được xác định theo công thức:

$$\mathbf{CV} = \left| \boldsymbol{\sigma_X} / \mathbf{E(X)} \right| (\%)$$

- Đo lường mức độ quan trọng tương đối của độ phân tán
- So sánh độ phân tán giữa các hiện tượng có đơn vị tính khác nhau hoặc giữa các hiện tượng cùng loại và có số trung bình bằng nhau

2.3.3 Các tham số đặc trưng cho dạng phân phối xác suất

Hệ số bất đối xứng, Hệ số nhọn

- **Định nghĩa:** Hệ số bất đối xứng, ký hiệu α_3, được xác định bằng công thức: $\boldsymbol{\alpha_3 = \mu_3 / \sigma^3}$

  $$\mu_3 = E[X-E(X)]^3 \text{ và } \sigma^3 = (\sigma_X)^3$$

- Nếu $\boldsymbol{\alpha_3 < 0}$**,** phân phối bất đối xứng, đồ thị xuôi về bên trái nhiều hơn
- Nếu $\boldsymbol{\alpha_3 = 0}$**,** phân phối đối xứng
- Nếu $\boldsymbol{\alpha_3 > 0}$**,** phân phối bất đối xứng, đồ thị xuôi về bên phải nhiều hơn
- **Định nghĩa:** Hệ số nhọn, ký hiệu $\boldsymbol{\alpha_4}$**,** được xác định bằng công thức:

  $$\alpha_4 = \mu_4 / \sigma^4$$

  $$\mu_4 = E[X-E(X)]^4 \text{ và } \sigma^4 = (\sigma_X)^4$$

- Khi phân phối xác suất được tập trung ở mức bình thường $\boldsymbol{\alpha_4 = 3}$**,** nếu tập trung mức cao $\boldsymbol{\alpha_4 > 3}$ còn phân phối tập trung mức thấp $\boldsymbol{\alpha_4 < 3}$

2.4 Các tham số đặc trưng cho hệ hai biến ngẫu nhiên

Hiệp phương sai

- Đối với hệ hai biến ngẫu nhiên, ngoài các tham số đặc trưng là kỳ vọng và phương sai các thành phần còn hai tham số quan trọng là **Hiệp phương sai và Hệ số tương quan.**
- **Hiệp phương sai:** Hiệp phương sai của 2 biến ngẫu nhiên X và Y, ký hiệu Cov(X,Y), được xác định theo công thức:

  Cov(X,Y) = E{[X-E(X)].[Y-E(Y)]}

  Cov(X,Y) = σ_{XY} = E(XY) - μ_X. μ_Y **Với E(X) = μ_X E(Y) = μ_Y**
- Đối với biến ngẫu nhiên rời rạc:

$$Cov(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(x_i, y_j) - E(X)E(Y)$$

- Đối với biến ngẫu nhiên liên tục:

$$Cov(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy - E(X)E(Y)$$

2.4 Các tham số đặc trưng cho hệ hai biến ngẫu nhiên

Hiệp phương sai

- **Một số tính chất của Hiệp phương sai:**
  - Nếu $Y = V + W$, $\mathbf{Cov}(X, Y) = \mathbf{Cov}(X, V) + \mathbf{Cov}(X, W)$
  - Nếu $Y = b$, b là hằng số, $\mathbf{Cov}(X, Y) = \mathbf{Cov}(X, b) = \mathbf{0}$
  - Nếu $Y = bZ$, b là hằng số

    $\mathbf{Cov}(X, Y) = \mathbf{Cov}(X, bZ) = b\mathbf{Cov}(X, Z)$
  - Hiệp phương sai và hệ số tương quan được dùng để **đặc trưng cho mức độ chặt chẽ của mối liên hệ phụ thuộc** giữa các biến ngẫu nhiên X và Y.
  - Hiệp phương sai có đơn vị đo lường bằng tích đơn vị đo lường của biến X và Y.
  - Hiệp phương sai có giá trị khác nhau tùy thuộc vào đơn vị đo lường của các biến X và Y.

2.4 Các tham số đặc trưng cho hệ hai biến ngẫu nhiên

Hệ số tương quan

- **Hệ số tương quan:** Hệ số tương quan của 2 biến ngẫu nhiên X và Y, ký hiệu ρ_{xy}, được xác định bằng công thức:

$$\boldsymbol{\rho_{xy} = Cov(X,Y)/\sigma_x\sigma_y}$$

- Hệ số tương quan không có đơn vị đo; $\mathbf{-1 < \rho < 1}$
- Hệ số tương quan đo lường **mối quan hệ tuyến tính giữa hai biến**.
- X và Y gọi là tương quan nếu $\mathbf{Cov(X,Y) \neq 0}$ hoặc $\boldsymbol{\rho_{xy} \neq 0}$
- X và Y gọi là không có tương quan nếu $\mathbf{Cov(X,Y) = 0}$ hoặc $\boldsymbol{\rho_{xy} = 0}$
- Nếu X, Y độc lập thì $\boldsymbol{\rho_{xy} = 0}$ (ngược lại chưa chắc đúng)
- Hệ số tương quan có tính đối xứng $\boldsymbol{\rho_{xy} = \rho_{yx}}$
- Nếu $\boldsymbol{\rho_{xy}} = \pm 1$: X và Y phụ thuộc hàm số với nhau
- Nếu $\boldsymbol{\rho_{xy} = -1}$: Mối quan hệ là nghịch biến hoàn hảo
- Nếu $\boldsymbol{\rho_{xy} = 1}$: Mối quan hệ là đồng biến hoàn hảo.

2.5 Một số quy luật phân phối xác suất thông dụng

- Phân phối đều liên tục (Uniform Distribution)
- Phân phối chuẩn (Normal Distribution)
- Phân phối chuẩn hoá (z-Distribution)
- Phân phối T (t-Distribution)
- Phân phối F (F-Distribution)
- Phân phối chi bình phương (Chi-Square Distribution)

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.1 Phân phối đều liên tục (Uniform Distribution)

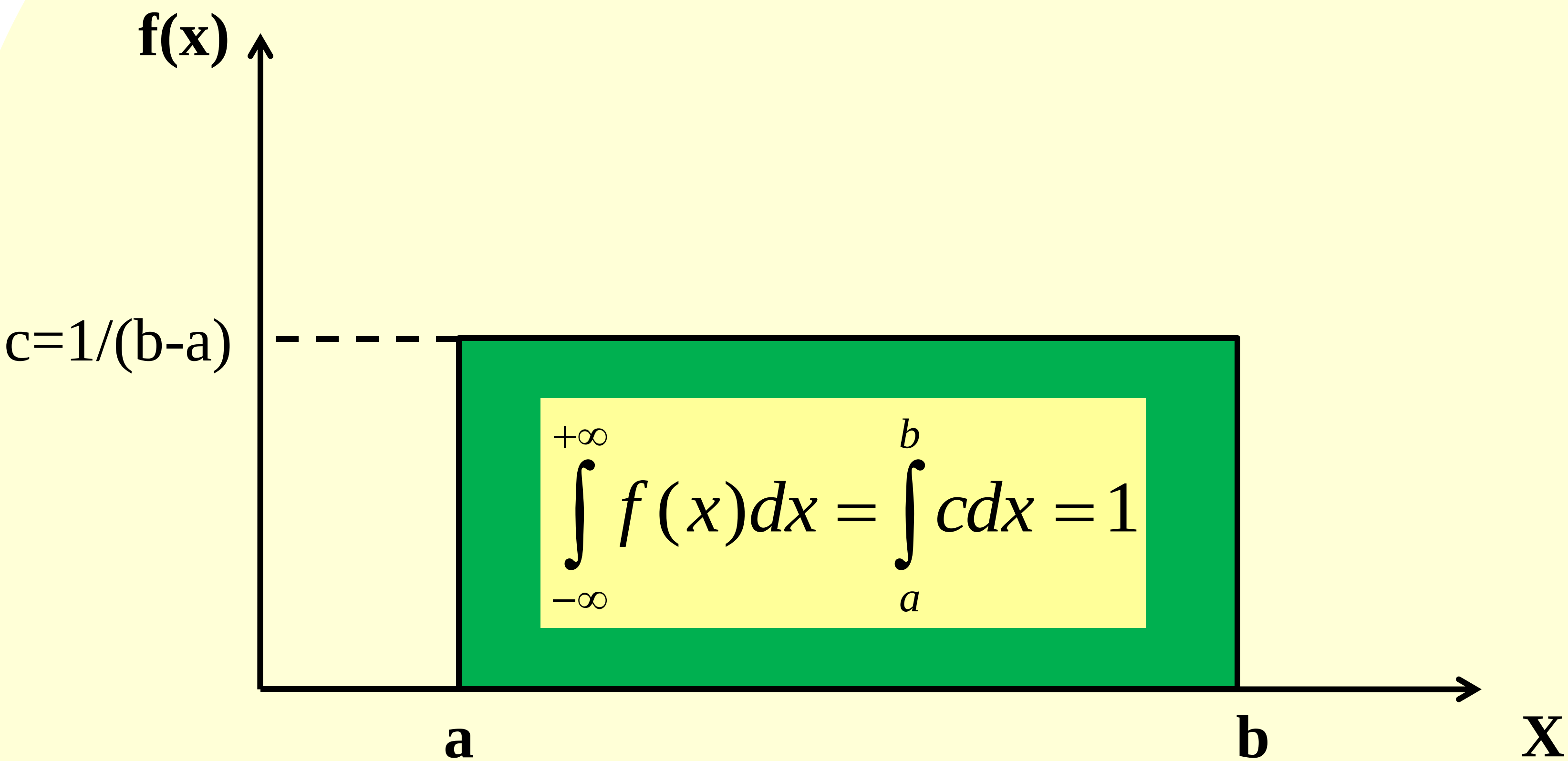
- **Biến ngẫu nhiên liên tục X có quy luật phân phối đều trong khoảng (a,b) nếu hàm mật độ xác suất của X có dạng**

$$f(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & x < a \text{ hoặc } x > b \end{cases}$$

- Phân phối đều liên tục là phân phối có xác suất xảy ra như nhau cho mọi kết cục của biến ngẫu nhiên liên tục.
- Phân phối đều liên tục còn gọi là phân phối hình chữ nhật.
- Giá trị kỳ vọng $\mathbf{E(X) = (a+b)/2}$
- Phương sai $\mathbf{V(X) = (b-a)^2/12}$

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.1 Phân phối đều liên tục (Uniform Distribution)



$$\int_{-\infty}^{+\infty} f(x)dx = \int_{a}^{b} c dx = 1$$

2.5 Một số quy luật phân phối xác suất thông dụng

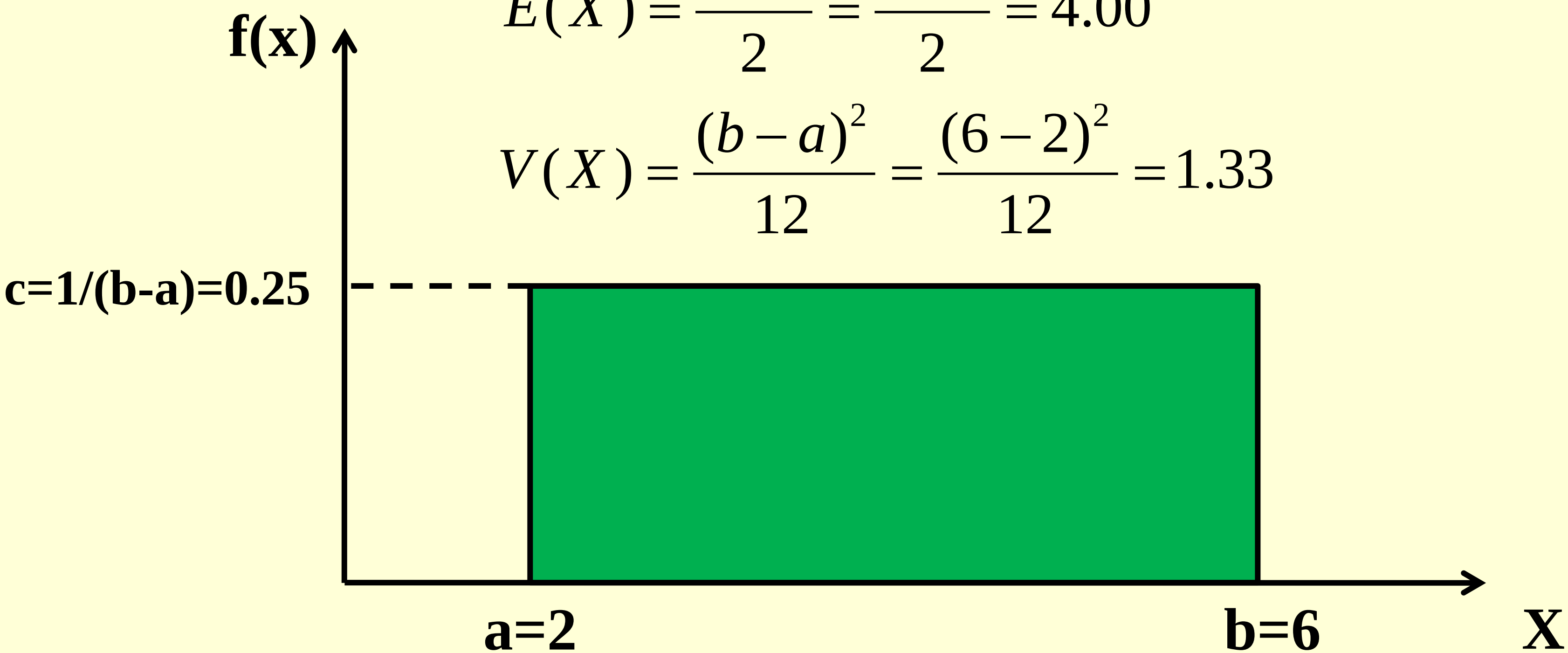
2.5.1 Phân phối đều liên tục (Uniform Distribution)

Ví dụ: X là biến ngẫu nhiên có phân phối xác suất đều liên tục trong khoảng $2 \le x \le 6$, hãy xác định các tham số đặc trưng cho phân phối E(X) và V(X)

$$f(x) = \frac{1}{b-a} = \frac{1}{6-2} = 0.25 \qquad 2 \le x \le 6$$

$$E(X) = \frac{a+b}{2} = \frac{2+6}{2} = 4.00$$

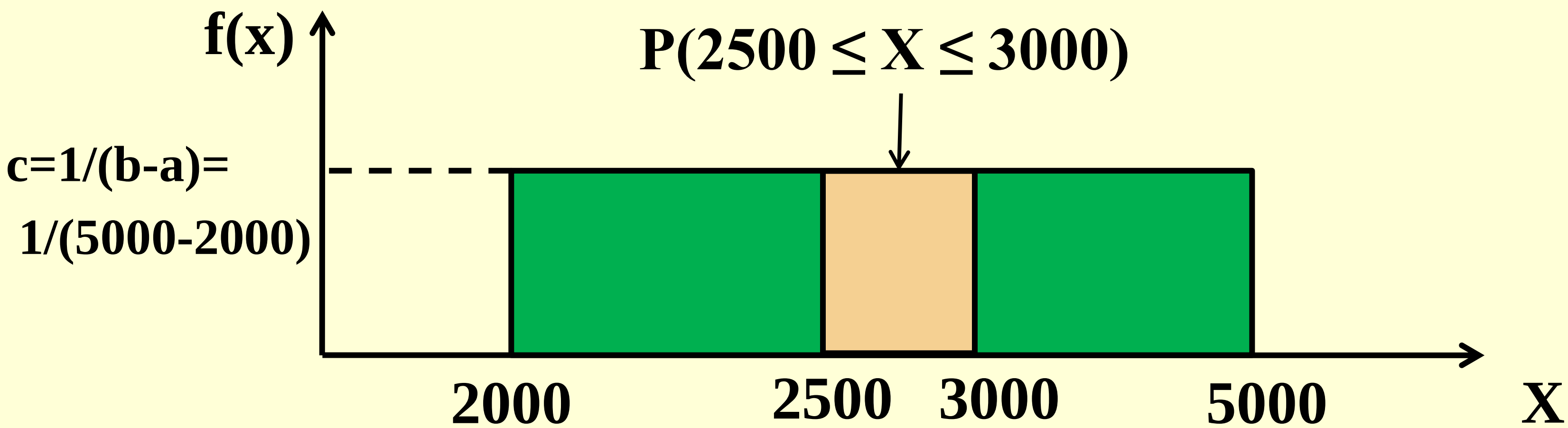
$$V(X) = \frac{(b-a)^2}{12} = \frac{(6-2)^2}{12} = 1.33$$



2.5 Một số quy luật phân phối xác suất thông dụng

2.5.1 Phân phối đều liên tục (Uniform Distribution)

- **Ví dụ:** Lượng xăng bán hàng ngày ở một cửa hàng tối thiểu là 2000 lít và tối đa là 5000 lít, Tìm xác suất bán trong ngày nằm trong khoảng 2500 lít đến 3000 lít. Giả thiết rằng lượng xăng bán trong ngày tuân theo quy luật phân phối đều liên tục.

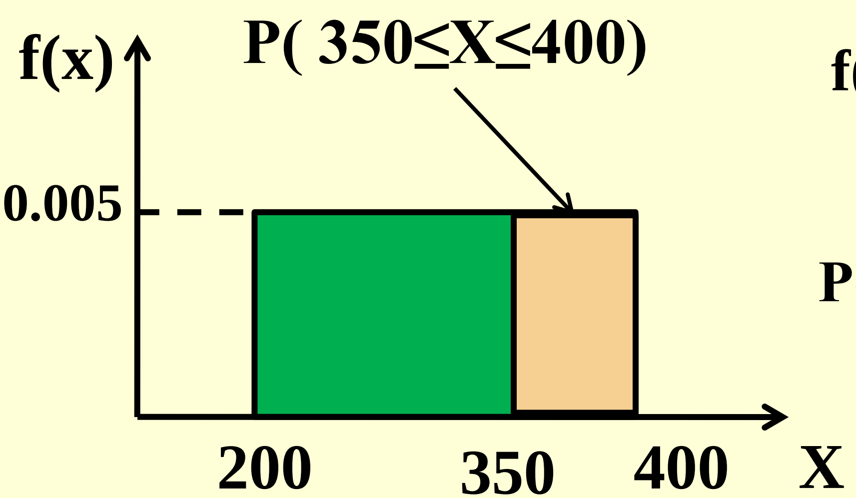


- $P(2500 \le X \le 3000) = (3000-2500)*1/3000 = 0.1667$
- Xác suất lượng bán trong ngày từ 2500-3000lít là 16.67%

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.1 Phân phối đều liên tục (Uniform Distribution)

- **Ví dụ:** Khi thâm nhập vào thị trường mới, một Doanh nghiệp dự kiến doanh số tối thiểu sẽ là 200triệu đồng/tháng và tối đa là 400triệu đồng/tháng. Tìm xác suất để DN đạt được doanh số tối thiểu là 350 triệu đồng/tháng. Giả thiết rằng doanh số của DN đạt hàng tháng là biến ngẫu nhiên có phân phối đều liên tục trong khoảng [200,400] triệu đồng.
- **Giải:** Doanh số hàng tháng có hàm mật độ xác suất



$$f(x) = \begin{cases} 1/(400-200) = 0.005 & 200 \le x \le 400 \\ 0 & x<200 \text{ hoặc } x>400 \end{cases}$$

$$P(350 \le X \le 400) = 0.005*(400-350) = 0.25$$

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

- Biến ngẫu nhiên liên tục **X** trong khoảng **(-∞, +∞)**, ký hiệu **X~N(μ,σ²)** gọi là phân phối chuẩn với các tham số **μ** và **σ²**, nếu hàm mật độ xác suất và hàm phân phối xác suất có dạng:

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

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Trong đó:

- **e = 2.71828**
- **π = 3.14159**
- **μ = Giá trị kỳ vọng = E(X)**
- **σ = Độ lệch chuẩn**
- **x = Giá trị bất kỳ của biến**

- ∞< x < +∞

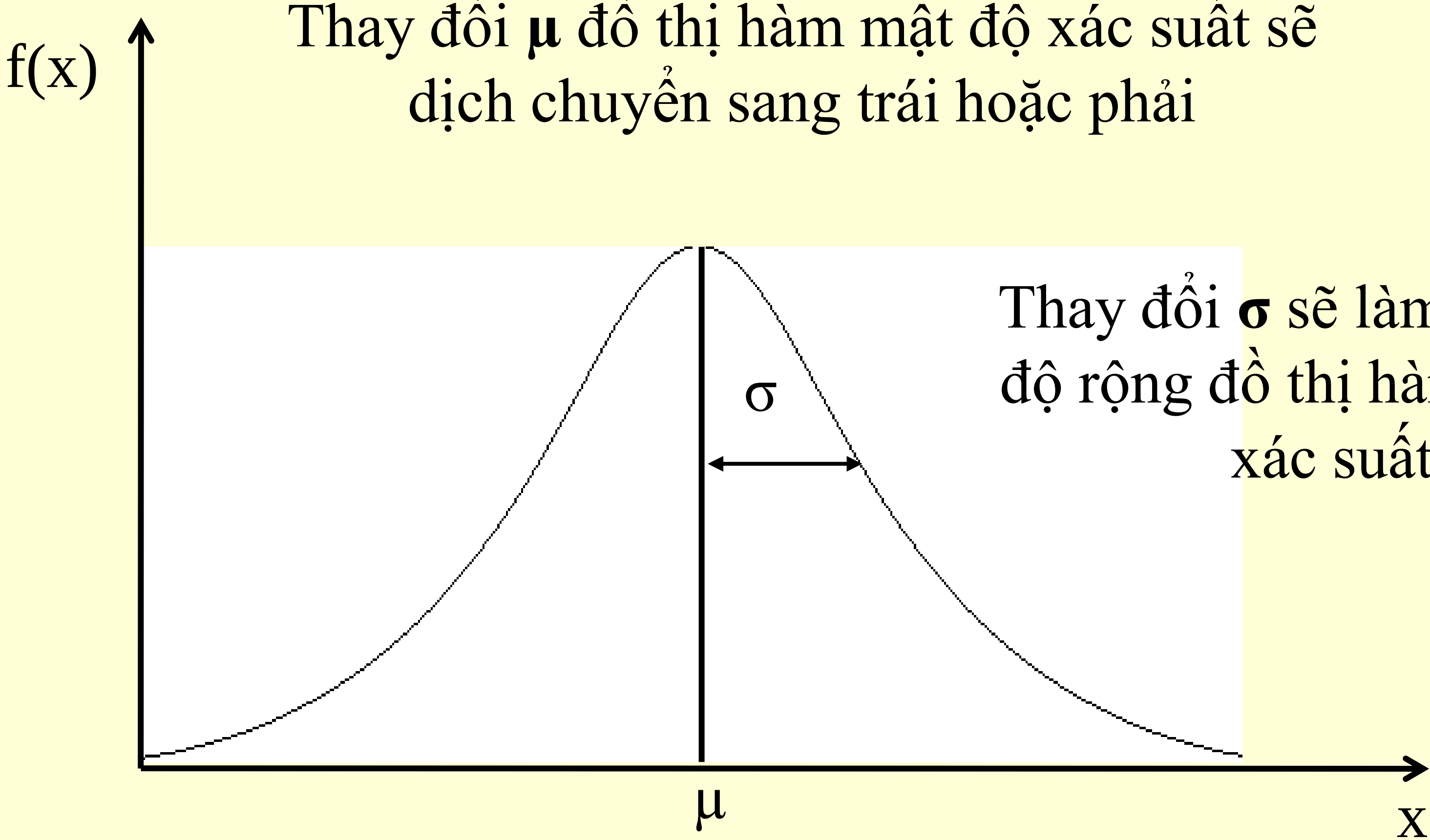
- **Phân phối chuẩn có dạng giống nhau chỉ khác nhau tham số μ và σ**

Khi tham số **μ** và **σ** **thay đổi** thì **vị trí** và **hình dáng** của đồ thị hàm mật độ phân phối xác suất sẽ thay đổi.

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

Thay đổi **μ** đồ thị hàm mật độ xác suất sẽ dịch chuyển sang trái hoặc phải



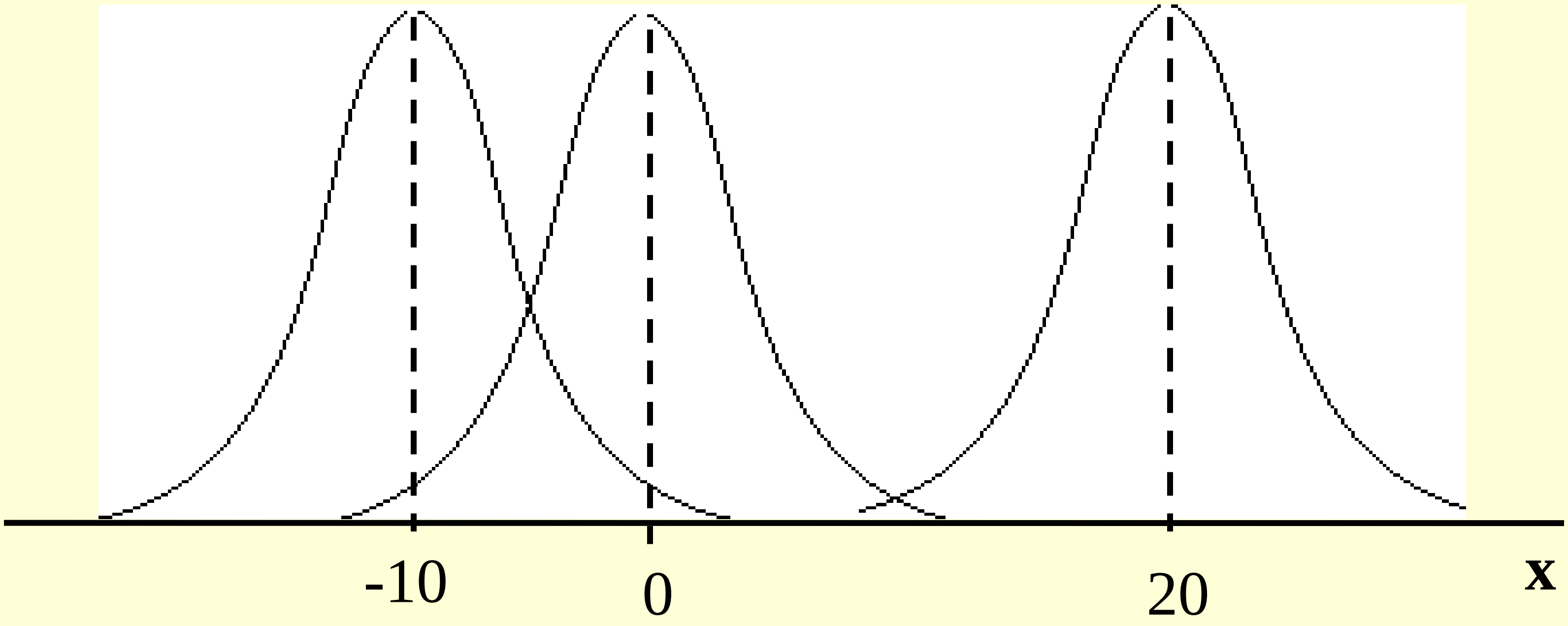
Thay đổi **σ** sẽ làm thay đổi độ rộng đồ thị hàm mật độ xác suất

Phân phối chuẩn $N(\mu,\sigma^2)$

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

Phân phối chuẩn có phương sai bằng nhau nhưng kỳ vọng khác nhau

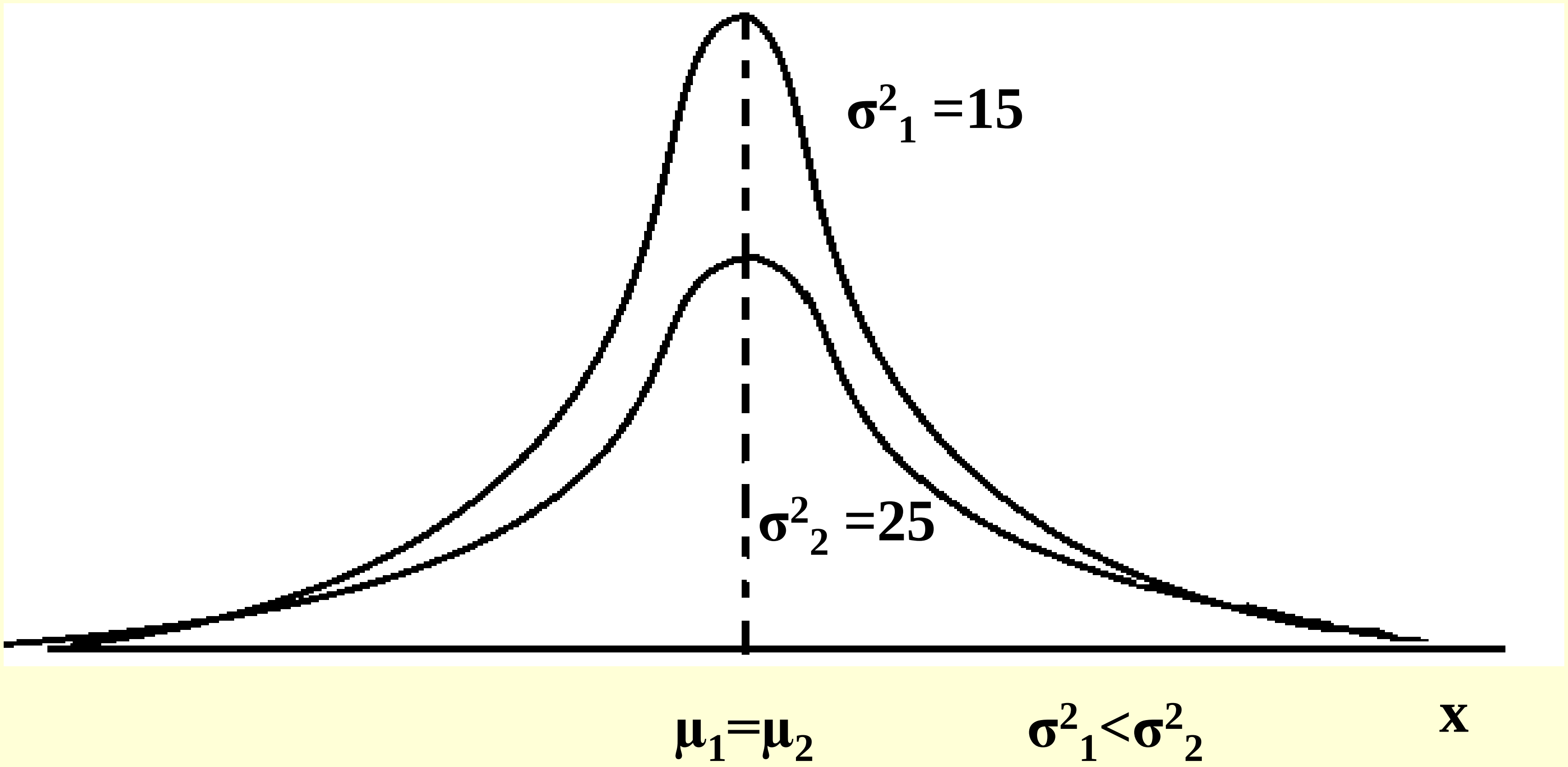


$$\sigma_1 = \sigma_2 = \sigma_3 \qquad \mu_1 < \mu_2 < \mu_3$$

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

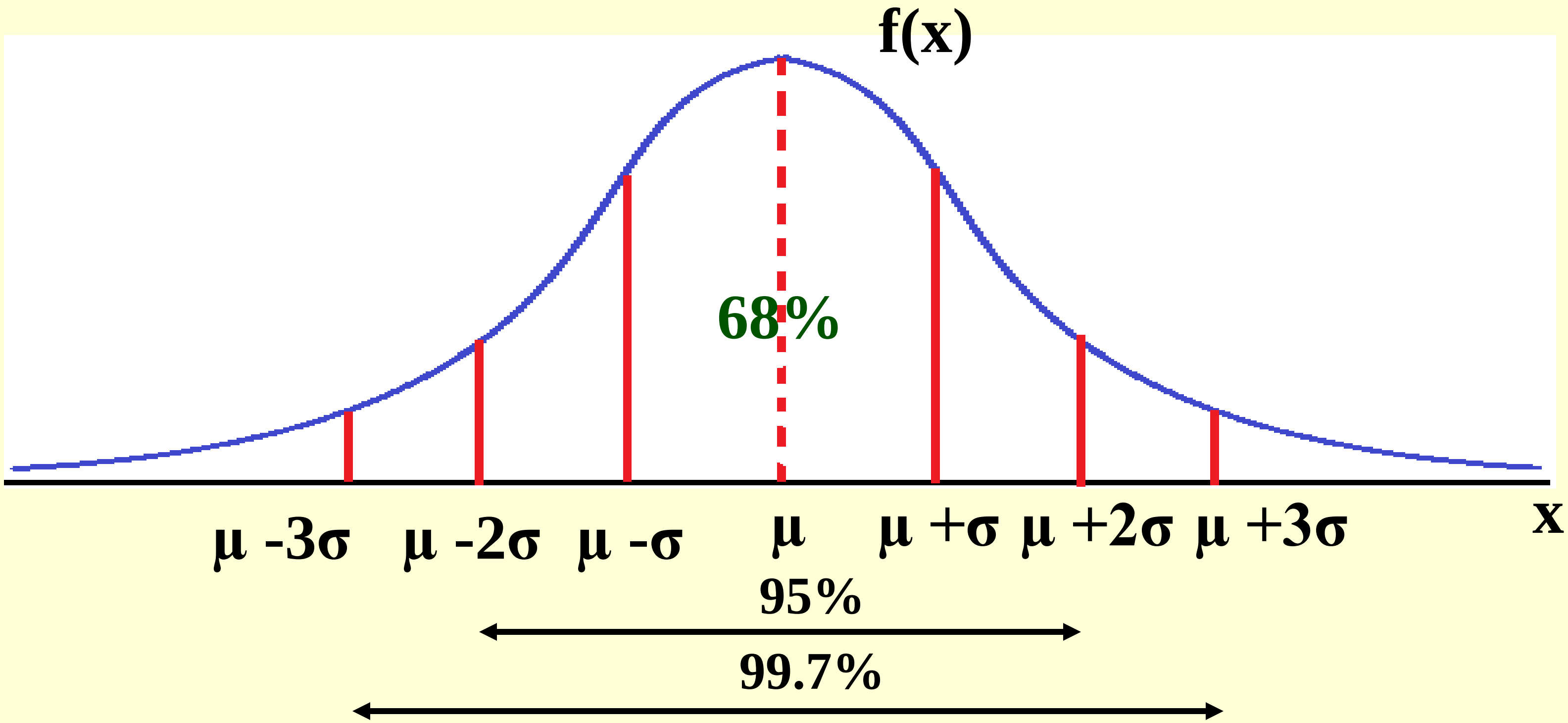
Phân phối chuẩn có kỳ vọng bằng nhau nhưng phương sai khác nhau



2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

- **Tính chất của phân phối chuẩn**
  - Hàm mật độ xác suất của PP chuẩn đối xứng quanh giá trị trung bình.
  - Xác suất để biến ngẫu nhiên có phân phối chuẩn nhận giá trị trong khoảng **$\mu \pm \sigma$** là xấp xỉ **68%**, **$\mu \pm 2\sigma$** là xấp xỉ **95%** và nhận giá trị **$\mu \pm 3\sigma$** xấp xỉ **99,7%**.



2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

- **Công thức tính xác suất để biến ngẫu nhiên (X) có phân phối chuẩn nhận giá trị trong khoảng (a,b):**

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

$$Z = \frac{x-\mu}{\sigma} \qquad \Phi_0(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{z^2}{2}} dZ$$

$$P(a < X < b) = \Phi_0\left(\frac{b-\mu}{\sigma}\right) - \Phi_0\left(\frac{a-\mu}{\sigma}\right)$$

- **$\Phi_0(-u) = -\Phi_0(u)$**
- **Với u > 5 thì $\Phi_0(u) = \Phi_0(5) = 0.5$**
- **Hàm $\Phi_0(u)$ được tra trong Bảng tính sẵn**

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

- **Ví dụ:** Một nhà sản xuất cần mua một loại gioăng cao su có độ dày từ 0.118cm đến 0.122cm. Có 2 cửa hàng cùng bán loại gioăng này với độ dày là các biến ngẫu nhiên phân phối chuẩn với các đặc trưng được cho trong bảng sau:

	Độ dày trung bình	**Độ lệch chuẩn**	**Giá bán $/hộp/1000cái**
Cửa hàng A	0.12	0.0010	3.0
Cửa hàng B	0.12	0.0015	2.6

  - Hỏi nhà sản xuất nên mua sản phẩm của cửa hàng nào?

- **Giải:** Trước hết phải xác định tỷ lệ gioăng đáp ứng được yêu cầu của nhà sản xuất trong mỗi hộp sản phẩm của 2 cửa hàng.
- Gọi $\mathbf{X_A}$ và $\mathbf{X_B}$ là độ dày của gioăng do cửa hàng **A** và **B** bán. Theo giả thiết $\mathbf{X_A}$ và $\mathbf{X_B}$ đều có phân phối chuẩn. Vì vậy tỷ lệ gioăng dùng được của 2 cửa hàng tương ứng:

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

$$P(0.118 < X_A < 0.122) = \Phi_0\left(\frac{b-\mu_A}{\sigma_A}\right) - \Phi_0\left(\frac{a-\mu_A}{\sigma_A}\right) =$$

$$= \Phi_0\left(\frac{0.122-0.12}{0.001}\right) - \Phi_0\left(\frac{0.118-0.12}{0.001}\right) = 2\Phi_0(2) = 0.9544$$

$$P(0.118 < X_B < 0.122) = \Phi_0\left(\frac{b-\mu_B}{\sigma_B}\right) - \Phi_0\left(\frac{a-\mu_B}{\sigma_B}\right) =$$

$$= \Phi_0\left(\frac{0.122-0.12}{0.0015}\right) - \Phi_0\left(\frac{0.118-0.12}{0.0015}\right) = 2\Phi_0(1.33) = 0.8164$$

- Vậy giá bán đối với mỗi chiếc gioăng dùng được của cửa hàng A:

  3/954.4 = 0.00314 $/chiếc

- Giá bán đối với mỗi chiếc gioăng dùng được của cửa hàng B:

  2.6/816.4 = 0.00318 $/chiếc

- **Vậy nhà sản xuất nên mua gioăng của cửa hàng A.**

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.2 Phân phối chuẩn (Normal Distribution)

- **Ví dụ:** Các vòng bi do máy tự động sản xuất ra được coi là đạt tiêu chuẩn nếu đường kính của nó sai lệch so với đường kính thiết kế không quá **0.7mm**. Biết rằng sai lệch này là biến ngẫu nhiên phân phối chuẩn với **μ = 0** và **σ =0.4mm**. Tìm tỷ lệ vòng bi đạt tiêu chuẩn của máy đó?

- **Giải:** Tỷ lệ vòng bi đạt tiêu chuẩn chính là xác suất để vòng bi đạt sai lệch kích thước trong phạm vi cho phép. **Gọi X là sai lệch giữa đường kính vòng bi được sản xuất và đường kính thiết kế**, xác suất cần xác định là xác suất xảy ra **$|X-\mu| < 0.7$**.

$$P(|X-\mu| < 0.7) = P(|X| < 0.7) = 2\Phi_0(0.7/0.4) = 2\Phi_0(1.75) = 2*0.4599 = 0.9198$$

Vậy tỷ lệ vòng bi đạt tiêu chuẩn của máy đó gần bằng 92%

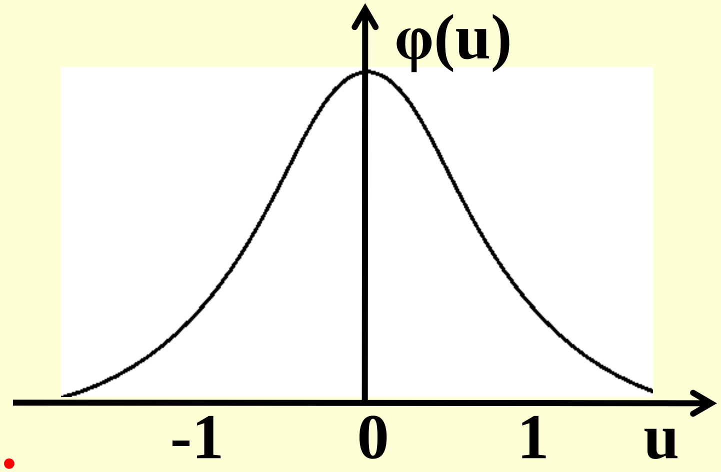
2.5 Một số quy luật phân phối xác suất thông dụng

2.5.3 Phân phối chuẩn hóa (Standard Normal Distribution)

- Biến ngẫu nhiên **(X)** có phân phối chuẩn với kỳ vọng bằng **μ** và độ lệch chuẩn bằng **σ** thì biến ngẫu nhiên **U = (X-μ)/σ**, **U** nhận các giá trị từ **(-∞,+∞)** tuân theo quy luật phân phối chuẩn hóa, ký hiệu **U~N(0,1)**, nếu hàm mật độ xác suất và hàm phân bố xác suất có dạng:

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

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{u^2}{2}} du$$



- **Đặc trưng của phân phối chuẩn hóa:**
  - **Đồ thị hàm mật độ xác suất đối xứng qua trục tung**
  - **E(U) = 0**
  - **V(U) = 1**

2.5 Một số quy luật phân phối xác suất thông dụng

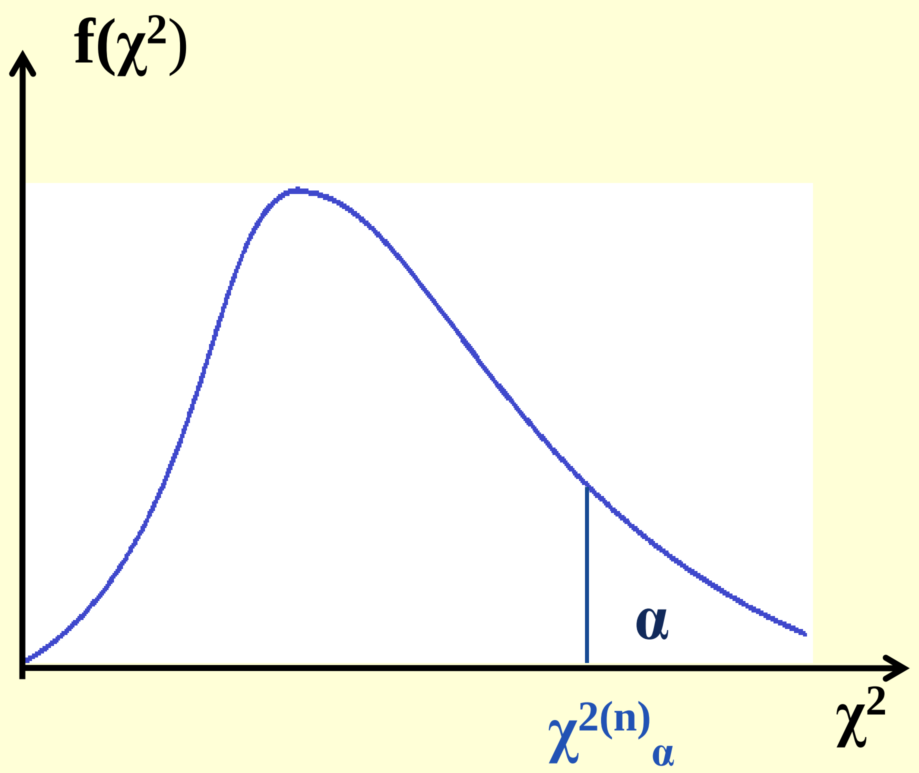
2.5.4 Phân phối Khi-bình phương (Chi-Square Distribution)

- Biến ngẫu nhiên liên tục χ^2 gọi là phân phối theo quy luật Khi-bình phương với **(n)** bậc tự do nếu hàm mật độ xác suất được xác định theo biểu thức:

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

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \qquad x > 0$$

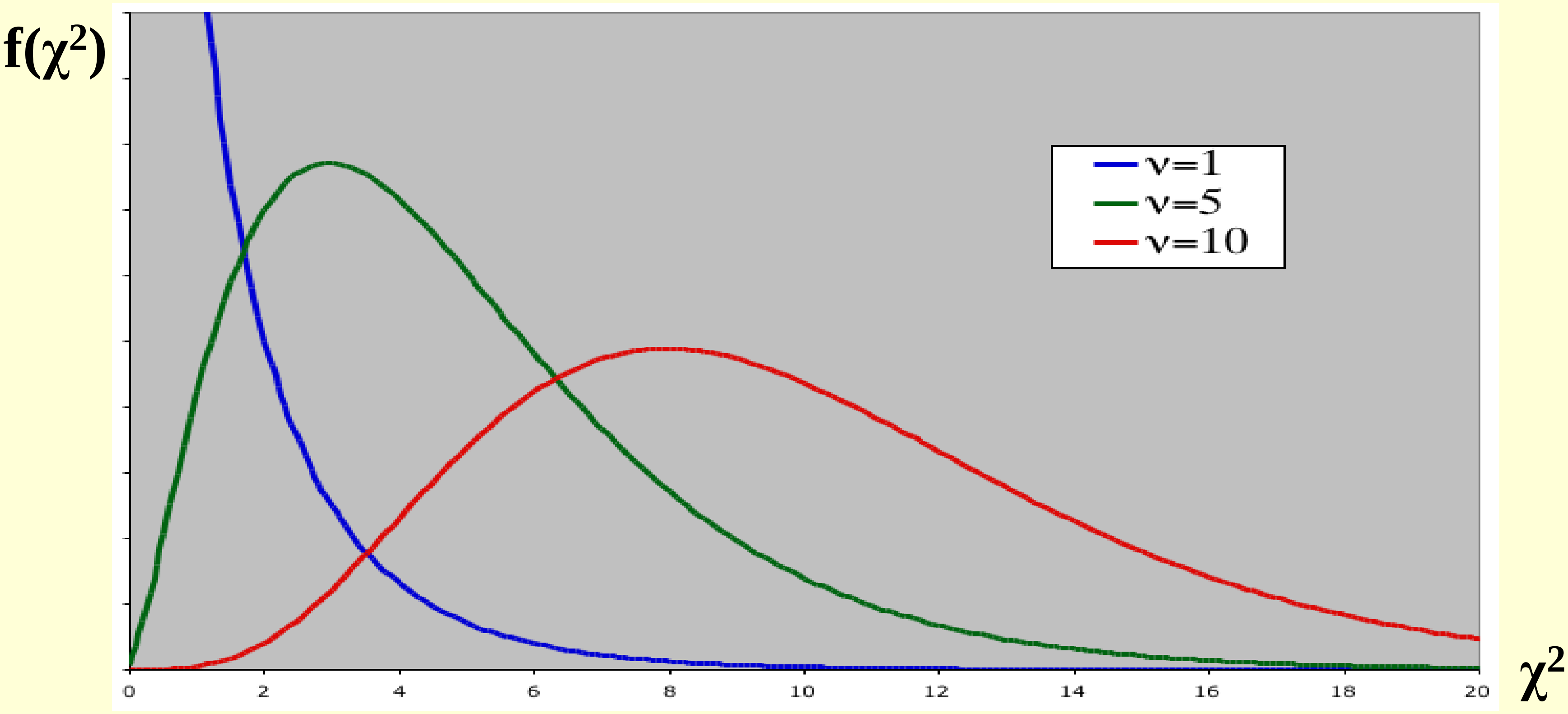
$$P(\chi^2 > \chi_\alpha^{2(n)}) = \alpha$$



2.5 Một số quy luật phân phối xác suất thông dụng

2.5.4 Phân phối Khi-bình phương (Chi-Square Distribution)

- Nếu $\mathbf{X_1, X_2,...X_n}$ là các biến ngẫu nhiên độc lập với nhau và cùng tuân theo quy luật chuẩn hóa N(0,1) thì $\boldsymbol{\chi^2 = \sum X_i^2}$ tuân theo quy luật phân phối Khi-Bình phương với n bậc tự do, ký hiệu $\boldsymbol{\chi^2(n)}$.



2.5 Một số quy luật phân phối xác suất thông dụng

2.5.4 Phân phối Khi bình phương (Chi-Square Distribution)

- **Tính chất của phân phối Khi-bình phương.**
  - Phân phối Khi-bình phương bắt đầu từ gốc tọa độ, lệch về phía bên trái và có đuôi dài vô tận về phía phải. Khi bậc tự do tăng dần thì phân phối χ^2 tiến gần đến phân phối chuẩn.
  - $\mathbf{E(\chi^2) = \mu = n}$
  - $\mathbf{V(\chi^2) = \sigma^2 = 2n}$
  - $\chi^2_1 + \chi^2_2 = \chi^2$ **và** $\mathbf{n_1 + n_2 = n}$
  - *Hai biến ngẫu nhiên độc lập cùng tuân theo quy luật phân phối Khi-bình phương có bậc tự do tương ứng là* $\mathbf{n_1}$, $\mathbf{n_2}$ *thì tổng của hai biến ngẫu nhiên đó cũng tuân theo quy luật phân phối Chi-bình phương với bậc tự do* $\mathbf{(n = n_1 + n_2)}$

2.5 Một số quy luật phân phối xác suất thông dụng

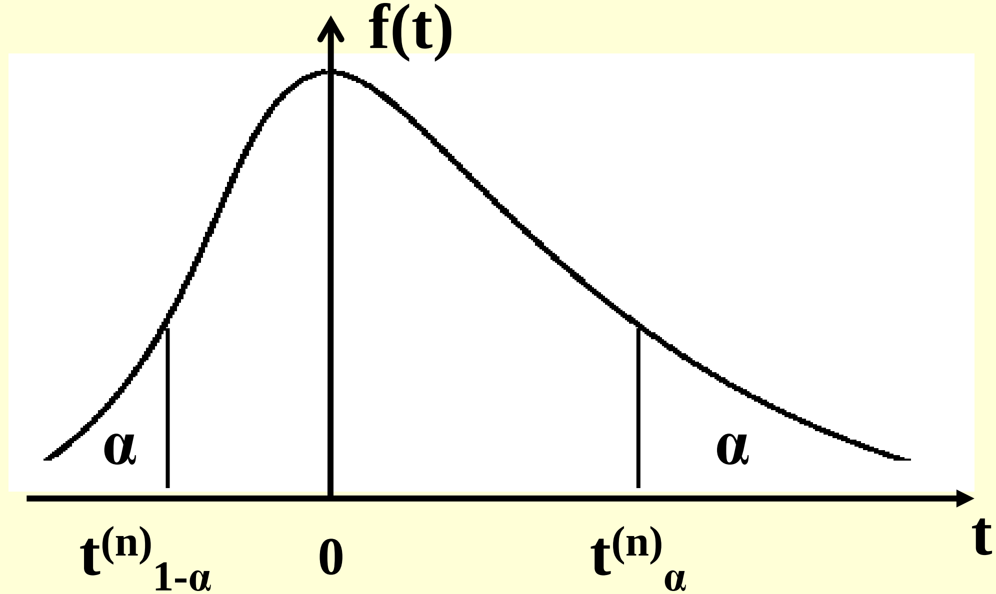
2.5.5 Phân phối Student (Student Distribution) - Phân phối T

- Biến ngẫu nhiên liên tục **(T)** gọi là phân phối theo quy luật Student với **(n)** bậc tự do nếu hàm mật độ xác suất được xác định bằng biểu thức:

$$f(t) = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi(n-1)}\,\Gamma(\frac{n-1}{2})}\left[1+\frac{t^2}{n-1}\right]^{-\frac{n}{2}} \qquad n > 0$$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

$$P(T > t_{\alpha}^{(n)}) = \alpha$$

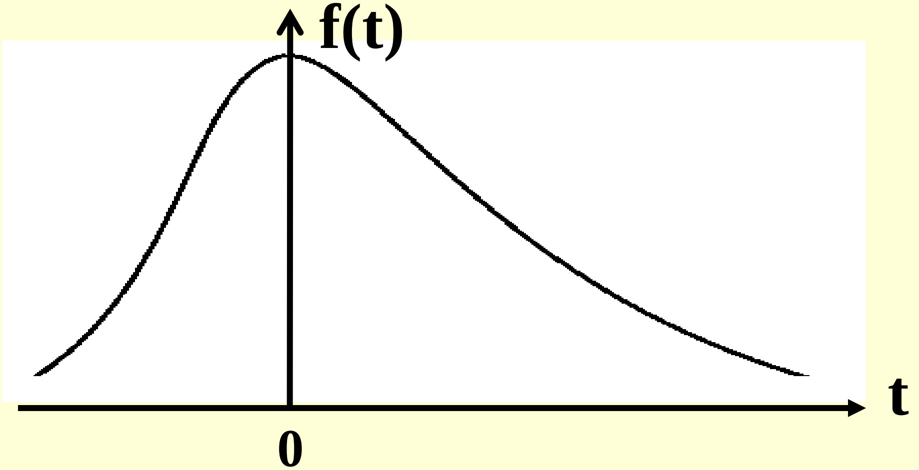


2.5 Một số quy luật phân phối xác suất thông dụng

2.5.5 Phân phối Student (Student Distribution) - Phân phối T

- Nếu **(U)** và **(V)** là hai biến ngẫu nhiên độc lập, **U** có phân phối chuẩn hóa **U~N(0,1)** và **V** có phân phối Khi-bình phương **V~ χ^2(n)**, thì biến ngẫu nhiên liên tục **(T)** xác định theo biểu thức:

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

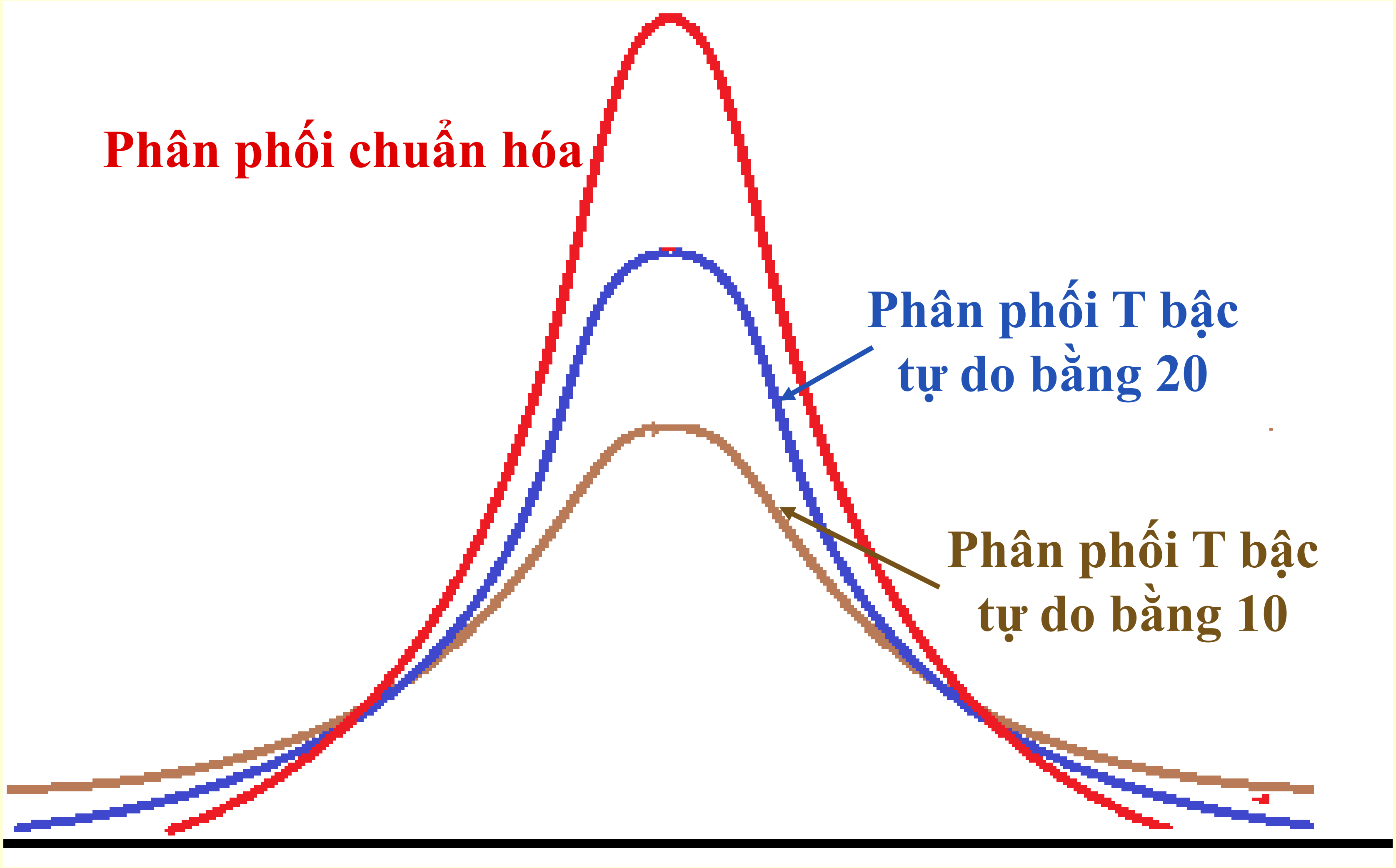


Gọi là tuân theo phân phối Student với (n) bậc tự do, ký hiệu T~T(n)

- Phân phối **(T)** có dạng như phân phối chuẩn hoá, khi **(n)** tiến đến vô hạn thì phân phối **(T)** tiến dần đến phân phối chuẩn hoá **(n>30)**.
- **E(T) = μ = 0**
- **V(T) = σ^2 = n/(n-2)**

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.5 Phân phối Student (Student Distribution) - Phân phối T

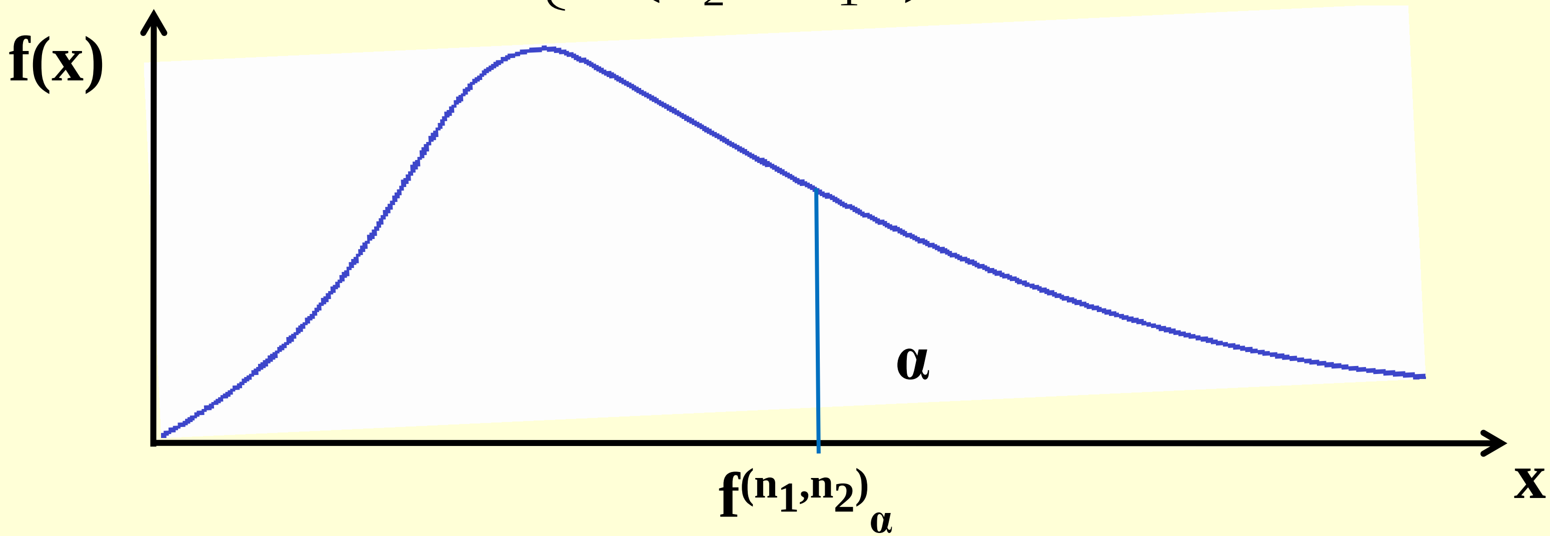


2.5 Một số quy luật phân phối xác suất thông dụng

2.5.6 Phân phối F (F- Distribution) - Phân phối Fisher

- Biến ngẫu nhiên liên tục **(F)** gọi là phân phối theo quy luật Fisher-Snedecor với $\mathbf{n_1}$, $\mathbf{n_2}$ bậc tự do nếu hàm mật độ xác suất được tính bằng biểu thức:

$$f(x) = \begin{cases} 0 & x \le 0 \\ C\dfrac{x^{\frac{n_1 - n_2}{2}}}{(n_2 + n_1 x)^{\frac{(n_1+n_2)}{2}}} & x > 0 \end{cases}$$

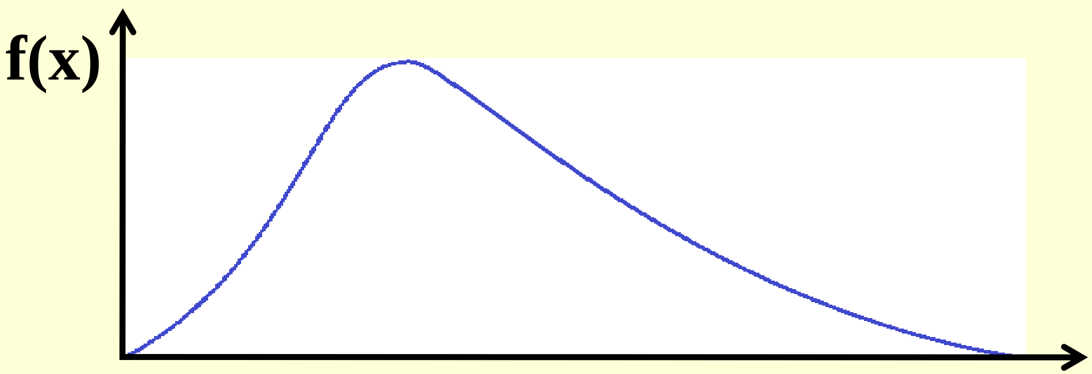


2.5 Một số quy luật phân phối xác suất thông dụng

2.5.6 Phân phối F (F- Distribution) - Phân phối Fisher

- Hai biến ngẫu nhiên độc lập **(U)** và **(V)** có phân phối Khi-bình phương với **n_1, n_2** bậc tự do **(U~(χ^2(n_1) và V~(χ^2(n_2))**, biến ngẫu nhiên độc lập **(F)** được xác định theo biểu thức:

$$F = \frac{\frac{U}{n_1}}{\frac{V}{n_2}} \sim F(n_1, n_2)$$



Biến ngẫu nhiên (F) được gọi là có phân phối Fisher-Snedecor với n_1, n_2 bậc tự do, ký hiệu **F~F(n_1, n_2).**

- **E(F) = μ = $n_2/(n_2-2)$**
- **V(F) = $2n^2_2(n_1+n^2_2-2)/[n_1(n_2-2)^2(n_2-4)]$**
- Khi bậc tự do **n_1** và **n_2** đủ lớn, phân phối **(F)** tiến đến phân phối chuẩn.

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.7 Phân phối mẫu

- **Tổng thể chung - N**
- **Tổng thể mẫu - n** (chọn có trả lại-chọn lặp) hoặc (chọn không trả lại-không lặp)

  Có trả lại có thể chọn $\mathbf{N^n}$ tổng thể mẫu khác nhau

  Không trả lại có thể chọn $C_N^n = \dfrac{N!}{n!(N-n)!}$ tổng thể mẫu khác nhau

- **Mỗi tổng thể mẫu chọn ngẫu nhiên, tham số trung bình mẫu, tỷ lệ mẫu, phương sai cũng là biến ngẫu nhiên có quy luật phân phối nhất định.**
- Tham số mô tả tổng thể chung và tổng thể mẫu

Tham số	Tổng thể chung (N)	Tổng thể mẫu (n)
Trung bình	$\mu=(x_1+x_2+...x_N)/N$	$\bar{x} = (x_1 + x_2 + ... + x_n)/n$
Tỷ lệ	**P = M/N**	**p = m/n**
Phương sai	$\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i-\mu)^2 n_i}{N}$	$S^2 = \dfrac{\sum_{i=1}^{N}(x_i-\bar{x})^2 * n_i}{n-1}$

- Tham số tổng thể chưa biết được xác định thông qua tham số tổng thể mẫu

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.7 Phân phối mẫu

- **Phân phối các đặc trưng mẫu như số trung bình, tỷ lệ, phương sai mẫu**
- **Từ đặc trưng phân phối mẫu suy đoán tham số tổng thể**
- **Tổng thể chung giả thiết tuân theo quy luật phân phối chuẩn**
- **Phân phối của trung bình mẫu:**
- Định lý giới hạn trung tâm (Central Limit Theorem)
  - Giả sử dãy $\mathbf{x_1, x_2 ... x_n}$ độc lập ngẫu nhiên cùng phân phối với **giá trị trung bình μ, độ lệch chuẩn σ.**
  - Trung bình $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ có phân phối xấp xỉ chuẩn với giá trị trung bình

$E(\bar{X}) = \mu$, độ lệch chuẩn $Se(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$ $\qquad G = U = \dfrac{\bar{X} - \mu}{Se(\bar{X})} = \dfrac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$

  - **X~N(μ,σ²) thì** $\bar{X}$ **~ N(μ,σ²/n)**
  - **n ≥ 30** $\bar{X}$ **có phân phối chuẩn**

2.5.7 Phân phối mẫu

- **Phân phối của trung bình mẫu:**
- **Ví dụ:** Trọng lượng trung bình 1 sản phẩm do 1 máy làm ra là **78.5g**, độ lệch chuẩn là **11.2g**. Chọn ngẫu nhiên **20** sản phẩm. Tính xác suất để trọng lượng trung bình của **20** sản phẩm đó lớn hơn **82g**?
- **Giải:** Gọi $\overline{X}$ là trọng lượng trung bình của 20 sản phẩm. $\overline{X}$ có phân phối xấp xỉ chuẩn với giá trị trung bình **78.5** và độ lệch chuẩn

$$Se(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{11.2}{\sqrt{20}} = 2.504$$

$$G = U = \frac{\overline{X} - \mu}{Se(\overline{X})} = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} = \frac{82 - 78.5}{2.504} = 1.398$$

$\Phi_0(1.4) = 0.4192$

$P(\overline{X} > 82) = 0.5 - 0.4192 = 0.0808$ (8%)

2.5 Một số quy luật phân phối xác suất thông dụng

2.5.7 Phân phối mẫu

- **Phân phối của tỷ lệ mẫu:**
  - Tổng thể chung có **N** đơn vị
  - Có **M** đơn vị mang dấu hiệu **A** cần quan tâm
  - Tỷ lệ đơn vị mang dấu hiệu **A** trong tổng thể gọi là tỷ lệ chung **P = M/N**
  - Mẫu **(n)** đơn vị từ **(N)** có **(m)** đơn vị có dấu hiệu **A**, tỷ lệ mẫu **p = m/n**
  - Đại lượng ngẫu nhiên, nếu **(n)** khá lớn **p** có phân phối chuẩn, trung bình là **(p)** và phương sai là **(pq/n)**
  - **p** ~**N(p, pq/n)**
- **Phân phối của phương sai mẫu:**
  - Tổng thể có phân phối chuẩn $X \sim N(\mu, \sigma^2)$
  - Kỳ vọng phương sai mẫu $E(S^2) = \sigma^2$
  - S^2 ước lượng không chệch cho σ^2 (Dùng S^2 ước lượng cho σ^2)

2.6 Ước lượng thống kê

- ***Tại sao cần phải ước lượng?***
- **Bài toán ước lượng tham số:**

Cho biến ngẫu nhiên **X** với quy luật phân phối xác suất đã biết, song chưa biết tham số **θ** nào của nó. Vấn đề đặt ra là phải ước lượng các giá trị **θ** (xác định gần đúng) từ các tham số của mẫu.

- **Phương pháp mẫu giải quyết bài toán ước lượng tham số**
  - Dựa vào mẫu $\{x_1, x_2 ..., x_n\}$ lập ra **Hàm ước lượng**:

  $\hat{\theta} = f(x_1, x_2 ... x_n)$ để ước lượng cho **θ**

  $\hat{\theta}$: Hàm ước lượng của θ (**ước lượng của θ**)
- **Ước lượng điểm và ước lượng khoảng.**

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- Phương pháp ước lượng điểm **dùng 1 giá trị để thay thế cho tham số θ** chưa biết của tổng thể.
- Giá trị được chọn là tham số thống kê $\hat{\theta}$ nào đó của mẫu ngẫu nhiên
- Có nhiều cách chọn $\hat{\theta}$ khác nhau tạo nên các phương pháp ước lượng điểm khác nhau
    - Phương pháp hàm ước lượng
    - Phương pháp ước lượng hợp lý tối đa
- **Phương pháp hàm ước lượng:**
    - ***Khái niệm:*** Giả sử cần ước lượng tham số **θ** của biến ngẫu nhiên **X** từ mẫu ngẫu nhiên kích thước **(n)**. Chọn tham số thống kê $\hat{\theta}$ đặc trưng mẫu tương ứng với tham số cần ước lượng **θ**.
    - Đối với 1 mẫu cụ thể tính được giá trị $\hat{\theta} = f(x_1, x_2 \ldots x_n)$
    - Giá trị của $\hat{\theta}$ tính được từ mẫu chính là giá trị **θ** của tổng thể mẫu

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- **Các tiêu chuẩn chọn hàm ước lượng:**

Ước lượng không chệch:

- Ước lượng $\hat{\theta}$ được gọi là ước lượng không chệch của tham số thống kê **θ** nếu kỳ vọng của $\hat{\theta}$ là **θ**

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

- Ước lượng không chệch không có nghĩa tất cả các giá trị trùng khít với tham số tổng thể mà chỉ có nghĩa trung bình của $\hat{\theta} = \theta$. Từng giá trị có thể sai lệch lớn.

Tham số sử dụng cho ước lượng không chệch

Trung bình mẫu	Kỳ vọng toán của biến ngẫu nhiên gốc
Tần suất mẫu	Xác suất của biến ngẫu nhiên gốc
Phương sai mẫu hiệu chỉnh	Phương sai của biến ngẫu nhiên gốc

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- ***Ước lượng hiệu quả:***

- Ước lượng $\hat{\theta}$ là ước lượng hiệu quả nhất của tham số θ của biến ngẫu nhiên gốc (X) nếu $\hat{\theta}$ là ước lượng không chệch và có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng mẫu đó.

- Để xác định $\hat{\theta}$ có phải là ước lượng hiệu quả nhất của θ hay không, cần tìm được giá trị nhỏ nhất có thể có của phương sai các hàm ước lượng.

- Bất đẳng thức Cramer-Crao: **θ* là ước lượng không chệch và hiệu quả nhất của θ**

$$V(\theta^*) \geq \frac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]}$$

- Gọi $\hat{\theta}_1, \hat{\theta}_2$ là 2 ước lượng không chệch của θ dựa trên số lượng của mẫu quan sát giống nhau $\hat{\theta}_1$ được gọi là hiệu quả hơn $\hat{\theta}_2$ nếu:

Hiệu quả tương đối

$$EF = \frac{V(\theta_2)}{V(\theta_1)} > 1$$

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- **Ví dụ:** Từ 1 mẫu ngẫu nhiên kích thước n= 2 xét 2 ước lượng của trung bình tổng thể (m) và phương sai (σ^2)

$$\overline{X} = \frac{1}{2}(X_1 + X_2)$$

$$X' = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

- Xem 2 ước lượng đó có là ước lượng không chệch của (m) hay không? Ước lượng nào hiệu quả hơn?
- **Giải:** Ước lượng không chệch là ước lượng có E($\hat{\theta}$)= θ
-

$$E(\overline{X}) = E(\frac{1}{2}X_1 + \frac{1}{2}X_2) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{1}{2}m + \frac{1}{2}m = m$$

$$E(X') = E(\frac{1}{3}X_1 + \frac{2}{3}X_2) = \frac{1}{3}E(X_1) + \frac{2}{3}E(X_2) = \frac{1}{3}m + \frac{2}{3}m = m$$

Cả 2 ước lượng đều là ước lượng không chệch

$$V(\overline{X}) = V(\frac{1}{2}X_1 + \frac{1}{2}X_2) = \frac{1}{4}V(X_1) + \frac{1}{4}V(X_2) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2$$

$$V(X') = V(\frac{1}{3}X_1 + \frac{2}{3}X_2) = \frac{1}{9}V(X_1) + \frac{4}{9}V(X_2) = \frac{1}{9}\sigma^2 + \frac{4}{9}\sigma^2 = \frac{5}{9}\sigma^2$$

$$V(\overline{X}) < V(X')$$

$$EF = \frac{V(X')}{V(\overline{X})} = \frac{5}{9}\sigma^2 / \frac{1}{2}\sigma^2 = \frac{10}{9} = 111\%$$

X' là ước lượng ít hiệu quả hơn

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- ***Ước lượng vững:***

Ước lượng $\hat{\theta}$ của mẫu được gọi là **ước lượng vững** của tham số thống kê θ của biến ngẫu nhiên gốc (X) nếu $\hat{\theta}$ hội tụ theo xác suất đến θ khi n→∞ $\lim_{n\to\infty} P(|\hat{\theta}-\theta| < \varepsilon) = 1$

- Điều này có nghĩa là mẫu càng lớn thì tham số ước lượng phải càng gần tham số cần ước lượng

$\overline{X}$ Ước lượng vững của kỳ vọng **μ** của biến ngẫu nhiên gốc (X)

$\hat{p}$ Ước lượng vững của xác suất **p** của biến ngẫu nhiên gốc (X)

Nếu $\hat{\theta}$ là ước lượng không chệch của θ và $\lim_{n\to\infty} V(\hat{\theta}) = 0$ thì $\hat{\theta}$ là ước lượng vững của θ

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- ***Ước lượng đủ:***
  - Một ước lượng $\hat{\theta}$ được gọi là đủ nếu nó chứa đựng toàn bộ thông tin trong mẫu về tham số θ cần ước lượng.
  - **Ví dụ**: Trung bình mẫu và số trung vị đều có thể là ước lượng không chệch của trung bình tổng thể. Song trung bình mẫu là ước lượng đủ còn số trung vị không phải là ước lượng đủ vì chỉ dùng giá trị ở giữa, còn trung bình mẫu sử dụng toàn bộ thông tin mẫu.
- Trong ước lượng sẽ chủ trương sử dụng các ước lượng đủ nếu điều đó cho phép.

2.6 Ước lượng thống kê

2.6.1 Ước lượng điểm

- **Phương pháp ước lượng hợp lý tối đa**
  - Giả sử biết quy luật phân phối xác suất tổng quát của biến ngẫu nhiên gốc X dưới dạng hàm mật độ f(x,θ). Cần ước lượng tham số θ của X.
  - Lập mẫu ngẫu nhiên kích thước (n), **W = (X_1, X_2...X_n).** Xây dựng hàm của đối số θ tại 1 giá trị cụ thể của mẫu:

  $L(x_1, x_2...x_n, \theta) = f(x_1, \theta).f(x_2, \theta)...f(x_n, \theta)$

  - **Hàm L được gọi là hàm ước lượng hợp lý của tham số θ**
- Giá trị hàm hợp lý là xác suất hay mật độ xác suất tại điểm **(x_1, x_2...x_n, θ).** Còn giá trị tham số θ tại điểm đó $\hat{\theta} = f(x_1, x_2, ...x_n)$ được gọi là ước lượng hợp lý tối đa của θ nếu ứng với giá trị này của θ, hàm hợp lý đạt cực đại.

Cách tìm: 1. Tìm đạo hàm bậc nhất của **lnL** (L và lnL cùng đạt giá trị cực đại) theo θ

2. Giải phương trình **dlnL/dθ = 0** Giả sử có nghiệm $\theta = \hat{\theta} = f(x_1, x_2, ...x_n)$

3. Tìm đạo hàm bậc 2: **$d^2lnL/d\theta^2$** Tại điểm $\theta = \hat{\theta}$

đạo hàm bậc 2<0 thì **lnL đạt cực đại (L cũng đạt cực đại)**

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

2.6.2 Ước lượng khoảng

- **Ước lượng bằng khoảng tin cậy**
  - Ước lượng điểm có thể sai lệch nhiều so với giá trị tham số cần ước lượng nếu kích thước mẫu nhỏ - sai số lớn
  - Ước lượng điểm không có khả năng đánh giá được khả năng mắc sai lầm là bao nhiêu.

→Ước lượng bằng khoảng tin cậy (Ước lượng khoảng).

- Từ một tham số thống kê nào đó của mẫu, xây dựng khoảng **(G_1, G_2)** sao cho với 1 xác suất cho trước tham số θ của biến ngẫu nhiên gốc **(X)** sẽ rơi vào khoảng **(G_1, G_2)**.

$$P(G_1 < \theta < G_2) = 1-\alpha$$

$I = G_2 - G_1$ Khoảng tin cậy (Khoảng ước lượng)

$1-\alpha$ Độ tin cậy của ước lượng (**α:** Mức ý nghĩa)

Ước lượng khoảng khắc phục các nhược điểm ước lượng điểm nhưng chứa đựng khả năng mắc sai lầm.

2.6.2 Ước lượng khoảng

- ***Ước lượng kỳ vọng biến ngẫu nhiên phân phối chuẩn N(μ, σ²)***

Cho mẫu ngẫu nhiên kích thước (n) **(X_1, X_2, ..., X_n)**

Tổng thể biến ngẫu nhiên gốc (X) có phân phối chuẩn N(μ, σ²), cần ước lượng tham số μ?

- **Trường hợp đã biết σ² của biến ngẫu nhiên gốc (X):**

$$U = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} = \frac{(\overline{X} - \mu)}{Se(\overline{X})}$$

U~N(0,1); $\overline{X}$: Trung bình mẫu, $Se(\overline{X})$: Độ lệch chuẩn của TB mẫu

- Với độ tin cậy **(1-α)** cho trước, xác suất để kỳ vọng toán **(μ)** của biến ngẫu nhiên phân phối chuẩn gốc **(X)** nằm trong khoảng:

$$P(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{\alpha 2} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n}} u_{\alpha 1}) = 1 - \alpha$$

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Trường hợp đã biết σ^2 của biến ngẫu nhiên gốc (X):**
  - ***Khoảng tin cậy đối xứng:*** $\alpha_1=\alpha_2=\alpha/2$

$$(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2};\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2})=(\overline{X}-\varepsilon;\overline{X}+\varepsilon)$$

  - ***Khoảng tin cậy bên phải (Ước lượng giá trị tối thiểu của μ):*** $\alpha_1=0;\ \alpha_2=\alpha$

$$(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha};+\infty)$$

  - ***Khoảng tin cậy bên trái (Ước lượng giá trị tối đa của μ):*** $\alpha_1=\alpha\ ;\ \alpha_2=0$

$$(-\infty;\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha})$$

  - *Xác định kích thước mẫu tối thiểu* **(n)** *với độ tin cậy* **(1-α)** cho trước *và độ dài khoảng tin cậy không vượt quá giá trị* $I_0=2\varepsilon_0$ cho trước:

n là số nguyên dương nhỏ nhất $\geq \dfrac{\sigma^2}{\varepsilon_0^2}u_{\alpha/2}^2$

Tăng kích thước mẫu (n), giữ nguyên độ tin cậy (1-α), tăng độ chính xác

Tăng độ tin cậy (1-α), giữ nguyên kích thước mẫu (n), giảm độ chính xác

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Trường hợp đã biết σ^2 của biến ngẫu nhiên gốc (X):**

- **Ví dụ:** Trọng lượng một loại sản phẩm là biến ngẫu nhiên tuân theo quy luật phân phối chuẩn, với độ lệch chuẩn là 1g. Cân thử 25 sản phẩm loại này thu được kết quả sau:

Trọng lượng (g)	**18**	**19**	**20**	**21**
Số sản phẩm tương ứng	**3**	**5**	**15**	**2**

Với độ tin cậy 0.95, hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình của loại sản phẩm nói trên?

- **Giải:**

Gọi (X) là trọng lượng sản phẩm, **$X \sim N(\mu,\sigma^2)$** với **$\sigma = 1$. Trọng lượng trung bình chính là tham số μ** chưa biết cần ước lượng. Trường hợp là bài toán **ước lượng kỳ vọng biến ngẫu nhiên phân phối chuẩn bằng khoảng tin cậy đối xứng khi đã biết phương sai**.

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Giá trị trung bình mẫu:** $\overline{x} = \dfrac{3\times 18 + 5\times 19 + 15\times 20 + 2\times 21}{3+5+15+2} = 19.64$
- Với độ tin cậy **1-α = 0.95 →α/2 = 0.025.**

Tra **Bảng giá trị tới hạn chuẩn** $\mathbf{u_{0.025} = 1.96}$ hoặc

Hàm Excel $\mathbf{u_{\alpha/2}}$**= NORMINV(1-α/2,0,1)**

$\mathbf{u_{0.025}}$**=NORMINV(1-0.025, 0,1) =1.96**

- Khoảng tin cậy đối xứng của mẫu với khoảng tin cậy 0.95:

$$(\overline{X} - \frac{1}{\sqrt{25}}1.96; \overline{X} + \frac{1}{\sqrt{25}}1.96) = (\overline{X} - 0.392; \overline{X} + 0.392)$$

- Với mẫu cụ thể này khoảng tin cậy đối xứng của μ với độ tin cậy 0.95:

$$19.64 - 0.392 < \mu < 19.64 + 0.392 = 19.248 < \mu < 20.032$$

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Trường hợp chưa biết σ^2 của biến ngẫu nhiên gốc X, kích thước mẫu nhỏ hơn 30:**
- Chọn thống kê T~T(n-1)

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S}$$

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

- Khoảng ước lượng của tham số **(μ)** với độ tin cậy **(1-α)**:

$\alpha_1+\alpha_2=\alpha$ $\quad (\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha 2}^{n-1}; \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha 1}^{n-1})$

$\alpha_1=\alpha_2=\alpha/2$ $\quad (\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}^{n-1}; \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}^{n-1})$

$\alpha_1=0; \alpha_2=\alpha$ $\quad (\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha}^{n-1}; +\infty)$

$\alpha_1=\alpha; \alpha_2=0$ $\quad (-\infty; \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha}^{n-1})$

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Trường hợp chưa biết σ^2 của biến ngẫu nhiên gốc (X):**
- **Ví dụ:** Để xác định trọng lượng trung bình của các bao bột trong kho, người ta đem cân ngẫu nhiên **15 bao** của kho đó và tìm được **trọng lượng trung bình 1 bao là 39.8kg; phương sai mẫu là 0.144**. Hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình các bao bột trong kho với yêu cầu độ tin cậy của việc ước lượng là **99%**. Giả thiết trọng lượng đóng bao của các bao bột là biến ngẫu nhiên tuân theo phân phối chuẩn.
- **Giải:** Gọi X là trọng lượng bao bột của kho **$X \sim N(\mu,\sigma^2)$.** Trọng lượng đóng bao trung bình chính là giá trị **(μ).** Trường hợp này là **Bài toán ước lượng khoảng tin cậy đối xứng giá trị tham số μ của phân phối chuẩn $X \sim N(\mu,\sigma^2)$ khi chưa biết σ^2.** Sử dụng phương sai mẫu và phân phối Student để ước lượng.

2.6.2 Ước lượng khoảng

Khoảng tin cậy: $(\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}^{n-1}; \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}^{n-1})$

Độ tin cậy 1-α = 0.99→ α/2=0.005

Tra Bảng giá trị tới hạn Student $t_{0.005}(14) = 2.977$ hoặc

Hàm Excel $t_{\alpha/2}(n-1)$ = TINV(α,n-1) $t_{0.005}(14)$ = TINV(0.01,14)

Với độ tin cậy 0.99 khoảng tin cậy đối xứng của μ là:

$$(\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}^{n-1}; \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}^{n-1}) = (\overline{X} - \frac{S}{\sqrt{15}} 2.977; \overline{X} + \frac{S}{\sqrt{15}} 2.977)$$

Với mẫu cụ thể có $\overline{x} = 39.8 \quad S^2 = 0.144 \rightarrow S = 0.379$

Khoảng tin cậy của trọng lượng trung bình các bao bột trong kho:

$$(39.8 - \frac{0.379}{\sqrt{15}} 2.977; 39.8 + \frac{0.379}{\sqrt{15}} 2.977) = (39.51 < \mu < 40.09)$$

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Ước lượng phương sai của biến ngẫu nhiên phân phối chuẩn X~N(μ,σ²)**
- ***Trường hợp đã biết kỳ vọng gốc μ***

$$P\left(\frac{nS^{*2}}{\chi^{2(n)}_{\alpha 2}} < \sigma^2 < \frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha 1}}\right) = 1-\alpha$$

$$\chi^2 = \frac{nS^2}{\sigma^2}$$

$\chi^2 \sim \chi^2(n)$

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 n_i$$

$\alpha_1+\alpha_2=\alpha$ $\quad \left(\frac{nS^{*2}}{\chi^{2(n)}_{\alpha 2}};\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha 1}}\right)$ **Khoảng tin cậy không đối xứng**

$\alpha_1=\alpha_2=\alpha/2$ $\quad \left(\frac{nS^{*2}}{\chi^{2(n)}_{\alpha/2}};\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha/2}}\right)$ **Khoảng tin cậy đối xứng**

$\alpha_1=0;\ \alpha_2=\alpha$ $\quad \left(\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha}};+\infty\right)$ **Khoảng tin cậy bên phải**

$\alpha_1=\alpha;\ \alpha_2=0$ $\quad \left(0;\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha}}\right)$ **Khoảng tin cậy bên trái**

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Ước lượng phương sai của biến ngẫu nhiên phân phối chuẩn X~N(μ,σ²)**
- ***Trường hợp đã biết kỳ vọng gốc μ***
- **Ví dụ:** Mức hao phí nguyên vật liệu cho một đơn vị sản phẩm là biến ngẫu nhiên phân phối chuẩn với mức trung bình là **20g**. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử **25 sản phẩm** và thu được kết quả sau:

Hao phí nguyên liệu (g)	**19.5**	**20.0**	**20.5**
Số sản phẩm tương ứng	**5**	**18**	**2**

Với độ tin cậy **90%** hãy ước lượng phương sai của mức hao phí nguyên vật liệu cho một đơn vị sản phẩm bằng cách khoảng tin cậy có $\alpha_1 = \alpha_2 = \alpha/2$?

- **Giải:**

Gọi **X** là mức hao phí nguyên liệu cho 1 đơn vị sản phẩm **X~N(μ,σ²) và μ = 20**. Đây là bài toán **ước lượng phương sai của phân phối chuẩn** N(**μ,σ²**) **khi đã biết μ** và **khoảng tin cậy có dạng** $\alpha_1 = \alpha_2 = \alpha/2$.

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Khoảng tin cậy của phương sai tổng thể σ^2:** $\left(\frac{nS^{*2}}{\chi^{2(n)}_{\alpha/2}};\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha/2}}\right)$
- **Tra Bảng giá trị tới hạn χ^2:** $\chi^{2(n)}_{\alpha/2}=\chi^{2(25)}_{0.05}=37.65$

$$\chi^{2(n)}_{1-\alpha/2}=\chi^{2(25)}_{0.95}=14.61$$

- **Phương sai S^2:**

X_i	n_i	$(X_i-\mu)$	$(X_i-\mu)^2$	$n_i(X_i-\mu)^2$
19.5	5	-0.5	0.25	1.25
20.0	18	0	0	0
20.5	2	0.5	0.25	0.5
	n=25			∑=1.75

- **Phương sai: $S^{*2} = (1/n)\sum n_i(x_i-\mu)^2= 1.75/25 = 0.07$**
- **Với độ tin cậy 90%, qua mẫu cụ thể, khoảng tin cậy của phương sai tổng thể σ^2**

$$\left(\frac{nS^{*2}}{\chi^{2(n)}_{\alpha/2}};\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha/2}}\right)=\left(\frac{25\times 0.07}{37.65};\frac{25\times 0.07}{14.61}\right)=(0.0464<\sigma^2<0.1198)$$

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Ước lượng phương sai của biến ngẫu nhiên phân phối chuẩn X~N(μ,σ²)**
- ***Trường hợp chưa biết kỳ vọng gốc μ***

  $\chi^2 \sim \chi^2(n-1)$

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

$$P\left[\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha 2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha 1}}\right] = 1-\alpha$$

- Với độ tin cậy (1-α), khoảng tin cậy của σ²:

$$\left[\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha 2}};\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha 1}}\right] \qquad \alpha_1+\alpha_2=\alpha$$

$$\left[\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha/2}};\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha/2}}\right] \qquad \alpha_1=\alpha_2=\alpha/2$$

$$\left[\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha}};+\infty\right] \qquad \alpha_1=0;\ \alpha_2=\alpha$$

$$\left[0;\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha}}\right] \qquad \alpha_1=\alpha;\ \alpha_2=0$$

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Ước lượng phương sai của biến ngẫu nhiên phân phối chuẩn X~N(μ,σ²)**
- ***Trường hợp chưa biết kỳ vọng gốc μ***
- **Ví dụ:** Với độ tin cậy **0.95** hãy ước lượng phương sai của kích thước các chi tiết trên cơ sở các số liệu mẫu cho ở Bảng, biết $\alpha_1 = \alpha_2 = \alpha/2$?

Kích thước chi tiết	Số chi tiết tương ứng
54.795-54.805	6
54.805-54.815	14
54.815-54.825	33
54.825-54.835	47
54.835-54.845	45
54.845-54.855	33
54.855-54.865	15
54.865-54.875	7
	n=200

- **Giải:**

Gọi **X** là kích thước các chi tiết do máy sản xuất, **X~N(μ,σ²)**. Đây là bài toán **ước lượng phương sai của phân phối chuẩn N(μ,σ²) khi chưa biết μ** và **khoảng tin cậy có dạng $\alpha_1 = \alpha_2 = \alpha/2$.**

2.6 Ước lượng thống kê

2.6.2 Ước lượng khoảng

- **Khoảng tin cậy của phương sai tổng thể:**

$$\left[\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha/2}};\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha/2}}\right]$$

- **Qua mẫu cụ thể tìm được: S^2= 0.000257224 và n= 200**
- **Tra Bảng χ^2 :**

$$\chi^{2(n-1)}_{\alpha/2} = \chi^{2(199)}_{0.025} = 241.1$$

$$\chi^{2(n-1)}_{1-\alpha/2} = \chi^{2(199)}_{0.975} = 162.7$$

- **Với độ tin cậy 0.95 khoảng tin cậy của phương sai tổng thể σ^2:**

$$\left[\frac{(n-1)S^2}{\chi^{2(n-1)}_{\alpha/2}};\frac{(n-1)S^2}{\chi^{2(n-1)}_{1-\alpha/2}}\right] = \left[\frac{(199)\times 0.000257224}{241.1};\frac{(199)\times 0.000257224}{162.7}\right]$$

$$= (0.000212 < \sigma^2 < 0.000315)$$

2.7 Kiểm định giả thuyết thống kê

2.7.1 Nguyên lý cơ bản

- **Giả thuyết thống kê** là giả thuyết về **dạng phân phối xác suất** của biến ngẫu nhiên, về **các tham số đặc trưng** của biến ngẫu nhiên hoặc về **tính độc lập** của biến ngẫu nhiên.
- Giả thuyết thống kê muốn kiểm định được đưa ra, ký hiệu là H_0 **(Giả thuyết gốc-Giả thuyết không)**.
- Giả thuyết ngược lại với giả thuyết gốc, ký hiệu H_1, **(Giả thuyết ngược-Giả thuyết đối)**.
- Giả thuyết thống kê **có thể đúng hoặc sai nên cần kiểm định** (kết luận chấp nhận hay không chấp nhận giả thuyết đó).
- **(H_0, H_1)** tạo nên **một cặp giả thuyết thống kê.** Việc **bác bỏ H_0** sẽ dẫn đến việc **chấp nhận giả thuyết H_1**
- ***Ví dụ:***

  H_0: Nhu cầu thị trường của mặt hàng X phân phối theo quy luật chuẩn;

  H_1: Không phân phối theo quy luật chuẩn.

  $H_0 : \mu = 0.5 \quad H_1 : \mu \neq 0.5$

  H_0: Nhu cầu X của thị trường và thu nhập Y của khách hàng độc lập nhau;
  H_1: X và Y phụ thuộc nhau.

2.7 Kiểm định giả thuyết thống kê

2.7.1 Nguyên lý cơ bản

- **Giả thuyết đơn:** Giả thuyết chỉ chứa 1 mệnh đề. $\mathbf{H_0 : \mu = 0.5}$
- **Giả thuyết kép:** Giả thuyết chứa một số hữu hạn hoặc vô hạn các giả thuyết đơn. $\mathbf{H_0 : \mu > 0.5}$ (Bao gồm một số vô hạn các giả thuyết đơn dạng H_0: $\boldsymbol{\mu = b_i}$, trong đó $\mathbf{b_i}$ là mọi số lớn hơn **0.5**)
- **Kiểm định một phía và kiểm định hai phía**

Các giá trị có thể có của tham số thống kê trong kiểm định có thể chia làm 2 miền: **Miền bác bỏ và Miền chấp nhận**. Điểm giới hạn phân chia 2 miền được gọi là **Giá trị tới hạn**.

- *Miền bác bỏ là miền chứa các giá trị làm cho giả thuyết H_0 bị bác bỏ.*

- *Miền chấp nhận là miền chứa các giá trị giúp cho giả thuyết H_0 không bị bác bỏ.*

2.7 Kiểm định giả thuyết thống kê

2.7.1 Nguyên lý cơ bản

- **Việc kiểm định được thực hiện theo các bước như sau:**

1. Lập 1 mẫu ngẫu nhiên $(X_1, X_2, X_3, \ldots, X_n)$ cho biến ngẫu nhiên X.

2. Tìm một hàm $G = f(X_1, X_2, X_3, \ldots, X_n, \theta_0)$, θ_0 là tham số liên quan đến giả thuyết cần kiểm định; Nếu H_0 đúng, quy luật phân phối xác suất của G là xác định. **G là tiêu chuẩn kiểm định.**

3. Tìm một miền W_α sao cho với điều kiện giả thuyết H_0 đúng thì xác suất để G nhận giá trị thuộc miền W_α đúng bằng α, với $0< \alpha <1$ và đủ bé để sao cho trong một phép thử rất khó có thể xảy ra giá trị G rơi vào miền W_α. Giá trị α gọi là **mức ý nghĩa** của kiểm định, W_α **miền bác bỏ** giả thuyết H_0 với mức ý nghĩa α

$$P(G \in W_\alpha/H_0) = \alpha$$

4. Lấy một mẫu cụ thể $(x_1, x_2, x_3 \ldots x_n)$ tính giá trị của hàm G cho mẫu

$G_{qs} = G_0\ (x_1, x_2, x_3 \ldots x_n)$; G_{qs} : Giá trị quan sát của tiêu chuẩn kiểm định

5. So sánh giá trị G_{qs} với miền bác bỏ W_α

G_{qs} thuộc $W_\alpha \rightarrow$ **Bác bỏ giả thiết H_0 ở mức ý nghĩa α**

G_{qs} không thuộc $W_\alpha \rightarrow$ **Chưa có cơ sở bác bỏ giả thiết H_0 (Chấp nhận)**

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- **Kiểm định giả thuyết về kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn và đã biết phương sai**
- **Giả thuyết:** - Tổng thể có phân phối chuẩn **X~N(μ, σ²)**

  - Giả thuyết **H_0: $\mu = \mu_0$ hoặc $\mu \le \mu_0$**   **H_1: $\mu > \mu_0$**

  H_0: $\mu = \mu_0$ hoặc $\mu \ge \mu_0$ **H_1: $\mu < \mu_0$**

  H_0: $\mu = \mu_0$ **H_1: $\mu \ne \mu_0$**

  - Phương sai đã biết **(σ^2 đã biết)**

  -Kiểm định với **mức ý nghĩa α (Giá trị α cho trước)**
- **Thống kê kiểm định:** sử dụng phân phối **U~N(0,1)**

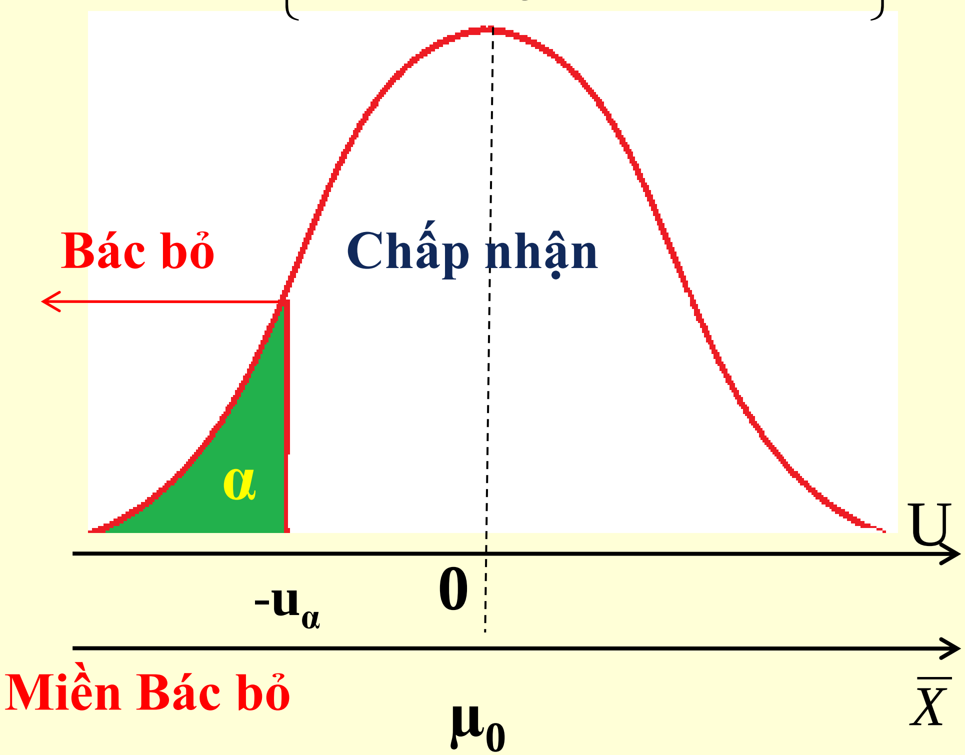
2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

Kiểm định một phía

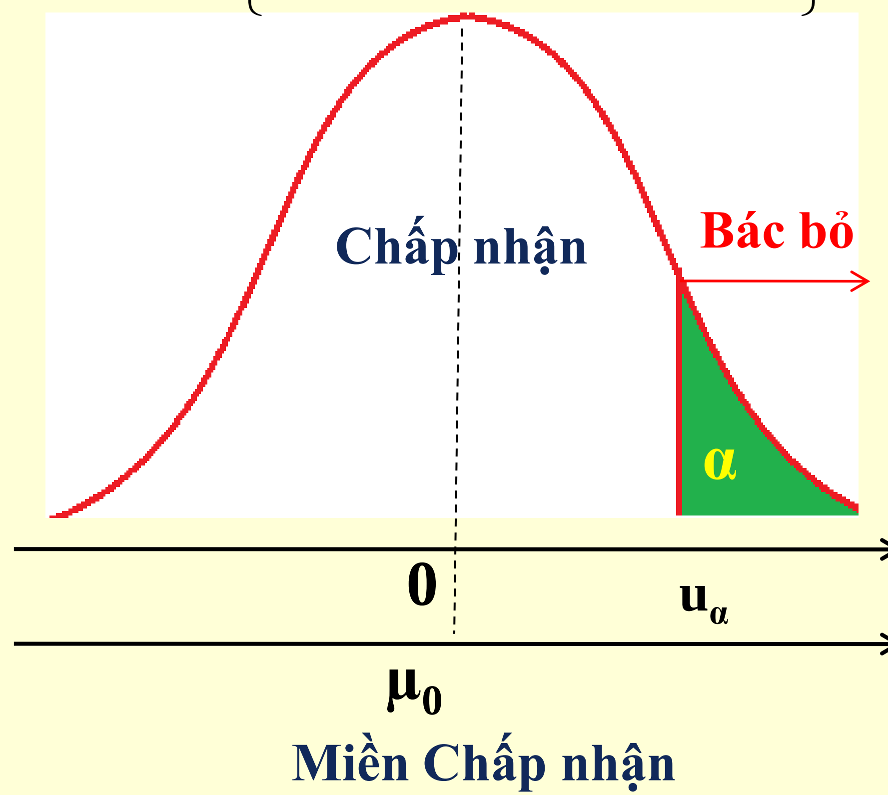
H_0: $\mu \ge \mu_0$ H_1: $\mu < \mu_0$

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}; U < -u_\alpha \right\}$$



H_0: $\mu \le \mu_0$ H_1: $\mu > \mu_0$

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}; U > u_\alpha \right\}$$



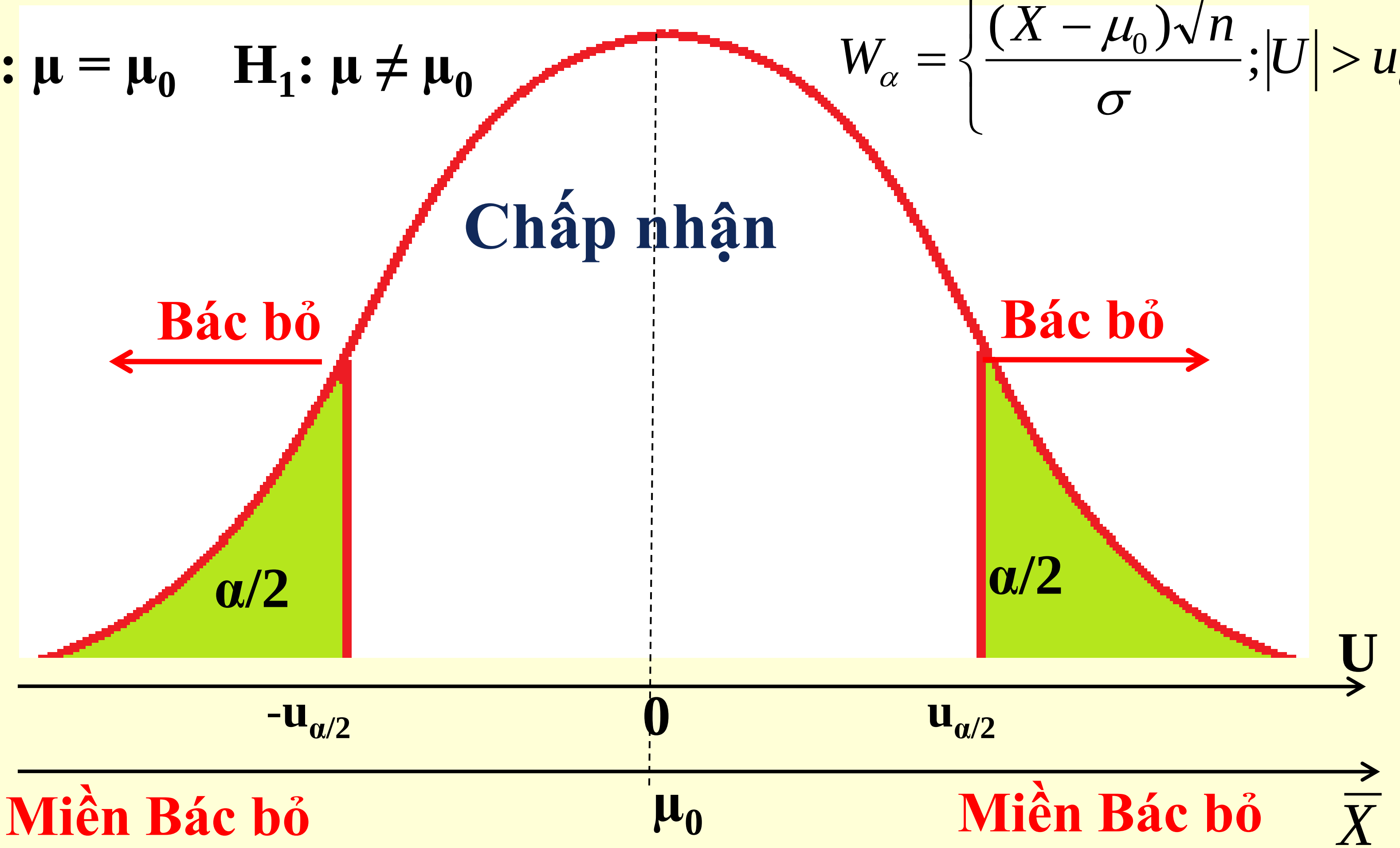
2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

Kiểm định hai phía

$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$

$$W_\alpha = \left\{ \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}; |U| > u_{\alpha/2} \right\}$$



2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

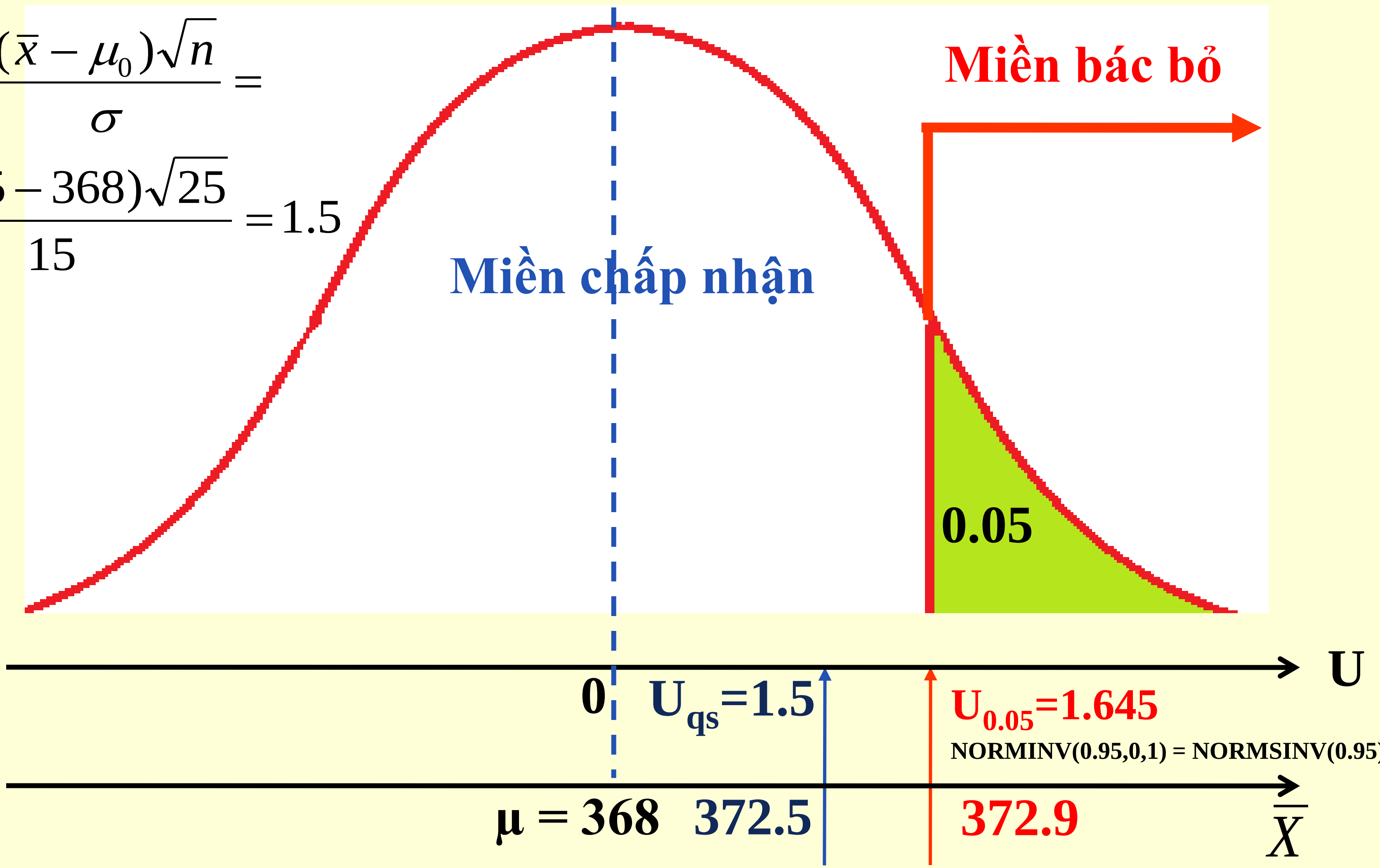
- **Kiểm định 1 phía cho kỳ vọng tổng thể:**
- **Ví dụ:** Để kiểm tra xem trọng lượng trung bình của hộp ngũ cốc có nhiều hơn **368g** hay không, người ta lấy mẫu **25** hộp và thấy rằng trọng lượng trung bình bằng **372.5g**. Công ty xác định độ lệch chuẩn cho phép là $\boldsymbol{\sigma = 15}$**g**. Hãy thực hiện kiểm định giả thuyết với $\boldsymbol{\alpha = 0.05}$. Biết rằng trọng lượng trung bình của hộp ngũ cốc có quy luật phân phối chuẩn.
- **Giải:** Giả thuyết $\mathbf{H_0: \mu \leq 368}$

  $\mathbf{H_1: \mu > 368}$

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

$$U_{qs} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} = \frac{(372.5 - 368)\sqrt{25}}{15} = 1.5$$



2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

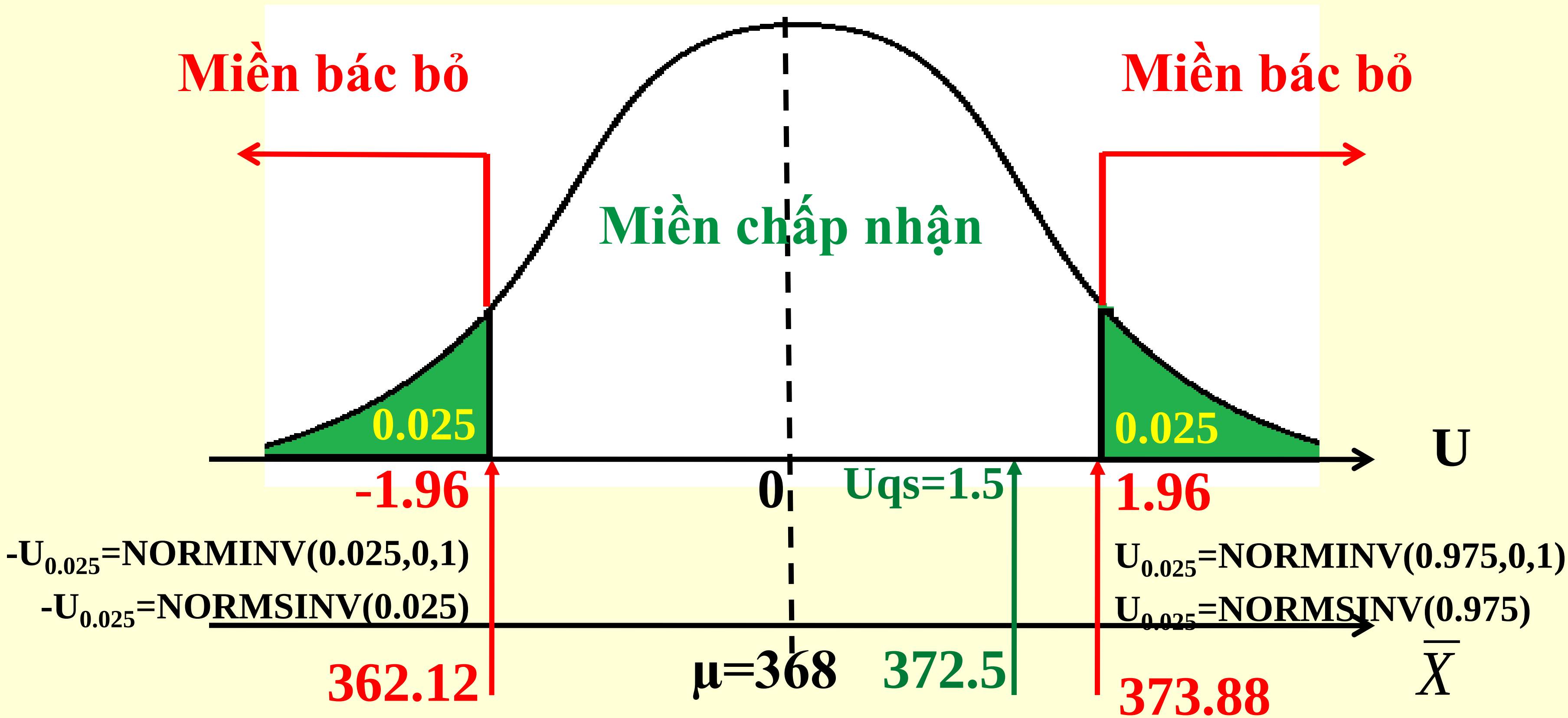
- **Kiểm định 2 phía cho kỳ vọng tổng thể:**
- **Ví dụ:** Để kiểm tra xem trọng lượng trung bình của hộp ngũ cốc có **đúng bằng 368g** hay không, người ta lấy mẫu **25** hộp và thấy rằng trọng lượng trung bình bằng **372.5g**. Công ty xác định độ lệch chuẩn cho phép là $\boldsymbol{\sigma = 15}$**g**. Hãy thực hiện kiểm định giả thuyết với $\boldsymbol{\alpha = 0.05}$. Biết rằng trọng lượng trung bình của hộp ngũ cốc có quy luật phân phối chuẩn.
- **Giải:** Giả thuyết $\mathbf{H_0: \mu = 368}$

  $\mathbf{H_1: \mu \neq 368}$

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

$$U_{qs} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} = \frac{(372.5 - 368)\sqrt{25}}{15} = 1.5$$



2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- **Giá trị xác suất của kiểm định (P-value):**
- Cách kiểm định khác thay cho phương pháp tiếp cận truyền thống

Thay vì kiểm định với giá trị **α** cho trước, thì định rõ các giả thuyết cơ sở $\mathbf{H_0}$ và giả thuyết đối $\mathbf{H_1}$, thu thập mẫu và xác định mức độ khẳng định việc bác bỏ giả thuyết $\mathbf{H_0}$. **Mức độ khẳng định này được gọi là giá trị P (P-value) của kiểm định.**

- Công thức tính giá trị P cho kiểm định giả thuyết thống kê:

Nếu $\mathbf{H_1: \mu>\mu_0}$ **thì P-value=P(U>Uqs)**

Nếu $\mathbf{H_1: \mu<\mu_0}$ **thì P-value=P(U<Uqs)**

Nếu $\mathbf{H_1: \mu\neq\mu_0}$ **thì P-value=P(U>|Uqs|)**

$$U_{qs} = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma}$$

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- **Nguyên tắc kiểm định:**

Nếu P-value >0.1	**Chấp nhận H_0**
Nếu 0.05<P-value <0.1	**Cân nhắc cẩn thận trước khi bác bỏ H_0**
Nếu 0.01<P-value <0.05	**Nghiêng về hướng bác bỏ H_0 nhiều hơn**
Nếu 0.001<P-value<0.01	**Ít băn khoăn khi bác bỏ H_0**
Nếu P-value <0.001	**Hoàn toàn yên tâm bác bỏ H_0**

- **Với α cho trước:**

  P -value < α Bác bỏ H_0, chấp nhận H_1

  P-value > α Chưa có cơ sở bác bỏ H_0 (Chấp nhận H_0)

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- **Ví dụ:** Một máy đóng mì gói tự động quy định trọng lượng trung bình là $\mu_0 = 75g$, độ lệch chuẩn là $\sigma = 15g$. Sau một thời gian sản xuất, kiểm tra **80** gói có trọng lượng trung bình mỗi gói là **72g**. Cho kết luận về tình hình sản xuất với mức ý nghĩa $\alpha = 10\%$. Giả thiết rằng trọng lượng trung bình của mỗi gói mỳ tuân theo quy luật phân phối chuẩn.

- **Giải:** Gọi trọng lượng trung bình của gói mỳ là $X \sim N(\mu,\sigma^2)$. Trọng lượng trung bình quy định là $\mu_0 = 75g$, độ lệch chuẩn là $\sigma = 15g$. Trọng lượng trung bình thực tế chưa biết. **Trọng lượng trung bình mẫu là 72g**. Kiểm định giả thuyết kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn, biết phương sai.

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- Giả thuyết H_0: **$\mu = \mu_0 = 75$**
- Giả thuyết H_1: **$\mu \neq \mu_0$**
- Giá trị kiểm định:

$$U_{qs} = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$$

- Với mức ý nghĩa **α = 10%**, kiểm định 2 phía **$U_{\alpha/2} = U_{0.1/2} = 1.645$**
- **$U_{qs} = -1.79 < -1.645$**
- **Bác bỏ H_0:** Sản xuất không đạt được yêu cầu, trọng lượng trung bình các gói mỳ ít hơn quy định.

- **Kiểm định giả thuyết theo P-value**

 + **P-value = $P(U>|U_{qs}|) = P(U>|-1.79|) = 0.0367$**

 + Tra Bảng $\Phi(1.79)=0.4633 \rightarrow \alpha/2 = 0.5 - 0.4633 = 0.0367$)

 + Excel (1-NORMSDIST(1.79))

 0.01< P-value <0.05 nghiêng về hướng bác bỏ H_0

 * ***$\alpha/2 = 0.5 - 0.4633 = 0.0367 \rightarrow \alpha = 0.0367*2 = 7.34\%$***

 Giả thuyết H_0 sẽ bị bác bỏ ở bất kỳ mức ý nghĩa (α) nào lớn hơn 7.34%

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- **Kiểm định giả thuyết về kỳ vọng toán của biến ngẫu nhiên phân phối theo quy luật chuẩn khi chưa biết phương sai**
- **Giả thuyết:** - Tổng thể có phân phối chuẩn $X \sim N(\mu, \sigma^2)$

  - Giả thuyết $H_0: \mu = \mu_0$ hoặc $\mu \le \mu_0$ $\quad H_1: \mu > \mu_0$

  $H_0: \mu = \mu_0$ hoặc $\mu \ge \mu_0$ $\quad H_1: \mu < \mu_0$

  $H_0: \mu = \mu_0$ $\quad H_1: \mu \neq \mu_0$

  - Phương sai chưa biết **(σ^2 chưa biết)**

  - Kiểm định với **giá trị α cho trước**

Thống kê kiểm định: sử dụng phân phối Student $T \sim T(n-1)$

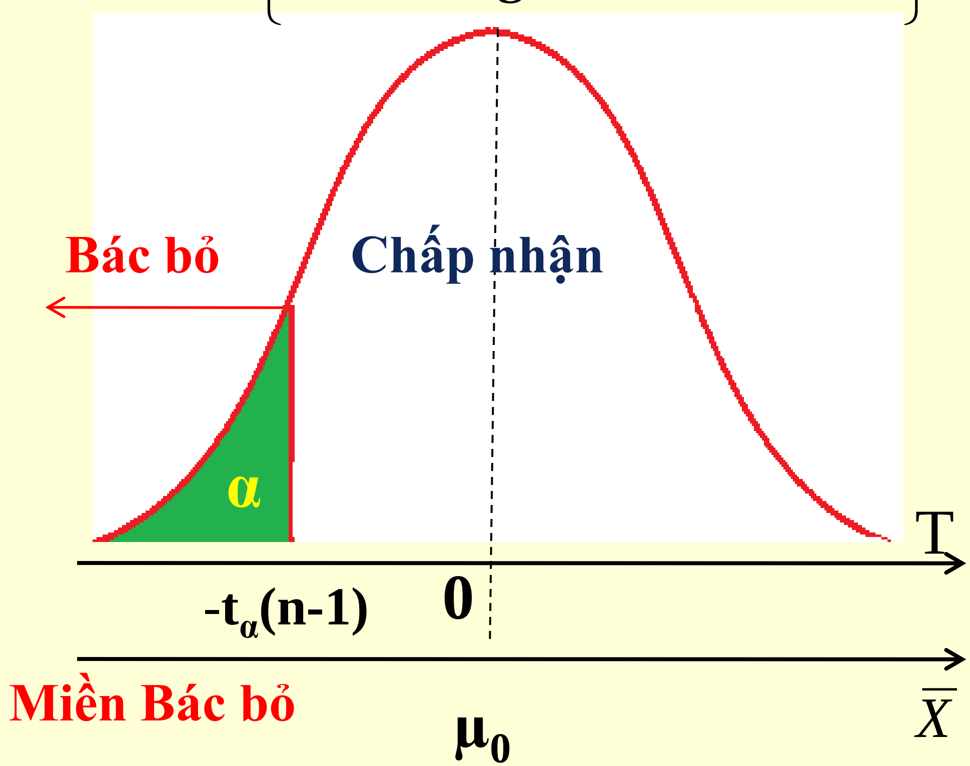
2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

Kiểm định một phía

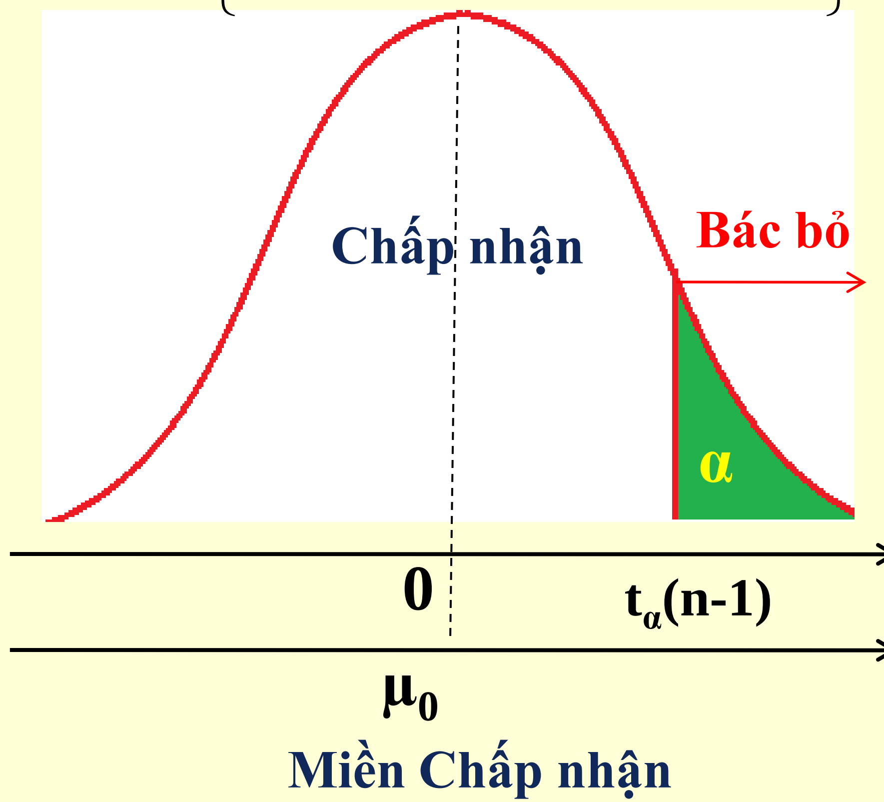
H_0: $\mu \geq \mu_0$ $\quad$ H_1: $\mu < \mu_0$

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; T < -t_\alpha^{(n-1)} \right\}$$



H_0: $\mu \leq \mu_0$ $\quad$ H_1: $\mu > \mu_0$

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; T > t_\alpha^{(n-1)} \right\}$$



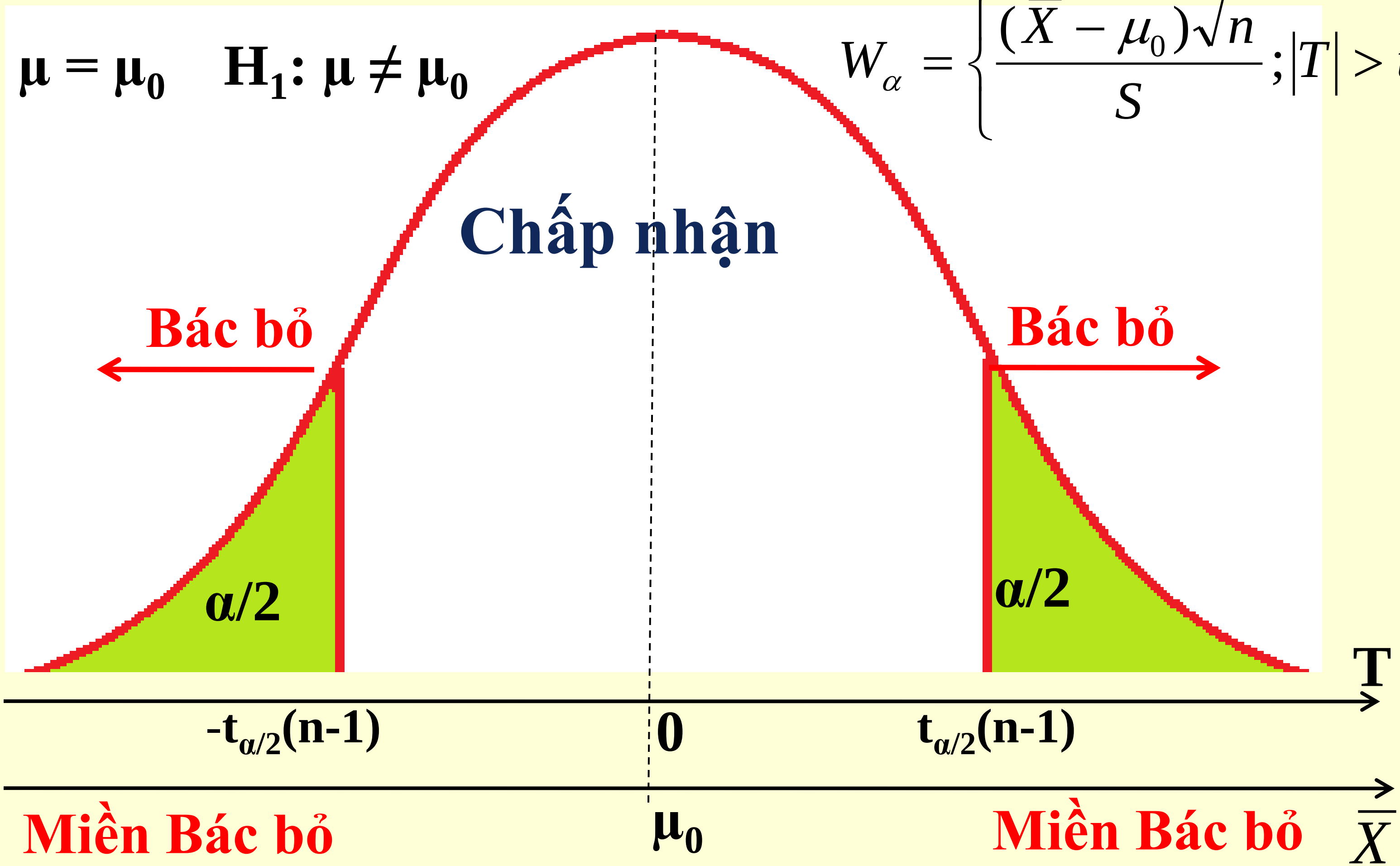
2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

Kiểm định hai phía

$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$

$$W_\alpha = \left\{ \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}; |T| > t_{\alpha/2} \right\}$$



2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- **Ví dụ:** Trọng lượng đóng bao của các bao gạo trong kho là biến ngẫu nhiên **phân phối chuẩn** với trọng lượng trung bình theo quy định là **50kg**. Nghi ngờ gạo bị đóng thiếu, người ta đem cân ngẫu nhiên 25 bao gạo và thu được các số liệu sau:

Trọng lượng bao (Kg)	Số bao gạo tương ứng
48.0 - 48.5	2
48.5 – 49.0	5
49.0 – 49.5	10
49.5 – 50.0	6
50.0 – 50.5	2
Tổng số	**25**

- Với mức ý nghĩa $\boldsymbol{\alpha = 0.01}$, hãy kết luận về điều nghi ngờ nói trên.

2.7 Kiểm định giả thuyết thống kê

2.7.2 Kiểm định kỳ vọng toán của biến ngẫu nhiên phân phối chuẩn

- Gọi **X** là trọng lượng đóng bao. Theo giả thuyết X phân phối chuẩn. Vậy trọng lượng đóng bao trung bình là tham số **μ**. Đây là bài toán kiểm định giả thuyết về kỳ vọng của phân phối chuẩn **N(μ,σ²)** khi chưa biết **σ²**.
- Giả thuyết thống kê: $\mathbf{H_0}$**:** $\boldsymbol{\mu = 50}$ $\mathbf{H_1}$**:** $\boldsymbol{\mu < 50}$
- Tiêu chuẩn kiểm định:

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \qquad S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 n_i}$$

- Giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs} = \frac{(49.27 - 50)\sqrt{25}}{0.53} = -6.887$$

- Miền bác bỏ $\mathbf{W_\alpha}$ (bên trái)

$-T_\alpha(n-1) = -T_{0.01}(25-1) = -2.492$ -TINV(0.01*2, 25-1)

Miền bác bỏ $W_\alpha(-\infty, -2.492)$

- **Vậy T_{qs} thuộc W_α → Bác bỏ H_0, thừa nhận H_1, tức là qua mẫu cụ thể này thừa nhận gạo bị đóng thiếu với mức ý nghĩa 0.01.**

2.7 Kiểm định giả thuyết thống kê

2.7.3 Kiểm định phương sai của biến ngẫu nhiên phân phối chuẩn

- **Kiểm định giả thuyết về phương sai của biến ngẫu nhiên phân phối theo quy luật chuẩn khi chưa biết phương sai (Có cơ sở giả thiết giá trị phương sai là σ^2_0)**
- **Giả thuyết: -** Tổng thể có phân phối chuẩn **$X \sim N(\mu, \sigma^2)$**

  - Giả thuyết $H_0: \sigma^2 \le \sigma^2_0 \,; \sigma^2 \ge \sigma^2_0 \,; \sigma^2 = \sigma^2_0$

$$H_1: \sigma^2 > \sigma^2_0$$

$$H_1: \sigma^2 < \sigma^2_0$$

$$H_1: \sigma^2 \neq \sigma^2_0$$

  - Phương sai đã biết **(σ_0^2 đã biết)**
  - Kiểm định với **giá trị α cho trước**

Thống kê kiểm định: sử dụng phân phối $\chi^2 \sim \chi(n-1)$

2.7 Kiểm định giả thuyết thống kê

2.7.3 Kiểm định phương sai của biến ngẫu nhiên phân phối chuẩn

- **Miền bác bỏ:**

$H_0: \sigma^2 \le \sigma^2_0$	$H_1: \sigma^2 > \sigma^2_0$	$W_\alpha = \{\chi^2 > \chi^2_\alpha\ (n-1)\}$
$H_0: \sigma^2 \ge \sigma^2_0$	$H_1: \sigma^2 < \sigma^2_0$	$W_\alpha = \{\chi^2 < \chi^2_{1-\alpha}\ (n-1)\}$
$H_0: \sigma^2 = \sigma^2_0$	$H_1: \sigma^2 \ne \sigma^2_0$	$W_\alpha = \{\chi^2 < \chi^2_{1-\alpha/2}\ (n-1)$ **hoặc** $\chi^2 > \chi^2_{\alpha/2}(n-1)\}$

Tiêu chuẩn kiểm định:

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

2.7 Kiểm định giả thuyết thống kê

2.7.3 Kiểm định phương sai của biến ngẫu nhiên phân phối chuẩn

- **Ví dụ:** Để kiểm tra độ chính xác của 1 máy, người ta đo ngẫu nhiên kích thước của **15 chi tiết** do máy đó sản xuất và tính được phương sai mẫu **14.6**. Với mức ý nghĩa **0.01** hãy kết luận máy đó có hoạt động bình thường không, biết rằng kích thước chi tiết là biến ngẫu nhiên phân phối chuẩn có dung sai theo thiết kế là **12**.
- **Giải:**

Gọi **X** là kích thước chi tiết, theo giả thiết **X** phân phối chuẩn. Bài toán kiểm định giả thuyết:

$$\mathbf{H_0: \sigma^2 \leq 12}$$

$$\mathbf{H_1: \sigma^2 > 12}$$

2.7 Kiểm định giả thuyết thống kê

2.7.3 Kiểm định phương sai của biến ngẫu nhiên phân phối chuẩn

- **Tiêu chuẩn kiểm định dạng:** $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

$\alpha = 0.01 \rightarrow \chi_\alpha^{2(n-1)} = \chi_{0.01}^{2(14)} = 29.14$

- **Miền bác bỏ dạng: $W_\alpha = \{\chi^2 > \chi^2_\alpha(n-1)\} = \{29.14; +\infty\}$**
- **Với mẫu cụ thể giá trị quan sát của tiêu chuẩn kiểm định:**

$$\chi^2_{qs} = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(15-1)14.6}{12} = 17.033$$

- χ^2_{qs} **không thuộc miền bác bỏ $W_\alpha = \{29.14; +\infty\}$**. Không có cơ sở bác bỏ giả thuyết **H_0**, nói cách khác **Máy hoạt động bình thường.**