Chương II: BIẾN NGẪU NHIÊN

(ĐẠI LƯỢNG NGẪU NHIÊN)

II.1. Định nghĩa và phân loại.

II.2. Biểu diễn các phân phối xác suất của biến ngẫu nhiên.

II.2.1 Bảng phân phối XS của BNN rời rạc.

II.2.2 Hàm phân phối XS của BNN.

II.2.3 Hàm mật độ XS của BNN liên tục.

II.3 Một số tham số đặc trưng của BNN.

II.3.1 Kỳ vọng toán

II.3.2 Phương sai và độ lệch

II.3.3 Mốt

II.3.4 Trung vị

II.3.5 Mômen, Hệ số bất đối xứng,Hệ số nhọn *(tham khảo)*.

II.3.7 Sử dụng máy tính bỏ túi để tính 1 số tham số đặc trưng.

II.4. Một số phân phối xác suất thông dụng.

II.4.1 Phân phối Bernoulli**.**

II.4.2 Phân phối nhị thức.

II.4.3 Phân phối hình học

II.4.4 Phân phối siêu bội.

II.4.5 Phân phối Poisson.

II.4.6 Phân phối đều.

II.4.7 Phân phối mũ.

II.4.8 Phân phối chuẩn.

II.4.9 Phân phối Student.

II.4.10 Phân phối Khi Bình phương.

II.4.11 Phân phối Fisher.

II.5. Các định lý giới hạn. *(Từ II.5.1 đến II.5.4 : tham khảo)*

II.6. Hàm của Biến ngẫu nhiên. *(phần đọc thêm ở file word kèm theo)*

II.1. Định nghĩa và phân loại

Định nghĩa:

Một biến số được gọi là ***biến ngẫu nhiên*** (hay còn gọi là *biến số ngẫu nhiên – random variable*, *đại lượng ngẫu nhiên*) nếu trong kết quả của mỗi 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 yếu tố ngẫu nhiên .

Kí hiệu cho biến ngẫu nhiên: $X, Y, Z, X_1, X_2, \ldots, X_n, \ldots$

Các giá trị có thể có của chúng được kí hiệu bằng chữ cái in thường $x, x_1, x_2, \ldots, x_n, \ldots y_1, y_2, \ldots$

Biến X nào đó được gọi là *ngẫu nhiên* vì trước khi tiến hành phép thử ta chưa thể biết chắc chắn nó sẽ nhận giá trị là bao nhiêu, chỉ có thể dự đoán điều đó với một xác suất nhất định.

Biến ngẫu nhiên được phân làm 2 loại:

* Biến ngẫu nhiên gọi là <u>rời rạc </u>nếu ta có thể đếm được các giá trị có thể có của nó (hữu hạn hoặc vô hạn).

VD: - Số chấm xuất hiện khi tung 1 con xúc xắc là một BNN rời rạc.

- Một người quyết định mua vé số thường xuyên cho đến khi trúng được giải đặc biệt thì thôi. Gọi X là số tờ vé số không trúng giải đặc biệt của người đó, thì X là BNN rời rạc.

* Biến ngẫu nhiên gọi là <u>liên tục </u>nếu các giá trị có thể có của nó lấp đầy một hay nhiều khoảng trên trục số.

Như vậy đối với biến ngẫu nhiên liên tục , người ta không thể đếm được các giá trị có thể có của nó.

Chiều cao của trẻ em ở một địa phương, mực nước mưa đo được sau mỗi trận mưa… là một ví dụ về biến ngẫu nhiên liên tục.

Nếu kí hiệu $\{ x_i, i \in I \}$ là tập các giá trị có thể có của X thì việc X nhận một giá trị nào đó như "$X= x_1$", "$X=x_2$"… thực chất là các biến cố ngẫu nhiên. Hơn nữa, khi thực hiện một phép thử, X nhất định sẽ nhận một và chỉ một trong các giá trị có thể có trong tập $\{x_i, i \in I\}$, do đó tập tất cả các biến cố {"$X= x_i$", $i \in I$} tạo nên một nhóm biến cố đầy đủ.

Lưu ý: cần phân biệt khái niệm "Biến cố" và "Biến ngẫu nhiên".

II.2 Biểu diễn các phân phối xác suất của BNN

- Quy luật phân phối xác suất của biến ngẫu nhiên là sự tương ứng giữa các giá trị có thể có của nó với các XS tương ứng.
- Người ta thường dùng 3 hình thức mô tả quy luật phân phối xác suất của BNN là:

  - *Bảng phân phối xác suất* (chỉ dùng cho BNN rời rạc)
  - *Hàm mật độ xác suất* (chỉ dùng cho BNN liên tục)
  - *Hàm phân phối xác suất* (dùng cho cả 2 loại BNN).

II.2.1 Bảng phân phối xác suất của BNN rời rạc

Bảng phân phối xác suất của BNN rời rạc đặc trưng cho phân phối xác suất của BNN X tại mỗi điểm, nó có dạng:

X	x_1	x_2	….	x_n	(…)
P	p_1	p_2	….	p_n	(…)

ở đây:

$x_1 < x_2 < \ldots < x_n$ (…) ; x_i là các giá trị có thể có của X.

$p_i = P(\text{"}X = x_i\text{"})$, $\forall i$.

Các tính chất : * $0 \le p_i \le 1$

* $\sum_i p_i = 1$

II.2.2 Hàm mật độ xác suất của BNN liên tục

Để *biểu thị mức độ tập trung xác suất* của biến ngẫu nhiên liên tục *trong lân cận của một điểm*, người ta đưa vào khái niệm hàm mật độ xác suất .

Ta nói f(x) là hàm mật độ xác suất của biến ngẫu nhiên liên tục nào đó $\Leftrightarrow \begin{cases} f(x) \geq 0, \forall x \\ \int_{-\infty}^{+\infty} f(x)dx = 1 \end{cases}$

Các tính chất:

* $P(a \leq X \leq b) = \int_a^b f(x)dx$
* $P(X = x_0) = 0 , \ \forall x_0$;
* $P(a \leq X < b) = P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b)$

II.2.3 Hàm phân phối xác suất

Giả sử X là một biến ngẫu nhiên , còn x là một số thực bất kỳ. Khi x thay đổi thì xác suất của biến cố " X < x " cũng thay đổi theo.

Ta định nghĩa $\boxed{F(x) = P(X < x), \quad x \in \mathbb{R}}$ (*)

là *hàm phân phối xác suất* của X, (còn gọi là hàm phân bố tích lũy – cumulative distribution function).

Về mặt ý nghĩa, giá trị hàm phân phối xác suất của biến X tại điểm x_0 phản ánh mức độ tập trung xác suất của BNN X ở về phía bên trái của số thực x_0 .

(*): trong 1 số tài liệu khác, người ta định nghĩa $F(x) = P(X \le x)$, $x \in \mathbb{R}$

Các tính chất của hàm phân phối xác suất :

* $0 \le F(x) \le 1, \quad \forall x \in \mathbb{R}$ $\quad F(-\infty) = 0 \quad F(+\infty) = 1$
* Nếu $x_1 < x_2$ thì $F(x_1) \le F(x_2) \Rightarrow F(x)$ là hàm tăng trên $\mathbb{R}$.
* $P(a \le X < b) = F(b) - F(a)$
* Nếu X là BNN rời rạc thì $F(x) = \sum_{x_i < x} p_i$
* Nếu X là BNN liên tục thì $F(x) = \int_{-\infty}^{x} f(t)dt$;

  khi đó $f(x) = F'(x)$ và $P(a \le X \le b) = F(b) - F(a)$.
* F(x) là hàm khả vi trên $\mathbb{R}$ (hoặc có thể trừ một số đếm được các điểm). Hàm phân phối của BNN liên tục là liên tục trên $\mathbb{R}$.

Ví dụ 1

Một hộp gồm 7 bi trắng và 3 bi xanh cùng cỡ .

Lấy ra ngẫu nhiên 3 viên bi từ hộp.

Gọi X là biến ngẫu nhiên chỉ số bi xanh trong các bi được lấy ra.

a) Lập bảng phân phối XS của X.

b) Gọi F(x) là hàm phân phối XS của X.

Tìm F(-1); F(2); F(2,3) và biểu thức F(x).

c) Vẽ đồ thị hàm phân phối XS của X.

d) Tính E(X); E(X^2); D(X); Mod(X); Med(X),

e) Tính E(2X+1); E(3X^2+5).

(câu d) và e) xem phần Lý thuyết II.3 ở phía sau).

Hướng dẫn: a) Các giá trị X có thể nhận được là { 0; 1; 2; 3}.

P(X=0) ≡ Xác suất KHÔNG CÓ bi xanh nào trong 3 bi được lấy ra.

$$=\frac{C_7^3}{C_{10}^3}=\frac{7}{24}$$

P(X=1) ≡ XS CÓ 1 bi xanh trong 3 bi được lấy ra $=\frac{C_3^1C_7^2}{C_{10}^3}=\frac{21}{40}$

P(X=2) ≡ XS CÓ 2 bi xanh trong 3 bi được lấy ra $=\frac{C_3^2C_7^1}{C_{10}^3}=\frac{7}{40}$

P(X=3) ≡ Xác suất cả 3 bi lấy ra đều có màu xanh $=\frac{C_3^3}{C_{10}^3}=\frac{1}{120}$

Bảng phân phối xác suất của X:

x_i	0	1	2	3
P_i	$\frac{7}{24}$	$\frac{21}{40}$	$\frac{7}{40}$	$\frac{1}{120}$

b) F(x) là hàm phân phối xác suất của X .

F(-1) = P(X< 0) = 0;

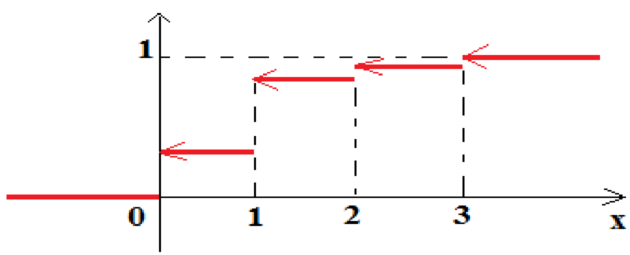
F(2)= P(X< 2) = 7/24 + 21/40.

F(2,3) = P(X<2,3) ≈ 0,9917

Tổng quát hơn:

$$F(x) = P(X < x) = \sum_{x_i < x} p_i = \begin{cases} 0 & khi \quad x \le 0 \\ \dfrac{7}{24} \approx 0,2917 & khi \quad 0 < x \le 1 \\ \dfrac{7}{24} + \dfrac{21}{40} \approx 0,8167 & khi \quad 1 < x \le 2 \\ \dfrac{7}{24} + \dfrac{21}{40} + \dfrac{7}{40} \approx 0,9917 & khi \quad 2 < x \le 3 \\ \dfrac{7}{24} + \dfrac{21}{40} + \dfrac{7}{40} + \dfrac{1}{120} = 1 & khi \quad 3 < x \end{cases}$$

c) Đồ thị hàm phân phối XS của X:



Ví dụ 2

Một người tung cùng lúc 2 con xúc xắc cho đến khi được tổng số chấm trên 2 con xúc xắc lớn hơn 10 thì dừng lại. Gọi Y là số lần người đó đã tung xúc xắc.

a) Hãy lập bảng phân phối xác suất của Y.

b) Tìm $P(2 < Y^2 < 10)$.

c) Trung bình người đó phải tung bao nhiêu lần để được tổng số chấm trên 2 con xúc xắc lớn hơn 10?

(câu c) xem phần Lý thuyết II.3 ở phía sau).

Hướng dẫn:

a) Gọi B_i là b/c lần tung thứ i được tổng số chấm > 10; i=1,2..

$$\Rightarrow \quad P(B_i) = \frac{3}{36} = \frac{1}{12}; \qquad P(\overline{B}_i) = \frac{11}{12}.$$

Ta tính được:

$$\bullet P(Y=1) = P(B_1) = \frac{1}{12} \qquad \bullet P(Y=2) = P(\overline{B}_1.B_2) = \frac{11}{12}\times\frac{1}{12}$$

$$\bullet P(Y=3) = P(\overline{B}_1.\overline{B}_2.B_3) = \left(\frac{11}{12}\right)^2\times\frac{1}{12} \qquad .....$$

Y	1	2	3	.....	k	.....
P	$\frac{1}{12}$	$\frac{11}{12}\times\frac{1}{12}$	$\left(\frac{11}{12}\right)^2\times\frac{1}{12}$	.....	$\left(\frac{11}{12}\right)^{k-1}\times\frac{1}{12}$	....

b) $$P(2<Y^2<10) = P(\sqrt{2}<Y<\sqrt{10}) = P(Y=2)+P(Y=3) = \frac{11}{12}\times\frac{1}{12}+\left(\frac{11}{12}\right)^2\times\frac{1}{12}$$

Ví dụ 3 Cho biến ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} 0 & x \notin \left(-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right) \\ a.\cos(x) & x \in \left(-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right) \end{cases}$$

a) Tìm hệ số a.

b) Tính $P\left(-\frac{\pi}{6} < X < \frac{\pi}{4}\right)$ *và* $P\left(0 < X < \frac{3\pi}{4}\right)$ *(2 công thức)*

c) Tìm xác suất trong 5 lần thực hiện phép thử ngẫu nhiên thì có ít nhất 3 lần X nhận giá trị trong khoảng $\left(-\frac{\pi}{6}, \frac{\pi}{4}\right)$

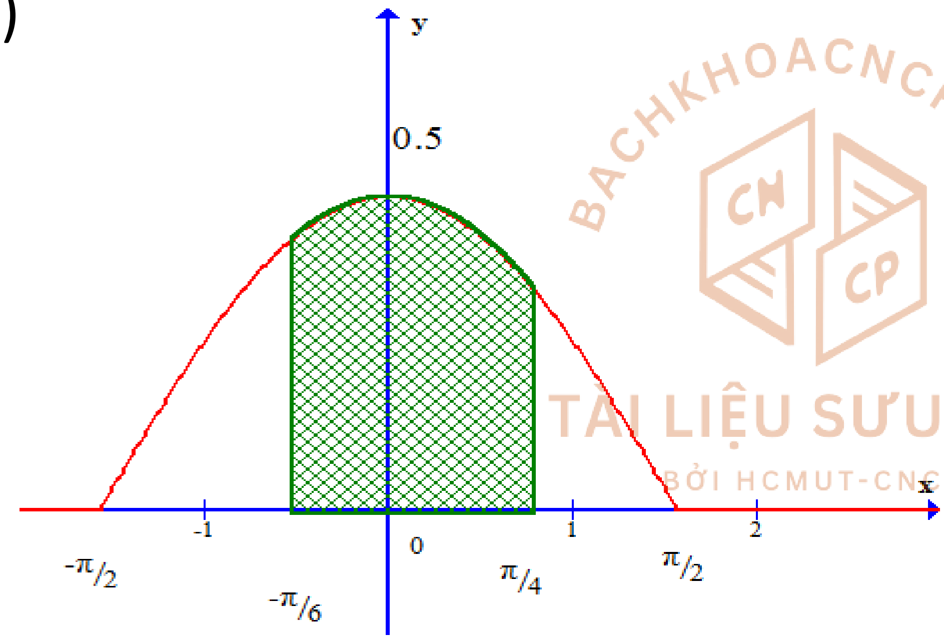
d) Tìm hàm phân phối F(x) của biến ngẫu nhiên X.

e) Tính E(X); D(X).

Hướng dẫn: a) * Điều kiện $f(x) \geq 0, \forall x \Rightarrow a \geq 0$.

$$* \text{ Đk } \int_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow \int_{-\pi/2}^{\pi/2} a.\cos(x)dx = 1 \Rightarrow a = \frac{1}{2}$$

b)



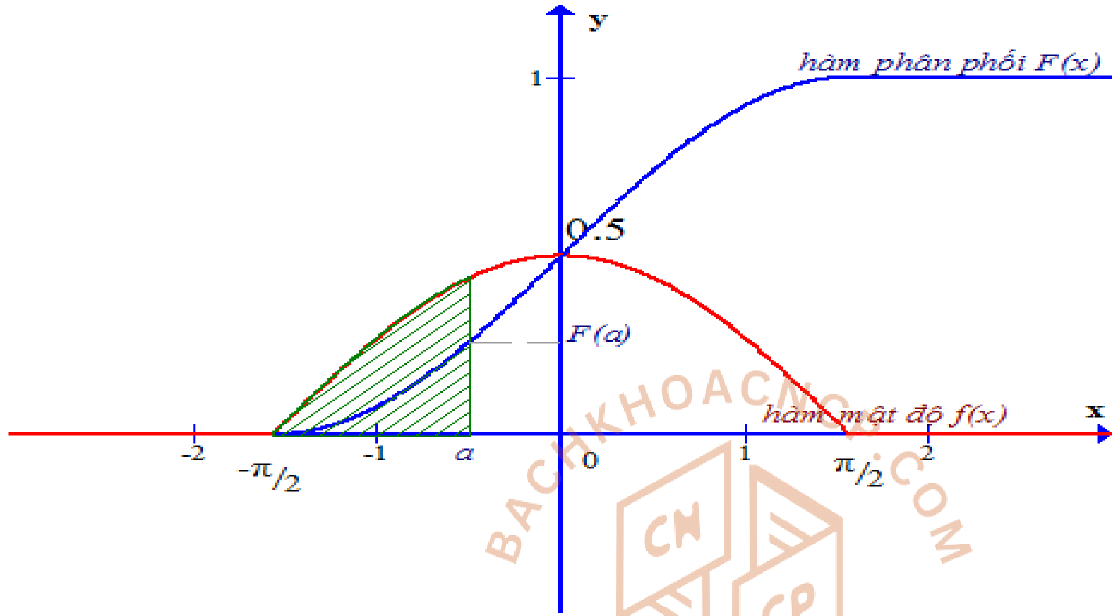
$$* P\left(-\frac{\pi}{6} < X < \frac{\pi}{4}\right) = \int_{-\pi/6}^{\pi/4} \frac{1}{2}\cos x\,dx \approx 0,6036$$

$$* P\left(0 < X < \frac{3\pi}{4}\right) = \int_{0}^{\pi/2} \frac{1}{2}\cos x\,dx = \frac{1}{2}$$

c) Đây là bt Bernoulli với n=5; p=0,6036; $3 \leq k \leq 5$. XS cần tìm:

$$C_5^3(0,6036)^3(1-0,6036)^2 + C_5^4(0,6036)^4(1-0,6036) + C_5^5(0,6036)^5 \approx 0,6888$$

d)



$$F(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} 0.dt = 0 & khi\ x < -\pi/2 \\ \int_{-\infty}^{-\pi/2} 0.dt + \int_{-\pi/2}^{x} \frac{1}{2}\cos t.dt = \frac{1}{2}\sin x + \frac{1}{2} & khi\ -\pi/2 < x < \pi/2 \\ \int_{-\infty}^{-\pi/2} 0.dt + \int_{-\pi/2}^{\pi/2} \frac{1}{2}\cos t.dt + \int_{\pi/2}^{x} 0.dt = 1 & khi\ x > \pi/2 \end{cases}$$

II.3 Một số tham số đặc trưng của BNN.

II.3.1 Kỳ vọng toán:

Kỳ vọng toán (*Expectation/Mean*, còn gọi là vọng số) của BNN X là giá trị trung bình theo xác suất của X, kí hiệu E(X) hay M(X).

Công thức tính :

- Đối với *BNN rời rạc* :

$$E(X) = \sum_i x_i p_i$$

- Đối với *BNN liên tục* :

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

Tìm kỳ vọng toán của các BNN trong các ví dụ : VD1 $\longrightarrow$ VD3 .

Các tính chất :

* $E(C) = C$ C là một BNN đặc biệt nhận giá trị C với xác suất =1.
* $E(a.X+b.Y) = a.E(X) + b.E(Y)$, với X,Y là các BNN; $a,b \in \mathbf{R}$
* $E(XY) = E(X) . E(Y)$ nếu các BNN X, Y là *độc lập,*

(X,Y độc lập tức là quy luật phân phối của biến ngẫu nhiên này không phụ thuộc vào biến ngẫu nhiên kia nhận giá trị là bao nhiêu, có thể xem thêm ở chương III).

* *Tham khảo:* Giả sử X, Y là các biến ngẫu nhiên và $Y = \varphi(X)$, thì:

+ $E(Y)=\sum_i \varphi(x_i).p_i$ *nếu X là BNN rời rạc có* $E(X)=\sum_i x_i.p_i$

+ $E(Y)=\int_{-\infty}^{+\infty} \varphi(x).f(x)dx$ *nếu X có hàm mật độ f(x).*

Các BNN X+Y; X.Y sẽ được nhắc về mặt lý thuyết ở chương III

Ví dụ 4: *Dưới đây là bảng điểm của 2 nhóm SV.*

Điểm nhóm 1:

6 7 4 3 10 4 9 6 6 7

X1	3	4	6	7	9	10
P	0.1	0.2	0.3	0.2	0.1	0.1

Điểm nhóm 2:

5 2 10 8 2 3 10 9 3 10

X2	2	3	5	8	9	10
P	0.2	0.2	0.1	0.1	0.1	0.3

a) Hãy kiểm tra lại kết quả: E(X1) = E(X2) = 6,2.

b) Dưới đây là bảng PPXS của các BNN $(X1\text{-}6,2)^2$ *và* $(X2\text{-}6.2)^2$.

$(X1\text{-}6.2)^2$	$(3\text{-}6,2)^2$	$(4\text{-}6,2)^2$	$(6\text{-}6,2)^2$	$(7\text{-}6,2)^2$	$(9\text{-}6,2)^2$	$(10\text{-}6,2)^2$
P	0.1	0.2	0.3	0.2	0.1	0.1

$(X2\text{-}6,2)^2$	$(2\text{-}6,2)^2$	$(3\text{-}6,2)^2$	$(5\text{-}6,2)^2$	$(8\text{-}6,2)^2$	$(9\text{-}6,2)^2$	$(10\text{-}6,2)^2$
P	0.2	0.2	0.1	0.1	0.1	0.3

Hãy tính $E([X1\text{-}6,2)^2)$ và $E([X2\text{-}6,2)^2)$? (4.36 và 11.16)

$\Rightarrow$ *Từ đó có thể so sánh gì về sự phân tán của các BNN X1, X2 quanh giá trị trung bình là 6.2 ?*

II.3.2 Phương sai và độ lệch chuẩn:

Phương sai (*Variance/Dispersion*, còn gọi là *Tán số*) của biến ngẫu nhiên X được định nghĩa bằng trung bình của bình phương sai lệch giữa biến ngẫu nhiên với kỳ vọng toán của nó.

Kí hiệu bởi D(X) hay V(X).

Công thức tính: $D(X) = E[X-E(X)]^2$ hay $D(X) = E(X^2) - [E(X)]^2$

- Nếu X là *BNN rời rạc* thì:

$$D(X) \overset{CT1}{=} \sum_i [x_i - E(X)]^2 p_i \overset{CT2}{=} \sum_i x_i^2 p_i - [E(X)]^2$$

- Nếu X là BNN liên tục thì:

$$D(X) \overset{CT1}{=} \int_{-\infty}^{+\infty} [x-E(X)]^2 f(x)dx \overset{CT2}{=} \int_{-\infty}^{+\infty} x^2 f(x)dx - [E(X)]^2$$

* Phương sai của biến ngẫu nhiên X phản ánh mức độ phân tán của các giá trị của X xung quanh giá trị kỳ vọng của nó.

Phương sai càng nhỏ thì giá trị của X càng tập trung gần E(X).

- Trong kỹ thuật, phương sai thường đặc trưng cho mức độ phân tán của kích thước các chi tiết gia công hay sai số của thiết bị. Phương sai cho biết sự ổn định của thiết bị. Trong nông nghiệp, phương sai đặc trưng cho mức độ đồng đều của vật nuôi hay cây trồng. Trong quản lý và kinh doanh, nó đặc trưng cho mức độ rủi ro của các quyết định.

Độ lệch chuẩn:

- *Độ lệch chuẩn (standard deviation)* của biến ngẫu nhiên X , kí hiệu σ_X , là căn bậc hai của phương sai :

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

Khi cần đánh giá mức độ phân tán của biến ngẫu nhiên theo đơn vị đo của nó, người ta thường dùng độ lệch chuẩn chứ không phải phương sai vì độ lệch chuẩn có cùng đơn vị đo với biến ngẫu nhiên cần nghiên cứu, còn đơn vị đo của phương sai bằng bình phương đơn vị đo của biến ngẫu nhiên .

Các tính chất :

* $D(X) \geq 0$; $D(C) = 0$

* $D(CX) = C^2.D(X)$

* $D(X+Y) = D(X) + D(Y)$, với X,Y là độc lập.

$D(X-Y) = D(X) + D(Y)$, với X,Y là độc lập.

* $D(C+X) = D(X)$

HQ: Nếu $X_1, X_2, \ldots, X_n$ là các BNN độc lập; $E(X_i)=a$; $D(X_i)=\sigma^2$; $\forall i$, thì:

- BNN $U = X_1 + X_2 + \ldots + X_n$ có $E(U) = n.a$ và $D(U) = n.\sigma^2$;
- BNN $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ có $E\left(\overline{X}\right) = a$; $D\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$.

II.3.3 Mốt: *Mốt của BNN X (kí hiệu mod(X))* là giá trị của biến ngẫu nhiên X tương ứng với xác suất lớn nhất nếu X là biến ngẫu nhiên rời rạc và tương ứng với cực đại của hàm mật độ xác suất nếu X là biến ngẫu nhiên liên tục .

II.3.4 Trung vị: *Trung vị (median)* của biến ngẫu nhiên X , kí hiệu med(X), là một giá trị thực mà:

$$\boxed{P(X < med(X)) \le \tfrac{1}{2} \quad \text{và} \quad P(X > med(X)) \le \tfrac{1}{2}} \; .$$

$\Rightarrow$ *Trung vị của X là giá trị nằm chính giữa tập các giá trị của X.*

*Trong **Ví dụ 4**:*

$\sigma(X1) = 2{,}0881$	$Mod(X1) = 6$	$Med(X1) = 6$
$\sigma(X2) = 3{,}3407$	$Mod(X2) = 10$	$Med(X2) = [5;8]$

Quay lại các VD1 $\to$ VD3:

Ví dụ 1: d) E(X)=0,9 E(X²)=1,3 D(X)=0,49 Mod(X)=Med(X)=1.

e) $E(2X+1) = 2E(X)+1 = 2,8$ $E(3X^2+5)= 3.E(X^2)+5 = 8,9$

Ví dụ 2: c) Số lần tung trung bình là E(Y)

$$E(Y) = \sum_{k=1}^{+\infty} k.\left(\frac{11}{12}\right)^{k-1}\frac{1}{12} = \frac{1}{12}\sum_{k=1}^{+\infty} k.\left(\frac{11}{12}\right)^{k-1} = \frac{1}{12}\sum_{k=1}^{+\infty} k.q^{k-1}; \quad q = \frac{11}{12}$$

$$= \frac{1}{12}\sum_{k=1}^{+\infty}\left[q^k\right]' = \frac{1}{12}\left[\sum_{k=1}^{+\infty} q^k\right]' = \frac{1}{12}\left[\frac{q}{1-q}\right]' = \frac{1}{12}.\frac{1}{(1-q)^2} = 12$$

Ví dụ 3: e)

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\pi/2}^{\pi/2} x\cos(x)dx = 0$$

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - E(X)^2 = \int_{-\pi/2}^{\pi/2} x^2\cos(x)dx \approx 0,9348$$

Ví dụ 5:

Theo điều tra , tỉ lệ người bị mắc bệnh sốt rét ở một vùng là 10%. Người ta cần làm xét nghiệm cho 5000 người ở vùng đó để tìm kí sinh trùng sốt rét. Có 2 phương án đưa ra:

Phương án 1: Làm các xét nghiệm máu cho từng người một.

Phương án 2: Lấy máu từng 10 người một trộn lẫn với nhau rồi làm xét nghiệm. Nếu xét nghiệm là âm tính (vô trùng) thì thông qua. Nếu xét nghiệm là dương tính (có trùng) thì chứng tỏ trong 10 người đó có ít nhất một người bệnh, khi đó phải làm thêm 10 xét nghiệm lẻ cho mỗi người để tìm người bệnh.

Hỏi làm theo phương án nào thì phải thực hiện ít xét nghiệm hơn?

(Đây là VD minh họa ý nghĩa của kỳ vọng toán)

Hướng dẫn:

* Nếu thực hiện theo phương án I cần 5000 xét nghiệm.

* Gọi X là số xét nghiệm cần thực hiện đối với mỗi nhóm 10 người theo phương pháp II.

X là biến ngẫu nhiên có bảng phân phối xác suất :

X	1	11
P	$0,9^{10}$	$1-0,9^{10}$

$$\Rightarrow E(X) = 1\times 0,9^{10} + 11\times(1-0,9^{10}) = 7,5132$$

Suy ra số xét nghiệm trung bình cho 500 nhóm như vậy là:

$$500 \times 7,5132 \approx 3\ 757 \text{ xét nghiệm.}$$

* Như vậy có thể nói số xét nghiệm cần thực hiện theo phương án II là ít hơn phương án I.

Ví dụ 6:

Thống kê về tai nạn giao thông cho thấy hàng năm, tỉ lệ một người bị tai nạn xe máy theo mức độ nhẹ và nặng tương ứng là 0,001 và 0,005.

Một công ty bảo hiểm xe máy có mức phí thu hàng năm là 30.000 đồng/người; số tiền chi trung bình cho mỗi người trong một vụ tai nạn giao thông ở mức độ nhẹ là 1 triệu đồng và nặng là 3 triệu đồng.

Hỏi lợi nhuận trung bình hàng năm công ty thu được đối với mỗi người mua bảo hiểm là bao nhiêu, biết rằng ngoài thuế doanh thu phải nộp 10% thì tổng tất cả các chi phí khác chiếm 15% doanh thu.

*Hướng dẫn: 30 ngàn -16 ngàn - (10%+15%)*30 ngàn (đồng)*

Ví dụ 7:

Cho biết tuổi thọ X (đơn vị: tháng) của một loại côn trùng là một ĐLNN có hàm phân phối xác suất :

$$F(x)=\begin{cases} 0 & x<0 \\ k\left(\dfrac{4x^3}{3}-\dfrac{x^4}{4}\right) & x\in[0;4] \\ 1 & x>4 \end{cases}$$

a) Tìm hệ số k và tính xác suất để côn trùng chết trước khi nó được 1 tháng tuổi.

b) Tìm hàm mật độ xác suất của X.

c) Hãy tính tuổi thọ trung bình của côn trùng đó.

d) Hãy tìm mức tuổi thọ mà 1 nửa số côn trùng không sống qua được mức đó.

Hướng dẫn: a) Do X là ĐLNN liên tục nên hàm F(x) liên tục trên R , suy ra F(x) liên tục tại x=4 $\Rightarrow$ F(4+) = F(4-) =F(4).

$$\Rightarrow k\left(\frac{4x^3}{3}-\frac{x^4}{4}\right)\Bigg|_{x=4}=1 \qquad \Rightarrow k=\frac{3}{64}$$

(cách lập luận này chưa chặt)

$$*\ P(\ 0<X<1\ )\ =\ F(1)-F(0)=\frac{3}{64}\left(\frac{4\times 1^3}{3}-\frac{1^4}{4}\right)-0=\frac{3}{256}$$

b) $$f(x)=[F(x)]'=\begin{cases}\frac{3}{64}*x^2(4\text{-}x) & x\in[0;4]\\ 0 & x\notin[0;4]\end{cases}$$

c) $$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^4\frac{3}{64}*x^3(4\text{-}x)dx=\frac{12}{5}$$

d) Tìm giá trị m mà P(X < m) =0,5 hay F(m) =0,5. m chính là trung vị của X.

II.3.7 HD Sử dụng MTBT tìm 1 số đặc trưng của BNN rời rạc:

Các bước thực hiện	Máy CASIO fx 570 ES	Máy CASIO fx 500 MS....
Xóa nhớ bài cũ	....	......
Mở cột tần số (nếu máy chưa mở)	SHIFT -- MODE (SETUP) – -- ▼ -- 4 (STAT) -- 1 (ON)	
Vào chế độ thống kê một biến.	MODE -- 3 (STAT) --1 (1-VAR)	MODE -- MODE --...-- 1 (SD)
Nhập dữ liệu	X FREQ 1 x1 p1 2 x2 p2 … … … … xn pn AC	Nhập lần lượt trên dòng theo thứ tự: X_i ; p_i M+ Xi là các giá trị của BNN, p_i là các XS tương ứng,
Đọc kết quả E(X)	SHIFT – 1 (STAT)- 4 (VAR) – --- 2 ($\overline{x}$) -- =	SHIFT – 2 (SVAR) -1 ($\overline{x}$)-- =
Đọc kết quả $\sqrt{D(X)}$	SHIFT – 1 (STAT)- 4 (VAR) – --- 3 (σX) -- =	SHIFT – 2 (SVAR)- 2 (xσn)-- =
Tham khảo các KQ trung gian	SHIFT – 1 (STAT)- 3 (SUM) -	SHIFT – 1 (SSUM) - ......

II.4 MỘT SỐ QUY LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

II.4.1 Phân phối Bernoulli. *(tham khảo)*

Định nghĩa: BNN rời rạc **X** gọi là có phân phối Bernoulli, *(Bernoulli Distribution)*, kí hiệu $X \sim B(1, p)$ với tham số $p \in (0,1)$ nếu X có bảng PPXS dạng:

X	*0*	*1*
P_i	*1-p*	*p*

Tính chất:

* Nếu $X \sim B(1, p)$ thì $E(X) = p$ và $D(X) = pq$; $q=1- p$.

- Phân phối này còn gọi là phân phối Không – Một, k.h $X\sim A(p)$.
- Nếu $X_1; X_2;…;X_n \sim B(1,p)$ thì biến ngẫu nhiên $X = X_1+ X_2+…+ X_n$ có phân phối Nhị thức B(n,p)(xem mục tiếp sau).

II.4.2 Phân phối nhị thức:

Định nghĩa: BNN rời rạc X gọi là có phân phối nhị thức *(Binomial Distribution)*, kí hiệu **X ~ B(n, p)**, với 2 tham số $n \in \mathbb{N}$; $p \in (0,1)$; ($q=1-p$) nếu X có bảng PPXS dạng:

X	0	1	…	k	…	n
P_i	q^n	$n.p^1.q^{n-1}$	…	$C_n^k p^k q^{n-k}$	…	p^n

Tính chất:

* Xét dãy n phép thử Bernoulli với xác suất "thành công" trong mỗi phép thử là p. Ký hiệu X là số lần "thành công" xuất hiện trong dãy n phép thử thì X ~ B(n, p).

* Nếu X ~ B(n, p) thì E(X) = np và D(X) = npq , với q=1-p .

* Mod (X) chính là số lần "thành công" có khả năng nhất.

* Khi số *n* khá lớn, p không quá gần 0 hay quá gần 1, người ta thường *xấp xỉ phân phối Nhị thức B(n,p) với phân phối Chuẩn N(np,npq) (xem phân phối chuẩn ở mục II.4.8)*, cụ thể:

$$+\ C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} f\left(\frac{k-np}{\sqrt{npq}}\right) \quad (1) \text{ với } f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (\text{hàm Gauss})$$

$$+\sum_{k=k_1}^{k=k_2} C_n^k p^k q^{n-k} \approx \Phi(x_2) - \Phi(x_1) \quad (2) \text{ ở đây } \quad \Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

là hàm tích phân Laplace và $x_i = \dfrac{k_i - np}{\sqrt{npq}}$; $i = 1, 2$

* Khi số *n* khá lớn và *p* rất gần với 0 hoặc 1 thì người ta thường dùng *xấp xỉ phân phối Nhị thức B(n,p) với phân phối Poisson P(np) (xem phân phối Poisson ở mục II.4.5)*, cụ thể:

$$+\ C_n^k p^k q^{n-k} \approx \frac{e^{-\lambda} \lambda^k}{k!}, \quad \lambda = n.p \qquad (3)$$

Một số lưu ý thêm:

* Xấp xỉ ở công thức (2) là tốt nhất nếu n lớn và np>5; nq>5 hoặc npq>20. Công thức (2) là kết quả định lý giới hạn Moivre – Laplace.

* Công thức (2) có thể hiệu chỉnh để có kết quả tốt hơn bằng cách tăng *k2* lên nửa đơn vị và k1 giảm đi nửa đơn vị , khi tính *x1*, *x2*. Cụ thể :

$$\sum_{k=k_1}^{k=k_2} C_n^k p^k q^{n-k} \approx \Phi(x_2^*) - \Phi(x_1^*)$$

$$x_1^* = \frac{k_1 - \frac{1}{2} - np}{\sqrt{npq}} \qquad x_2^* = \frac{k_2 + \frac{1}{2} - np}{\sqrt{npq}}$$

* Công thức (3) xấp xỉ tốt khi n> 20 và p< 5%. (Kết quả xấp xỉ rất chính xác khi n>100 và λ=np <10). Vì p nhỏ nên người ta còn gọi nó là công thức của định luật số hiếm.

Ví dụ 8

Một tổng đài nội bộ của một cơ quan phục vụ 100 máy điện thoại. Xác suất để trong một phút mỗi máy điện thoại gọi đến tổng đài là 0,02.

a) Tìm xác suất để trong 1 phút có 3 máy gọi đến tổng đài.

b) Gọi X là BNN chỉ số máy điện thoại gọi đến tổng đài trong 1 phút. Hãy cho biết X có phân phối gì?

c) Tìm số máy gọi đến tổng đài trung bình trong một phút.

d) Tìm xác suất trong một phút có từ 3 đến 10 máy gọi đến tổng đài *(tính bằng các công thức xấp xỉ rồi so sánh với cách tính trực tiếp).*

Ví dụ 9

Tỷ lệ nảy mầm của một loại hạt giống là 75%. Nếu gieo ngẫu nhiên 120 hạt giống thì xác suất có được từ 80 hạt nảy mầm trở lên là bao nhiêu?

Ví dụ 10

Xác suất 1 sản phẩm bị lọt qua khâu kiểm tra chất lượng ban đầu là 8%.

a) Tính xác suất trong 900 sản phẩm có 70 sản phẩm không được kiểm tra chất lượng ban đầu.

b) Tính xác suất trong 9000 sản phẩm có từ 700 đến 800 sản phẩm không được kiểm tra chất lượng ban đầu.

Ví dụ 8 Đây là bài toán Bernoulli với n=100; p=0,02.

a) $C_{100}^{3}\times 0,02^{3}0,98^{97}$ b) X ~ B(n=100;p=0,02) c) E(X)= np=2

d) * Tính trực tiếp: $\sum_{k=3}^{10} C_{100}^{k}\times 0,02^{k}0,98^{100-k}\approx 0,3233087$

* Xấp xỉ pp Chuẩn: $\Phi(\frac{10-100\times 0,02}{\sqrt{100\times 0,02\times 0,98}})-\Phi(\frac{3-100\times 0,02}{\sqrt{100\times 0,02\times 0,98}})\approx 0,237525$

* Xấp xỉ pp Poisson: $\sum_{k=3}^{10} e^{-2}\frac{2^{k}}{k!}\approx 0,3233153;$ $\lambda = 100\times 0,02 = 2$

Ví dụ 9

$$\sum_{k=80}^{120} C_{120}^{k}(0.75)^{k}(0.25)^{120-k}\approx \Phi(\frac{120-90}{\sqrt{90\times 0,25}})-\Phi(\frac{80-90}{\sqrt{90\times 0,25}})\approx 98,249\%$$

Ví dụ 10

a) $C_{900}^{70}(0,08)^{70}(0,92)^{830}\approx \frac{1}{\sqrt{900\times 0,08\times 0,92}}f(\frac{70-900\times 0,08}{\sqrt{900\times 0,08\times 0,92}})\approx 0,122868\times 0,38708\approx 0,0476$

b) $\Phi(\frac{800-9000\times 0,08}{\sqrt{9000\times 0,08\times 0,92}})-\Phi(\frac{700-9000\times 0,08}{\sqrt{9000\times 0,08\times 0,92}})\approx \Phi\left(3,1083\right)-\Phi\left(-0,7771\right)=0,7805$

II.4.3 Phân phối Hình học *(tham khảo)*

Định nghĩa: BNN rời rạc X gọi là có phân phối Hình học (*Geometric Distribution*), kí hiệu **X ~ G(p)** với tham số $p \in (0;1)$ nếu bảng phân phối xác suất của X có dạng:

X	*1*	*2*	*3*	…	*k*	…
P_i	p	$q^1.p$	$q^2.p$	…	$(1-p)^{k-1}.p$	…

Tính chất:

* Nếu $X \sim P(\lambda)$ thì: $E(X) = \dfrac{1}{p}$; $D(X) = \dfrac{1}{p^2} - \dfrac{1}{p}$

* X được coi là số thử nghiệm Bernoulli cần thiết để có được sự thành công.

* Trong Ví dụ 2, BNN Y có phân phối hình học.

Ví dụ 11:

Một băng chuyền tự động sản xuất ra các sản phẩm với xác suất phế phẩm là p và được dừng ngay để điều chỉnh khi xuất hiện *một* phế phẩm. Tìm kỳ vọng của số sản phẩm được sản xuất giữa 2 lần điều chỉnh kề nhau. ĐS: 1/p.

Ví dụ 12:

Hai cầu thủ bóng rổ lần lượt ném bóng vào rổ cho đến khi ném lọt mới thôi. Xác suất ném trúng rổ của người thứ nhất là a và người thứ 2 là b.

a) Lập bảng phân phối xác suất cho biến ngẫu nhiên X là số lần ném của người thứ nhất.

b) Tìm số lần ném trung bình của người thứ nhất và tìm D(X).

HD: $X \sim G(p = a+(1-a).b)$

II.4.4 Phân phối siêu bội.

Định nghĩa: BNN X gọi là có phân phối Siêu bội (*Hypergeometric Distribution*), kí hiệu **X ~ H(N,M,n)**, với tham số là các số tự nhiên n, N, M , n≤ M≤ N ; nếu bảng PPXS của X có dạng:

X	0	1	...	k	...	n
P_i	$\frac{C_{N-M}^{n}}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	...	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	...	$\frac{C_M^n}{C_N^n}$

Tính chất:

* Nếu X ~ H(N,M,n) thì E(X) = np

và $D(X) = npq\dfrac{N-n}{N-1}$ với $p = \dfrac{M}{N}$ và $q = 1-p$.

- Khi n << N (n rất nhỏ so với N) thì người ta thường *xấp xỉ phân phối Siêu bội với phân phối Nhị thức* , tức là coi như

$$p_k = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k q^{n-k};\ \ k = 0,1,\dots,n.$$ *(xem mục II.5.5)*

Ví dụ 13

Một lô hàng có N= 50 bóng đèn, trong đó lẫn 10 bóng hỏng. Lấy ngẫu nhiên 5 bóng để kiểm tra.
Gọi X là BNN chỉ số bóng hỏng trong 5 bóng được lấy ra.

a) Tìm P(X =2).

b) Hãy cho biết X có phân phối gì ?

c) Tính số bóng hỏng trung bình trong các bóng được lấy ra và phương sai của X.

Lưu ý 1: Nếu đổi giả thiết là "Một lô hàng có 50 bóng đèn, mỗi bóng đèn có tỉ lệ hỏng là 20% ..." $\Rightarrow$ **X có PP Nhị thức.**

Lưu ý 2: Nếu đổi giả thiết N=500, ta thấy n=5 << N=500 nên

$$P(X=2) \overset{C1}{=} \frac{C_{10}^{2} C_{490}^{3}}{C_{500}^{5}} = 0.003436 \overset{C2}{\approx} C_5^2 \left(\frac{10}{500}\right)^2 \left(\frac{490}{500}\right)^3 = 0.003765$$

II.4.5 Phân phối Poisson.

Định nghĩa: Biến ngẫu nhiên rời rạc X gọi là có phân phối Poisson (*Poisson Distribution*), kí hiệu **X ~ P(λ)** ; với tham số λ>0 , nếu bảng phân phối XS của X có dạng:

X	0	1	2	…	k	…
P_i	$e^{-\lambda}$	$e^{-\lambda}.\lambda$	$\frac{e^{-\lambda}.\lambda^2}{2!}$	…	$\frac{e^{-\lambda}\lambda^k}{k!}$	…

Tính chất:

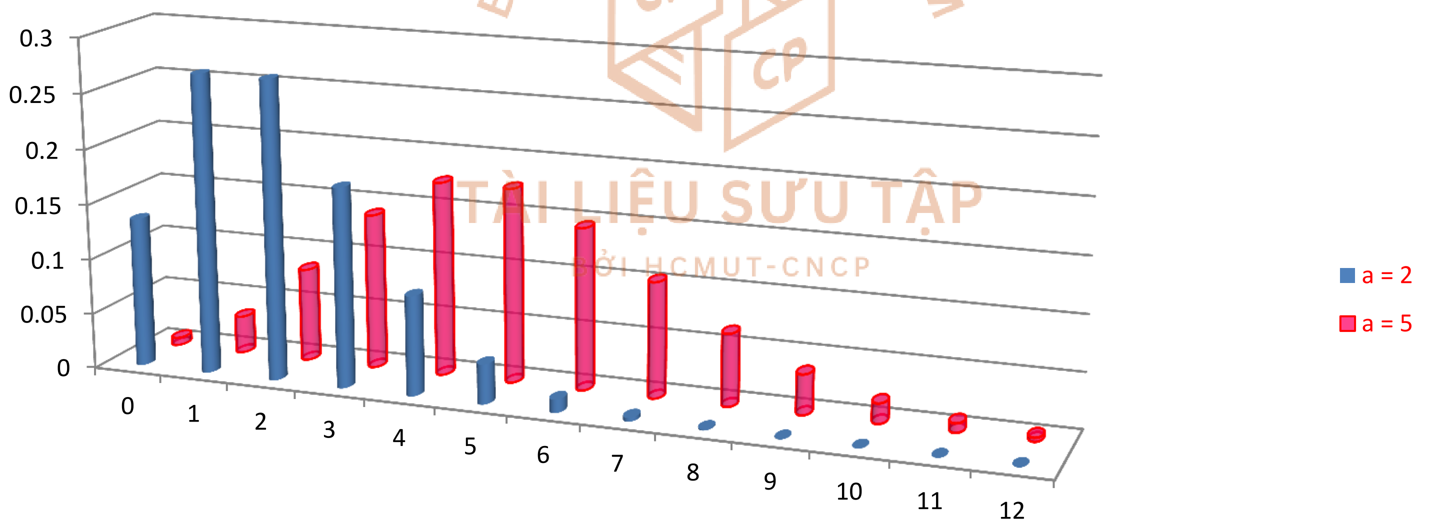
* Nếu X ~ P(λ) thì **E(X) = D(X) = λ** * λ-1 ≤ Mode X ≤ λ; Mode X ∈ $\mathbb{N}$

* Trong thực tiễn, công thức Poisson có thể dùng thay cho công thức Bernoulli khi số phép thử rất lớn và xác suất thành công rất nhỏ. Có nhiều BNN tuân theo luật phân phối Poisson, chẳng hạn như số người vào các trạm phục vụ công cộng trong một đơn vị thời gian; hay số lỗi trong mỗi trang của một quyển sách; số hoa nở trong ngày của một loại hoa; số lần truy cập vào máy chủ web trong mỗi phút; số lượng ngôi sao trong 1 thể tích không gian vũ trụ,…

Dưới đây là biểu đồ minh họa cho phân phối Poisson,

khi λ (hay a)=2 ; và khi λ = 5.

k=	0	1	2	3	4	5	6	7	8	9	10	11	12
$\lambda = 2$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0034	0.0009	0.0002	3.82E-05	6.944E-06	1.157E-06
$\lambda = 5$	0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462	0.1044	0.0653	0.0363	0.0181	0.0082	0.0034



Ví dụ 14

Biết rằng số người vào một đại lý bưu điện trong một khoảng thời gian xác định là một BNN có phân phối Poisson.

Quan sát thấy cứ 5 phút có 12 người ghé vào một đại lý bưu điện.

Tìm xác suất trong một phút có 3 người vào đại lý bưu điện đó.

Hướng dẫn:

Gọi X là số người vào một đại lý bưu điện trong khoảng thời gian 1 phút. Khi đó $X \sim P(\lambda)$.

Từ giả thiết ta thấy số người trung bình vào ĐLBĐ này trong 1 phút là $12/5 = 2,4 \Rightarrow E(X) = 2,4 \Rightarrow \lambda = 2,4$.

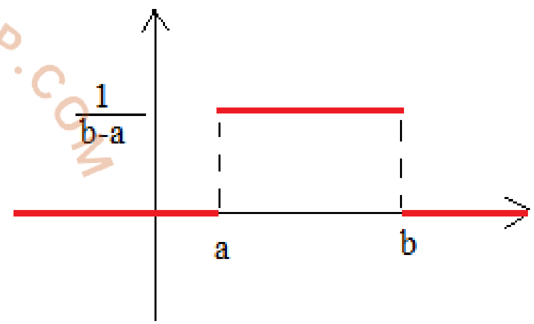
Vậy xác suất cần tìm là $P(X=3) = e^{-2,4}\dfrac{2,4^3}{3!} \approx 0,2090$

II.4.6 Phân phối đều.

a) Phân phối đều liên tục:

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối đều (*Uniform Distribution*) trên đoạn [a, b] , kí hiệu **X~ U(a,b)** nếu hàm mật độ của nó có dạng:

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



Tính chất:

- Nếu X ~ U(a,b) thì $E(X) = \dfrac{a+b}{2}$ và $D(X) = \dfrac{(b-a)^2}{12}$

b) Phân phối đều rời rạc: Nếu X chỉ nhận n giá trị với xác suất như nhau thì ta nói X có phân phối đều rời rạc.

* Trong thống kê người ta thường có quy ước: nếu ta không biết gì về giá trị của tham số cần ước lượng thì coi như mỗi giá trị có thể có của tham số là đồng khả năng. Điều đó dẫn đến việc quan niệm giá trị tham số cần ước lượng như một BNN tuân theo quy luật phân phối đều.

Ví dụ 15

Giả thiết rằng cứ 15 phút có 1 chuyến xe buýt đi qua trạm. Chuyến đầu tiên bắt đầu lúc 5g00. Sinh viên A tới trạm vào một thời điểm bất kỳ trong khoảng từ 5g00 đến 5g30. Tìm xác suất sinh viên A phải đợi trên 10 phút.

Ví dụ 16

Một đoạn thẳng AB dài 12 cm được chia thành 2 đoạn bởi một điểm M lấy ngẫu nhiên trên AB. Người ta dùng 2 đoạn AM và BM để làm 2 cạnh của 1 hình chữ nhật. Tính diện tích trung bình của hình chữ nhật đó.

Hướng dẫn:

Ví dụ 15 :

Gọi X là thời điểm sinh viên A có mặt ở bến xe buýt. Từ giả thiết có thể xem X có phân phối đều trên [5g00; 5g30]. X có hàm mật độ:

$$f(x)=\begin{cases}\dfrac{1}{30} & x\in[5g00;5g30]\\ 0 & x\notin[5g00;5g30]\end{cases}$$

XS cần tìm = P(5g00<X<5g05)+ P(5g15<X<5g20)= $^1/_3$.

Ví dụ 16:

Gọi X là độ dài đoạn AM, Bnn X có phân phối đều trên(0;12).

Diện tích HCN = X(12-X) = 12X – X^2 .

E(X) = E(12X – X^2) = 12.E(X) - E(X^2) = 24 ; do:

$$E(X)=\int_0^{12} x.\frac{1}{12}dx=6 \qquad E(X^2)=\int_0^{12} x^2.\frac{1}{12}dx=48$$

II.4.7 Phân phối mũ.

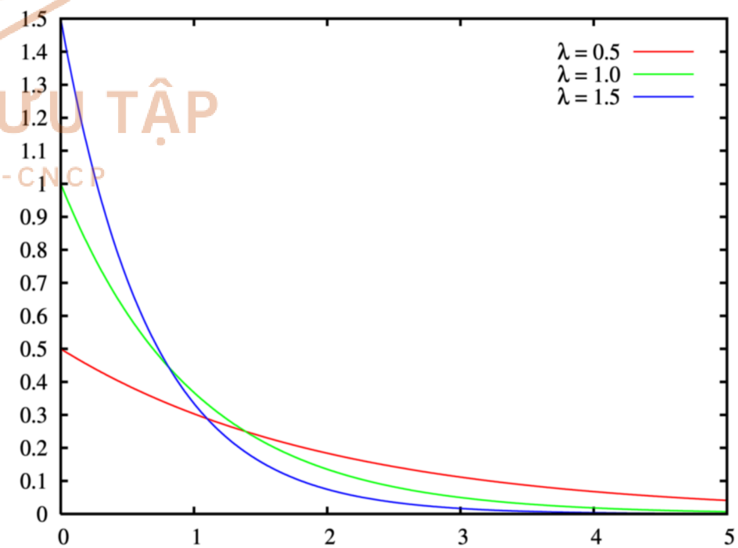
Định nghĩa: Biến nn liên tục X được gọi là có phân phối mũ (*Exponential Distribution*) với tham số $\lambda > 0$, kí hiệu **X ~ E(λ)**, nếu hàm mật độ của nó có dạng :

$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

Tính chất:

* Nếu X ~ E(λ) thì $E(X) = \frac{1}{\lambda}$ và $D(X) = \frac{1}{\lambda^2}$

Phân phối mũ có ứng dụng trong nhiều lĩnh vực, nó thường được dùng trong việc đánh giá quá trình sản xuất và cung cấp dịch vụ. Trong các hệ thống kỹ thuật, thời gian làm việc liên tục của máy móc thiết bị giữa 2 lần sửa chữa cũng tuân theo phân phối mũ.



Ví dụ 17 Thời gian chờ được phục vụ của khách hàng ở một cửa hàng là 1 BNN X có hàm mật độ xác suất sau:

(đơn vị x: phút)

$$\varphi(x) = \begin{cases} Ae^{-5x} & khi\ x > 0 \\ 0 & khi\ x \le 0 \end{cases}$$

a) Tìm hệ số A và tính xác suất thời gian chờ được phục vụ của một khách hàng nào đấy nằm trong khoảng từ 0,4 đến 1 phút.

b) Hãy tìm thời gian khách phải chờ trung bình và phương sai của X.

Ví dụ 18

Khoảng thời gian (phút) giữa 2 người kế tiếp nhau đến 1 máy ATM là một ĐLNN mà hàm mật độ xác suất có dạng:

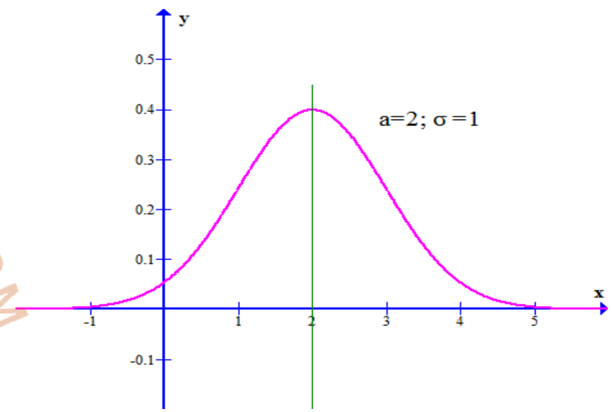
$$f(x) = \begin{cases} ke^{-\frac{1}{2}x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

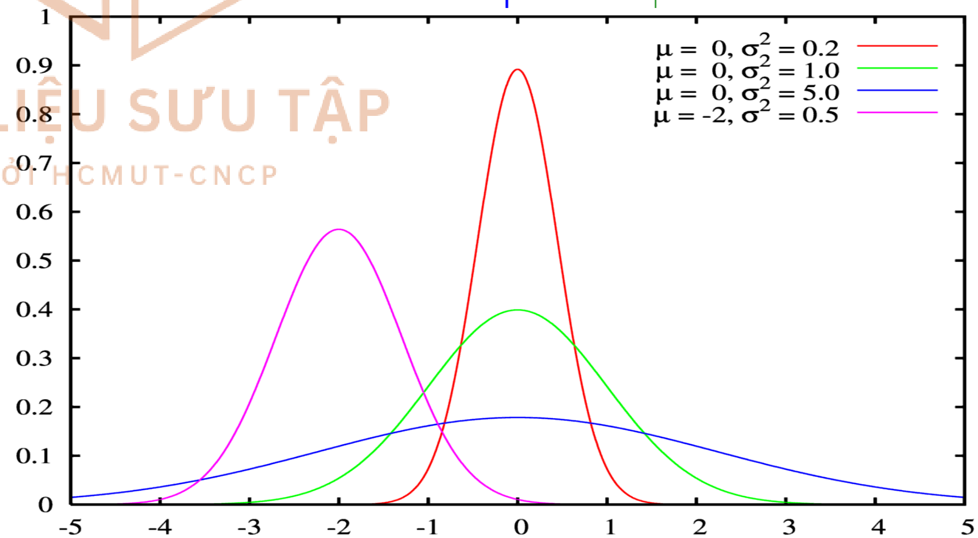
Nếu có một người vừa đến máy ATM thì xác suất sẽ có người kế tiếp đến máy này trong vòng 2 phút tiếp theo là bao nhiêu?

II.4.8 Phân phối chuẩn.

Định nghĩa: BNN X được gọi là có *phân phối chuẩn* *(Normal Distribution)*, ký hiệu *X~N(a, σ^2)*, nếu hàm mật độ xác suất của X có dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \sigma > 0, \; x \in \mathbb{R}$$



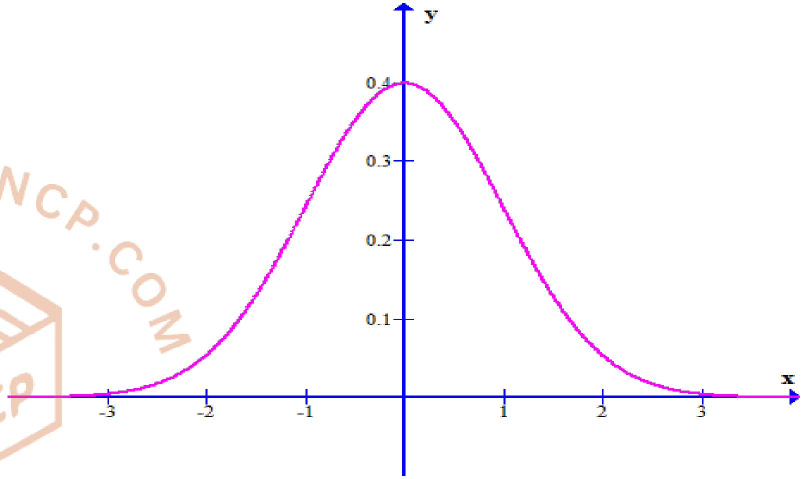


Tính chất:

- Nếu X~N(a, σ^2) thì E(X) = a, D(X) = σ^2

- Trường hợp a = 0 và σ =1 thì $X \sim N(0, 1)$, còn gọi là phân phối (chuẩn) chuẩn tắc. Hàm mật độ của phân phối chuẩn tắc được gọi là hàm mật độ Gauss.

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



- Nếu $X \sim N(a, \sigma^2)$ thì

$$Y = \frac{X-a}{\sigma} \sim N(0;\ 1)$$

- Giá trị hàm mật độ Gauss có thể tính trực tiếp bằng MTBT hay tra từ bảng I trong phần Phụ lục phía sau giáo trình. Khi tra bảng ta lưu ý hàm Gauss là hàm chẵn, và với các giá trị x mà $|x| > 3$ thì f(x) xấp xỉ bằng 0.
- VD: $f(1{,}24) \approx 0{,}1849$. *(Tìm số nằm ở dòng 1,2; cột 4 trong PL1, thêm 0,.. ở trước kết quả tìm được).* $f(-1{,}24) \approx 0{,}1849$.

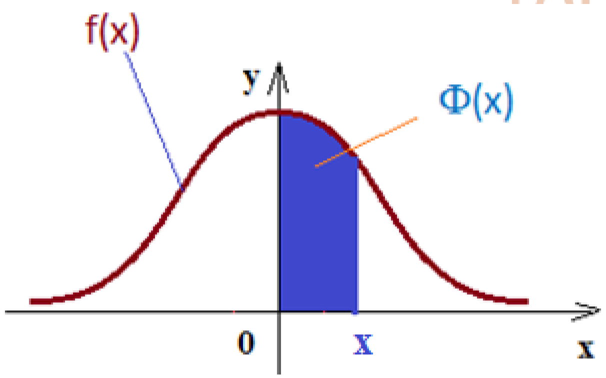
- Hàm phân phối xác suất của phân phối chuẩn tắc là một hàm không biểu diễn được ở dạng hàm sơ cấp.

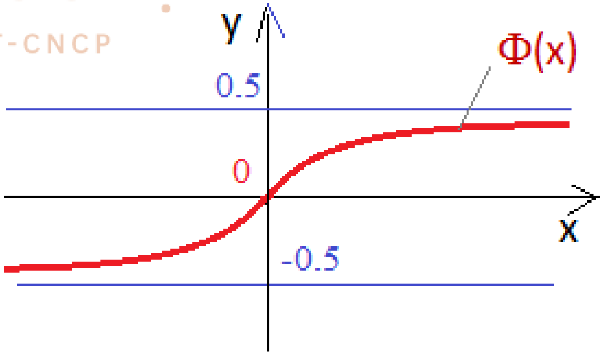
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- Trong tính toán người ta còn thường dùng hàm tích phân Laplace:

$$\boxed{\Phi(x) = \int_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}$$

Hàm Φ là hàm lẻ , tăng trên R. Dễ thấy F(x) = Φ(x) + 0,5.





Ta có thể tìm giá trị $\Phi(x)$ theo các cách:

Cách 1: *Dùng bảng II để tìm gần đúng giá trị của hàm.*

Ví dụ: $\Phi(1,24) \approx 0,39251$. $\Phi(-1,24) \approx -0,39251$

Khi x >5 , hàm tăng rất chậm về 0,5 nên nếu có thể bỏ qua sai số thì $\Phi(x) \approx 0,5$ khi $x > 5$. Tương tự, $\Phi(x) \approx -0,5$ khi $x < -5$.

Cách 2: Sử dụng MTBT CASIO fx 570 ES PLUS để tính $|\Phi(x)|$:

Vào chế độ thống kê 1 biến **MODE -- 3 (STAT) -- 1 (1-VAR)**

Nhấn phím **AC** để bỏ qua bước nhập số liệu.

Bấm **SHIFT -- 1 (STAT) -- 5 (Distr) -- 2 (Q( ) -- x (*Nhập x*) -- =**

Nếu $x > 0$ thì $\Phi(x)=Q(x) > 0$; nếu $x< 0$ thì $\Phi(x)= - Q(x) < 0$

Tham khảo: Nếu chọn **1** (**R**() thì ta được giá trị hàm F(x).

Cách 3: Dùng MTBT bấm trực tiếp theo công thức hàm $\Phi(x)$.

* Nếu $X \sim N(a, \sigma^2)$ thì :

$$+P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$

$$+P(|X - a| < \varepsilon) = 2.\Phi\left(\frac{\varepsilon}{\sigma}\right) \qquad \varepsilon > 0.$$

Cm:

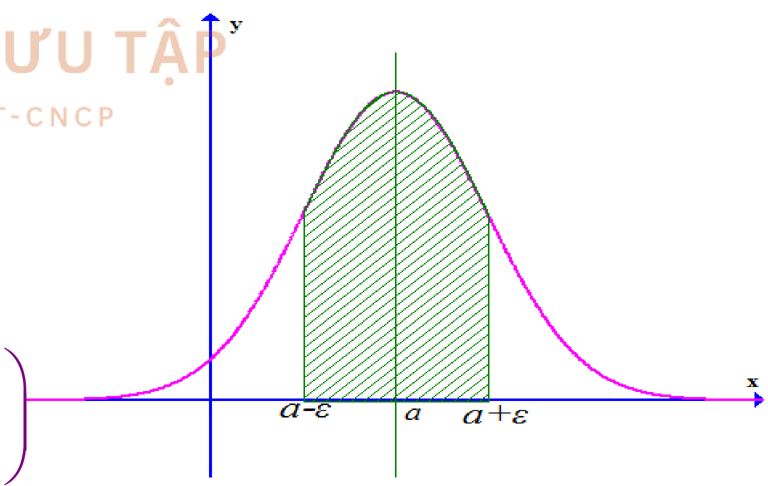
$$+P(\alpha < X < \beta) = \int_{\alpha}^{\beta} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \overset{đb}{=} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \quad y = \frac{x-a}{\sigma}$$

$$= \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$

$$+P(|X - a| < \varepsilon) = P(-\varepsilon < X - a < \varepsilon)$$

$$= P(a - \varepsilon < X < a + \varepsilon)$$

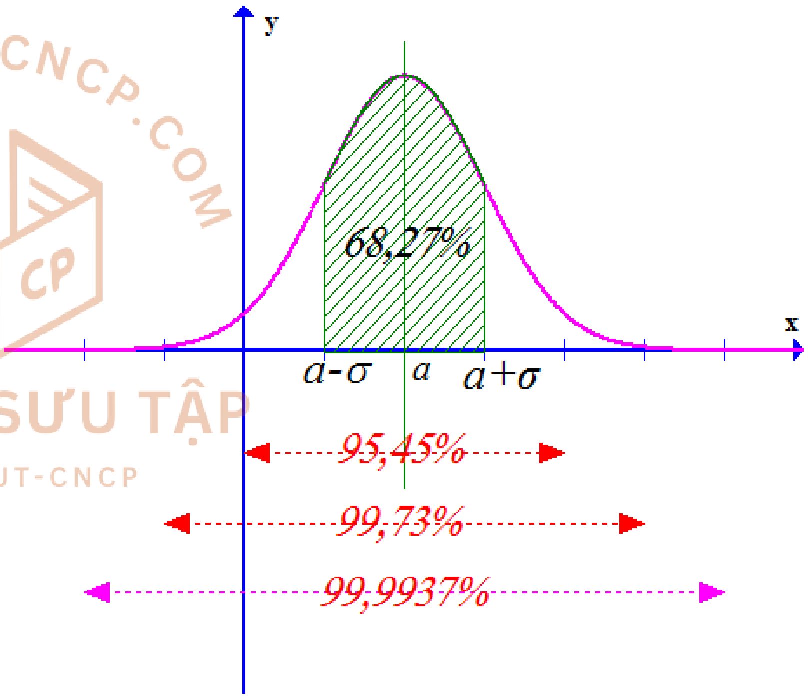
$$= \Phi\left(\frac{a + \varepsilon - a}{\sigma}\right) - \Phi\left(\frac{a - \varepsilon - a}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$



* Người ta thường nhắc đến các quy tắc 2- sigma và 3- sigma:

+ $P(|X-a| < 2\sigma) = 2.\Phi(2) \approx 95{,}45\ \%$
+ $P(|X-a| < 3\sigma) = 2.\Phi(3) \approx 99{,}73\ \%$

Một công ty thực phẩm đang chuẩn bị đưa một loại bánh mới ra thị trường.



Người ta nhận thấy rằng số ngày sử dụng tốt của mỗi chiếc bánh trong điềukiện khuyến cáo là bnn có phân Phối chuẩn mà trung bình là 40 ngày Và phương sai 4 ngày^2. *(dùng thống kê, sẽ học sau).*

Công ty nên công bố thời hạn sử dụng của loại bánh này là bao nhiêu ngày?

- Phân phối chuẩn là 1 quy luật phân phối rất thường gặp vì có nhiều phân bố xác suất trong tự nhiên và trong thực tế đời sống có hình dáng khá giống phân phối chuẩn.
- Trong công nghiệp, người ta đã xác định được rằng kích thước của các chi tiết do các nhà máy sản xuất ra sẽ có phân phối chuẩn nếu quá trình sản xuất diễn ra bình thường.
- Trong nông nghiệp, năng suất của cùng một loại cây trồng tại các thửa ruộng khác nhau cũng có phân phối chuẩn.
- Các chỉ số về thể lực và trí tuệ con người cũng tuân theo phân phối chuẩn…

Ví dụ 19 Trọng lượng sản phẩm X do một máy tự động sản xuất là một BNN tuân theo qui luật chuẩn với E(X) = 100 g và độ lệch chuẩn 1g. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó nằm từ 99 đến 101,2 g .

a) Tìm tỉ lệ sản phẩm đạt tiêu chuẩn do máy sản xuất.

b) Cho máy sản xuất liên tiếp 200 sản phẩm. Tìm xác suất có được ít nhất 100 sản phẩm đạt tiêu chuẩn.

Hướng dẫn

$$a)\ P\left(99 < X < 101{,}2\right) = \Phi\left(\frac{101{,}2-100}{1}\right) - \Phi\left(\frac{99-100}{1}\right) = 0{,}38493 + 0{,}34134 = 72{,}63\%$$

b) Gọi Y là số sản phẩm đạt tiêu chuẩn trong 200 sản phẩm.

Y~ B(n=200;p=0,7263). Xác suất cần tìm:

$$P\left(100 \le Y \le 200\right) \approx \Phi\left(\frac{200-200\times 0{,}7263}{\sqrt{200\times 0{,}7263\times 0{,}2737}}\right) - \Phi\left(\frac{100-200\times 0{,}7263}{\sqrt{200\times 0{,}7263\times 0{,}2737}}\right) = 0{,}5+0{,}5=1$$

Ví dụ 20

Tuổi thọ của một loại sản phẩm là biến ngẫu nhiên phân phối chuẩn với trung bình 11 năm và độ lệch chuẩn là 2 năm.

a) Nếu quy định thời gian bảo hành sản phẩm là 10 năm thì tỉ lệ sản phẩm phải bảo hành là bao nhiêu?

b) Nếu muốn tỉ lệ sản phẩm phải bảo hành chỉ là 10% thì người ta cần quy định thời gian bảo hành là bao lâu ?

c) Nếu một sản phẩm đã hoạt động tốt qua thời gian bảo hành là 10 năm (câu a) thì xác suất nó vẫn hoạt động tốt trong 3 năm tiếp theo là bao nhiêu ?

Tuổi thọ sản phẩm trong bài này được quy ước là khoảng thời gian từ khi người dùng mua sản phẩm cho đến khi sản phẩm cần đem đến bảo hành.

Gọi X là tuổi thọ của sản phẩm loại này. $X \sim N(11, (2 \text{ năm})^2)$.

a) $P(X \le 10) = P(0 \le X \le 10) =$

$$= \Phi\left(\frac{10-11}{2}\right) - \Phi\left(\frac{0-11}{2}\right) = \Phi(-0,5) + 0,5 = -0,1915 + 0,5 = 0,3085$$

b) Gọi A là thời hạn bảo hành cần tìm theo yêu cầu.

Từ giả thiết suy ra $P(X \le A) = 10\%$, hay $\Phi\left(\frac{A-11}{2}\right) + 0,5 = 0,1$

$$\Rightarrow \Phi\left(\frac{A-11}{2}\right) = -0,4 = \Phi(-1,28) \Rightarrow \frac{A-11}{2} = -1,28$$ (do hàm Φ đơn điệu).

Suy ra A= 8,4 năm.

c) A là b/c sản phẩm đã hoạt động tốt qua thời gian bảo hành,

B là biến cố sản phẩm vẫn hoạt động tốt trong 3 năm tiếp.

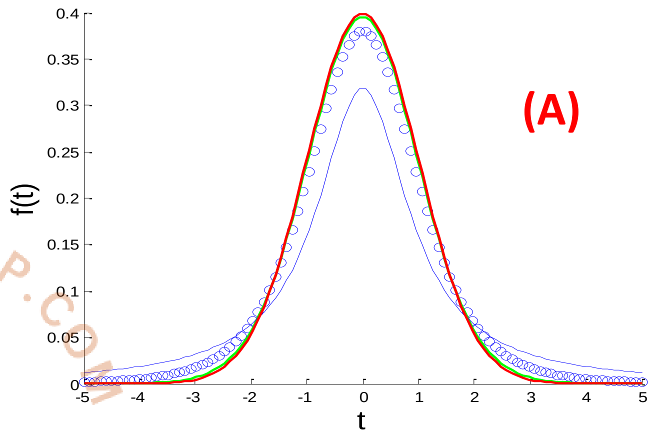
Xác suất cần tìm:

$$P(B/A) = \frac{P(AB)}{P(A)} = \frac{P(X>13)}{P(X>10)} = \frac{0,5-\Phi(\frac{13-11}{2})}{0,5-\Phi(\frac{10-11}{2})} = 0,5142$$

SV đọc giáo trình để tìm hiểu thêm về các dạng phân phối sau:

II.4.9 Phân phối Student (A)

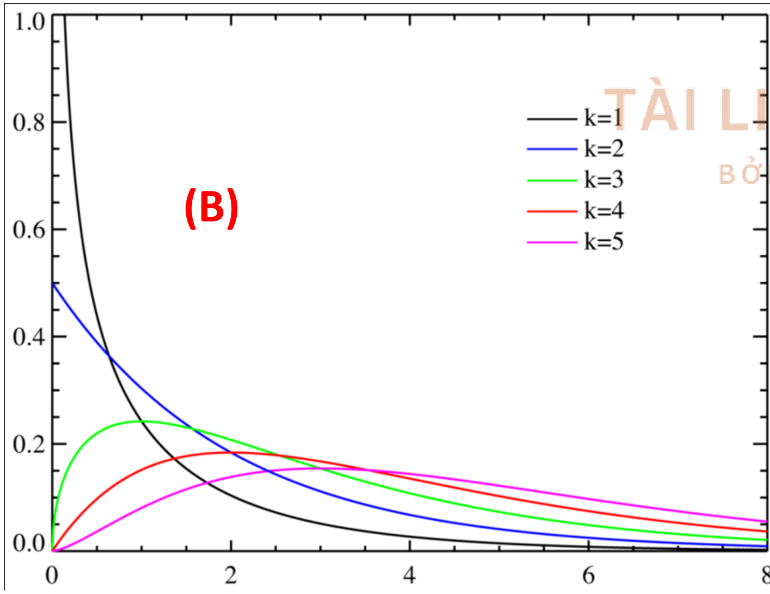
Khi $n \geq 30$, PP Student coi như xấp xỉ PP Chuẩn chuẩn tắc.

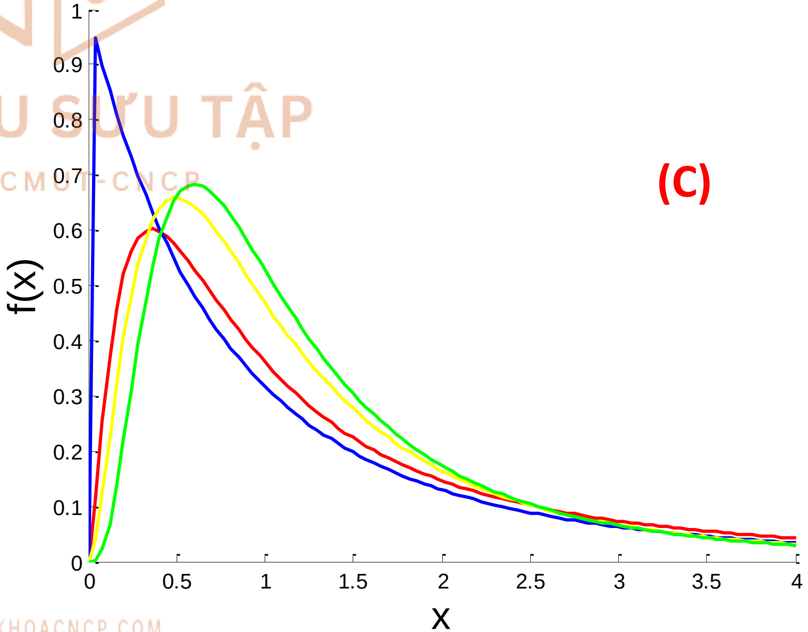


II.4.10 Phân phối Chi bình phương (B)

(hay Khi bình phương)

II.4.11 Phân phối Fisher (C)





II.5 CÁC ĐỊNH LÝ GIỚI HẠN

II.5.1 Bất đẳng thức Chebyshev:

Nếu X là biến ngẫu nhiên có kỳ vọng toán và phương sai hữu hạn thì với mọi số dương ε tùy ý ta luôn có:

$$\boxed{P\left(\left|X - E(X)\right| < \varepsilon\right) \geq 1 - \frac{D(X)}{\varepsilon^2}}$$

Ví dụ 21:

Thu nhập trung bình hàng năm của dân cư ở 1 vùng là 18 triệu và độ lệch chuẩn là 3,2 triệu đồng. Hãy tìm một khoảng thu nhập hàng năm xung quanh giá trị trung bình của ít nhất 95% cư dân vùng đó.

HD: Gọi X là BNN chỉ mức thu nhập hàng năm của dân cư trong vùng. Ta chưa biết phân phối xác suất của X nhưng biết E(X)=18 triệu, D(X)= (3,2 triệu đồng)^2.

Theo bđt Chebyshev thì : $P\left(\left|X - 18\right| < \varepsilon\right) \geq 1 - \dfrac{3,2^2}{\varepsilon^2}$

Cho vế phải =0,95 $\Rightarrow$ ε =14,3108.

Khoảng cần tìm (18 -14,3108; 18+14,3108)

II.5.2 Định lý Chebyshev:

- *Nếu các BNN $X_1, X_2, \ldots, X_n$ độc lập từng đôi, có các kỳ vọng toán hữu hạn và các phương sai đều bị chặn trên bởi hằng số C thì với mọi số dương ε tùy ý ta luôn có:*

$$\lim_{n\to+\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\frac{E(X_1)+E(X_2)+\ldots+E(X_n)}{n}\right|<\varepsilon\right)=1$$

- Trường hợp riêng, nếu *các BNN $X_1, X_2, \ldots, X_n$ độc lập từng đôi,* có cùng kỳ vọng $E(X_i) = m$, $i=1,2,\ldots,n$; và các phương sai cùng bị chặn trên thì *với mọi số dương ε tùy ý ta luôn có:*

$$\lim_{n\to+\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-m\right|<\varepsilon\right)=1$$

Định lý này còn gọi là luật số lớn của Chebyshev.

- Định lý Chebyshev chứng minh sự hội tụ theo xác suất của trung bình số học của 1 số lớn BNN về trung bình số học của các kỳ vọng toán tương ứng, mặc dù từng BNN độc lập có thể nhận giá trị khác nhiều so với kỳ vọng toán của chúng.
- Định lý Chebyshev có nhiều ứng dụng rộng rãi trong các lĩnh vực khác nhau. Ví dụ:

+ Trong việc đo lường 1 đại lượng vật lý, người ta thường tiến hành đo nhiều lần và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Định lý Chebyshev chỉ ra rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng cần đo.

+ Định lý Chebyshev cho phép dự đoán giá trị của trung bình số học các BNN. Nó là cơ sở của phương pháp mẫu trong thống kê: dựa vào mẫu ngẫu nhiên khá nhỏ có thể kết luận về toàn bộ tập hợp tổng quát của các đối tượng được nghiên cứu.

II.5.3 Định lý Bernoulli:

Nếu $\boldsymbol{f_n} = \frac{m}{n}$ *là tần suất xuất hiện biến cố A trong n phép thử độc lập và* ***p*** *là xác suất xuất hiện biến cố đó trong mỗi phép thử thì với mọi số dương* ε *tùy ý ta luôn có:*

$$\boxed{\lim_{n\to+\infty} P\left(\left|f_n - p\right| < \varepsilon\right) = 1}$$

- Định lý này còn gọi là luật số lớn của Bernoulli, được xem là cơ sở toán học của định nghĩa xác suất theo thống kê.
- Ở đây sự hội tụ theo xác suất của tần suất **f** = m/n $\rightarrow$ **p** khác với sự hội tụ theo nghĩa giải tích cổ điển. Theo nghĩa giải tích, với ε >0 cho trước, luôn tồn tại số tự nhiên N để với mọi n> N thì |m/n – p|< ε. Sự hội tụ hiểu theo nghĩa xác suất ở chỗ dù n lớn bao nhiêu đi nữa thì vẫn có thể xảy ra trường hợp cá biệt mà biểu thức |m/n – p|< ε không được thỏa mãn.

II.5.4 Định lý Giới hạn trung tâm:

(Trường hợp riêng)

Giả sử X_1, X_2,…,X_n là các BNN độc lập cùng tuân theo một quy luật phân phối xác suất nào đó. Kí hiệu $E(X_i)= a$ và $D(X_i)= \sigma^2$, $\forall i$. Khi $n \to \infty$, chúng ta có sự hội tụ theo xs của các BNN sau:

a) Bnn $X = X_1+ X_2+ \ldots + X_n$ hội tụ về phân phối chuẩn $N(n.a,n.\sigma^2)$

b) Bnn $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ hội tụ về phânphối chuẩn $N(a,\dfrac{\sigma^2}{n})$,

hay Bnn $U = \dfrac{\overline{X} - a}{\frac{\sigma}{\sqrt{n}}}$ hội tụ về phân phối chuẩn tắc N(0,1).

Nói cách khác, khi $n \to \infty$, ta có: $P(U < x) \approx \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$

Trong phần thống kê, khi n>30 ta sử dụng công thức xấp xỉ này.

Ví dụ 22:

Tung 1 con xúc xắc 200 lần. Tính xác suất tổng số chấm thu được trong các lần tung nhận giá trị từ 300 đến 650.

Gọi X_i là số chấm xuất hiện trên con xúc xắc ở lần tung thứ i, i=1,2,...,200.

Các X_i độc lập, ta tính được $E(X_i) = 3,5$ và $D(X_i) \approx 2,9167$.

Đặt $X = X_1 + X_2 + \dots + X_{200}$.

$E(X) = 200 \times 3,5 = 700$; $D(X) = 200 \times 2,9167 = 583,3333$.

Theo định lý giới hạn trung tâm: $X \sim N(700; 583,3333)$

Do đó xác suất cần tìm:

$$P(300 \le X \le 650) \approx \Phi\left(\frac{650-700}{\sqrt{583,3333}}\right) - \Phi\left(\frac{300-700}{\sqrt{583,3333}}\right) = 1,92\%$$

Ví dụ 23:

Chọn ngẫu nhiên 500 số trong đoạn [1; 2]. Tính xác suất giá trị trung bình của các số đó nằm trong khoảng (1,45; 1,55).

Gọi Bnn X_i là giá trị số thứ i được chọn, i=1,2,…,500.

Ta xem như X_i có phân phối đều liên tục trên [1; 2].

Các X_i độc lập và ta tính được $E(X_i) = 1,5$; $D(X_i) = 1/12 \approx 0,0833$.

Đặt: $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_{500}}{500}$

Theo định lý giới hạn trung tâm: $\overline{X} \sim N\left(1,5;\ \dfrac{0,8333}{500} = 0,000167\right)$

Do đó xác suất cần tìm:

$$P(700 \le X \le 800) \approx \Phi\left(\frac{1,55 - 1,5}{\sqrt{0,000167}}\right) - \Phi\left(\frac{1,45 - 1,5}{\sqrt{0,000167}}\right) = 99,99\%$$

II.5.5 Các công thức gần đúng: *(trích từ GT, đã sử dụng ở mục II.4.2; II.4.4)*

1- Xấp xỉ Phân phối Siêu bội với phân phối Nhị thức

Định lý 4.18 Cho $X \sim H(N, M, n)$. Nếu n cố định và $\lim\limits_{N\to\infty} \frac{M}{N} = p$ thì với $k = \overline{0,n}$, ta có

$$\lim_{N\to\infty} P(X = k) = C_n^k p^k q^{n-k}$$

Theo định lý 4.18, nếu N khá lớn so với n thì có thể coi $X \sim B\left(n, \frac{M}{N}\right)$

tức là ta có công thức gần đúng $\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k \left(\frac{M}{M}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}, k = \overline{0,n}$

2- Xấp xỉ Phân phối Nhị thức với phân phối Poisson

Định lý 4.19 Cho $X \sim B(n, p)$. Nếu $p \to 0$ và $np \to \lambda$ khi $n \to \infty$ thì với $k = \overline{0,n}$ ta có

$$\lim_{n\to\infty} P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Theo định lý 4.19, nếu p khá bé và n khá lớn thì có thể coi $X \sim P(np)$, tức là ta có công thức gần đúng: $C_n^k p^k q^{n-k} \approx e^{-np} \frac{(np)^k}{k!}, k = \overline{0,n}$

3- Phân phối Nhị thức và phân phối Chuẩn

Định lý 4.20 (Định lý Moivre - Laplace địa phương). Cho $X \sim B(n, p)$. Nếu $n, k \in N$ sao cho $x = \dfrac{k-np}{\sqrt{npq}}$ bị chặn khi $n \to \infty$ thì $P(X=k) = \dfrac{1}{\sqrt{npq}}.f\left(\dfrac{k-np}{\sqrt{npq}}\right).\lambda_n$

trong đó $\lambda_n \to 1$ khi $n \to \infty$ và f là hàm mật độ Gauss.

Theo định lý 4.20, khi n khá lớn ta có công thức gần đúng

$$C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} f\left(\frac{k-np}{\sqrt{npq}}\right), k = \overline{0, n}$$

Định lý 4.21 (Định lý Moivre - Laplace tích phân). Với giả thiết như trong định lý 4.20, ta có

$$P(k_1 \le X \le k_2) = \lambda_n \left[\Phi\left(\frac{k_2-np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1-np}{\sqrt{npq}}\right)\right]$$

trong đó: $\lambda_n \to 1$ khi $n \to \infty$ và Φ là tích phân Laplace.

Theo định lý 4.21, khi n khá lớn ta có công thức gần đúng:

$$P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} \approx \Phi\left(\frac{k_2-np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1-np}{\sqrt{npq}}\right)$$

Hai công thức gần đúng sau cùng này thường chỉ sử dụng khi p không quá gần 0 hoặc 1, vì trong trường hợp đó sai số là lớn.

II.6 HÀM CỦA BIẾN NGẪU NHIÊN

Dạng bài: Cho biết quy luật phân phối xác suất của biến ngẫu nhiên X, hãy tìm quy luật phân phối xác suất của biến ngẫu nhiên Y=f(X).

Ví dụ 24

Cho X có bảng phân phối xác suất:

X	1	2	...	n	...
P	$\frac{1}{2}$	$\frac{1}{2^2}$	...	$\frac{1}{2^n}$	...

Tìm bảng phân phối xác suất của BNN $Y = \cos\left(\frac{\pi}{2}X\right)$

Hướng dẫn:

X	1	2	3	4	5	.. n …		
P	$\frac{1}{2}$	$\frac{1}{2^2}$	$\frac{1}{2^3}$	$\frac{1}{2^4}$	$\frac{1}{2^5}$	…	$\frac{1}{2^n}$	…
Y	0	-1	0	1	0	…	.. …….	

$$P(Y=0) = \frac{1}{2} \times \frac{1}{1-\left(\frac{1}{2}\right)^2} = \frac{2}{3}; \; P(Y=-1) = \frac{1}{2^2} \times \frac{1}{1-\left(\frac{1}{2}\right)^4} = \frac{4}{15}$$

$$P(Y=1) = \frac{1}{2^4} \times \frac{1}{1-\left(\frac{1}{2}\right)^4} = \frac{1}{15}. \Rightarrow$$

Y	-1	0	1
p_i	4/15	2/3	1/15

Ví dụ 25

Cho X là biến ngẫu nhiên có phân phối đều trên đoạn [1, 4]. Tìm hàm phân phối xác suất của biến ngẫu nhiên Y = lnX + 1 .

Hướng dẫn:

Hàm mật độ xác suất của X: $f_X(x) = \begin{cases} \dfrac{1}{3} & x \in [1;4] \\ 0 & x \notin [1;4] \end{cases}$

Hàm phân phối XS của X : $F_X(x) = \begin{cases} 0 & x < 1 \\ \dfrac{1}{3}x - \dfrac{1}{3} & 1 \le x \le 4 \\ 1 & x > 4 \end{cases}$

Kí hiệu $F_Y(y)$ là hàm phân phối xác suất của Y. Ta thấy:

$F_Y(y) = P(Y < y) = P(\ln X + 1 < y) = P(X < e^{y-1}) = F_X(e^{y-1})$ với $y \in R$

$$= \begin{cases} 0 & e^{y-1} < 1 \\ \dfrac{1}{3}e^{y-1} - \dfrac{1}{3} & 1 \le e^{y-1} \le 4 \\ 1 & e^{y-1} > 4 \end{cases}$$

$$= \begin{cases} 0 & y < 1 \\ \dfrac{1}{3}e^{y-1} - \dfrac{1}{3} & 1 \le y \le \ln 4 + 1 \\ 1 & y > \ln 4 + 1 \end{cases}$$

Từ đây có thể tìm thêm được hàm mật độ XS của Y:

$$f_Y(y) = \begin{cases} \dfrac{1}{3}e^{y-1} & 1 \le y \le \ln 4 + 1 \\ 0 & y \notin [1;\ \ln 4 + 1] \end{cases}$$