Chương 2
BIẾN NGẪU NHIÊN VÀ LUẬT PHÂN PHỐI XÁC SUẤT

Khái niệm biến ngẫu nhiên (BNN) (còn được gọi là đại lượng ngẫu nhiên) và các đặc trưng của chúng là những khái niệm rất quan trọng của lí thuyết xác suất. Định nghĩa chính xác mang tính toán học thuần tuý về BNN vượt khỏi yêu cầu của giáo trình này. Định nghĩa được trình bày ở đây mang tính mô tả, tuy nhiên nó cũng giúp cho người học hiểu được thế nào là BNN, BNN rời rạc, BNN liên tục. Đối với BNN ta chỉ quan tâm đến vấn đề BNN này nhận một giá trị nào đó hoặc nhận giá trị trong một khoảng nào đó với xác suất bao nhiêu. Nói cách khác BNN X có thể được khảo sát thông qua hàm phân phối xác suất của nó $F(x) = P\{X < x\}$. Như vậy khi ta biết quy luật phân phối xác suất của một BNN thì ta đã nắm được toàn bộ thông tin về BNN này. Ngoài phương pháp sử dụng hàm phân phối để xác định BNN, trong nhiều trường hợp bài toán chỉ đòi hỏi cần khảo sát những đặc trưng cơ bản của BNN như kì vọng, phương sai và độ lệch chuẩn. Trong chương này ta cũng xét một số quy luật phân phối xác suất quan trọng như: phân phối nhị thức (thường gặp trong dãy phép thử Bernoulli); phân phối Poisson (thường gặp trong bài toán về quá trình đếm sự xuất hiện sự kiện A nào đó, quá trình đếm của các hệ phục vụ); phân phối mũ; phân phối đều; phân phối chuẩn (thường gặp trong các bài toán về sai số khi đo đạc). Trong thực tế, nhiều BNN tuân theo phân phối chuẩn hoặc tiệm cận chuẩn (định lí giới hạn trung tâm), chẳng hạn: trọng lượng, chiều cao của một nhóm người nào đó, điểm thi của thí sinh, mức lãi suất của một công ty, nhu cầu tiêu thụ của một mặt hàng... Ngoài ra, khái niệm BNN nhiều chiều cũng được giới thiệu một cách sơ lược trong chương này.

2.1. BIẾN NGẪU NHIÊN VÀ PHÂN PHỐI XÁC SUẤT

2.1.1. Khái niệm biến ngẫu nhiên

a) Khái niệm

"Biến" là cái có thể thay đổi. "Ngẫu nhiên" nghĩa là ta chưa xác định được. Một biến có thể là ngẫu nhiên với người này, nhưng không ngẫu nhiên với người khác, tùy theo lượng thông tin nhận được. Ví dụ, số tuổi của ông A là một số xác định, không ngẫu nhiên đối với ông A, nhưng nó là một số không xác định, ngẫu nhiên với ông B nào đó.

Khái niệm biến số (đại lượng biến thiên) đã rất thông dụng trong toán giải tích. Chính vì thế ta tìm cách đưa vào khái niệm biến số ngẫu nhiên như là một đại lượng phụ thuộc vào kết cục của một phép thử ngẫu nhiên.

Ví dụ 2.1. Khi nghiên cứu BNN "nhiệt độ" của một phản ứng hóa học trong một khoảng thời gian nào đó, ta thấy nhiệt độ đó nhận giá trị trong một khoảng $[t; T]$, trong đó t và T tương ứng là các nhiệt độ thấp nhất và cao nhất của phản ứng trong khoảng thời gian trên.

Về mặt hình thức, có thể định nghĩa BNN như là một hàm số có giá trị thực xác định trên không gian các sự kiện sơ cấp (sao cho nghịch ảnh của một khoảng số là một sự kiện), tức là

$$X : \mathbf{\Omega} \to \mathbb{R},$$

trong đó $\mathbf{\Omega}$ là không gian các sự kiện sơ cấp. Để phân biệt sau này ta kí hiệu X, Y,... là các BNN, còn x, y,... là giá trị của các BNN đó. Như vậy, X mang tính ngẫu nhiên, còn x là giá trị cụ thể quan sát được khi phép thử đã tiến hành. Trong giáo trình này, ta chỉ xét những BNN thỏa mãn điều kiện $X(\omega) < +\infty$, với mọi $\omega \in \mathbf{\Omega}$.

Tương tự như với các số và các hàm số, ta có thể làm nhiều phép toán khác nhau với các BNN: cộng, trừ, nhân, chia, lấy giới hạn, tích phân, hàm hợp... Qua các phép toán như vậy, chúng ta có thể suy ra các BNN mới từ các BNN cho trước.

Ví dụ 2.2. Một học sinh thi vào đại học phải thi 3 môn. Điểm của mỗi môn có thể coi là BNN. Tổng điểm cũng là một BNN, và là tổng của 3 BNN trước.

Ví dụ 2.3. (xem [1]) Tốc độ V của một xe ô tô đang chạy trên đường có thể coi là một BNN. Nếu xe đang chạy mà phải phanh gấp lại vì phía trước có nguy hiểm, thì từ thời điểm người lái xe đạp phanh cho đến thời điểm xe dừng lại, xe phải chạy thêm được một quãng đường có độ dài D nữa. D cũng có thể coi là một BNN. Nó không phải là tỷ lệ thuận với V, mà là tỷ lệ thuận với V^2. Tức là BNN D có thể được sinh ra từ BNN V theo công thức: $D = k.V^2$. Hệ số k ở đây phụ thuộc vào điều kiện của đường và điều kiện của xe; nó có thể coi là xác định nếu ta biết các điều kiện này, còn nếu không thì có thể coi là một BNN khác. Ví dụ, trong điều kiện bình thường, thì $k = 0,08s^2/m$, một xe đang chạy với tốc độ 36km/h (10m/s) thì từ lúc bóp phanh đến lúc dừng hẳn sẽ chạy thêm $0,08 \times 10^2 = 8$ mét, nhưng nếu xe đang chạy với tốc độ 72km/h (2×36km/h) thì từ lúc đạp phanh đến lúc dừng hẳn sẽ chạy thêm những $8 \times 2^2 = 32$ mét.

b) Phân loại

BNN được gọi là *rời rạc*, nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử. Ví dụ: số điểm thi của một học sinh, số cuộc điện thoại gọi đến một tổng đài trong một đơn vị thời gian, số viên đạn trúng đích khi bắn liên tiếp n viên đạn độc lập vào một mục tiêu,...

BNN được gọi là *liên tục*, nếu tập giá trị của nó lấp kín một khoảng trên trục số (số phần tử của tập giá trị là vô hạn không đếm được). Ví dụ: độ cao của một cây tại thời gian t nào đó, nhiệt độ không khí ở mỗi thời điểm nào đó, độ dài của chi tiết máy, tuổi thọ của một loại thiết bị điện tử đang hoạt động,...

Như vậy miền giá trị của một BNN rời rạc sẽ là một dãy số $x_1, x_2, \cdots, x_n, \cdots$ có thể hữu hạn hoặc vô hạn. Miền giá trị của một BNN liên tục sẽ là một đoạn $[a, b] \subset \mathbb{R}$ hoặc là chính $\mathbb{R} = (-\infty, +\infty)$.

2.1.2. Phân phối xác suất của biến ngẫu nhiên

a) Bảng phân phối xác suất và hàm xác suất

Đối với BNN rời rạc, mỗi giá trị của nó được gắn với một xác suất đặc trưng cho khả năng BNN nhận giá trị đó $p_i = P(X = x_i)$. Bảng phân phối xác suất của X có dạng sau

$X = x$	x_1	x_2	$\cdots$	x_n	$\cdots$
$p(x)$	p_1	p_2	$\cdots$	p_n	$\cdots$

trong đó $\{x_1, x_2, \cdots, x_n, \cdots\}$ là tập các giá trị của X (đã sắp xếp theo thứ tự tăng), $p_n = p(x_n) = P(X = x_n)$.

Ví dụ 2.4. Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Lời giải. Số đạn cần bắn, kí hiệu là X, là một BNN rời rạc. Từ yêu cầu của bài toán sẽ có 3 giá trị của X là 1, 2 và 3.

$X = 1$ là sự kiện viên thứ nhất trúng và $p_1 = P(X = 1) = 0,6$.

$X = 2$ là sự kiện viên thứ nhất trượt, còn viên thứ hai trúng và do độc lập nên $p_2 = P(X = 2) = 0,4.0,6 = 0,24$.

Cuối cùng, nếu viên thứ hai vẫn trượt thì chắc chắn phải bắn viên thứ ba, do đó $p_3 = P(X = 3) = (0,4)^2 = 0,16$.

Từ đó bảng phân phối cần tìm là:

X	1	2	3
$p(x)$	0,6	0,24	0,16

Hàm số $p(x) = P(X = x)$ với x thuộc tập giá trị của X, thường được gọi là *hàm xác suất* của X và có 2 tính chất cơ bản gồm:

(i) $p(x) \geq 0, \forall x$.

(ii) $\sum\limits_{x} p(x) = 1$.

Ví dụ 2.5. Một xạ thủ bắn 3 phát, xác suất bắn trúng mục tiêu của mỗi phát là 0,6. Hãy lập bảng phân phối xác suất của số đạn trúng mục tiêu.

Lời giải. Gọi X là số đạn bắn trúng mục tiêu, khi đó tập các giá trị của X sẽ là $\{0, 1, 2, 3\}$. Xem việc bắn 3 viên đạn độc lập vào một mục tiêu như tiến hành dãy 3 phép thử Bernoulli với xác suất bắn trúng đích của mỗi viên đạn là 0,6. Như vậy $p(k) = P(X = k) = C_3^k.(0,6)^k.(0,4)^{3-k}$, với $k = \overline{0,3}$. Từ đó bảng phân phối xác suất cần tìm là:

X	0	1	2	3
$p(x)$	0,064	0,288	0,432	0,216

b) Hàm phân phối xác suất

Bảng phân phối xác suất có một hạn chế cơ bản là chưa đủ tổng quát để đặc trưng cho một BNN tùy ý, nhất là trường hợp BNN liên tục. Vì vậy người ta đưa ra khái niệm sau.

Định nghĩa 2.1. Hàm phân phối xác suất của BNN X, kí hiệu là $F(x)$, được xác định như sau

$$F(x) = P(X < x), \quad x \in \mathbb{R}. \tag{2.1}$$

Từ định nghĩa trên, $F(x)$ phản ánh độ tập trung xác suất ở bên trái của số thực x. Trong trường hợp BNN rời rạc, *(2.1)* cho ta một hàm còn được gọi là hàm phân phối tích lũy (hay xác suất tích lũy), tức là

$$F(x) = P(X = x_1) + \cdots + P(X = x_{i-1}), \text{ với } x_{i-1} < x \le x_i. \tag{2.2}$$

Ví dụ 2.6. Giả sử BNN X có bảng phân phối xác suất như sau

X	1	2	3
$p(x)$	0,5	0,2	0,3

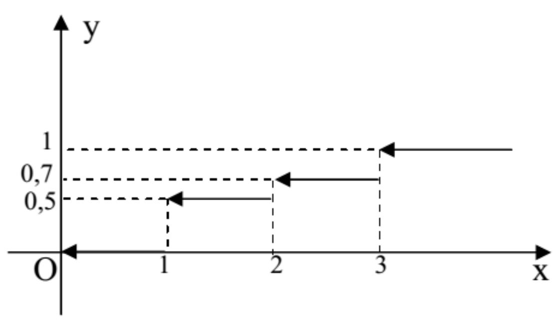
Từ bảng phân phối trên và *(2.2)*, ta có hàm phân phối

$$F(x) = \begin{cases} 0 & \text{nếu } x \le 1, \\ 0,5 & \text{nếu } 1 < x \le 2, \\ 0,7 & \text{nếu } 2 < x \le 3, \\ 1 & \text{nếu } x > 3. \end{cases}$$

Để ý là số giá trị của X bằng số điểm gián đoạn loại 1 của $F(x)$ (đồ thị dạng bậc thang).

Hàm phân phối xác suất có vai trò quan trọng khi nghiên cứu các BNN liên tục. Nếu ta biết được hàm phân phối xác suất thì ta sẽ xác định hoàn toàn BNN. Tuy nhiên trong thực tế, việc tìm được $F(x)$ là rất khó, nếu không nói là hầu như không thể làm được.

Đồ thị F(x)



Một số **tính chất** của hàm phân phối $F(x)$:

(i) $0 \le F(x) \le 1$;

(ii) $F(x)$ là một hàm không giảm, tức là nếu $x_1 < x_2$ thì $F(x_1) \le F(x_2)$;

(iii) $P(a \le X \le b) = F(b) - F(a)$;

(iv) $\lim\limits_{x \to -\infty} F(x) = 0$ và $\lim\limits_{x \to +\infty} F(x) = 1$.

Ví dụ 2.7. Cho hàm phân phối xác suất của một BNN liên tục X có dạng:

$$F(x) = \begin{cases} 0, & x \le 2, \\ k(x-2)^2, & 2 < x \le 4, \\ 1, & x > 4. \end{cases}$$

Xác định hằng số k và tính $P(2 \le X < 3)$.

Lời giải. Vì $F(x)$ liên tục nên tại $x = 4$ ta có $k(4-2)^2 = 1$, từ đó $k = 1/4$. Hơn nữa từ tính chất (iii) và (iv) ta được

$$P(2 \le X < 3) = F(3) - F(2) = \frac{1}{4}(3-2)^2 - 0 = \frac{1}{4}.$$

c) Hàm mật độ xác suất

Hàm phân phối xác suất $F(x)$ không cho biết rõ phân phối xác suất ở lân cận một điểm nào đó trên trục số. Vì vậy đối với các BNN liên tục, có $F(x)$ khả vi, người ta đưa ra khái niệm sau đây.

Định nghĩa 2.2. Hàm mật độ xác suất của BNN liên tục X, kí hiệu là $f(x)$, có hàm phân phối $F(x)$ khả vi (trừ ở một số hữu hạn điểm gián đoạn bị chặn), được xác định bằng

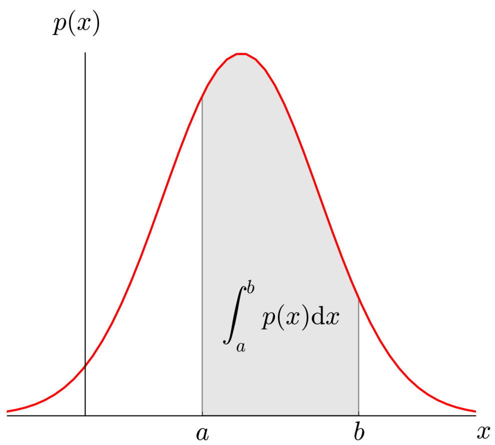
$$f(x) = F'(x). \tag{2.3}$$

Từ định nghĩa trên ta thấy nếu X là BNN rời rạc thì hàm phân phối xác suất của X là hàm gián đoạn nên $F(x)$ không khả vi tại những điểm gián đoạn. Vì vậy BNN rời rạc không có hàm mật độ xác suất.

Ngoài ra, từ *(2.3)* ta có

$$F(x) = \int_{-\infty}^{x} f(x)dx \quad \text{và} \quad P(a \le X < b) = \int_a^b f(x)dx.$$

Về mặt hình học, xác suất để BNN X nhận giá trị trong (a, b) (hoặc $[a, b]$) là diện tích phần mặt phẳng chắn bởi đường cong $p(x)$, trục Ox và các đường thẳng tương ứng (xem H. 2.1)



Hình 2.1: Xác suất $\mathbf{P(a \le X < b)}$

Giả sử X là BNN liên lục, khi đó $F(x)$ liên tục. Ta có

$$P(X = a) = \lim_{b \to a^+} P(a \le X < b) = \lim_{b \to a^+} [F(b) - F(a)] = 0.$$

Như vậy, từ $P(X = a) = 0$, ta thấy rằng: Nếu X là BNN liên tục thì

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

Hàm mật độ xác suất của một BNN liên tục có hai tính chất cơ bản giống như hàm xác suất là:

(i) $f(x) \ge 0, \forall x.$

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

Từ tính chất (ii), ta thấy diện tích của hình giới hạn bởi hàm mật độ xác suất và trục Ox bằng 1. Người ta cũng chứng minh được rằng nếu một hàm thực không âm $f(x)$ thỏa mãn $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ thì nó cũng là hàm mật độ xác suất của một BNN X nào đó.

Ví dụ 2.8. Cho $f(x) = \begin{cases} 0, & \text{nếu } x \notin [0, \pi], \\ k \sin x, & \text{nếu } x \in [0, \pi]. \end{cases}$

a) Tìm k để $f(x)$ là hàm mật độ xác suất.

b) Tìm hàm phân phối xác suất tương ứng.

c) Tính xác suất để X nhận giá trị trong khoảng $\left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$.

Lời giải.

a) Để $f(x)$ là hàm mật độ xác suất thì $f(x) \geq 0$ và $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$. Từ $f(x) \geq 0$ ta suy ra $k \geq 0$. Đồng thời

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Leftrightarrow \int_0^\pi k \sin x dx = 1 \Leftrightarrow k = \frac{1}{2}.$$

b) Với $x \leq 0$ thì $\int\limits_{-\infty}^{x} f(t)dt = 0$.

Với $0 < x \leq \pi$ thì

$$\int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_0^x \frac{1}{2} \sin t dt = \frac{1}{2}(1 - \cos x).$$

Với $x > \pi$ thì

$$\int_{-\infty}^{x} f(t)dt = \int_0^\pi \frac{1}{2} \sin t dt + \int_\pi^x 0dt = 1.$$

Vậy $F(x) = \begin{cases} 0, & \text{nếu } x \leq 0, \\ \frac{1}{2}(1 - \cos x), & \text{nếu } 0 < x \leq \pi, \\ 1, & \text{nếu } x > \pi. \end{cases}$

c) Xác suất để X nhận giá trị trong khoảng $\left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$ là

$$P\left(\frac{\pi}{2} < X < \frac{3\pi}{2}\right) = F\left(\frac{3\pi}{2}\right) - F\left(\frac{\pi}{2}\right) = 1 - \frac{1}{2}\left(1 - \cos\frac{\pi}{2}\right) = \frac{1}{2}.$$

2.2. CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

Khi biết bảng phân phối xác suất đối với BNN rời rạc hay biết hàm mật độ xác suất đối với BNN liên tục thì ta hoàn toàn xác định được quy luật xác suất của BNN. Tuy nhiên, trong thực tế, để giải quyết một vấn đề nào đó nhiều khi không cần thông tin một trong các loại hàm nêu trên mà chỉ cần biết một số giá trị đặc trưng tương ứng với BNN đang xét. Các giá trị đặc trưng này được chia thành hai nhóm: một nhóm đặc trưng cho vị trí và một nhóm đặc trưng cho mức phân tán của BNN.

2.2.1. Kì vọng của biến ngẫu nhiên

a) Định nghĩa

Định nghĩa 2.3. Kì vọng của BNN X, kí hiệu là $E(X)$, xác định như sau:
- Nếu X là BNN rời rạc có hàm xác suất $p(x_i) = p_i$, $i = 1, 2, \cdots, n$ thì

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

- Nếu X là BNN rời rạc nhận vô hạn đếm được các giá trị x_i với hàm xác suất tương ứng $p(x_i) = p_i$ và nếu $\sum\limits_{i=1}^{\infty} |x_i| p_i$ hội tụ thì $E(X) = \sum\limits_{i=1}^{\infty} x_i p_i$;

- Nếu X là BNN liên tục có hàm mật độ xác suất $f(x)$, $x \in \mathbb{R}$ và $\int\limits_{-\infty}^{+\infty} |x| f(x) dx$ hội tụ thì $E(X) = \int\limits_{-\infty}^{+\infty} x f(x) dx$.

Nhận xét:

(i) Các BNN rời rạc nhận một số hữu hạn các giá trị luôn có kì vọng.

(ii) Các BNN rời rạc nhận một số vô hạn đếm được hoặc không đếm được các giá trị có thể không có giá trị kì vọng.

(iii) Kì vọng của BNN X là giá trị đặc trưng cho vị trí (trung tâm) của BNN, tức là các giá trị cụ thể của X sẽ tập trung quanh kì vọng.

Định nghĩa 2.4. BNN X được gọi là quy tâm nếu $E(X) = 0$. Đối với BNN X bất kì, BNN $Y = X - E(X)$ là quy tâm và được gọi là BNN quy tâm hóa của X.

Ví dụ 2.9. Hàm mật độ xác suất của BNN X là $f(x) = \dfrac{1}{\pi(1+x^2)}$. Tính $E(X)$?

Lời giải. Vì X là BNN liên tục nên kì vọng của X là

$$\int_{-\infty}^{+\infty} |x| f(x) dx = \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx.$$

Rõ ràng tích phân trên phân kì nên X không có kì vọng.

Ví dụ 2.10. Giả sử một cái bình đựng 10 quả cầu giống nhau nhưng khác nhau về trọng lượng: 5 quả nặng 1 kg, 2 quả nặng 2 kg, 3 quả nặng 3 kg. Lấy ngẫu nhiên từ bình ra 1 quả cầu và gọi X là trọng lượng của quả cầu đó. Tính $E(X)$ và so sánh $E(X)$ với trọng lượng trung bình của 1 quả cầu trong hộp.

Lời giải. X là trọng lượng của quả cầu được lấy ra. Khi đó ta có bảng phân phối xác suất của X:

X	1	2	3
$p(x)$	5/10	2/10	3/10

Kì vọng $E(X) = \sum\limits_{i=1}^{3} x_i p_i = 1.\dfrac{5}{10} + 2.\dfrac{2}{10} + 3.\dfrac{3}{10} = \dfrac{18}{10} = 1,8.$

Gọi M là trọng lượng trung bình của các quả cầu trong bình, ta có

$$M = \frac{5.1 + 2.2 + 3.3}{10} = \frac{18}{10} = 1,8.$$

Vậy $E(X) = M$.

Tính chất:

(i) $E(C) = C$ với C là hằng số;

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

(iii) $E(X + Y) = E(X) + E(Y)$;

(iv) Nếu X, Y độc lập thì $E(X.Y) = E(X).E(Y)$;

(v) Nếu $Y = \varphi(X)$ thì $E(Y) = \sum_i \varphi(x_i)p(x_i)$ hoặc $E(Y) = \int_{-\infty}^{+\infty} \varphi(x)f(x)dx$, trong đó $p(x)$ và $f(x)$ là các hàm xác suất và mật độ xác suất tương ứng.

b) Ý nghĩa của kì vọng

Tiến hành n phép thử. Giả sử X là BNN nhận các giá trị có thể $x_1, x_2, \cdots, x_N$ với số lần xuất hiện tương ứng $k_1, k_2, \cdots, k_N$. Giá trị trung bình (trung bình cộng) của BNN X trong n phép thử là

$$\overline{X} = \frac{x_1k_1 + x_2k_2 + \cdots + x_Nk_N}{n} = \frac{k_1}{n}x_1 + \cdots + \frac{k_N}{n}x_N = \sum_{i=1}^{N} x_if_i$$

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

Sử dụng định nghĩa xác suất theo lối thống kê ta có $\lim\limits_{n\to\infty} f_i = p_i$. Vì vậy với n đủ lớn ta có

$$\overline{X} = \sum_{i=1}^{N} x_ip_i = E(X).$$

Như vậy, kì vọng chính là tổng có trọng số của tất cả các giá trị của X, hay còn là trị trung bình (trung bình có trọng số) của BNN (phân biệt với trung bình cộng của các giá trị). Trong thực tế, nếu quan sát các giá trị của X nhiều lần và lấy trung bình cộng (trung bình số học), thì khi số quan sát càng lớn, số trung bình đó càng gần tới kì vọng $E(X)$, vì vậy kì vọng còn được gọi là *trung bình số học* của biến X mà không sợ nhầm lẫn.

Ví dụ 2.11. Một người dùng 10 000 đồng để mua 1 con số nằm trong dãy số từ 00 đến 99. Người mua sẽ thắng 700 000 đồng nếu số mua trùng với 2 số cuối của giải đặc biệt do Nhà nước phát hành trong ngày hôm đó. Hãy tìm số tiền thắng trung bình của một lần chơi như vậy.

Lời giải. Gọi X là số tiền thắng của một lần chơi, rõ ràng X nhận các giá trị 0 và 700 000 với các tần suất (và coi luôn là xác suất) tương ứng là 99/100 và 1/100. Từ đó số tiền thắng trung bình trong mỗi lần chơi chính là kì vọng

$$E(X) = 0.0,99 + 700\ 000.0,01 = 7\ 000 \text{ đồng}.$$

Mặc dù $E(X) > 0$, nhưng anh ta đã bỏ ra 10 000 đồng để mua 1 con số. Như vậy trong thực tế mỗi lần chơi anh ta mất trung bình 3 000 đồng.

Ví dụ 2.12. Một gia đình chi tiêu trong một tháng ở 2 mức: 10 triệu hoặc 8 triệu đồng, trong đó có 11 tháng ở mức 10 triệu, chỉ có 1 tháng ở mức 8 triệu. Hỏi trung bình mỗi tháng gia đình này tiêu hết bao nhiêu tiền? Như vậy có 2 mức chi tiêu nên trung bình số học (trung bình cộng) sẽ là $(10+8)/2 = 9$ triệu đồng. Nhưng trung bình có trọng số (kì vọng) sẽ là $10.\frac{11}{12} + 8.\frac{1}{12} \approx 9,83$ triệu đồng. Rõ ràng kì vọng phản ánh chính xác hơn, hợp lí hơn trung bình số học.

c) Giá trị kì vọng hình học

Giá trị kì vọng ứng với trung bình cộng, còn giá trị kì vọng hình học ứng với trung bình nhân. Một ví dụ đơn giản sau đây cho thấy sự quan trọng của trung bình nhân trong thực tế.

Ví dụ 2.13. (xem [1]) Giả sử giá nhà dao động trong 4 năm như sau. Năm đầu tiên giảm 15%, năm thứ hai tăng 35%, năm thứ ba giảm 20%, năm thứ tư tăng 20%. Hỏi xem trong 4 năm đó giá nhà tăng lên (giảm đi) trung bình mỗi năm bao nhiêu %?

Nếu ta lấy trung bình cộng thì được $(-15\% + 35\% - 20\% + 20\%)/4 = 5\%$ một năm. Nhưng con số đó có phản ánh chính xác sự đi lên của giá nhà trong 4 năm?

Nếu gọi giá lúc đầu là X, thì sau năm đầu giá là $(1-15\%)X$, sau năm thứ hai giá là $(1+35\%)(1-15\%)X$, sau năm thứ ba giá là $(1-20\%)(1+35\%)(1-15\%)X$, sau 4 năm giá là $(1+20\%)(1-20\%)(1+35\%)(1-15\%)X = 1,1016.X$. Tức là sau 4 năm giá nhà chỉ tăng lên 10,16%, chứ không phải 20%. Để có cái nhìn chính xác về độ tăng trưởng trung bình hàng năm trong giai đoạn 4 năm, cần tính như sau

$$\big[(1+0,2)(1-0,2)(1+0,35)(1-0,15)\big]^{1/4} - 1 \approx 0,02449,$$

tức là giá nhà tăng 2,449% một năm.

Trung bình nhân có thể được định nghĩa qua trung bình cộng

$$\left(\prod_i x_i\right)^{1/n} = \exp\left(\frac{1}{n}\sum_i \ln x_i\right). \qquad (2.4)$$

Mặt khác hàm $f(x) = \ln x$ là hàm lõm trên nửa đường thẳng dương (đạo hàm bậc hai của nó bằng $-1/x^2$ là một hàm âm), do đó

$$\frac{1}{n}\sum_i \ln x_i \le \ln\left(\frac{1}{n}\sum_i x_i\right).$$

Lấy exp hai vế của bất đẳng trên ta được

$$\left(\prod_i x_i\right)^{1/n} \le \frac{1}{n}\sum_i x_i, \qquad (2.5)$$

dấu "=" xảy ra khi và chỉ khi tất cả các số x_i bằng nhau.

Từ *(2.4)* gợi ý cho ta định nghĩa giá trị kì vọng hình học của X sau đây.

Định nghĩa 2.5. Nếu X là một BNN chỉ nhận các giá trị dương thì giá trị kì vọng hình học của X, kí hiệu là $G(X)$, được cho bởi công thức sau

$$G(X) = \exp(E(\ln X)).$$

Từ *(2.5)*, ta được $G(X) \leq E(X)$. Dấu "=" xảy ra khi X là hằng số.

Ví dụ 2.14. Giả sử có một cơ hội đầu tư trong đó khả năng thắng/thua là 50-50, sau 1 tháng sẽ biết kết quả. Nếu thắng thì lãi 100%, nếu thua thì lỗ 50% số tiền bỏ ra. Hỏi đối với người đầu tư thì có nên đầu tư vào những cơ hội như vậy không? Theo bạn nên đầu tư với nhiều nhất bao nhiêu % vốn đầu tư (để đạt kì vọng lợi nhuận cao nhất, giả sử là không có các cơ hội đầu tư khác)?

Lời giải. Trước hết, ta có thể tính giá trị kì vọng lợi nhuận của đầu tư theo cơ hội trên với 1 đơn vị vốn bỏ ra. Gọi L là biến "lợi nhuận", ta có 2 khả năng: hoặc $L = 1$ hoặc $L = -1/2$, mỗi khả năng có xác suất 50%. Như vậy kì vọng lợi nhuận trên 1 đơn vị vốn bỏ ra là:

$$E(L) = 0,5.1 + 0,5.(-1/2) = 0,25.$$

Kì vọng lợi nhuận ở đây là dương (bằng 25% vốn bỏ ra trong một tháng), nên đây là cơ hội nên đầu tư, trừ khi có những cơ hội khác tốt hơn.

Câu hỏi thứ hai là nhà đầu tư nên đầu tư vào đó nhiều nhất là bao nhiêu phần trăm vốn đầu tư? Nếu giả sử đầu tư toàn bộ 100% vốn. Khi đó có 2 khả năng, hoặc là tổng số vốn tăng lên gấp đôi, hoặc là giảm đi còn 1 nửa, với xác suất của mỗi khả năng là 50%. Nhưng nếu một nhà đầu tư mà làm như vậy 2 lần liên tiếp, 1 lần thắng và 1 lần thua, thì sau hai lần số vốn lại về như cũ, không tăng trưởng được gì. Muốn đảm bảo cho vốn tăng trưởng "về lâu về dài", cái cần tính đến không phải là kì vọng của vốn sau mỗi lần đầu tư, mà là giá trị kì vọng hình học. Nếu giả sử chỉ có 1 cơ hội đầu tư duy nhất như trên, thì giá trị kì vọng hình học của vốn có được sau khi đầu tư Y tiền vào đó trên tổng số X tiền sẽ là $\sqrt{(X - Y/2)(X + Y)}$. Để tối đa hóa giá trị kì vọng hình học tức là tìm Y sao cho $\sqrt{(X - Y/2)(X + Y)}$ đạt cực đại, ta cần chọn $Y = X/2$ (với X cho trước), và khi đó giá trị kì vọng hình học của vốn sau khi đầu tư là $\sqrt{(X - X/4)(X + X/2)} = 1,061.X$. Như vậy, kì vọng lợi nhuận của một cơ hội đầu tư như trên, tính trên toàn bộ vốn của nhà đầu tư, chỉ có không quá 6,1% chứ không phải 25%.

2.2.2. Phương sai và độ lệch chuẩn

a) Định nghĩa

Xét BNN X có kì vọng $E(X)$.

Định nghĩa 2.6. Phương sai của BNN X, kí hiệu $D(X)$, được định nghĩa như sau:

$$D(X) = E\Big[\big(X - E(X)\big)^2\Big]. \qquad (2.6)$$

Sử dụng tính chất của kì vọng, ta có thể biến đổi công thức *(2.6)* thành

$$D(X) = E(X^2) - [E(X)]^2.$$

Trong tính toán, phụ thuộc vào X là rời rạc (với hàm xác suất $p(x)$) hay liên tục (với hàm mật độ $f(x)$), ta có hai công thức tính phương sai:

$$D(X) = \sum_{\forall i}[x_i - E(X)]^2 p_i = \sum_{\forall i} x_i^2 p_i - \left(\sum_{\forall i} x_i p_i\right)^2,$$

$$D(X) = \int_{-\infty}^{+\infty}[x - E(X)]^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} xf(x)dx\right)^2.$$

Ví dụ 2.15. Theo thống kê, việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 99,2%, còn xác suất để người đó chết trong vòng một năm tới là 0,8%. Một công ty bảo hiểm đề nghị người đó tham gia chương trình bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1 000 USD, còn tiền đóng là 10 USD. Tính phương sai của lợi nhuận mà công ty bảo hiểm nhận được.

Lời giải. Gọi X là lợi nhuận của công ty bảo hiểm trong 1 năm. Khi đó X nhận 2 giá trị là 10 USD (nếu khách hàng không chết) và -990 USD (nếu khách hàng qua đời) với xác suất tương ứng là 0,008 và 0,992. Như vậy kì vọng (lợi nhuận của công ty trên 1 khách hàng trong 1 năm) là

$$E(X) = (-990).0,008 + 10.0,992 = 2.$$

Ta có $E(X^2) = (-990)^2.0,008 + 10^2.0,992 = 7940$ và phương sai

$$D(X) = E(X^2) - [E(X)]^2 = 7940 - 4 = 7936.$$

Để ý rằng phương sai $D(X)$ luôn là một số không âm (xem công thức *(2.6)*). Từ định nghĩa ta cũng thấy rằng về mặt vật lí $D(X)$ không cùng thứ nguyên (cùng đơn vị đo) đối với X, vì vậy ta đưa vào khái niệm sau đây.

Định nghĩa 2.7. Độ lệch chuẩn của BNN X, kí hiệu là $\sigma(X)$, được định nghĩa là $\sigma(X) = \sqrt{D(X)}$.

Phương sai còn được kí hiệu là $\sigma^2(X)$ hoặc σ^2 (nếu đã biết rõ là phương sai của BNN X nào đó). Độ lệch chuẩn được dùng thường xuyên hơn phương sai do có cùng đơn vị đo với chính BNN X.

Tính chất:

(i) $D(C) = 0$ với C là hằng số,

(ii) $D(CX) = C^2 D(X)$, $\sigma(CX) = |C|\sigma(X)$,

(iii) Nếu X, Y độc lập và $D(X), D(Y)$ hữu hạn thì $D(X \pm Y) = D(X) + D(Y)$ và

$$\sigma(X \pm Y) = \sqrt{D(X) + D(Y)} \leq \sigma(X) + \sigma(Y).$$

b) Ý nghĩa và ứng dụng thực tế của phương sai

Phương sai chính là độ lệch bình phương trung bình quanh giá trị kì vọng $E(X)$. Vậy phương sai đặc trưng cho độ phân tán của BNN quanh trị trung bình của BNN đó. Cũng theo ý nghĩa đó, phương sai càng lớn thì độ bất định của biến tương ứng càng lớn. Trong công nghiệp, phương sai biểu thị độ chính xác của sản xuất; trong kỹ thuật, phương sai đặc trưng cho mức độ phân tán của các chi tiết gia công hay sai số của thiết bị; trong quản lí và kinh doanh thì nó đặc trưng cho mức độ rủi ro của các quyết định; trong chăn nuôi, phương sai biểu thị độ đồng đều của các con gia súc; trong trồng trọt, nó biểu thị mức độ ổn định của năng suất...

Ví dụ 2.16. Trong Ví dụ 2.15, đầu tư của công ty bảo hiểm cho những người 25 tuổi là có lãi nhưng rủi ro là rất lớn với $\sigma(X) = \sqrt{7936} \approx 89,08$.

Phương sai của trung bình cộng n biến ngẫu nhiên độc lập cùng phân phối sẽ bé hơn n lần phương sai của các biến thành phần, tức là nếu $D(X_i) = \sigma^2, \forall i = \overline{1,n}$ thì

$$D(\overline{X}) = D\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{\sigma^2}{n}.$$

Đây chính là lí do khi đo đạc các đại lượng vật lí, người ta thường đo nhiều lần rồi lấy trung bình cộng các kết quả. Khi đó sai số của kết quả đo đạc sẽ rất nhỏ.

2.2.3. Một số tham số đặc trưng khác

Định nghĩa 2.8. Mode của BNN X là giá trị của BNN X, kí hiệu $\mathrm{Mod}X$, mà tại đó BNN X nhận xác suất lớn nhất (nếu X rời rạc) hoặc tại đó hàm mật độ đạt cực đại (nếu X liên tục).

Định nghĩa 2.9. Trung vị là giá trị của BNN X chia phân phối thành hai phần có xác suất giống nhau, tức là nếu kí hiệu trung vị là $\mathrm{Med}X$ thì

$$P(X < \mathrm{Med}X) = P(X \geq \mathrm{Med}X) = \frac{1}{2}.$$

Từ định nghĩa hàm phân phối, để tìm trung vị, ta chỉ cần giải $F(x) = 1/2$. Trong nhiều trường hợp ứng dụng, trung vị là đặc trưng vị trí rất tốt, nhiều khi tốt hơn cả kì vọng, nhất là khi trong số liệu có những sai sót thái quá.

Định nghĩa 2.10. Phân vị mức α của BNN X là số X_α được xác định bởi

$$P(X < X_\alpha) \leq \alpha \leq P(X \leq X_\alpha).$$

Nếu $\alpha = 0,5$ thì $X_{0,5}$ chính là trung vị của BNN X.

Từ định nghĩa hàm phân phối, ta có $F(X_\alpha) = \alpha$. Thông thường người ta hay xét các phân vị 25%, 50% (trung vị), 75%, 95%,...

Ví dụ 2.17. Giả sử BNN X có hàm mật độ

$$f(x) = \begin{cases} 0, & \text{nếu } x \leq 0, \\ \dfrac{x}{2}\exp\left(-\dfrac{x^2}{4}\right), & \text{nếu } x > 0. \end{cases}$$

Xác định $\mathrm{Mod}X$ và trung vị.

Lời giải. $\mathrm{Mod}X$ sẽ là nghiệm của phương trình

$$f'(x) = \frac{1}{2}\exp\left(-\frac{x^2}{4}\right) - \frac{x^2}{4}\exp\left(-\frac{x^2}{4}\right) = 0.$$

Từ đó $\mathrm{Mod}X$ là nghiệm của $1 - \dfrac{x^2}{2} = 0$. Nhưng do $x > 0$ nên $\mathrm{Mod}X = \sqrt{2}$.

Trung vị là nghiệm của phương trình

$$\int_0^{\mathrm{Med}X} \frac{x}{2}\exp\left(-\frac{x^2}{4}\right)dx = 0,5 \quad \text{hay} \quad 1 - \exp\left(-\frac{(\mathrm{Med}X)^2}{4}\right) = 0,5.$$

Từ đây suy ra $\mathrm{Med}X = 1,665$.