# 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.

II.3.6 Sử dụng máy tính bỏ túi để tính 1 số tham số đặc trưng.

II.3.7 Hàm của biến ngẫu nhiên.

# 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 , $X_1$ , $X_2$ ..., $X_n$, ...

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

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 rời rạc 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.*

*- Có một người mỗi ngày mua 1 tờ vé số cho đến khi trúng được giải đặc biệt thì thôi. Gọi X là số vé người đó đã mua cho đến khi trúng giải đặc biệt, thì X là BNN rời rạc.*

* Biến ngẫu nhiên gọi là liên tục nếu các giá trị có thể có của nó lấp đầy ít nhất một 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∈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∈I} , do đó tập tất cả các biến cố {"X= $x_i$" ,i∈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 và hàm XS* (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.

$* \; p_i = P(\text{"}X = x_i\text{"}), \forall i$ $\qquad * \; 0 \le p_i \le 1$ $\qquad * \sum_i p_i = 1$

* **Hàm xác suất của X:**

(*probability mass function*)

còn gọi là hàm khối xác suất

$$f_X(x) = \begin{cases} p_1 & khi\; x = x_1 \\ p_2 & khi\; x = x_2 \\ .. & \\ p_n & khi\; x = x_n \\ (\ldots) & \\ 0 & khi\; x \notin \{x_1;\; x_2; \ldots; x_n; (\ldots)\} \end{cases}$$

## 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 *(probability density function)*.

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

*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 trên R thì xác suất của biến cố “ X < $x$ ” cũng thay đổi theo.

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

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

Trong trường hợp không sợ nhầm lẫn, người ta có thể chỉ ký hiệu hàm phân phối xác suất là F(x) .

Về mặt ý nghĩa, giá trị $\mathbf{F_X}(x_0)$ phản ánh mức độ tập trung xác suất của BNN X ở về phía bên trái của số thực $x_0$ .

## Các tính chất của hàm phân phối xác suất :

* $0 \le F(x) \le 1, \quad \forall x \in \mathbb{R}$ F(-∞) = 0 F(+∞) = 1
* Nếu $x_1 < x_2$ thì $F(x_1) \le F(x_2) \Rightarrow$ F(x) là hàm tăng trên $\mathbb{R}$.
* $P(a < X \le b) = F(b) - F(a)$

* Nếu X là BNN rời rạc thì $F(x) = \sum_{x_i \le x} p_i$

* 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*

Hàm khối xác suất của biến ngẫu nhiên X có dạng:

$$f(x)=\begin{cases}\dfrac{x+c}{10} & khi\; x\in\{1;2;3\}\\ 0 & khi\;\; x\notin\{1;2;3\}\end{cases}$$

Tìm giá trị phù hợp của c .

Đáp số: c = 4/3

***Ví dụ 2***

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 và viết hàm khối xác suất của X.

b) Gọi $F_X(x)$ là hàm phân phối XS của X. Tính các giá trị F(-1); F(2); F(2.3); và tìm biểu thức tổng quát $F_X(x)$.

c) Vẽ đồ thị hàm phân phối XS của X.

*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.

$$= \frac{C_7^3}{C_{10}^3} = \frac{7}{24} \quad \text{hoặc} \quad = \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8} = \frac{7}{24}$$

P(X=1) ≡ XS có 1 bi xanh trong 3 bi được lấy ra

$$= \frac{C_3^1 C_7^2}{C_{10}^3} \overset{hay}{=} \frac{3}{10} \times \frac{7}{9} \times \frac{6}{8} \times 3 = \frac{21}{40}$$

P(X=2) $= \frac{C_3^2 C_7^1}{C_{10}^3} \overset{hay}{=} \frac{3}{10} \times \frac{2}{9} \times \frac{7}{8} \times 3 = \frac{7}{40}$ P(X=3)$= \frac{C_3^3}{C_{10}^3} \overset{hay}{=} \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} = \frac{1}{120}$

*Bảng phân phối xác suất của X:*

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

Hàm khối xác suất:

| X | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f(x)$ | $\frac{7}{24}$ | $\frac{21}{40}$ | $\frac{7}{40}$ | $\frac{1}{120}$ |

hoặc:

$$f(x)=\begin{cases} 7/24 & khi\ x=0 \\ 21/40 & khi\ x=1 \\ 7/40 & khi\ x=2 \\ 1/120 & khi\ x=3 \\ 0 & x\notin\{0;1;2;3\} \end{cases}$$

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

$F_X(-1) = P(X \le -1) = 0;$

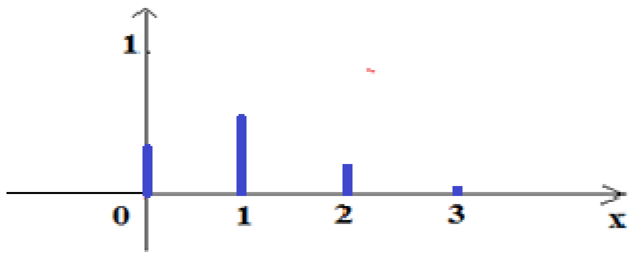
$F_X(1) = P(X \le 1) = 7/24 + 21/40.$

$F_X(2.3) = P(X \le 2.3) = 119/120$

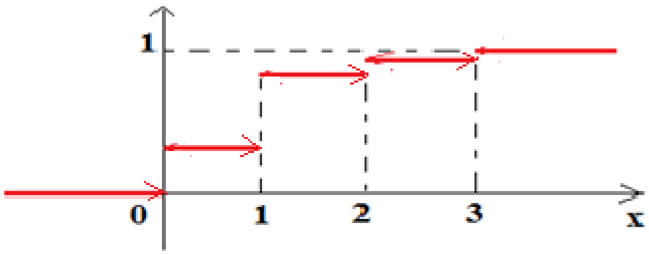
*Tổng quát hơn:*

$$F_X(x) = P(X \le x) = \sum_{x_i \le x} p_i = \begin{cases} 0 & khi \quad x<0 \\ \frac{7}{24} & khi \quad 0 \le x < 1 \\ \frac{7}{24}+\frac{21}{40}=\frac{49}{60} & khi \quad 1 \le x < 2 \\ \frac{7}{24}+\frac{21}{40}+\frac{7}{40}=\frac{119}{120} & khi \quad 2 \le x < 3 \\ \frac{7}{24}+\frac{21}{40}+\frac{7}{40}+\frac{1}{120}=1 & khi \quad x \ge 3 \end{cases}$$

c) $F_X(x)$ là hàm phân phối xác suất của X .




Đồ thị hàm khối XS f(x)




Đồ thị hàm phân phối XS F(x)

**Bài tập 1** Tung 2 con xúc xắc. Hãy lập bảng phân phối xác suất cho biến ngẫu nhiên X chỉ số chấm lớn nhất ở mặt trên của các con xúc xắc.

**Bài tập 2** Một nhóm có 3 sinh viên A; B ; C. Giả sử trong 1 buổi học, xác suất các sinh viên đi trễ lần lượt là 0,1; 0,1; 0,2. Hãy lập bảng phân phối xác suất và hàm phân phối xác suất cho biến ngẫu nhiên Y chỉ số sinh viên của nhóm đi trễ trong một buổi học. Nếu một ngày có sinh viên trong nhóm đi trễ thì xác suất ngày đó bạn C trễ là bao nhiêu?

**Bài tập 3** Một cô gái muốn mua 5 dây đèn trang trí và kiểm tra từng cái một, nếu gặp dây đèn có lỗi sẽ dừng lại và chỉ mua nếu cả 5 dây đèn đều đạt. Giả sử xác suất mỗi dây đèn đạt yêu cầu đều là 0,9. Hãy lập hàm xác suất cho số dây đèn đã được kiểm tra.

**Bài tập 4** Một lớp có 50 sinh viên, trong đó có 20 sinh viên yêu thích bóng đá. Chọn ngẫu nhiên 4 sinh viên từ danh sách lớp. Lập bảng phân phối số người yêu thích môn bóng đá trong các sinh viên được chọn ra.

**Bài tập 5** Tỉ lệ thanh niên yêu thích môn bóng đá ở một vùng là 40%. Chọn ngẫu nhiên 4 thanh niên trong vùng. Lập bảng phân phối số thanh niên yêu thích môn bóng đá trong những người được chọn ra.

**Bài tập 6** Hai cầu thủ luân phiên ném bóng vào rổ cho đến khi có người ném lọt mới thôi. Xác suất ném trúng rổ trong mỗi lần ném của người đầu là *0,6* và người sau là *0,5*.

a) Hãy lập bảng phân phối xác suất cho biến ngẫu nhiên X là số lần ném bóng của người đầu.

b) Tìm xác suất người đầu thắng cuộc (ném trúng rổ) trong trường hợp không có ai ném quá 2 lần.

***Ví dụ 3*** 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) Tìm xác suất người đó tung đúng 7 lần nếu biết người đó đã không dừng lại trước lần tung thứ 4.

*Hướng dẫn:*

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

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

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

$$\bullet P\left(Y=3\right)=P(\overline{B_1}.\overline{B_2}.B_3)\ =\left(\frac{11}{12}\right)^2\times\frac{1}{12}\qquad .....$$

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

c) Xác suất cần tìm là:

$$P(Y=7/Y\geq 4)=\frac{P(Y=7)}{P(Y\geq 4)}=\frac{\left(\frac{11}{12}\right)^6\times\frac{1}{12}}{1-\sum_{k=1}^{3}\left(\frac{11}{12}\right)^{k-1}\times\frac{1}{12}}\approx 0,0642$$

***Ví dụ 4*** Cho biến ngẫu nhiên X có hàm mật độ xác suất :

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

a) Tìm hệ số k .

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

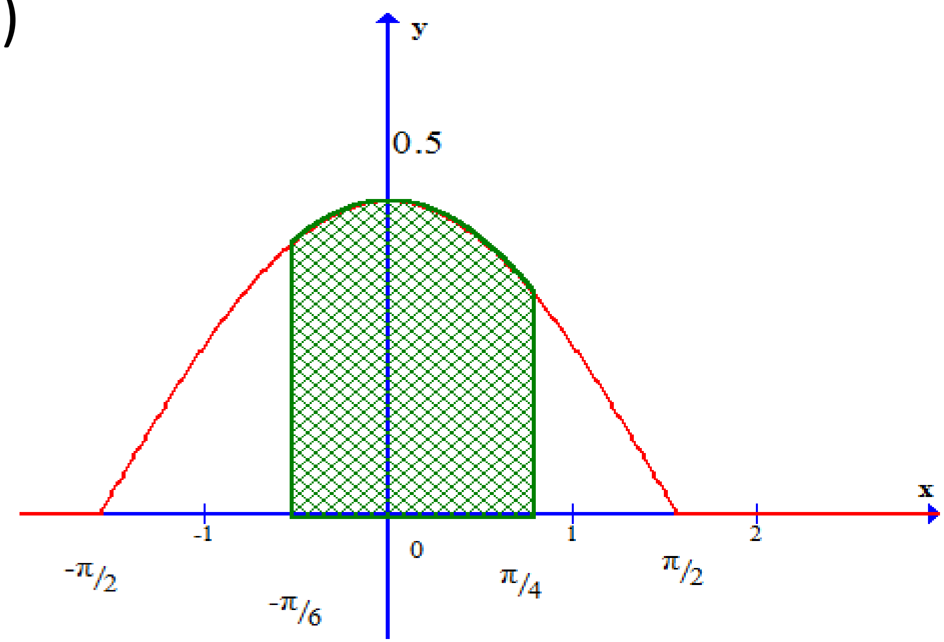
c) Tìm xác suất trong 5 lần thực hiện phép thử ngẫu nhiên độc lập thì có 3 lần X nhận giá trị trong khoảng $\left(-\frac{\pi}{6}, \frac{\pi}{4}\right)$

d) Tìm hàm phân phối $F_X$ của biến ngẫu nhiên X.

***Hướng dẫn:*** a) * Điều kiện f(x) ≥ 0, ∀x ⇒ k ≥ 0.

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

b)




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

$$\approx 0,6036$$

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

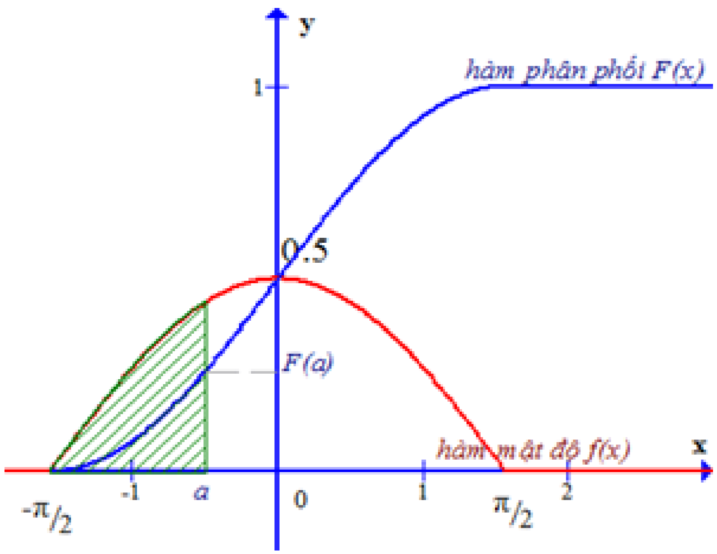
$$= \frac{1}{2}$$

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

$$C_5^3(0,6036)^3(1-0,6036)^2 \approx 0,3456$$

d)

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




*Có thể kiểm tra lại kết quả câu b) theo công thức dùng hàm phân phối :*

$$P\left(-\frac{\pi}{6} < X < \frac{\pi}{4}\right) = F_X\left(\frac{\pi}{4}\right) - F_X\left(-\frac{\pi}{6}\right)$$

$$P\left(0 < X < \frac{3\pi}{4}\right) = F_X\left(\frac{3\pi}{4}\right) - F_X(0)$$

***Ví dụ 5*** Cho biến ngẫu nhiên X có hàm mật độ xác suất:

$$f(x)=\begin{cases}0 & x\notin(-2;9)\\ k & x\in(-2;9)\end{cases}$$

a) Tìm hệ số *k* và tính $P(X^2 < 9)$.

b) Tìm xác suất X nhận giá trị dương nếu biết $X^2 < 9$.

c) Tìm xác suất trong 12 phép thử độc lập ít nhất 10 lần X nhận giá trị dương.

d) Ký hiệu $F_X$ hàm phân phối xác suất của X. Tìm giá trị biểu thức $A = F_X(0) + 3*F_X(1)$ .

e) Xét biến ngẫu nhiên $Y = X^3$.

Tính giá trị hàm phân phối $F_Y$ tại Y= 4: $F_Y(4)$

*Hướng dẫn:*

a) $k = \dfrac{1}{11}$ $\quad P(X^2 < 9) = \dfrac{5}{11}$

b) $P(X>0 \,/\, X^2 < 9) = \dfrac{3}{5}$

c) $\sum_{k=10}^{12} C_{12}^{k} \left(\dfrac{9}{11}\right)^{k} \left(\dfrac{2}{11}\right)^{12-k} \approx 0{,}6233$

d) $A = P(X \le 0) + 3 \times P(X \le 1) = \dfrac{2}{11} + 3 \times \dfrac{3}{11} = 1$

e) $F_Y(4) \overset{dn}{=} P(Y \le 4) = P(X^3 \le 4)$

$$= P(X \le \sqrt[3]{4}) = \frac{\sqrt[3]{4} + 2}{11}$$

Bài toán tìm biểu thức tổng quát $F_Y(y)$ hay $f_Y(y)$ gọi là bài toán tìm phân phối xác suất hàm (Y) của biến ngẫu nhiên X.

***Ví dụ 6***

Một chi tiết máy được tạo thành từ 3 linh kiện hoạt động độc lập. Tuổi thọ (*đơn vị: giờ*) của mỗi linh kiện là biến ngẫu nhiên liên tục có hàm mật độ xác suất:

$$f(x) = \begin{cases} k.e^{-\frac{x}{5000}} & khi\ \ x \ge 0 \\ 0 & khi\ \ x < 0 \end{cases}$$

Chi tiết bị hỏng khi có ít nhất 2 linh kiện bị hỏng. Tìm xác suất chi tiết bị hỏng trong 1000 giờ hoạt động đầu tiên.

*Hướng dẫn:*

* $k = 1/5000$

* Xác suất linh kiện thứ i bị hỏng trong 1000 giờ đầu tiên:

$$p_i = P(X_i < 1000)\ ,\ i=1; 2 ; 3$$

* Bài toán đưa về dạng Bernoulli. Đáp số: 0,0867

***Ví dụ 7***

Trong nhiên liệu rắn của tên lửa, kích thước hạt là một yếu tố quan trọng. Các nghiên cứu từ dữ liệu sản xuất trước đây đã xác định kích thước hạt (đơn vị $\mu$m) có phân bố xác suất đặc trưng bởi hàm mật độ:

$$f(x) = \begin{cases} 3x^{-4} & x > 1 \\ 0 & x \le 1 \end{cases}$$

a) Tìm xác suất một hạt ngẫu nhiên được lấy từ trong nguyên liệu sản xuất có kích thước lớn hơn 4 $\mu$m?

b) Tìm hàm phân phối xác suất $F_X(x)$.

c) Tìm kích thước hạt trung bình và phương sai của kích thước hạt *( xem định nghĩa ở các slide 23-26).*

*Hướng dẫn:* $a)\ P(X > 4) = 1/64$ $\quad$ $b)\ F_X(x) = \begin{cases} 1 - x^{-3} & x > 1 \\ 0 & x \le 1 \end{cases}$

$c)\ E(X) = 3/2;\ D(X) = 3/4$

## 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).

*Công thức tính :*

- Đối với *BNN rời rạc* : $$E(X) = \sum_i x_i p_i$$

- Đối với *BNN liên tục* : $$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$$

*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 $\in$ **R** (1)
* E( XY) = E(X) . E(Y) nếu các BNN X, Y là *độc lập,* (2)
* Giả sử Y là hàm của biến ngẫu nhiên X và viết 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).*

(1) (2) *Các BNN X+Y; X .Y và khái niệm độc lập sẽ được trình bày ở chương 3 về Véc tơ ngẫu nhiên .*

**Ví dụ 8:** *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* $(X1-6.2)^2$ *và* $(X2-6.2)^2$ .

| $(X1-6.2)^2$ | $(3-6,2)^2$ | $(4-6,2)^2$ | $(6-6,2)^2$ | $(7-6,2)^2$ | $(9-6,2)^2$ | $(10-6,2)^2$ |
|---|---|---|---|---|---|---|
| **P** | **0.1** | **0.2** | **0.3** | **0.2** | **0.1** | **0.1** |

| $(X2-6,2)^2$ | $(2-6,2)^2$ | $(3-6,2)^2$ | $(5-6,2)^2$ | $(8-6,2)^2$ | $(9-6,2)^2$ | $(10-6,2)^2$ |
|---|---|---|---|---|---|---|
| **P** | **0.2** | **0.2** | **0.1** | **0.1** | **0.1** | **0.3** |

Hãy tính $E([X1-6.2)^2)$ và $E([X2-6.2)^2)$ ? ( 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 của chúng 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:* $V(X) = E[X-E(X)]^2$ hay $V(X) = E(X^2) - [E(X)]^2$

- Nếu X là BNN rời rạc thì:

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

- Nếu X là BNN liên tục thì:

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

* 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 - sd)* của biến ngẫu nhiên X, kí hiệu $\sigma_X$ , là căn bậc hai của phương sai :

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

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 :* 

∗ $V(X) \geq 0$ ; $V(C) = 0$

∗ $V(CX) = C^2.V(X)$

∗ $V(C+X) = V(X)$

∗ $V(X+Y) = V(X) + V(Y)$ , với X,Y là độc lập. **(1)**

∗ $V(X-Y) = V(X) + V(Y)$ , với X,Y là độc lập. **(2)**

**(1) (2)** *Các BNN X+Y; X - Y và khái niệm độc lập sẽ được trình bày ở chương 3 về Véc tơ ngẫu nhiên .*

**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 nhỏ nhất trong các giá trị $x_0$ mà $F(x_0) \geq 0.5$.

Trong **Ví dụ 8**: $\sigma(X1) = 2.0881$ Mod(X1) = 6 Med(X1) = 6

$\sigma(X2) = 3.3407$ Mod(X2)= 10 Med(X2)= 5

* Định nghĩa trung vị của biến ngẫu nhiên X xác định qua tập số liệu: xem ở chương 5- Lý thuyết mẫu.

## *Quay lại các VD2 $\rightarrow$ VD4:*

Ví dụ 2: E(X)=0.9 $E(X^2)=1.3$ V(X) =0.49 Mod(X)=Med(X)=1.

E(2X+1) = 2E(X)+1 = 2.8 $E(3X^2+5)= 3.E(X^2)+5 = 8.9$

Ví dụ 3: Số lần tung trung bình là E(Y)

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

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

Ví dụ 4:

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

$$V(X)=\int_{-\infty}^{+\infty} x^2 f(x)dx-E(X)^2=\int_{-\pi/2}^{\pi/2} x^2\cos(x)dx\approx 0.9348$$

## ***Ví dụ 9:***

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:

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

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

*Lưu ý: Kết quả của bài toán có ý nghĩa khi số nhóm là lớn.*

## ***Ví dụ 10:***

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 11:*

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 loại côn trùng đó.

d) 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).

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

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

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

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

*d) Gợi ý:* Tìm giá trị m $\in$R mà P(X $\leq$ m) =0,5 hay $F_X$(m) =0.5.

( m chính là trung vị của X )

# II.3.6 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 580 VN.... |
|---|---|---|
| Xóa nhớ bài cũ | ....... | ...... |
| Vào chế độ thống kê một biến. | MODE -- 3 (STAT) --1 (1-VAR) | MENU – 6 (Statistics) – 1 (1-variable) |
| Mở cột tần số (nếu máy chưa mở) | SHIFT -- MODE (SETUP) –<br>-- ▼ -- 4 (STAT) -- 1 (ON) | SHIFT -- MENU –<br>-- ▼ -- 3 (Statistics)-- 1 (ON) |
| Nhập dữ liệu | X FREQ<br>1 x1 p1<br>2 x2 p2<br>... ... ...<br>... xn pn<br>AC | |
| Đọc kết quả E(X) | SHIFT – 1 (STAT)- 4 (VAR) –<br>--- 2 ( $\overline{x}$ ) -- = | OPTN --- ▼ -- 2 – 1 ( $\overline{x}$ )<br>hoặc OPTN - 2 |
| Đọc kết quả $\sqrt{V(X)}$ | SHIFT – 1 (STAT)- 4 (VAR) –<br>--- 3 ( σX ) -- = | OPTN --- ▼ -- 2 – 3 ( σ )<br>hoặc OPTN - 2 |
| Tham khảo các tổng trung gian | SHIFT – 1 (STAT)- 3 (SUM) - .... | OPTN --- ▼ -- 1 – 2....<br>hoặc OPTN - 2 |

## II.3.7 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 12:***

Biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

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

Tìm bảng phân phối xác suất của biến ngẫu nhiên $Y = \cos\left(\frac{\pi}{2}.X\right)$ và tính E(Y); V(Y).

*Hướng dẫn:* $Y = \cos\left(\frac{\pi}{2}X\right)$

| X | 1 | 2 | 3 | 4 | 5 | 6 | …. n … |
|---|---|---|---|---|---|---|---|
| P | $\frac{1}{2}$ | $\frac{1}{2^2}$ | $\frac{1}{2^3}$ | $\frac{1}{2^4}$ | $\frac{1}{2^5}$ | $\frac{1}{2^6}$ | … $\frac{1}{2^n}$ … |
| Y | 0 | -1 | 0 | 1 | 0 | - 1 | . …… |

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

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

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

$E(Y) = -0.2; \quad V(Y) = 0.293\overline{3}$

## *Ví dụ 13*

Cho X là biến ngẫu nhiên có phân phối đều [(*)] trên đoạn [1; 4].

a) Tìm hàm phân phối xác suất của biến ngẫu nhiên Y = lnX +1.

b) Tìm E(Y).

*Hướng dẫn:*

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

( *: theo định nghĩa ở chương 4)

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

Đặt Y = lnX + 1.

Kí hiệu $F_Y(y)$ là hàm phân phối xác suất của Y ; y $\in$ R.

Biến đổi theo định nghĩa:

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

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

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

b) Tìm E(Y)

Cách 1: Tìm qua hàm mật độ xác suất của Y.

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

Thay công thức:

$$E(Y) = \int_{-\infty}^{+\infty} y.f(y)dy = \int_{1}^{\ln 4+1} y.\frac{e^{y-1}}{3}dy \approx 1.8484$$

Cách 2: Sử dụng tính chất kỳ vọng của hàm theo biến X.

$$E(Y) = \int_{-\infty}^{+\infty} \varphi(x) f(x)dx = \int_{1}^{4} (\ln x + 1) \times \frac{1}{3}\, dx \approx 1.8484$$