Bài số 7

MỘT SỐ PHÂN PHỐI XÁC SUẤT THƯỜNG GẶP

Thông thường, các quan sát thu được từ những thí nghiệm mang tính thống kê khác nhau đều có cùng kiểu đặc tính chung. Do đó các biến ngẫu nhiên liên kết với các phép thử, về bản chất, có thể được mô tả bởi cùng một phân phối xác suất vì thế có thể được biểu thị bởi chỉ một công thức. Như vậy, trong tính toán, ta chỉ cần một số phân phối xác suất quan trọng để mô tả nhiều biến ngẫu nhiên.

A. MỘT SỐ PHÂN PHỐI XÁC XUẤT RỜI RẠC

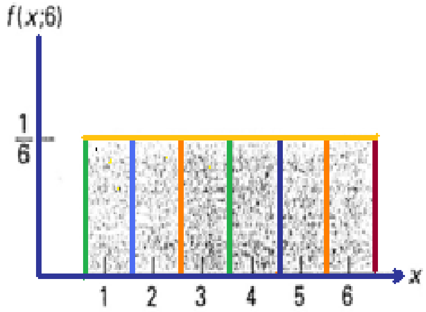
I. Phân phối đều rời rạc

1. Định nghĩa. Giả sử biến ngẫu nhiên rời rạc X với miền giá trị là $\{x_1, x_2, \ldots, x_k\}$ và xác suất để X nhận mỗi giá trị có thể của nó là bằng nhau: $P(X = x_i) = P(X = x_j), \forall i \neq j$. Khi đó ta nói rằng biến ngẫu nhiên X có phân phối đều rời rạc, và ta có:

$$f(x;k) = \frac{1}{k},\ x \in \{x_1, x_2, \ldots, x_k\}.$$

Ví dụ 1. Tung một con xúc xắc, mỗi một phần tử của không gian mẫu $S = \{1,2,3,4,5,6\}$ xuất hiện với xác suất đều bằng 1/6. Do đó, ta có phân phối đều với: $f(x;6) = \frac{1}{6},\ x = 1,2,3,4,5,6.$

Minh hoạ hình học phân phối đều là các biểu đồ chỉ gồm các hình chữ nhật với chiều cao bằng nhau.



Hình 1 :Phân phối cho BNN trong phép thử tung xúc sắc

2. Các tham số đặc trưng.

Cho BNN rời rạc X với miền giá trị là $\{x_1, x_2, \ldots, x_k\}$. Giá trị trung bình (kỳ vọng) và phương sai của phân phối đều rời rạc $f(x;k)$ là

$$E(X) = \mu = \frac{\sum_{i=1}^{k} x_i}{k} \quad \text{và} \quad \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}.$$

Ví dụ 2. Trong phép thử tung xúc sắc như trong Ví dụ 1, ta tìm được

$$E(X) = \mu = \frac{1+2+3+4+5+6}{6} = 3,5$$

Và: $$\sigma^2 = \frac{(1-3,5)^2 + (2-3,5)^2 + \cdots + (6-3,5)^2}{6} = \frac{35}{12}$$

II. Phân phối nhị thức và phân phối đa thức

1.Phân phối nhị thức.

a. Phép thử Bernoulli.

Một thí nghiệm thường bao gồm nhiều **phép thử được lặp đi lặp lại**, mỗi phép thử với hai biến cố, ta có thể đặt tên cho chúng là **biến cố thành công** và **biến cố thất bại.** Chẳng hạn, khi rút (theo phương thức hoàn lại) các quân bài liên tiếp từ một bộ bài tú lơ khơ và mỗi lần rút được coi là thành công hay thất bại tuỳ thuộc vào việc quân bài rút được có phải là quân cơ hay không. Khi mỗi quân bài được hoàn lại và xáo cỗ bài trước khi rút quân tiếp theo, cả hai lần rút quân bài đều có các tính chất tương tự nhau, đó là *các phép thử độc lập và xác suất thành công trong mỗi phép thử đều bằng nhau*. Quá trình vừa được đề cập đến được gọi là **quá trình Bernoulli** và mỗi phép thử trong quá trình đó được gọi là **phép thử Bernoulli.**

Định nghĩa. ***Phép thử Bernoulli*** là một quá trình thỏa mãn đồng thời các tính chất sau:

1. Một thí nghiệm gồm n phép thử ***cùng loại được lặp đi lặp lại***.

2. Mỗi biến cố của một phép thử được **phân loại theo biến cố thành công hoặc biến cố thất bại**.

3. ***Xác suất thành công trong mỗi phép thử đều bằng nhau*** và được kí hiệu là p.

4. Các phép thử là **độc lập**.

b. Phân phối nhị thức.

Định nghĩa. Số lần thành công X trong n phép thử Bernoulli được gọi là **biến ngẫu nhiên nhị thức**. Phân phối xác suất của BNN rời rạc này được gọi là **phân phối nhị thức**. Xác suất được kí hiệu là $b(x; n; p)$ - bởi vì nó phụ thuộc vào số phép thử và xác suất thành công trong mỗi phép thử

Công thức tính: Cho phép thử Bernoulli với xác suất thành công là p và thất bại là $q = 1 - p$. Phân phối xác suất của biến ngẫu nhiên nhị thức X (số lần thành công trong n phép thử độc lập), là

$$b(x;n,p) = P(X = x) = C_n^x p^x q^{n-x},\ x = 0,1,2,\ldots,n.$$

Ví dụ 3. Xác suất để một loại bộ phận nào đó còn sống sau một cú va chạm là 3/4. Tìm xác suất để có đúng 2 trong 4 thành phần còn sống sau cú va chạm.

Giải: + Bốn bộ phận tham gia vào cú va chạm đó cho ta bốn phép thử độc lập với xác suất $p = {}^3\!/_4$,

+ Đây là một phép thử Bernoulli với $n = 4;\ p = {}^3\!/_4$ và BNN rời rạc X có phân phối nhị thức:

$$b(2;4,\frac{3}{4}) = C_4^2 \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = \frac{4!}{2!2!}.\frac{3^2}{4^2} = \frac{27}{128}.$$

Chú ý:

1. Do $p + q = 1$ nên ta được: $\sum_{x=0}^{n} b(x;n,p) = 1$

2. Nhiều khi ta cần tính $P(X < r)$ và $P(a \le X \le b)$. Khi đó ta cần các kết quả đã được tính sẵn, các tổng nhị thức: $B(r;n,p) = \sum_{x=0}^{r} b(x;n,p)$ đã được tính sẵn và ghi trong Bảng A.1 trong phần phụ lục, với $n = 1,2,\ldots,20$ và các giá trị xác suất p từ 0,1 đến 0,9.

Ví dụ 4. Xác suất để một bệnh nhân sống sót sau khi mắc một loại bệnh hiếm thấy về máu là 0,4. Nếu biết rằng đã có 15 người mắc loại bệnh này, tìm xác suất để:

a. Có ít nhất 10 người sống sót;
b. Có từ 3 đến 8 người sống sót;
c. Có đúng 5 người sống sót.

Giải: Gọi X là số người sống sót.

a. Xác suất để có ít nhất 10 người sống sót là:

$$P(X \ge 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x;15,\ 0,4) = 1 - 0,9662$$
$$= 0,0338.$$

b. Xác suất để có từ 3 đến 8 người sống sót là:

$$P(3 \le X \le 8) = \sum_{x=0}^{8} b(x;15,\ 0,4) - \sum_{x=0}^{2} b(x;15,\ 0,4) = 0,9050 - 0,0271$$
$$= 0,8779.$$

c. Xác suất để có đúng 5 người sống sót là:

$$P(X = 5) = b(5;15,\ 0,4) = \sum_{x=0}^{5} b(x;15,\ 0,4) - \sum_{x=0}^{4} b(x;15,\ 0,4)$$
$$= 0,4032 - 0,2173 = 0,1859.$$

c.Tham số đặc trưng.

Giá trị trung bình (kỳ vọng) và phương sai của phân phối nhị thức $b(x;\ n,\ p)$ được xác định bởi:

$$\mu = np \text{ và } \sigma^2 = npq$$

2. Phân phối đa thức.

a. Định nghĩa. Phép thử nhị thức trở thành **phép thử đa thức** nếu mỗi phép thử có ***nhiều hơn hai kết quả***. Khi đó phân phối xác suất của phép thử đa thức được gọi là ***phân phối đa thức***.

Ví dụ: + Sự phân loại sản phẩm của một dây chuyền sản xuất dựa vào việc sản phẩm nặng, nhẹ.

+ Việc rút lần lượt từng quân bài từ một bộ bài tú lơ khơ theo phương thức *có hoàn lại* và ta quan tâm đến việc rút được chất nào (rô, cơ, bích, nhép). Đó là các phép thử đa thức.

Nếu một phép thử có *k* kết cục E_1, $E_{2, \dots,}E_k$ với xác suất tương ứng là p_1, $p_{2,\dots,}p_k$, thì **phân phối đa thức** sẽ cho ta xác suất để E_1 xuất hiện x_1 lần, E_2 xuất hiện x_2 lần,…,E_k xuất hiện x_k lần trong *n* phép thử độc lập, trong đó : $x_1 + x_2 + \cdots + x_k = n$

Ta kí hiệu phân phối xác suất đồng thời này là : $f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n)$

Dễ thấy, $p_1 + p_2 + \cdots + p_k = 1$, vì các kết quả của phép thử phải là một trong *k* kết cục có thể.

<u>Công thức tính</u>: Nếu một phép thử có *k* kết cục E_1, $E_{2, \dots,}E_k$ với xác suất tương ứng là p_1, $p_{2,\dots,}p_k$, thì phân phối xác suất của các biến ngẫu nhiên $X_1, X_2, \dots, X_k$ biểu thị số lần xuất hiện của E_1, $E_{2,\dots,}E_k$ tương ứng, trong dãy *n* phép thử độc lập là

$$f(x_1, x_2, .., x_k; p_1, p_2, \dots, p_k, n) = C_n^{x_1, x_2, \dots, x_k} p_1^{x_1}.p_2^{x_2}...p_k^{x_k}$$

trong đó $\sum_{i=1}^{k} x_i = n$, $\sum_{i=1}^{k} p_i = 1$.

Ví dụ 5. Tung một cặp xúc xắc 6 lần, tính xác suất để: tổng số chấm xuất hiện là 7 hoặc 11 xuất hiện hai lần, số chấm trên hai con là như nhau xuất hiện một lần, và các trường hợp còn lại xuất hiện 3 lần.

<u>Giải:</u> + Ta có các biến cố,

E_1 : tổng số chấm xuất hiện là 7 hoặc 11, khi đó $p_1 = P(E_1) = {}^2\!/_9$

E_2: số chấm xuất hiện bằng nhau, khi đó $p_2 = P(E_2) = {}^1\!/_6$

E_3: tổng số chấm xuất hiện không bằng 7, không bằng 11 và số chấm xuất hiện cũng không bằng nhau, khi đó $p_3 = P(E_3) = {}^2\!/_9$.

+ Các giá trị này là như nhau ở cả 6 lần tung.

+ Dùng phân phối đa thức với $x_1 = 2$, $x_2 = 1$, và $x_3 = 3$, ta được xác suất cần tìm

$$f\left(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6\right) = C_6^{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$$
$$= \frac{6!}{2!1!3!}\frac{2^2}{9^2}\frac{1}{6}\frac{11^3}{18^3} = 0{,}1127.$$

III. <u>Phân phối siêu bội.</u>

Các kiểu ứng dụng phân phối siêu bội tương tự như phân phối nhị thức: quan tâm đến việc tính các xác suất của số lượng các kết cục rơi vào một kiểu đặc biệt.

Trong trường hợp phân phối nhị thức: các phép thử là độc lập. Như vậy, chỉ dùng được phân phối nhị thức khi mẫu được lấy từ tổng thể có số lượng đông đảo (bộ bài, các sản phẩm từ một dây chuyền sản xuất), việc lấy mẫu phải được tiến hành theo phương thức **có hoàn lại**.

Trong khi đó, phân phối siêu bội không đòi hỏi tính độc lập của các phép thử và do đó việc lấy mẫu là theo phương thức **không hoàn lại**.

Phân phối siêu bội được áp dụng trong nhiều lĩnh vực, đặc biệt được sử dụng rộng rãi trong việc lấy mẫu chấp nhận, thí nghiệm điện tử, lấy mẫu kiểm tra để đảm bảo chất lượng. Rõ ràng, với nhiều lĩnh vực việc kiểm định đánh giá chất lượng rất tốn kém và phung phí khi ta phải phá bỏ đối tượng được kiểm tra và do đó không thể hoàn trả lại tổng thể. Vì thế, việc lấy mẫu không hoàn lại là cần thiết.

Ví dụ: Rút quân bài từ bộ bài tú lơ khơ. Ta muốn tính xác suất để rút được 3 quân đỏ trong 5 quân bài được rút ngẫu nhiên từ bộ bài gồm 52 quân, phân phối nhị thức không thể áp dụng được, trừ khi mỗi quân bài được hoàn lại sau khi rút và quan sát sau đó xáo lại bài trước khi rút quân tiếp theo.

1. Định nghĩa. Khi chọn ngẫu nhiên một mẫu cỡ n từ N phần tử, ta quan tâm đến xác suất để chọn được x phần tử thành công. Phép thử kiểu này được gọi là **phép thử siêu bội,** nếu nó thỏa mãn hai tính chất sau:

1. Một mẫu cỡ n được chọn ngẫu nhiên theo **phương thức không hoàn lại** từ N phần tử.
2. Trong N phần tử đã định rõ k phần tử là thành công và $N - k$ phần tử còn lại là thất bại.

Số phần tử thành công X trong phép thử siêu bội được gọi là **biến ngẫu nhiên siêu bội.**
Phân phối xác suất của biến ngẫu nhiên siêu bội được gọi là **phân phối siêu bội** và các giá trị của nó được kí hiệu là $P(X = x) = h(x; N, n, k)$.

Ví dụ 6. Một uỷ ban gồm 5 người được chọn ngẫu nhiên từ 3 nhà hóa học và 5 nhà vật lí. Tìm phân phối xác suất của số lượng nhà hoá học được chọn.

Giải: + Gọi biến ngẫu nhiên X là số nhà hoá học trong uỷ ban nói trên: đây là BNN siêu bội.

+ Do đó: $h(x;8,5,3) = P(X = x) = \dfrac{C_3^x C_5^{5-x}}{C_8^5}, \quad x = 0,1,2,3.$

+ Cụ thể:

$$P(X=0) = h(0;8,5,3) = \frac{C_3^0 C_5^5}{C_8^5} = \frac{1}{56}; \quad P(X=1) = h(1;8,5,3) = \frac{C_3^1 C_5^4}{C_8^5} = \frac{15}{56}$$

$$P(X=2) = h(2;8,5,3) = \frac{C_3^2 C_5^3}{C_8^5} = \frac{30}{56}; \quad P(X=3) = h(3;8,5,3) = \frac{C_3^3 C_5^2}{C_8^5} = \frac{10}{56}$$

+ Dạng bảng của phân phối siêu bội của X là:

x	0	1	2	3
$h(x;8,5,3)$	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

2.Công thức tính: Phân phối xác suất của biến ngẫu nhiên siêu bội X (biểu thị số thành công trong mẫu cỡ n được chọn ngẫu nhiên từ N phần tử) trong đó có k phần tử là thành công và $N - k$ phần tử được đặt thất bại, được xác định bởi công thức:

$$h(x; N, n, k) = \frac{C_k^x C_{N-k}^{n-x}}{C_N^n}, \; x = 0,1,2,\ldots,n.$$

Ví dụ 7. Mỗi lô gồm 40 sản phẩm được coi là chấp nhận được nếu chứa không quá 3 phế phẩm. Biện pháp lấy mẫu là chọn ngẫu nhiên 5 sản phẩm trong lô để kiểm tra và loại lô hàng nếu có một phế phẩm. Tính xác suất để có đúng 1 phế phẩm được tìm thấy trong mẫu nếu trong lô hàng có 3 phế phẩm?

Giải: + Gọi X là số phế phẩm tìm thấy: đây là BNN siêu bội.

+ Sử dụng phân phối siêu bội với n = 5, N = 40, k = 3 và x =1 ta tìm được xác suất nhận được một phế phẩm là : $h(1;40,5,3) = \dfrac{C_3^1 C_{37}^4}{C_{40}^5} = 0,3011$.

3. Các tham số đặc trưng

Trung bình (kỳ vọng) và phương sai của phân phối siêu bội $h(x;\, N,\, n,\, k)$ được xác định bởi:

$$\mu = \frac{nk}{N} \quad \text{và} \quad \sigma^2 = \frac{N-n}{N-1}\, n\frac{k}{N}\left(1 - \frac{k}{N}\right).$$

Ví dụ 8. Tìm trung bình và phương sai của biến ngẫu nhiên trong Ví dụ 7.

Giải: + Trong Ví dụ 7 ta gọi X là số phế phẩm tìm thấy: đây là BNN siêu bội.

+ Khi đó áp dụng công thức với N = 40, n = 5, và k = 3 ta có: $\mu = \dfrac{(5)(3)}{40} = \dfrac{3}{8} = 0,375$

và : $\sigma^2 = \left(\dfrac{40-5}{39}\right)(5)\left(\dfrac{3}{40}\right)\left(1 - \dfrac{3}{40}\right) = 0,3113$

Chú ý: Nếu N đã được chia thành k tập đôi một không giao nhau $A_1, A_2, \ldots, A_k$ với số lượng phần tử tương ứng là $a_1, a_2, .., a_k$, thì phân phối xác suất đồng thời của các biến ngẫu nhiên $X_1, X_2, \ldots, X_k$ - số lượng phần tử của $A_1, A_2, .., A_k$ tương ứng trong mẫu cỡ n - là

$$f(x_1, x_2, \ldots, x_k; a_1, a_2, .., a_k, N, n) = \frac{C_{a_1}^{x_1} C_{a_2}^{x_2} \ldots C_{a_k}^{x_k}}{C_N^n}$$

trong đó $\sum_{i=1}^{k} x_i = n$ và $\sum_{i=1}^{k} a_i = N$.

IV. Phân phối nhị thức âm và phân phối hình học.

1.Phân phối nhị thức âm.

Xét một phép thử có các tính chất tương tự như các tính chất của phép thử nhị thức, nhưng số phép thử được lặp lại (độc lập) cho đến khi số lượng biến cố thành công xuất hiện là một con số được ấn định trước. Khi đó, ta quan tâm đến xác suất để có được k lần thành công và dừng lại ở lần thực hiện phép thử thứ x. Dãy phép thử kiểu này được gọi là **dãy phép thử nhị thức âm.**

a. Định nghĩa. Số phép thử X để có được k biến cố thành công trong phép thử nhị thức âm được gọi là **biến ngẫu nhiên nhị thức âm**, và phân phối xác suất của nó được gọi là **phân phối nhị thức âm**, kí hiệu các xác suất là $b^*(x;k,p)$.

b. Công thức. Nếu một phép thử được lặp đi lặp lại một cách độc lập với biến cố thành công xuất hiện trong một lần thực hiện có xác suất là p và biến cố thất bại xuất hiện với xác suất là $q = 1 - p$, thì phân phối xác suất của biến ngẫu nhiên X (biểu thị số phép thử cần phải thực hiện để có được lần thứ k biến cố thành công xuất hiện) là

$$b^*(x;k,p) = C_{x-1}^{k-1}p^k q^{x-k}, \qquad x = k, k+1, k+2, \ldots$$

Ví dụ 9. Một người tung ba đồng xu nhiều lần, tìm xác suất để người này hai lần thu được toàn mặt ngửa hoặc toàn mặt sấp ngay ở lần tung thứ năm.

Giải: + Gọi X là số lần tung để hai lần để đạt được hai lần thu được toàn mặt ngửa hoặc toàn mặt sấp: đây là BNN nhị thức âm.

+ Sử dụng phân phối nhị thức âm với $x = 5$, $k = 2$ và $p = 1/4$, ta có

$$b^*(5;2,\frac{1}{4}) = C_4^1\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = \frac{4!}{1!3!}\frac{3^3}{4^5} = \frac{27}{256}.$$

2. Phân phối hình học. Trường hợp đặc biệt của phân phối nhị thức âm là **phân phối hình học** và kí hiệu các giá trị của nó là $g(x;p)$.

a. Định nghĩa. Nếu một phép thử được **lặp đi lặp lại một cách độc lập** và xác suất xuất hiện biến cố thành công trong mỗi phép thử là p và biến cố thất bại là $q = 1 - p$, thì phân phối xác suất của biến ngẫu nhiên X (biểu thị số phép thử phải thực hiện đến khi một biến cố thành công xuất hiện) là

$$g(x;p) = pq^{x-1}, \quad x = 1, 2, 3, \ldots$$

2. Tham số đặc trưng. Giá trị trung bình (kỳ vọng) và phương sai của biến ngẫu nhiên tuân theo luật phân phối hình học, là: $\mu = \dfrac{1}{p}$, $\sigma^2 = \dfrac{1-p}{p^2}$

V. Phân phối Poisson và quá trình Poisson.

1. Định nghĩa.

Định nghĩa 1. Các phép thử cho kết quả là các **giá trị bằng số của biến ngẫu nhiên** X, biểu thị số biến cố sơ cấp xuất hiện trong suốt một khoảng thời gian cho trước (thể có độ dài bất kỳ) hoặc một miền xác định, được gọi là **phép thử Poisson**.

Phép thử Poisson được xác định bởi **quá trình Poisson** - quá trình có những tính chất sau đây:

1. Số biến cố sơ cấp xuất hiện trong một khoảng thời gian hoặc một vùng nào đó là độc lập với số biến cố sơ cấp xuất hiện trong bất kì khoảng thời gian nào hoặc vùng nào không có phần chung với khoảng thời gian hay vùng đang xét của không gian. Vì lý do này, ta nói Quá trình Poisson là quá trình không có trí nhớ.
2. Xác suất để một biến cố sơ cấp xuất hiện trong mỗi khoảng thời gian ngắn hoặc trong một vùng nhỏ thì tỉ lệ với chiều dài của khoảng thời gian hoặc cỡ của vùng và không phụ thuộc vào số biến cố sơ cấp đang xuất hiện ngoài khoảng thời gian này hoặc vùng này.

3. Xác suất để có nhiều hơn một biến cố sơ cấp xuất hiện trong một khoảng thời gian ngắn hoặc một vùng như trên là không đáng kể.

Định nghĩa 2. Số biến cố sơ cấp X xuất hiện trong phép thử Poisson được gọi là **biến ngẫu nhiên Poisson** và phân phối xác suất của nó được gọi là **phân phối Poisson**.

2. Công thức. Phân phối xác suất của biến ngẫu nhiên Poisson *X*, biểu thị số biến cố sơ cấp xuất hiện trong một khoảng thời gian cho trước hoặc một vùng định trước được ký hiệu bởi *t*, là

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0,1,2,\ldots$$

trong đó λ là số biến cố xuất hiện trung bình trong một đơn vị thời gian hoặc vùng, và $e \approx 2.71828\ldots$

Chú ý. Bảng A.2 chứa các tổng xác suất Poisson : $P(r;\lambda t) = \sum_{x=0}^{r} p(x;\lambda t)$

với một số giá trị của λt thay đổi từ 0,1 đến 18.

Ví dụ 10. Số thùng dầu trung bình được đưa đến một cảng của thành phố nào đó trong mỗi ngày là 10. Năng lực bốc dỡ của cảng nhiều nhất là 15 thùng một ngày. Tính xác suất để một ngày cho trước có thùng dầu phải chở quay lại.

Giải: + Gọi *X* là số thùng dầu được đưa đến cảng trong một ngày: đây là BNN Poisson
+ Theo Bng A.2, ta có

$$P(X > 15) = 1 - P(X \le 15) = 1 - \sum_{x=0}^{15} p(x;10) = 1 - 0.9513 = 0.0487.$$

3.Tham số đặc trưng.

Giá trị trung bình(Kỳ vọng) và phương sai của phân phối Poisson $p(x;\lambda t)$ đều bằng λt .

Chú ý: ***Phân phối Poisson là dạng giới hạn của phân phối nhị thức***
Cho *X* là biến ngẫu nhiên nhị thức với phân phối xác suất $b(x;\, n,\, p)$. Khi $n \to +\infty$, $p \to 0$, và np là hằng số, thì ta có: $b(x;\, n,\, p) \to p(x;\, \mu)$

Ví dụ 11. Trong một quy trình sản xuất kính, khi sản xuất ra những sản phẩm bị lỗi hoặc bị sủi bọt, người ta vẫn phải đưa những sản phẩm không mong muốn này cho bộ phận tiếp thị. Biết rằng, trung bình trong 1000 sản phẩm có 1 sản phẩm bị sủi bọt. Xác suất để một mẫu ngẫu nhiên gồm 8000 sản phẩm có chứa ít hơn 7 sản phẩm bị sủi bọt, là bao nhiêu?

Giải: Đây chính là phép thử nhị thức với $n = 8000$ và $p = 0.001$. Do *p* rất gần số 0 và *n* khá lớn, ta có thể xấp xỉ bởi phân phối nhị thức bằng cách sử dụng: $\mu = (8000)(0.001) = 8$
Do vậy, nếu X là số sản phẩm bị sủi bọt, ta có:

$$P(X < 7) = \sum_{x=0}^{6} b(x;8000,0.001) \simeq \sum_{x=0}^{6} p(x;8) = 0.3134.$$

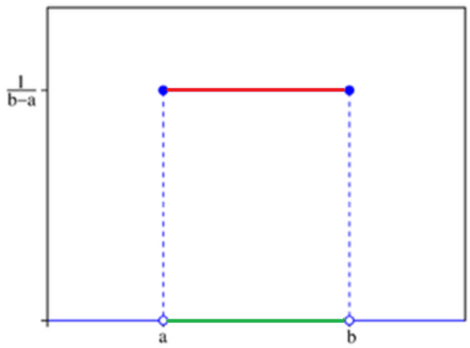
B. MỘT SỐ PHÂN PHỐI XÁC SUẤT LIÊN TỤC

I. Phân phối liên tục đều. Một trong những phân phối đơn giản nhất trong thống kê là **phân phối liên tục đều.** Phân phối này được đặc trưng bởi một hàm mật độ, nó rất "bằng phẳng", và do đó xác suất là không thay đổi trên một khoảng đóng $\left[a;b\right]$.

1. Định nghĩa. Biến ngẫu nhiên liên tục X được gọi là có phân phối liên tục đều trên khoảng $\left[a;b\right]$ nếu hàm mật độ của nó trên khoảng đó được xác định bởi:

$$f(x;a,b) = \begin{cases} \dfrac{1}{b-a} & \text{khi } \ a \le x \le b \\ 0 \ \text{ khi } & x \notin \left[a;b\right] \end{cases}$$

Do hàm mật độ có dạng một hình chữ nhật với đáy là $b-a$ và **chiều cao không đổi** $\dfrac{1}{b-a}$ nên phân phối đều còn được gọi là **phân phối chữ nhật.**



Ví dụ 12. Giả sử rằng một phòng họp lớn của một công ty xác định chỉ có thể được sử dụng không quá 4h. Tuy nhiên, trên thực tế thời gian sử dụng phòng họp trong thời gian dài hay ngắn đều xảy ra khá thường xuyên. Sự thật, có thể coi thời gian sử dụng phòng họp X là biến ngẫu nhiên đều trên khoảng [0, 4].

a.Tìm hàm mật độ xác suất của X .

b. Tính xác suất để thời gian sử dụng phòng họp của một cuộc họp nào đó kéo dài ít nhất 3h.

Giải: a.Hàm mật độ của biến ngẫu nhiên X có phân phối đều trên [0, 4] là:

$$f(x) = \begin{cases} \dfrac{1}{4} & \text{khi } 0 \le x \le 4 \\ 0 & \text{khi } x \notin [0;4] \end{cases}$$

b. Xác suất để thời gian sử dụng phòng kéo dài ít nhất 3 gời là: $P[X \ge 3] = \int_3^4 \frac{1}{4}dx = \frac{1}{4}.$

Nhận xét: Nếu BNN liên tục X có phân phối đều trên $[a;b]$, thì hàm phân phối của X được xác định bởi:

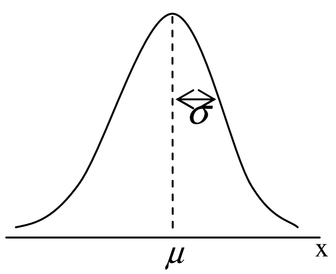
$$F(x)=\begin{cases}0\,, & x<a\\ \int_{-\infty}^{x} f(x)dx=\dfrac{x-a}{b-a}, & a\le x\le b\\ 1\,, & x>b\end{cases}$$

2. Tham số đặc trưng.

Kỳ vọng và phương sai của phân phối đều được xác định bởi: $\mu=\dfrac{a+b}{2}$ và $\sigma^2=\dfrac{(b-a)^2}{12}$.

II. Phân phối chuẩn.

Phân phối xác suất quan trọng nhất trong toàn bộ lĩnh vực thống kê là **phân phối chuẩn.** Đồ thị của nó, được gọi là một **đường cong chuẩn:** một đường cong có dạng hình quả chuông, nó miêu tả xấp xỉ nhiều hiện tượng xảy ra trong tự nhiên, trong công nghiệp và trong nghiên cứu. Các số đo vật lí trong một vùng như các thí nghiệm về khí tượng, các số đo về lượng mưa, và các số đo về các khâu của quá trình sản xuất, thông thường được giải thích bằng phân phối chuẩn.. Phân phối chuẩn thường được nhắc đến như là **phân phối Gaussian,** để tưởng nhớ đến Karl Friedrich Gauss (1777 - 1855), người cũng tìm thấy phương trình của đường cong chuẩn nhờ việc nghiên cứu các sai số trong việc lặp đi lặp lại các phép đo với một số lần như nhau.



Hình 2. Đường cong chuẩn

1.Định nghĩa.

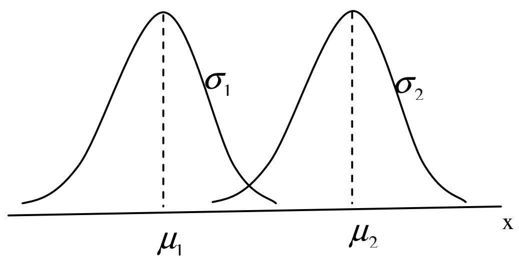
Một biến ngẫu nhiên liên tục X có phân phối hình quả chuông được gọi là một **biến ngẫu nhiên chuẩn.**

Mật độ của X được ký hiệu bởi $n(x\;;\mu,\sigma)$.

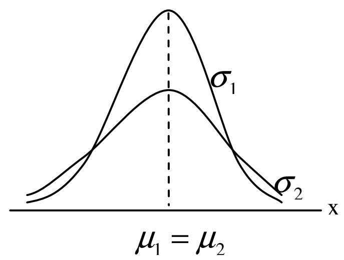
2. Phân phối chuẩn.

Cho biến ngẫu nhiên chuẩn X, với kỳ vọng μ và phương sai σ^2. Khi đó hàm mật độ của biến ngẫu nhiên X được xác định bởi:

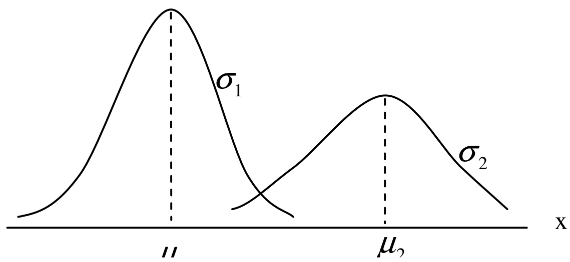
$$n(x\;;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty,\ \pi=3.14159\ldots\ \text{và}\ e=2.71828\ldots.$$



Hình 3. Hai đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 = \sigma_2$



Hình 4. Hai đường cong chuẩn với $\mu_1 = \mu_2$ và $\sigma_1 < \sigma_2$



Hình 5 Hai đường cong chuẩn với $\mu_1 < \mu_2$ **và** $\sigma_1 < \sigma_2$

3. Tính chất của đường cong chuẩn.

1. **Mode**, là điểm trên trục hoành mà **tại đó đường cong đạt giá trị lớn nhất**, xảy ra tại $x = \mu$.
2. Đường cong có trục đối xứng là đường thẳng đứng đi qua μ.
3. Đường cong có hai điểm uốn tại $x = \mu \pm \sigma$, lồi nếu $\mu - \sigma < X < \mu + \sigma$ và lõm nếu ngược lại.
4. Đường cong tiệm cận với trục hoành nếu chúng ta cho x di chuyển dần xa khỏi giá trị trung bình.
5. Tổng diện tích của phần bên dưới đường cong và bên trên trục hoành bằng 1.

4. Tham số đặc trưng.

Nếu X là BNN chuẩn có hàm mật độ $n(x;\mu,\sigma)$ thì $E(X) = \mu$, $\sigma_X^2 = \sigma$.

5. Xác suất của BNN chuẩn

a. Công thức. Cho X là BNN chuẩn có hàm mật độ $n(x;\mu,\sigma)$ khi đó:

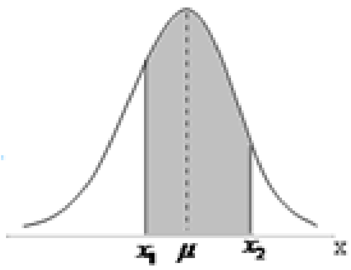
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x;\mu,\sigma)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-(1/2)[(x-\mu)/\sigma]^2} dx$$

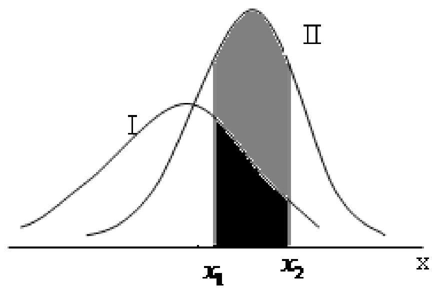
b. Diện tích bên dưới đường cong chuẩn.

Bất kỳ phân phối xác suất liên tục hay hàm mật độ nào đều có đồ thị là một đường cong, mà diện tích vùng dưới của đường cong chắn bởi hai trục $x = x_1$ và $x = x_2$ bằng xác suất để biến ngẫu nhiên X nhận giá trị nằm giữa $x = x_1$ và $x = x_2$. Do đó, với đường cong chuẩn

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x;\mu,\sigma)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-(1/2)[(x-\mu)/\sigma]2}dx$$

bằng diện tích của vùng bôi đen.





Hình 6. $P(x_1 < X < x_2)$ = diện tích phần bôi đen **Hình 7.** $P(x_1 < X < x_2)$ của hai đường cong chuẩn khác nhau

Nhận xét: Bây giờ ta thực hiện bởi phép chuyển: $Z = \dfrac{X-\mu}{\sigma}$. Khi đó X nếu nhận các giá trị trong khoảng (x_1, x_2) thì biến ngẫu nhiên Z sẽ nhận các giá trị trong khoảng $(z_1, z_2) : z_1 = \dfrac{x_1-\mu}{\sigma}, z_2 = \dfrac{x_2-\mu}{\sigma}$.

Từ đó, chúng ta có thể viết

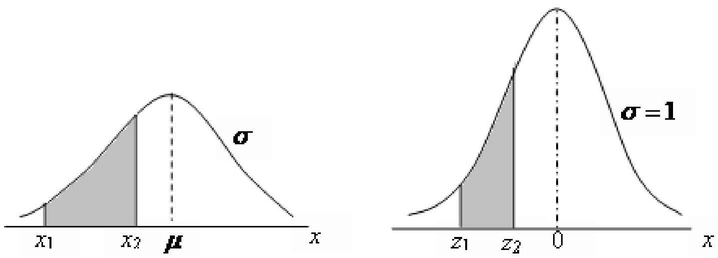
$$\begin{aligned} P(x_1 < X < x_2) &= \frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-(1/2)[(x-\mu)/\sigma]2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-z^2/2}dz = \int_{z_1}^{z_2} n(z;0,1)dz = P(z_1 < Z < z_2), \end{aligned}$$

ở đó chúng ta thấy Z là một phân phối chuẩn có trung bình bằng 0 và phương sai bằng 1.

c. Phân phối tiêu chuẩn.

Định nghĩa. Phân phối của biến ngẫu nhiên chuẩn có kỳ vọng bằng 0 và phương sai bằng 1 được gọi là **phân phối tiêu chuẩn:** $n(x;0,1) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, -\infty < x < +\infty$

Như vậy ta có thể quy gọn tất cả các giá trị cần tra của diện tích phần bên dưới của đường cong chuẩn bất kỳ về làm một, đó là bảng tra của phân phối tiêu chuẩn. Bảng A.3 chỉ ra diện tích phần bên dưới đường cong tiêu chuẩn ứng với $P(Z < z)$, với giá trị của z chạy từ -3.49 đến 3.49.



Hình 8. Phân phối chuẩn gốc và phân phối sau phép chuyển

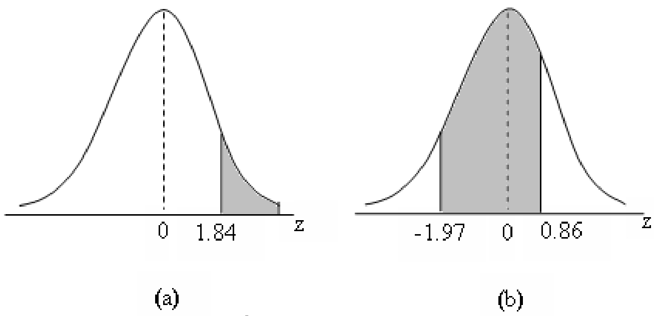
Ví dụ 13. Cho phân phối chuẩn, tìm diện tích phần nằm bên dưới đường cong ở hình

a. Ở bên phải số z = 1.84 và diện tích phần nằm bên dưới đường cong
b. Giữa hai số z = - 1.97 và z = 0.86.

Giải:

a. Diện tích phần nằm trong Hình 9 (a) ở bên phải số z = 1.84 bằng 1 trừ đi diện tích phần nằm bên trái số z = 1.84, bằng cách tra Bảng A.3 ta có diện tích phần cần tìm là 1 – 0.9671 = 0.0329.

b. Diện tích phần nằm trong Hình 9(b) ở giữa hai số z = - 1.97 và z = 0.86 bằng diện tích của phần nằm bên trái số z = 0.86 trừ đi diện tích phần nằm bên trái số z = -1.97. Từ Bảng A3 chúng ta tìm ra diện tích của phần hình cần tìm là 0.8051 – 0.0244 = 0.7807.



Hình 9. Phần diện tích trong Ví dụ 13

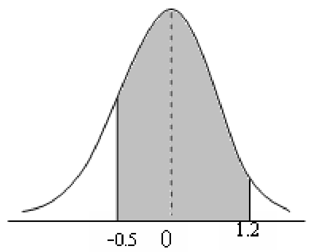
Ví dụ 14. Cho một phân phối chuẩn có $\mu = 50$ và $\sigma = 10$. Tìm xác suất để X nhận giá trị trong khoảng 45 và 62.

Giải: Các giá trị z tương ứng với $x_1 = 45$ và $x_2 = 62$ là

$$z_1 = \frac{45-50}{10} = -0.5 \text{ và } z_2 = \frac{62-50}{10} = 1.2.$$

Do đó,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2).$$



Hình 10. Phần diện tích trong Ví dụ 14

Xác suất $P(-0.5 < Z < 1.2)$ chính là diện tích của phần được tô đậm trong Hình 6.11.
Diện tích của phần này có thể được tìm bằng cách trừ diện tích của phần bên trái z = 1.2 cho diện tích của phần bên trái z = - 0.5. Dùng Bảng A.3 chúng ta có:

$$P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(-0.5) = 0.8849 - 0.3085 = 0.5764.$$

Ví dụ 15. Cho một phân phối chuẩn với $\mu = 40$ và $\sigma = 6$, tìm giá trị của x tương ứng với

a. Phần hình nằm ở bên trái có diện tích 45%
b. Phần hình nằm ở bên phải có diện tích 14%

Giải:

a. Phần có diện tích 0,45, nằm ở bên trái giá trị x được tô đậm trong Hình 11(a). Điều này đòi hỏi chúng ta tìm giá trị của z sao cho diện tích phần bên trái z bằng 0.45.

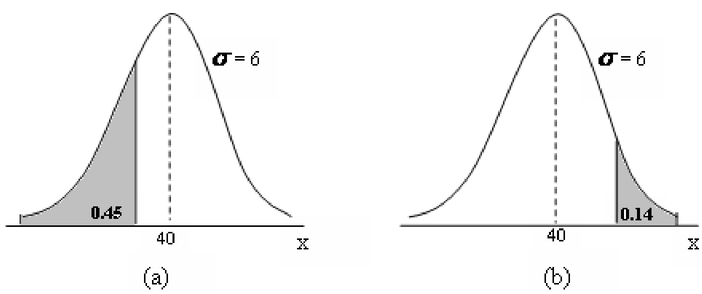
+ Từ Bảng A.3 chúng ta thấy $P(Z < -0{,}13) = 0{,}45$ vì thế z = - 0, 13.

+ Do đó: $x = (6).(-0{,}13) + 40 = 39{,}22$

b. Trong Hình 11(b) chúng ta bôi đen phần ở bên phải giá trị x có diện tích bằng 0.14. Điều này đòi hỏi chúng ta tìm giá trị của z sao cho diện tích phần bên phải z bằng 0.14 và do đó diện tích phần ở bên trái z bằng 0.86.

+ Từ bảng A.3 chúng ta thấy $P(Z < 1{,}08) = 0{,}86$.

+ Do đó: $z = 1{,}08$ và $x = (6)(1{,}08) + 40 = 46{,}48$.



Hình 11. Phần diện tích trong Ví dụ 15

6. Xấp xỉ chuẩn của phân phối nhị thức.

Ta đã biết phân phối Poisson có thể được dùng để xấp xỉ cho phân phối nhị thức khi n đủ lớn và p rất gần với 0 hoặc 1. Cả phân phối Poisson và phân phối nhị thức đều là các phân phối rời rạc. Phân phối chuẩn sẽ là một xấp xỉ tốt đối với phân phối rời rạc khi các số liệu của phân phối rời rạc tạo nên một đường gấp khúc có hình dáng gần giống với một quả chuông.

Định lý. Nếu X là một BNN nhị thức với trung bình $\mu = np$ và phương sai $\sigma^2 = npq$, thì phân phối giới hạn của

$$Z = \frac{X - np}{\sqrt{npq}}, \text{khi } n \to +\infty,$$

là phân phối tiêu chuẩn chuẩn $n(z;0,1)$.

Chú ý: 1. Phân phối chuẩn với $\mu = np$ và $\sigma^2 = np(1-p)$ không chỉ là một xấp xỉ rất chính xác của phân phối nhị thức khi n lớn và p không quá gần với 0 và 1 mà nó còn là một xấp xỉ khá tốt ngay cả khi n nhỏ và p gần bằng 1/2.

2. Một cách nữa để xác định xem khi nào nên dùng xấp xỉ chuẩn là tính giá trị của np và nq: Nếu cả hai giá trị này đều lớn hơn hoặc bằng 5 thì kết quả xấp xỉ là tốt.

3. Bảng 6.1 ở trang 164 như là một sự chỉ dẫn về chất lượng của xấp xỉ.

Ví dụ 16. Một kì thi trắc nghiệm có 200 câu hỏi với 4 phương án trả lời cho mỗi câu trong đó chỉ có một phương án đúng. Với xác suất bằng bao nhiêu thì một sinh viên không có kiến thức về phần đó có thể trả lời đúng từ 25 đến 30 câu hỏi trong 80 câu hỏi lấy từ 200 câu hỏi trên.

Giải: +Xác suất để sinh viên trả lời đúng 1 câu trong 80 câu hỏi là 1/4.

+ Gọi X là số câu trả lời đúng của sinh viên, khi đó

$$P(25 \le X \le 30) = \sum_{x=25}^{30} b(x;80,1/4).$$

+ Dùng đường cong chuẩn xấp xỉ với

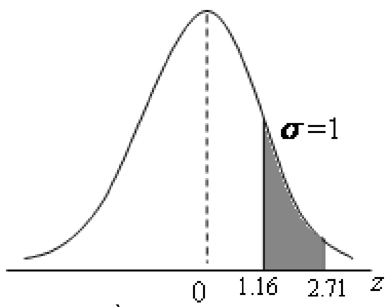
$$\mu = np = 80(1/4) = 20$$

và

$$\sigma = \sqrt{npq} = \sqrt{80(1/4)(3/4)} = 3.873,$$

+ Ta cần tính diện tích của hình nằm giữa $x_1 = 24.5$ và $x_2 = 30.5$. Các giá trị z tương ứng là

$$z_1 = \frac{24.5 - 20}{3.873} = 1.16 \quad \text{và} \quad z_2 = \frac{30.5 - 20}{3.873} = 2.71.$$



Hình 12. Phần diện tích trong Ví dụ 16

+ Xác suất trả lời đúng từ 25 đến 30 câu hỏi được cho bởi diện tích của phần bôi đen trong Hình 12.

+ Từ Bảng A.3 chúng ta có

$$P(25 \le X \le 30) = \sum_{x=25}^{30} b(x;80,1/4)$$
$$\cong P(1.16 < Z < 2.71) = P(Z < 2.71) - P(Z < 1.16)$$
$$= 0.9966 - 0.8770 = 0.1196.$$

III. Phân phối mũ và phân phối gamma

Phân phối mũ và gamma đóng một vai trò quan trọng trong lí thuyết xếp hàng và bài toán tin cậy. Thời gian giữa hai lần đến một nơi phục vụ nào đó, thời gian giữa hai lần hỏng của một bộ phận máy móc hay một hệ thống điện nào đó thường tuân theo mô hình của phân phối mũ. Mối liên hệ giữa phân phối mũ và phân phối gamma làm cho phân phối gamma cũng có liên quan đến các dạng bài toán tương tự trên.

1.Hàm Gamma

a.Định nghĩa. Hàm gamma là hàm thuộc lớp các hàm đặc biệt và được được định nghĩa bởi:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \qquad \text{trong đó } \alpha > 0.$$

b. Tính chất.

i. Bằng cách tích phân từng phần đặt: $u = x^{\alpha-1}$ và $dv = e^{-x}dx$, chúng ta nhận được công thức truy hồi sau:

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1).$$

ii. Lặp lại công thức truy hồi chúng ta có:

$$\Gamma(\alpha) = (\alpha-1)(\alpha-2)\Gamma(\alpha-2) = (\alpha-1)(\alpha-2)(\alpha-3)\Gamma(\alpha-3),$$

Đặc biệt: khi $\alpha = n,$ với *n* là số nguyên dương thì:

$$\Gamma(n) = (n-1)(n-2)....\Gamma(1).$$

iii. Ta có :

$$\Gamma(1) = \int_0^\infty e^{-x}dx = 1.$$

Do đó: $\Gamma(n) = (n-1)!$.

iv. Một tính chất quan trọng của phân phối gamma đó là: $\Gamma(1/2) = \sqrt{\pi}.$

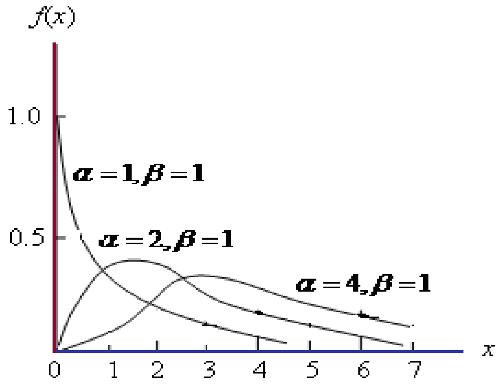
2. Phân phối Gamma.

a.Định nghĩa. Biến ngẫu nhiên liên tục X có **phân phối gamma,** với các tham số α và β, nếu hàm mật độ của nó được cho bởi

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta}, & \text{khi x} > 0 \\ 0 \end{cases}$$

trong đó $\alpha, \beta > 0$.

Đồ thị của một số phân phối gamma được minh họa trong Hình 13, trong đó α và β nhận các giá trị xác định.



Hình 13. Phân phối Gamma

b.Tham số đặc trưng. Kỳ vọng và phương sai của phân phối gamma là

$$\mu = \alpha\beta \quad \text{và} \quad \sigma^2 = \alpha\beta^2.$$

2. Phân phối mũ Một phân phối với $\alpha = 1$ **được gọi là phân phối mũ**

a.Định nghĩa. Biến ngẫu nhiên liên tục X có ***phân phối mũ***, với tham số β, nếu hàm mật độ của nó được cho bởi

$$f(x) = \begin{cases} \dfrac{1}{\beta}e^{-x/\beta}, & \text{khi x} > 0 \\ 0, & \text{khi } x \le 0 \end{cases}$$

trong đó $\beta > 0$.

b.Tham số đặc trưng. Kỳ vọng và phương sai của phân phối mũ là

$$\mu = \beta \quad \text{và} \quad \sigma^2 = \beta^2.$$

IV.Phân phối χ^2

Một trường hợp đặc biệt khác của phân phối gamma cũng rất quan trọng, đó là khi ta cho $\alpha = {}^{v}\!/_{2}$ và $\beta = 2$, trong đó v là một số nguyên dương. Kết quả được gọi là **phân phối** χ^2**.** Phân phối này có một tham số *v*: gọi là bậc tự do.

a.Định nghĩa. Biến ngẫu nhiên *X* có ***phân phối*** χ^2 **(*Khi bình phương*)** với ν bậc tự do nếu hàm mật độ của nó được cho bởi

$$f(x) = \begin{cases} \dfrac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2} dx, & x > 0 \\ 0 & x \le 0 \end{cases}$$

ở đó *v* là một số nguyên dương.

b.Tham số đặc trưng. Kỳ vọng và phương sai của phân phối χ^2 là

$$\mu = v \text{ và } \sigma^2 = 2v.$$

Về nhà:

Tự đọc Mục 6.4; 6.7; 6.9; 6.10; 8.3

Bài tập: Tr. **140; 148; 157; 178;186; 198**

Đọc trước các Mục 8.1; 8.2; 8.4; 8.5; 8.6; 8.7 chuẩn bị cho **Bài số :**

Mẫu ngẫu nhiên và phân phối của các thống kê cơ bản