Chương 2

BIẾN NGẪU NHIÊN VÀ MỘT SỐ LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

Chương này tìm hiểu về biến ngẫu nhiên với việc thiết lập luật phân phối xác suất và tính các tham số đặc trưng của nó. Các phân phối xác suất liên tục và rời rạc thông dụng cũng được khảo sát trong chương này.

1. BIẾN NGẪU NHIÊN VÀ LUẬT PHÂN PHỐI XÁC SUẤT

1.1 Khái niệm và phân loại biến ngẫu nhiên

a) Khái niệm

Trong chương 1, các vấn đề được xem xét xoay quanh biến cố ngẫu nhiên. Khi thực hiện một phép thử ta chỉ quan tâm đến khả năng xảy ra một biến cố nào đó của phép thử này. Trong chương 2 này, phép thử ngẫu nhiên được nghiên cứu một cách khái quát hơn qua khái niệm biến ngẫu nhiên. *Biến ngẫu nhiên là biến nhận các giá trị là các khả năng có thể của phép thử ngẫu nhiên* với một xác suất nhất định nào đó phụ thuộc vào kết quả của phép thử ngẫu nhiên.

Các biến ngẫu nhiên thường được ký hiệu bằng các chữ cái hoa như: X, Y, Z, … hoặc dạng chỉ số: $X_1, X_2, \ldots, X_n$; $Y_1, Y_2, \ldots, Y_n$;…

Các giá trị có thể có của biến ngẫu nhiên được ký hiệu là: $x_1, x_2, \ldots, x_m$; $y_1, y_2 \ldots, y_m, \ldots$

Một biến ngẫu nhiên coi như được xác định nếu biết được tập các giá trị của nó và các xác suất mà nó nhận giá trị thuộc tập đó.

b) Phân loại

Gọi tập giá trị của biến ngẫu nhiên X là $X(\Omega)$. Căn cứ vào tập $X(\Omega)$, biến ngẫu nhiên được chia thành 2 loại: rời rạc và liên tục.

Biến ngẫu nhiên rời rạc: Biến ngẫu nhiên được gọi là rời rạc nếu $X(\Omega)$ là hữu hạn hoặc vô hạn đếm được, cách quãng nhau.

■ **Ví dụ 2.1.** Các biến sau là biến ngẫu nhiên rời rạc:

Số chấm xuất hiện trên mặt con xúc xắc khi tung một con xúc xắc.

Số học sinh vắng mặt trong một buổi.

Số sản phẩm tốt khi mua một lô hàng.

Biến ngẫu nhiên liên tục: Biến ngẫu nhiên được gọi là liên tục nếu $X(\Omega)$ lấp đầy một khoảng hay một số khoảng hay toàn bộ trục số.

■ **Ví dụ 2.2.** Các biến sau là biến ngẫu nhiên liên tục:

Nhiệt độ không khí tại một thời điểm nào đó.

Sai số khi đo lường một đại lượng vật lý nào đó.

Khoảng thời gian giữa 2 ca cấp cứu của một bệnh viện.

1.2 Hàm mật độ xác suất

Luật phân phối xác suất của biến ngẫu nhiên là một cách biểu diễn quan hệ giữa các giá trị của biến ngẫu nhiên với xác suất tương ứng mà nó nhận các giá trị đó. Luật phân phối xác suất thường được thể hiện dưới hai hình thức: Hàm mật độ xác suất và hàm phân phối xác suất.

a) Đối với biến ngẫu nhiên rời rạc

Giả sử biến ngẫu nhiên rời rạc X có n giá trị có thể x_i, $i = 1, 2, \ldots, n$ với xác suất tương ứng $p_i = P(X = x_i) > 0$, khi đó hàm mật độ xác suất của X (ký hiệu $f(x)$) được xác định như sau:

$$f(x) = \begin{cases} p_i \text{ khi } x = x_i, i = 1,2,\ldots,n \\ 0 \text{ khi } x \neq x_i \end{cases}$$

Thông thường để thuận lợi trong đánh giá biến ngẫu nhiên rời rạc, hàm mật độ xác suất được biểu diễn dưới dạng bảng phân phối xác suất như sau:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Chú ý: *Nếu giá trị có thể của biến ngẫu nhiên X gồm hữu hạn giá trị $x_1, x_2, \ldots, x_n$ thì các biến cố $X = x_1$, $X = x_2$, …, $X = x_n$ lập thành một nhóm biến cố đầy đủ, vì vậy ta có* $\sum_{i=1}^{n} p_i = 1$.

■ **Ví dụ 2.3.** Trên một cái kệ có 6 cuốn toán và 4 cuốn lý, chọn ngẫu nhiên 3 cuốn sách. Lập bảng phân phối xác suất của số sách toán chọn được.

Giải

Gọi X là số cuốn sách toán chọn được.

Ta có X là biến ngẫu nhiên rời rạc có thể nhận các giá trị: 0, 1, 2, 3.

$$P(X=0)=\frac{C_6^0.C_4^3}{C_{10}^3}=\frac{4}{120}.$$

$$P(X=1)=\frac{C_6^1.C_4^2}{C_{10}^3}=\frac{36}{120}.$$

$$P(X=2)=\frac{C_6^2.C_4^1}{C_{10}^3}=\frac{60}{120}.$$

$$P(X=3)=\frac{C_6^3.C_4^0}{C_{10}^3}=\frac{20}{120}.$$

Vậy ta có bảng phân phối xác suất của biến ngẫu nhiên X là

X	0	1	2	3
P	$\frac{4}{120}$	$\frac{36}{120}$	$\frac{60}{120}$	$\frac{20}{120}$

b) Đối với biến ngẫu nhiên liên tục

Hàm số $y=f(x)$ xác định trên R được gọi là hàm mật độ xác suất của biến ngẫu nhiên liên tục X nếu nó thỏa 2 tính chất sau:

i) $f(x)$ là hàm không âm: $f(x)\geq 0\ \forall x$,

ii) $\int_{-\infty}^{+\infty} f(x)dx=1$.

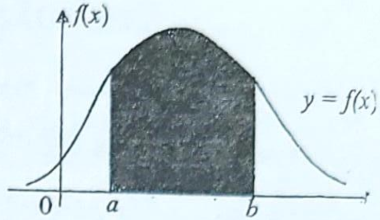
Về mặt hình học, việc tìm hàm mật độ $f(x)$ có thể xem là việc tìm hàm số $f(x)$ không âm mà diện tích hình phẳng giới hạn bởi đồ thị $y=f(x)$ và trục hoành bằng 1. Diện tích này đặc trưng cho tất cả khả năng xảy ra của phép thử. Từ ý nghĩa hình học này ta rút ra tính chất quan trọng của hàm mật độ xác suất như sau:

$$\boxed{P(a\leq X\leq b)=\int_a^b f(x)dx} \qquad (2.1)$$

Xác suất để biến ngẫu nhiên X nhận giá trị trong đoạn $[a, b]$ bằng diện tích của hình phẳng giới hạn bởi trục $0x$, đồ thị $y=f(x)$ và các đường thẳng $x=a, x=b$.

Từ ý nghĩa hình học của tính chất trên, ta dễ dàng rút ra một số kết quả sau:

i) $P(X=x_0)=0$, với x_0 là một giá trị tùy ý.



Hình 2.1. Xác suất để $(a\leq X\leq b)$

ii) $P(a\leq X\leq b)=P(a\leq X<b)=P(a<X\leq b)=P(a<X<b)$.

Ví dụ 2.4. Giả sử tuổi thọ của một loại côn trùng là biến ngẫu nhiên liên tục (đơn vị là tháng) có hàm mật độ xác suất như sau:

$$f(x)=\begin{cases}0 & \text{khi } x\notin[0;2]\\ m(x-2) & \text{khi } x\in[0;2]\end{cases}$$

a) Tìm tham số m.

b) Tính tỷ lệ côn trùng chết trước khi nó được 1 tháng tuổi.

Giải

a) Ta có

$$\int_{-\infty}^{+\infty} f(x)dx=1 \Leftrightarrow \int_{-\infty}^{0} f(x)dx+\int_0^2 f(x)dx+\int_2^{+\infty} f(x)dx=1$$

$$\Leftrightarrow \left[m(\frac{x^2}{2}-2x)\right]_0^2=1$$

$$\Leftrightarrow -2m=1$$

$$\Leftrightarrow m=-\frac{1}{2}.$$

Vậy $f(x)=\begin{cases}0 \quad \text{khi } x\notin[0,2]\\ -\frac{1}{2}(x-2) \quad \text{khi } x\in[0,2]\end{cases}$

Vì $f(x)\geq 0\ \ \forall x$, vậy $m=-\frac{1}{2}$ thỏa điều kiện bài toán.

b) Ta có $P(0\leq X<1)=\int_0^1 f(x)dx$

$$=\int_0^1 -\frac{1}{2}(x-2)dx$$

$$= \left[-\frac{1}{2}\left(\frac{x^2}{2} - 2x\right)\right]_0^1$$

$$= -\frac{1}{2}\left(-\frac{3}{2}\right) = 0,75.$$

Vậy tỷ lệ côn trùng chết trước khi nó được 1 tháng tuổi là 75%.

⋆ **Chú ý**: *Từ định nghĩa hàm mật độ ta có với* Δx *đủ bé ta có*

$$P(x \le X \le x + \Delta x) = f(x)\Delta x.$$

Do đó ta thấy xác suất để X nhận giá trị thuộc lân cận khá bé $(x, x + \Delta x)$ *gần như tỉ lệ với* $f(x)$.

1.3 Hàm phân phối xác suất

a) Định nghĩa

Hàm phân phối xác suất của biến ngẫu nhiên X (kí hiệu là $F(x)$) là hàm số được xác định như sau:

$$\boxed{F(x) = P(X < x)} \qquad (2.2)$$

Công thức trên có thể cụ thể hóa cho từng loại biến ngẫu nhiên như sau:

- Khi X là biến ngẫu nhiên rời rạc nhận các giá có thể có $x_1, x_2, \ldots, x_k$ với xác suất tương ứng $p_1, p_2, \ldots, p_k$ thì

$$\boxed{F(x) = \sum_{x_i < x} p_i} \qquad (2.3)$$

- Khi X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ thì

$$\boxed{F(x) = \int_{-\infty}^{x} f(t)dt} \qquad (2.4)$$

b) Ý nghĩa

Hàm phân phối xác suất $F(x)$ phản ánh mức độ tập trung xác suất về bên trái của điểm x.

c) Tính chất

Hàm phân phối xác suất có những tính chất sau:

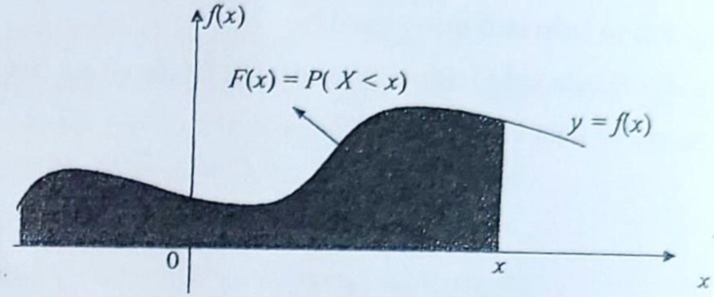
i) $0 \le F(x) \le 1 \quad \forall x.$

ii) $F(-\infty) = 0; \quad F(+\infty) = 1.$

iii) $F(x)$ là hàm số không giảm.

iv) $P(a \le X < b) = F(b) - F(a).$

v) Nếu X là biến ngẫu nhiên liên tục thì $F'(x) = f(x)$.



Hình 2.2. Hàm phân phối xác suất

■ **Ví dụ 2.5.** Một người hằng ngày từ nhà đến cơ quan phải qua 4 ngã tư. Xác suất gặp đèn đỏ ở mỗi ngã tư là 25%. Lập hàm phân phối xác suất số lần gặp đèn đỏ của người đó.

Giải

Gọi X là số lần gặp đèn đỏ của người đó.

Theo công thức Bernoulli ta có

$$p_k = P(X = k) = C_4^k (0,25)^k (0,75)^{4-k}, (k = \overline{0,4}).$$

Cụ thể ta có

$$P(X=0) = 0,3164; P(X=1) = 0,4219; P(X=2) = 0,2109,$$
$$P(X=3) = 0,0469; P(X=4) = 0,0039.$$

Khi đó ta có bảng phân phối xác suất của X:

X	0	1	2	3	4
P	0,3164	0,4219	0,2109	0,0469	0,0039

Ta có X là biến ngẫu nhiên rời rạc nên $F(x) = \sum_{x < x_i} p_i$, với $(p_i = P(X = x_i))$

Khi $x \le 0$, $F(x) = P(X < 0) = 0.$

Khi $0 < x \le 1$, $F(x) = p_0 = 0,3164.$

Khi $1 < x \le 2$, $F(x) = p_0 + p_1 = 0,3164 + 0,4219 = 0,7383.$

Khi $2 < x \le 3$, $F(x) = p_0 + p_1 + p_2 = 0,7383 + 0,2109 = 0,9492.$

Khi $3 < x \le 4$, $F(x) = p_0 + p_1 + p_2 + p_3 = 0,9492 + 0,0469 = 0,9961$.

Khi $x > 4$, $F(x) = p_0 + p_1 + p_2 + p_3 + p_4 = 1$.

Vậy ta có

$$F(x) = \begin{cases} 0 & \text{khi } x \le 0 \\ 0,3164 & \text{khi } 0 < x \le 1 \\ 0,7383 & \text{khi } 1 < x \le 2 \\ 0,9492 & \text{khi } 2 < x \le 3 \\ 0,9961 & \text{khi } 3 < x \le 4 \\ 1 & \text{khi } x > 4 \end{cases}$$

■ **Ví dụ 2.6.** Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x) = \begin{cases} 0 & \text{khi } x \le 0 \\ x & \text{khi } 0 < x \le 1 \\ 2 - x & \text{khi } 1 < x \le 2 \\ 1 & \text{khi } x > 2 \end{cases}$$

Tìm hàm phân phối xác suất của biến ngẫu nhiên X.

Giải

Ta có X là biến ngẫu nhiên liên tục nên $F(x) = \int\limits_{-\infty}^{x} f(x)dx$.

Khi $x \le 0$ thì $F(x) = 0$.

Khi $0 < x \le 1$ thì $F(x) = \int\limits_0^x xdx = \left[\frac{x^2}{2}\right]_0^x = \frac{x^2}{2}$.

Khi $1 < x \le 2$ thì

$$F(x) = \int\limits_0^1 xdx + \int\limits_1^x (2-x)dx = \left[\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^x = \frac{-x^2}{2} + 2x - 1.$$

Khi $x > 2$ thì $F(x) = 1$.

Vậy ta có hàm phân phối xác suất của biến ngẫu nhiên X là

$$F(x) = \begin{cases} 0 & \text{khi } x \le 0 \\ \dfrac{x^2}{2} & \text{khi } 0 < x \le 1 \\ \dfrac{-x^2}{2} + 2x - 1 & \text{khi } 1 < x \le 2 \\ 1 & \text{khi } x > 2 \end{cases}$$

2. THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

Khi nghiên cứu luật phân phối xác suất của biến ngẫu nhiên chúng ta có một cái nhìn tương đối khái quát về khả năng xảy ra của một phép thử. Để đánh giá một biến ngẫu nhiên, cũng như để so sánh các biến ngẫu nhiên với nhau ta cần phải có những con số cụ thể đặc trưng cho từng biến ngẫu nhiên. Các con số này được gọi là các tham số đặc trưng của biến ngẫu nhiên. Các tham số đặc trưng thông thường để đánh giá một biến ngẫu nhiên là mode, kỳ vọng, phương sai và độ lệch chuẩn.

2.1 Mode

Mode của một biến ngẫu nhiên là giá trị của biến ngẫu nhiên mà tại đó nó có nhiều khả năng xảy ra nhất.

Kí hiệu: Mode của biến ngẫu nhiên X được kí hiệu là Mod(X).

Cụ thể ta có thể định nghĩa mode cho từng loại biến ngẫu nhiên như sau:

- *Mode của biến ngẫu nhiên rời rạc*: Là giá trị của biến ngẫu nhiên mà tại đó nó có xác suất lớn nhất.
- *Mode của biến ngẫu nhiên liên tục:* Là giá trị của biến ngẫu nhiên mà tại đó hàm mật độ đạt giá trị cực đại.

■ **Ví dụ 2.7.** Trong một trò chơi, mỗi người chơi được bắn 2 viên đạn. Nếu bắn trúng mỗi viên sẽ được 20000 đồng, nếu bắn trật mỗi viên sẽ bị mất 10000 đồng. Giả sử xác suất bắn trúng mỗi viên đạn là 40%. Nếu bạn tham gia cuộc thi này thì số tiền mà bạn nhiều khả năng có nhất là bao nhiêu?

Giải

Gọi X là số tiền người tham gia có thể có.

Khi bắn 2 viên đạn có 3 khả năng xảy ra: bắn trật 2 viên, bắn trúng 1 viên trật 1 viên, bắn trúng 2 viên, do đó X có thể nhận các giá trị: -20000, 10000, 40000.

Ta có

$$P(X = -20000) = C_2^2(0.4)^0(0.6)^2 = 0,36.$$
$$P(X = 10000) = C_2^1(0,4)^1(0,6)^1 = 0,48.$$
$$P(X = 40000) = C_2^2(0,4)^2(0,6)^0 = 0,16.$$

Khi đó ta có bảng phân phối xác suất của X như sau:

X	-20000	10000	40000
P	0,36	0,48	0,16

Nhìn vào bảng phân phối xác suất này ta có $\text{Mod}(X) = 10000$. Vậy khi 1 người tham gia cuộc chơi này thì được 10000 ngàn đồng là khả năng xảy ra nhiều nhất.

★ **Chú ý:** *Một biến ngẫu nhiên có thể có nhiều giá trị Mode.*

2.2 Kỳ vọng

a) Định nghĩa

Ta định nghĩa kỳ vọng cho từng loại biến ngẫu nhiên như sau:

- Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

thì kỳ vọng của X (kí hiệu là $E(X)$) được xác định bằng công thức:

$$\boxed{E(X) = \sum_{i=1}^{n} x_i p_i} \qquad (2.5)$$

- Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất là $f(x)$ thì kỳ vọng của

X được xác định bằng công thức:

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x f(x)dx} \qquad (2.6)$$

b) Tính chất

Cho C là một hằng số, X và Y là hai biến ngẫu nhiên. Từ định nghĩa kỳ vọng ta rút ra được các tính chất sau:

i) $E(C) = C$.

ii) $E(C.X) = C.E(X)$.

iii) $E(X \pm Y) = E(X) \pm E(Y)$.

iv) $E(X.Y) = E(X).E(Y)$ nếu X, Y là 2 biến ngẫu nhiên độc lập.

c) Ý nghĩa

Tiến hành n phép thử. Giả sử X là biến ngẫu nhiên nhận các giá trị có thể có là $x_1, x_2, \ldots, x_n$ với số lần xuất hiện $k_1, k_2, \ldots, k_n$. Khi đó giá trị trung bình (kí hiệu $\overline{X}$) của biến ngẫu nhiên X trong n phép thử là

$$\overline{X} = \frac{k_1x_1 + k_2x_2 + \ldots + k_nx_n}{n} = \frac{k_1}{n} + \frac{k_2}{n} + \ldots + \frac{k_n}{n} = f_1x_1 + f_2x_2 + \ldots + f_nx_n.$$

Với $f_i = \dfrac{k_i}{n}$ là tần suất để X nhận giá trị x_i.

Theo định nghĩa xác suất bằng thống kê ta có $\lim\limits_{n\to\infty} f_i = p_i$. V

vậy với n đủ lớn ta có

$$\overline{X} = f_1x_1 + f_2x_2 + \ldots + f_nx_n = E(X).$$

Vì điều này ta có thể nói rằng:

> *Kỳ vọng của biến ngẫu nhiên là con số đặc trưng cho giá trị trung bình của biến ngẫu nhiên đó.*

Trong cơ học nếu coi $x_1, x_2, \ldots x_n$ là các chất điểm, còn khối lượng là $p_1, p_2, \ldots, p_n$ thì kỳ vọng $E(X) = \sum_{i=1}^{n} p_i x_i$ chính là trọng tâm của hệ chất điểm.

Nếu X là trọng lượng thì $E(X)$ là trọng lượng trung bình, nếu X là chiều cao thì $E(X)$ là chiều cao trung bình, …

Ví dụ 2.8. Xét lại ví dụ 2.7, hãy tìm số tiền trung bình mà một người có được qua cuộc thi trên.

Giải

Ta có X là số tiền mà 1 người tham gia có thể có qua cuộc thi trên.

Theo ý nghĩa của kỳ vọng thì số tiền trung bình mà một người tham gia có được chính là $E(X)$:

$$E(X) = \sum_{i=1}^{3} x_i p_i = -20000.0{,}36 + 10000.0{,}48 + 40000.0{,}16 = 4000 \text{ (đồng)}.$$

Ví dụ 2.9. Giả sử thời gian sống của một loài sinh vật là biến ngẫu nhiên X (đơn vị tính bằng giờ) có hàm mật độ xác suất $f(x) = \begin{cases} 0 & \text{khi } x \notin [0;2] \\ m(x-2) & \text{khi } x \in [0;2] \end{cases}$

Tính thời gian sống trung bình của loài sinh vật đó.

Giải

Trước hết ta xác định tham số m.

Ta có

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow \int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)dx + \int_{2}^{+\infty} f(x)dx = 1$$

$$\Leftrightarrow \int_0^2 m(x-2)dx = 1$$

$$\Leftrightarrow \left[m\left(\frac{x^2}{2}-2x\right)\right]_0^2 = 1$$

$$\Leftrightarrow m = -\frac{1}{2}.$$

Khi đó $f(x) = \begin{cases} 0 & \text{khi } x \notin [0;\,2] \\ -\frac{1}{2}(x-2) & \text{khi } x \in [0;\,2] \end{cases}$

Ta có $f(x)$ là hàm không âm. Vậy giá trị m tìm ở trên là thỏa điều kiện bài toán.

Thời gian sống trung bình của loài sinh vật đó chính là $E(X)$.

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = -\int_0^2 \frac{1}{2}x(x-2)dx = \left[-\frac{1}{2}\left(\frac{x^3}{3}-x^2\right)\right]_0^2 = \frac{2}{3} \text{ (giờ).}$$

2.3 Phương sai và độ lệch chuẩn

a) Phương sai

i) Định nghĩa

Phương sai của biến ngẫu nhiên X (kí hiệu $V(X)$) được xác định bằng biểu thức:

$$\boxed{V(X) = E[X - E(X)]^2} \qquad (2.7)$$

Ta có thể phát biểu bằng lời định nghĩa trên như sau: *Phương sai của biến ngẫu nhiên bằng trung bình của bình phương sự chênh lệch của những giá trị biến ngẫu nhiên so với trung bình của nó.*

ii) Công thức

Ta có

$$\begin{aligned} V(X) &= E(X-a)^2 \quad (a = E(X)) \\ &= E(X^2 - 2aX + a^2) \\ &= E(X^2) - 2aE(X) + E(a^2) \\ &= E(X^2) - a^2. \end{aligned}$$

Từ công thức này ta cụ thể cho từng biến ngẫu nhiên như sau:

- *Khi X là biến ngẫu nhiên rời rạc:*

$$\boxed{V(X) = \sum_{i=1}^{n} x_i^2 p_i - [E(X)]^2} \qquad (2.8)$$

- *Khi X là biến ngẫu nhiên liên tục:*

$$\boxed{V(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - [E(X)]^2} \qquad (2.9)$$

iii) Tính chất

Cho C là hằng số, X và Y là hai biến ngẫu nhiên, ta có

$V(C) = 0.$

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

Nếu X, Y độc lập thì $V(X \pm Y) = V(X) + V(Y)$.

b) Độ lệch chuẩn

Phương sai của một biến ngẫu nhiên là con số đặc trưng cho sự phân tán của biến ngẫu nhiên quanh kỳ vọng của nó. Tuy nhiên, nó không cùng đơn vị với biến ngẫu nhiên. Chính vì điều này, người ta đưa ra một tham số mới cũng có ý nghĩa giống như phương sai, nhưng cùng đơn vị với biến ngẫu nhiên. Đại lượng này được gọi là độ lệch chuẩn.

Kí hiệu: $\sigma(X)$.

$$\boxed{\sigma(X) = \sqrt{V(X)}} \qquad (2.10)$$

Ví dụ 2.10. Năng suất của hai máy tương ứng là các biến ngẫu nhiên X, Y (đơn vị: sản phẩm/phút) có bảng phân phối xác suất như sau:

X	0	1	2	3
P	0.4	0.1	0.15	0.35

Y	1	2	3
P	0.1	0.5	0.4

Nếu phải chọn mua một trong hai máy này ta nên chọn máy nào?

Giải

$E(X) = 0.0,4 + 1.0,1 + 2.0,15 + 3.0,35 = 1,45.$

$E(X^2) = 0^2.0,4 + 1^2.0,1 + 2^2.0,15 + 3^2.0,35 = 3,85.$

Do đó $V(X) = 3,85 - (1,45)^2 = 1,7475.$

$E(Y) = 1.0,1 + 2.0,5 + 3.0,4 = 2,3.$

$E(Y^2) = 1^2.0,1 + 2^2.0,5 + 3^2.0,4 = 5,7.$

Do đó $V(Y) = 5,7 - (2,3)^2 = 0,41.$

Ta có

$E(X) < E(Y)$: Năng suất trung bình của Y cao hơn X,

$V(X) > V(Y)$: Năng suất của Y ổn định hơn X.

Vậy chọn mua máy Y.

Ví dụ 2.11. Trọng lượng của một loại sản phẩm là X (đơn vị kg) có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & \text{khi } x \in [0,1] \\ 0 & \text{khi } x \notin [0,1] \end{cases}$$

Tìm phương sai và độ lệch chuẩn của X.

Giải

Ta có

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \frac{3}{2}\int_2^3 x(1-x^2)dx = 0,375.$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{3}{2}\int_2^3 x^2(1-x^2)dx = 0,2.$$

Do đó

$$V(X) = E(X^2) - [E(X)]^2 = 0,0594.$$

$$\sigma(X) = \sqrt{V(X)} = 0,2437.$$

3. LUẬT SỐ LỚN

3.1 Khái niệm hội tụ của dãy biến ngẫu nhiên

Cho dãy biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ **(kí hiệu** (X_n)**)** và biến ngẫu nhiên X, ta có các khái niệm hội tụ của (X_n) như sau:

a) Hội tụ hầu chắc chắn

(X_n) hội tụ hầu chắc chắn về X nếu

$$P\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

Kí hiệu: $X_n \xrightarrow{hcc} X$.

b) Hội tụ theo xác suất

(X_n) hội tụ theo xác suất về X nếu

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0, \forall \varepsilon > 0.$$

Kí hiệu: $X_n \xrightarrow{P} X$.

c) Hội tụ theo phân phối

(X_n) hội tụ theo phân phối về X (kí hiệu: $X_n \xrightarrow{F} X$) được xét trong hai trường hợp như sau:

- *Rời rạc*:

$$\lim_{n\to\infty} P(X_n = x) = P(X = x), \forall x \in G.$$

trong đó (X_n) và X đều rời rạc có cùng tập giá trị $G = \{x_1, x_2, \ldots, x_n, ..\}$.

- *Liên tục*:

$$\lim_{n\to\infty} P(X_n < x) = P(X < x), \forall x \in R$$

★ **Chú ý:**

i) *Nếu* $X_n \xrightarrow{F} X$ *thì với n đủ lớn thì* $F_{X_n}(x) = F_X(x)$.

ii) *Hội tụ hầu chắc chắn dẫn đến hội tụ theo xác suất, hội tụ theo xác suất dẫn đến hội tụ theo phân phối.*

3.2 Bất đẳng thức Chebyshev và luật số lớn

a) Bất đẳng thức Chebyshev

i) Bất đẳng thức Chebyshev thứ nhất

Nếu X là biến ngẫu nhiên nhận giá trị không âm, có kỳ vọng hữu hạn, thì $\forall \varepsilon > 0$ ta có

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Như vậy với ε cho trước, bất đẳng thức này cho ta biết cận trên của xác suất để biến ngẫu nhiên X nhận giá trị không nhỏ hơn ε.

ii) Bất đẳng thức Chebyshev thứ hai

Giả sử biến ngẫu nhiên X có kỳ vọng và phương sai hữu hạn, với $\varepsilon > 0$ tùy ý, ta có

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

hay

$$P(|X - E(X)| < \varepsilon) \leq 1 - \frac{V(X)}{\varepsilon^2}.$$

b) Luật số lớn

i) Định nghĩa

Dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ được gọi là tuân theo luật số lớn (LSL) nếu hầu như chắc chắn trung bình cộng của một số lượng đủ lớn các biến ngẫu nhiên xấp xỉ với trung bình cộng của các kỳ vọng những biến ngẫu nhiên đó, nghĩa là

$$\forall \varepsilon > 0, \lim_{n\to+\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right] = 1.$$

Hệ thức trên tương đương với hệ thức sau:

$$\forall \varepsilon > 0, \lim_{n\to+\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \geq \varepsilon\right] = 0.$$

ii) Điều kiện để dãy các biến ngẫu nhiên tuân theo luật số lớn

Các định lý sau chỉ ra những dấu hiệu đủ để dãy các biến ngẫu nhiên thỏa mãn luật số lớn.

¤ **Định lý 1 (Định lý Chebyshev):** *Các biến ngẫu nhiên độc lập từng đôi X_1, X_2, ... có kỳ vọng hữu hạn và có phương sai bị chặn đều, nghĩa là tồn tại hằng số C sao cho $V(X_i) \leq C, i = 1, 2, \ldots$ thì tuân theo LSL.*

Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị sai khác nhiều so với kỳ vọng của chúng, nhưng trung bình số học của một số lớn biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học của các kỳ vọng của chúng. Điều này cho phép chúng ta dự đoán trung bình số học của các biến ngẫu nhiên.

¤ **Định lý 2:** *Xét một lược đồ Bernoulli gồm n phép thử độc lập và biến cố A có xác suất xảy ra trong mỗi phép thử không đổi P(A) = p. Gọi f_n là tần số xuất hiện biến cố A trong n phép thử thì $\forall \varepsilon > 0$ ta có*

$$\lim_{n\to\infty} P(|f_n - p| < \varepsilon) = 1.$$

Như vậy, tần số xuất hiện biến cố trong *n* phép thử độc lập dần về xác suất xuất hiện biến cố trong mỗi phép thử khi số phép thử tăng lên vô hạn.

¤ **Định lý 3**: *Dãy các biến ngẫu nhiên $X_1, X_2, \ldots$ có phương sai thỏa:*

$$\lim_{i\to+\infty} V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \lim_{i\to+\infty} \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = 0.$$

thì tuân theo luật số lớn.

3.3 Định lý giới hạn trung tâm

¤ **Định lý 4 (Định lý giới hạn trung tâm Moivre Laplace):**

Xét một lược đồ Bernoulli và các biến ngẫu nhiên X_i, i = 1, 2, ..., n nói trong định lý 2. Đặt $S_n = \sum_{i=1}^{n} X_i$ thì S_n là biến ngẫu nhiên nhận giá trị bằng tần số của biến cố A trong n phép thử. Khi n rất lớn thì dãy các đại lượng ngẫu nhiên:

$$\frac{S_n - np}{\sqrt{npq}} = \frac{S_1 - p}{\sqrt{npq}} + \frac{S_2 - p}{\sqrt{npq}} + \ldots + \frac{S_n - p}{\sqrt{npq}}$$

hội tụ theo phân phối về biến ngẫu nhiên U có phân phối chuẩn tắc N(0; 1).

Định lý trên được mở rộng thành định lý 5: *Định lý giới hạn trung tâm.* Đây là định lý có ý nghĩa rất lớn trong lý thuyết thống kê.

¤ **Định lý 5 (Định lý giới hạn trung tâm):**

Cho dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập có cùng phân phối xác suất và có kỳ vọng $E(X_i) = a$, $\sigma(X_i) = b$, i = 1, 2, ..., n. Đặt $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ thì khi n lớn, dãy các biến ngẫu nhiên $Z_n = \frac{(\overline{X}_n - a)\sqrt{n}}{b}$ hội tụ theo phân phối về biến ngẫu nhiên Z có luật phân phối chuẩn tắc N(0, 1).

Tổng quát hơn, ta có định lý sau đây phát biểu điều kiện hội tụ cho dãy các biến ngẫu nhiên độc lập có luật phân phối, kỳ vọng, phương sai khác nhau.

¤ **Định lý 6 (Định lý giới hạn trung tâm Linderberg):**

Cho dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ độc lập có kỳ vọng $E(X_i) = a_i$, $\sigma(X_i) = b_i$, i = 1, 2, ..., n. Đặt $B_n = \sum_{i=1}^{n} b_i^2$. Nếu $\frac{1}{B_n^2}\sum_{i=1}^{n} E(X_i - a_i)^2 \xrightarrow[n\to+\infty]{} 0$ thì dãy $Z_n = \frac{1}{B_n^2}\sum_{i=1}^{n} E(X_i - a_i)$ hội tụ theo phân phối về biến ngẫu nhiên Z có luật phân phối chuẩn tắc N(0, 1).