Một số phân phối thông dụng

1. Phân phối nhị thức:

- Định nghĩa:
- **Phân phối nhị thức** là một dạng phân phối rời rạc thường dùng trong thống kê, ngược lại của các dạng phân phối liên tục như phân phối chuẩn.
- Điều này là vì phân phối nhị thức chỉ tính đến hai trường hợp, thường được thể hiện là 1 (cho thành công) hoặc 0 (cho thất bại) trong một số lượng lần thử.

- **Phân phối nhị thức** thể hiện xác suất để **x** thành công trong **n** phép thử.
- Xác suất thành công **p** của mỗi phép thử.
- Mỗi phép thử có hai khả năng thành công hoặc thất bại.
- Nếu p là xác suất thành công thì 1-p = q là xác suất thất bại.

- Tính chất:

Phân phối xác suất của phân phối nhị thức được tính toán như sau: Với x lần thành công trong n phép thử, ta được công thức

$$\mathrm{P(X = 1) = C_n^x}.\,p^{\mathrm{x}}.\,(1 - \mathrm{p})^{\mathrm{n-x}} = C_n^x p^x q^{n-x}$$

- Trung bình: $\mathrm{E(X)} = n\mathrm{p}$.
- Phương sai: $\mathrm{Var(X)} = n\mathrm{p}(1 - \mathrm{p})$

- Ví dụ:
- Tỷ lệ các thí nghiệm thành công trong một viện nghiên cứu là 25%. Tiến hành quan sát 5 cuộc thí nghiệm của viện nghiên cứu.
- Gọi X là số thí nghiệm thành công trong 5 cuộc thí nghiệm đó.

- Khi đó X nhận các giá trị: 0, 1, 2, 3, 4, 5 với xác suất :

$$P(X = x) = C_5^x(0{,}25)^x(0{,}75)^{5-x}$$

- Vậy X là biến ngẫu nhiên có phân phối nhị thức B(5, 0,25).

Bài tập:

- Biết tỉ lệ suy dinh dưỡng ở trẻ em dưới 5 tuổi là 20%. Nếu chúng ta khám 15 trẻ dưới 5 tuổi. Tính xác suất để có 3 em bị suy dinh dưỡng.

- Tỉ lệ suy dinh dưỡng trẻ em 20% (như ví dụ trên).Tính xác suất để có ít hơn 4 em bị suy dinh dưỡng.

2. Phân phối đa thức:

- Định nghĩa: **Phân phối đa thức** là tổng quát của phân phối nhị thức.
- Cho nhiều hơn hai giá trị.
- Giống như phân phối nhị thức, phân phối đa thức là một hàm phân phối cho các quá trình rời rạc trong đó xác suất cố định chiếm ưu thế cho mỗi giá trị được tạo độc lập.

- Mặc dù các quy trình liên quan đến phân phối đa thức có thể được nghiên cứu bằng cách sử dụng phân phối nhị thức bằng cách tập trung vào một kết quả quan tâm và kết hợp tất cả các kết quả khác thành một loại (đơn giản hóa phân phối thành hai giá trị), phân phối đa thức hữu ích hơn khi tất cả các kết quả là quan tâm.

- Công thức xác suất của phân phối đa thức được tính toán như sau:

❖Xét một quá trình có tập *k* kết quả có thể có $(\mathrm{X}_1, \mathrm{X}_2\ , \mathrm{X}_3\ ,..., \mathrm{X}_\mathrm{k}\ )$ với các xác suất liên quan $(\mathrm{p}_1, \mathrm{p}_2, \dots, \mathrm{p}_\mathrm{k})$

❖$\Sigma\ p_i = 1$

❖ Sau đó, với n lần thử lặp lại của quá trình, hãy đặt x_i cho biết số lần kết quả X_i xảy ra, với điều kiện $0 \leq \mathrm{x}_\mathrm{i} \leq \mathrm{n}$ và $\sum \mathrm{x}_\mathrm{i} = \mathrm{n}$.

- Khi đó ta có:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{\prod_{i=1}^{k} x_k!} \prod_{i=1}^{k} p_i^{x_i}$$

Ví dụ:

- Ví dụ: Ba người cùng chơi bài với nhau. Xác suất người chơi A thắng ở bất kỳ ván nào là 20%, xác suất người chơi B thắng là 30%, và xác suất người chơi C thắng là 50%. Nếu họ chơi 6 ván, xác suất người chơi A thắng 1 ván, người chơi B thắng 2 ván và người chơi C thắng 3 ván là bao nhiêu?

Ví dụ:

- n = 6
- $n_1 = 1$ (số ván người chơi A thắng).
- $n_2 = 2$ (số ván người chơi A thắng).
- $n_3 = 3$ (số ván người chơi A thắng).
- $p_1 = 0{,}2$ (xác suất người chơi A thắng).
- $p_2 = 0{,}3$ (xác suất người chơi A thắng).
- $p_3 = 0{,}5$ (xác suất người chơi A thắng).

Ví dụ:

- Như vậy ta sẽ có công thức:

$$P(A = 1, B = 2, C = 3) = \frac{6!}{1!\,2!\,3!}(0{,}2).(0{,}3)^2(0{,}5)^3 = 0{,}135$$

Bài tập:

- Ta bốc 5 lá bài bất kì từ một bộ bài Tây. Tìm xác suất bốc được 1 lá cơ, 1 lá rô, 1 lá chuồn và 2 lá bích.
- Giả sử ta có một hộp bi gồm 10 viên bi. Trong đó có 2 bi đỏ, 3 bi xanh và 5 bi vàng. Ta sẽ bốc 4 viên bi ra khỏi hộp bi trên cùng một lúc. Hỏi xác suất ta bốc được 2 bi xanh và 2 bi vàng là bao nhiêu.

3. Phân phối Poisson:

- Định nghĩa: phân phối Poisson là một phân phối xác suất rời rạc.
- là trung bình số lần xảy ra thành công của một sự kiện trong một khoảng thời gian nhất định.
- không phải là xác suất để một sự kiện xảy ra thành công trong một lần thử như trong phân phối Bernoulli hay là số lần mà sự kiện đó xảy ra trong n lần thử như trong phân phối nhị thức.

- Tiến hành n phép thử độc lập với biến cố A có hai trường hợp xảy ra (kí hiệu là A = 1) và không xảy ra (A = 0).
- Trong trường hợp này, ta cố định:

$$\lambda = \mathrm{np}$$

- Tham số λ chính là giá trị trung bình và cũng là phương sai.
- $E(X) = Var(X) = \lambda$

- Từ mô tả trên, ta được công thức:

$$P(k) = P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

- Công thức trên được xấp xĩ thành

$$\frac{\lambda^k}{k!}\, e^{-\lambda}$$

Ví dụ:

- Một trạm cho thuê xe ô tô có 3 xe. Mỗi chiếc xe cho thuê được với giá 100 nghìn/ngày. Giả sử yêu cầu thuê xe của trạm là biến ngẫu nhiên X có phân phối Poisson với tham số $\lambda = 3$.
- Tính xác suất trong một ngày có 3 khách thuê .
- Tính tiền lãi trung bình trạm thu được trong một ngày.

Ví dụ:

- Xác suất để một ngày có 3 khách thuê là:

$$P(X=3)=\frac{e^{-3}.3^{3}}{3!}\cong$$

- Tiền lãi trung bình: Gọi Y là biến ngẫu nhiên tiền lãi. Ta sẽ chia thành nhiều trường hợp để xét bài toán trên.

- Trường hợp không xe nào được thuê:

$$P(Y = 0) = P(X = 0) = \frac{e^{-3}.3^0}{0!}$$

- Trường hợp n xe được thuê với $1 \le n \le 3$:

$$P(Y = 100n) = P(X = n) = \frac{e^{-3}.3^n}{n!}$$

- Tiền lãi kì vọng (trung bình) của nhà xe là:

$$E(Y) = \sum_{i=0}^{4} P(Y = i.100).100i$$

Bài tập:

- Cũng là với bài toán trên, nếu ta thêm tiền thuế mỗi ngày cho quản lí thị trường là 50k/ ngày thì lợi nhuận trung bình của nhà xe sẽ là bao nhiêu ?

4. Phân phối chuẩn:

- Định nghĩa:
- Phân phối liên tục có đồ thị hình chuông.
- Là dạng phân phối liên tục trực quan nhất.
- Đối xứng xung quanh giá trị trung bình.
- Biểu diễn của phân phối chuẩn phụ thuộc vào giá trị trung bình và phương sai.
- Tạo nên hàm mật độ xác suất.

Công thức của phân phối chuẩn:

- Khi nhắc đến phân phối chuẩn, ta sẽ có công thức về hàm mật độ xác suất và hàm phân phối xác suất dựa trên hai đại lượng μ- trung bình và σ-phương sai như sau:

❖ Hàm mật độ:

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

❖Hàm phân phối:

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

- Dạng đặc biệt của phân phối chuẩn: Phân phối chuẩn tắc.
- Ta nói Z là một phân phối chuẩn tắc nếu $\mathrm{Z} \sim \mathrm{N}(0,1)$.
- Công thức phân phối xác suất được viết lại thành:

$$\Phi_0(\mathrm{x}) = \int_{-\infty}^{\mathrm{x}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(\mathrm{x})^2}{2}} \mathrm{dx}$$

Phép tính với kì vọng và phương sai:

- Cho biến ngẫu nhiên $\mathrm{X} \sim \mathrm{X}_1 + \mathrm{X}_2$ với $\mathrm{X}_\mathrm{i} \sim \mathrm{N}(\mu_\mathrm{i}, \sigma_\mathrm{i}{}^2)$, khi đó

$$\mathrm{X} \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- Cho biến ngẫu nhiên $\mathrm{X} \sim \mathrm{X}_1 - \mathrm{X}_2$ với $\mathrm{X}_\mathrm{i} \sim \mathrm{N}(\mu_\mathrm{i}, \sigma_\mathrm{i}{}^2)$, khi đó

$$\mathrm{X} \sim \mathrm{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

- Khi tính toán cụ thể với phân phối chuẩn thì ta có thể đưa về dạng phân phối chuẩn tắc như sau:

❖Cho $\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2)$, khi đó $\mathrm{Z} = \frac{\mathrm{X}-\mu}{\sigma} \sim \mathrm{N}(0,1)$ (tại sao ?)

❖$\mathrm{P}(\mathrm{a} \leq \mathrm{x} \leq \mathrm{b}) = \mathrm{P}(\frac{\mathrm{a}-\mu}{\sigma} \leq \frac{\mathrm{x}-\mu}{\sigma} \leq \frac{\mathrm{b}-\mu}{\sigma})$

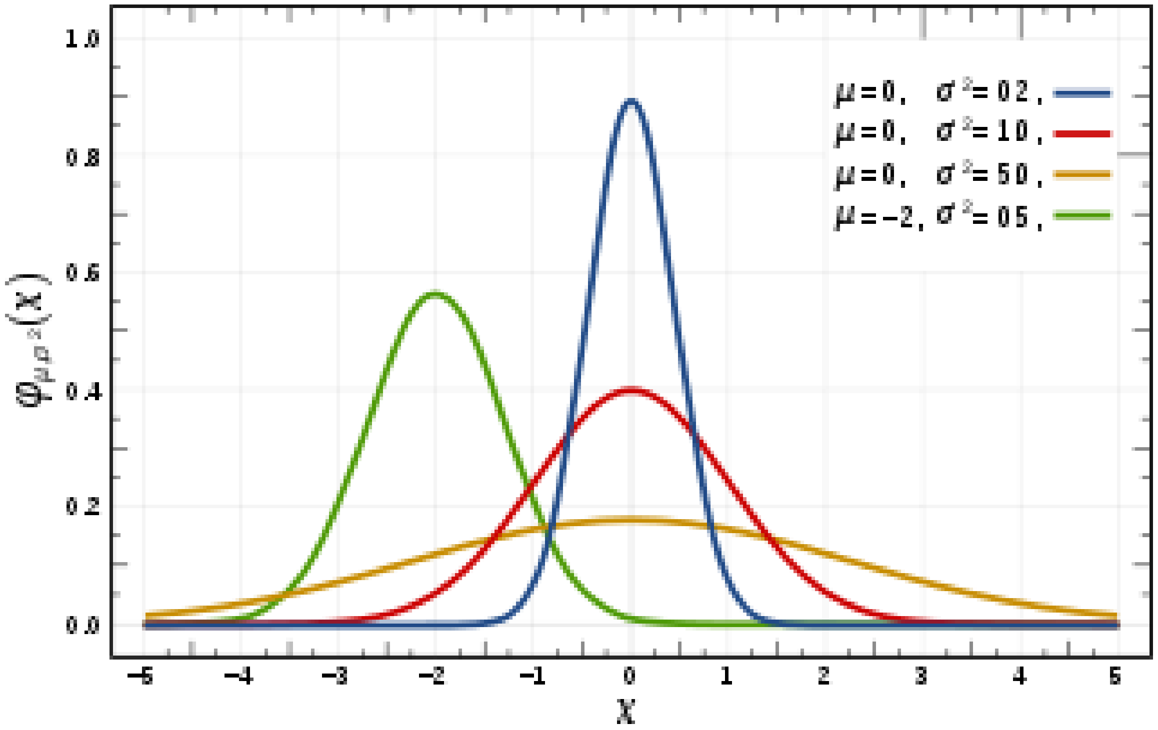
$$= \Phi_0\left(\frac{\mathrm{b}-\mu}{\sigma}\right) - \Phi_0(\frac{\mathrm{a}-\mu}{\sigma})$$

- $P(x \leq b) = P\left(\frac{x-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \frac{1}{2} + \Phi_0(\frac{b-\mu}{\sigma})$

- $P(x > a) = P\left(\frac{x-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = \frac{1}{2} - \Phi_0(\frac{b-\mu}{\sigma})$

- Từ những công thức tính toán trên, ta thấy sự thuận tiện của việc đưa về phân phối chuẩn tắc.
- Lưu ý: bảng giá trị phân phối chuẩn ta thường thấy sau sách là bảng giá trị của phân phối chuẩn tắc.

Đồ thị của Phân phối chuẩn:



Ví dụ:

- Trọng lượng của một loại sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn với họng lượng trung bình $\mu = 5$ kg và độ lệch tiêu chuẩn $\sigma = 0,1$.
- Tính tỷ lệ những sản phẩm có trọng lượng từ 4,9 kg đến 5,2 kg?

- Gọi X là trọng lượng của loại sản phẩm này.
- Ta có $\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2)$ với $\mu = 5$ và $\sigma = 0{,}1$
- Tỷ lệ sản phẩm có trọng lượng từ 4,9 đến 5,2 kg chính là:

$$\mathrm{P}(4{,}9 \leq \mathrm{X} \leq 5{,}2) = \Phi_0\left(\frac{5{,}2 - 5}{0{,}1}\right) - \Phi_0\left(\frac{4{,}9 - 5}{0{,}1}\right)$$

Bài tập:

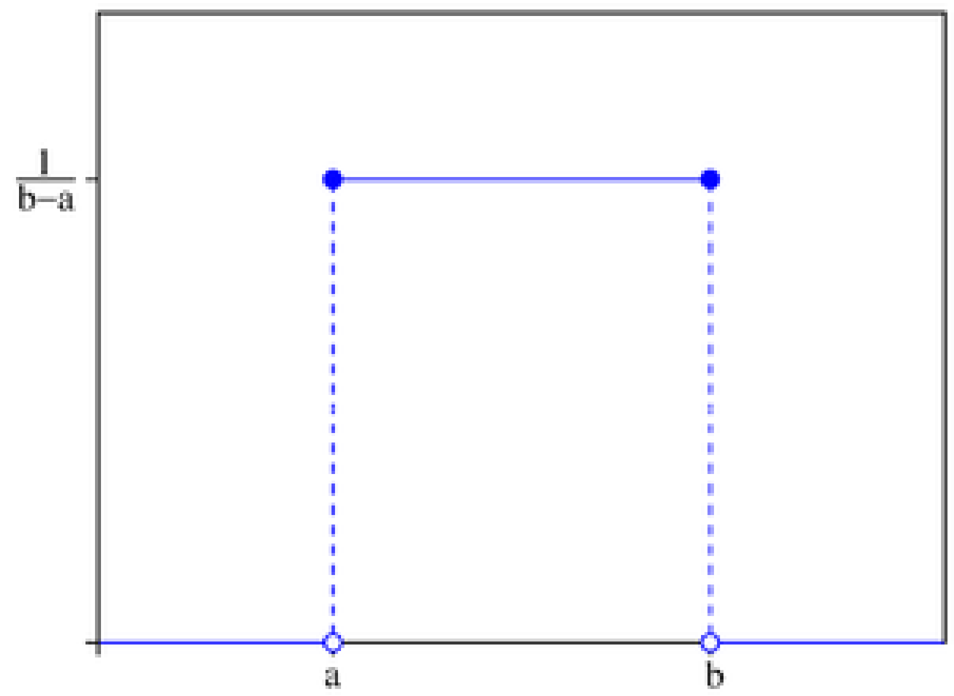
- Năng suất của một loại cây ăn quả là một biến ngẫu nhiên phân phối chuẩn với năng suất trung bình là 15 kg/cây và độ lệch chuẩn là 2 kg. Cây đạt tiêu chuẩn hàng hoá là cây có năng suất tối thiểu là 12 kg.
- Câu hỏi: Tính tỉ lệ cây đạt tiêu chuẩn hàng hoá.

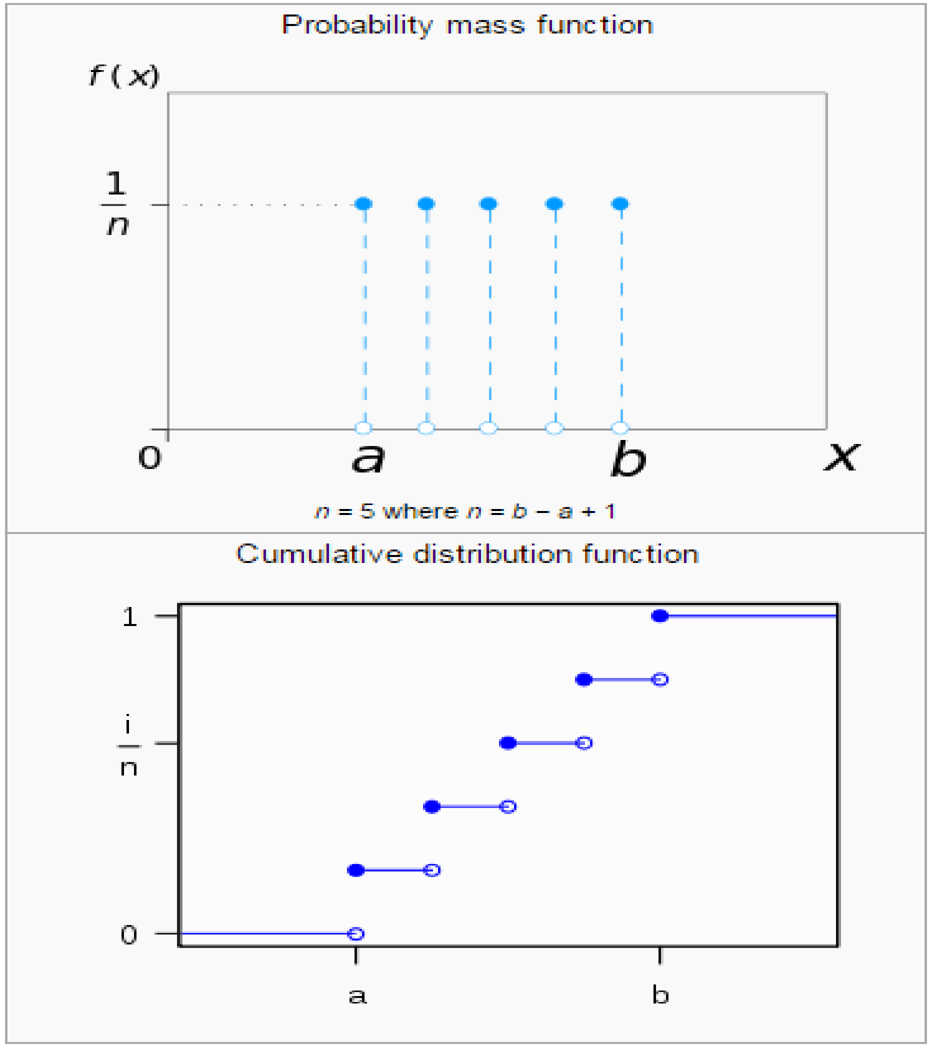
5. Phân phối đều:

Định nghĩa:

- **Phân phối đều liên tục** là một phân phối mà xác suất xảy ra như nhau cho mọi kết cục của biến ngẫu nhiên liên tục.
- Phân phối đều liên tục đôi khi còn được gọi là **phân phối hình chữ nhật** và khi biểu diễn bằng hình vẽ sẽ có dạng hình chữ nhật.
- Phân phối đều rời rạc là phân phối của biến ngẫu nhiên X trong đó X nhân giá trị trong một tập hữu hạn và X nhận giá trị bằng mỗi phần tử của tập đó với xác suất bằng nhau.

Đồ thị của hàm phân phối đều:





Probability mass function
f (x)
1/n
0
a
b
x
n = 5 where n = b − a + 1
Cumulative distribution function
1
i/n
0
a
b

Tính chất:

- Ta sẽ chỉ xét trường hợp liên tục.
- Hàm mật độ và phân phối xác suất cho phân phối đều trong đoạn [a,b] có dạng như sau:

$$\mathrm{f(x)} = \frac{1}{\mathrm{b-a}} \text{ với } \mathrm{x} \in [\mathrm{a,b}] \text{và } \mathrm{f(x)} = 0 \text{ với } \mathrm{x} \notin [\mathrm{a,b}]$$

$$\mathrm{F(x)} = \frac{1}{\mathrm{b-a}}\mathrm{x} \text{ với } \mathrm{x} \in [\mathrm{a,b}] \text{và } \mathrm{F(x)} = \mathrm{const} \text{ với } \mathrm{x} \notin [\mathrm{a,b}]$$

6. Phân phối chi bình phương:

- Định nghĩa: Phân phối chi bình phương (χ^2) là phân phối một phân phối liên tục.
- Được biểu diễn dưới dạng tổng hữu hạn các bình phương của phân phối chuẩn tắc.
- Ý tưởng của phân phối chi bình phương: Xét một phân phối chuẩn $=$ $\mathrm{X} = \mathrm{N}(\mathrm{a}, \mathrm{b})$, chuyển về phân phối chuẩn tắc $\mathrm{Z} =\ \mathrm{N}(0{,}1)$ ta có:

$$\mathrm{Z} = \frac{\mathrm{X} - \mathrm{a}}{\mathrm{b}}$$

- Khi đó, ta được:

$$\mathrm{Z}^2 =$$

- Đây là phân phối chi bình phương tương ứng với một lần thử.
- Giả sử ta thử nghiệm thêm nhiều lần.

- Khi đó ta sẽ được:

$$Z^2 = \sum_{i=1}^{n} \left(\frac{X_i - a_i}{b_i} \right)^2$$

Chính là công thức chi bình phương tổng quát.

- Nói cách khác:

$$\chi^2(\mathrm{n}) \sim \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X}_{\mathrm{i}}^2$$

Với X_{i} là các biến ngẫu nhiên theo phân phối chuẩn tắc.

- Một công thức chung cho phân phối chi bình phương có thể được viết như sau:

❖ Xét phân phối $\chi^2(\mathrm{n})$ với bậc tự do n, ta có hàm phân phối xác suất của phân phối chi bình phương là

$$\mathrm{f(x) = 0 \text{ nếu } x \leq 0 \text{ và } f(x) = } \frac{\mathrm{x}^{\frac{\mathrm{n}}{2}}.\mathrm{e}^{-\frac{\mathrm{x}}{2}}}{2^{\frac{\mathrm{n}}{2}}.\Gamma(\frac{\mathrm{k}}{2})}$$

- Ý nghĩa của phân phối chi bình phương:

❖Phân phối chi bình phương đại diện cho độ chênh lệnh giữa giá trị thực tế thu thập được và giá trị kì vọng.

❖Giá trị tự do n thể hiện cho n phép thử khác nhau và độc lập với nhau.

❖Các phân phối trong thực tiễn có thể không tuân theo quy luật của phân phối chuẩn tắc, nhưng phân phối chuẩn tắc và kéo theo phân phối chi bình phương sẽ trở thành một giá trị “đẹp” để làm tiêu chuẩn kiểm định.

Đồ thị của phân phối chi bình phương: