Bài số 5

KỲ VỌNG VÀ PHƯƠNG SAI

I. KỲ VỌNG

Nếu tung hai đồng xu 16 lần và X là số mặt ngửa xuất hiện trong mỗi lần tung, thì các giá trị của X có thể là 0, 1 và 2. Giả sử rằng khi thực hiện xong phép thử ta thu được:

+ Số lần mặt ngửa không xuất hiện là 4
+ Số lần chỉ có đúng một mặt ngửa xuất hiện là 7
+ Số lần cả hai mặt ngửa đều xuất hiện là 5.

Khi đó số mặt ngửa xuất hiện trung bình trong mỗi lần tung là

$$\frac{(0)(4)+(1)(7)+(2)(5)}{16}=1,06.$$

Đây là giá trị trung bình và không nhất thiết phải là kết quả của phép thử. Chẳng hạn, thu nhập trung bình một tháng của người bán hàng nào đó không hẳn bằng thu nhập ở một tháng cụ thể của người ấy.

Khi quan tâm tới biến ngẫu nhiên X nào đó, có những trường hợp ta muốn biết **giá trị trung bình của biến ngẫu nhiên X** là bao nhiêu? Giá trị trung bình của biến ngẫu nhiên ấy còn gọi là **kỳ vọng của** của nó.

1. Kỳ vọng của biến ngẫu nhiên.

Định nghĩa.

a. Cho X là một ***biến ngẫu nhiên rời rạc*** với hàm phân phối xác suất là $f(x)$. Khi đó **kỳ vọng** (***giá trị trung bình***) của X là một số thực ký hiệu là $E(X)$ (hoặc μ) được xác định bởi

$$\mu = E(X)=\sum_x xf(x)$$

b. Cho X là một ***biến ngẫu nhiên liên tục*** với hàm mật độ xác suất là $f(x)$. Khi đó **kỳ vọng** (***giá trị trung bình***) của X là một số thực ký hiệu là $E(X)$ được xác định bởi:

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

Chú ý:

a. Trường hợp biến ngẫu nhiên rời rạc: nếu tập giá trị $\{x_1,x_2,\dots,x_n,\dots\}$ của biến ngẫu nhiên là đếm được nhưng có vô hạn phần tử thì kỳ vọng sẽ tồn tại nếu chuỗi $\sum_{i=1}^{\infty}|x_i|f(x_i)$ hội tụ.

b. Trường hợp biến ngẫu nhiên liên tục: kỳ vọng sẽ tồn tại nếu tích phân $\int_{-\infty}^{+\infty} xf(x)dx$ hội tụ.

c. Từ định nghĩa ta nhận thấy: Hai biến ngẫu nhiên có cùng phân phối xác suất thì sẽ có cùng kỳ vọng.

Ví dụ 1. Một vị thanh tra chất lượng kiểm tra một lô hàng gồm 7 sản phẩm trong đó có chứa 4 chính phẩm và 3 phế phẩm. Ông ta lấy ra một mẫu gồm 3 sản phẩm. Hãy tìm giá trị trung bình của số chính phẩm trong mẫu.

Giải: + Đặt X là số sản phẩm tốt trong mẫu, khi đó phân phối xác suất của X là:

$$f(x) = \frac{C_4^x.C_3^{3-x}}{C_7^3}; \quad x = 0,1,2,3.$$

+ Ta có: $f(0) = 1/35, f(1) = 12/35, f(2) = 18/35, f(3) = 4/35$.

+ Do đó:

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1,7.$$

Ví dụ 2. Tung ngẫu nhiên ba đồng xu. Người chơi sẽ nhận được 5 USD nếu tất cả các đồng xu đều sấp hoặc đều ngửa, và người chơi sẽ mất 3 USD nếu trong ba đồng xu có cả đồng xu xuất hiện mặt sấp và đồng xu xuất hiện mặt ngửa. Người chơi hi vọng sẽ kiếm được bao nhiêu tiền?

Giải: + Không gian biến cố sơ cấp cho phép thử tung ba đồng xu một cách đồng thời (hay tương tung một đồng xu ba lần) là:

S = {NNN, NNS, NSN, SNN, NSS, SNS, SSN, SSS}

+ Các biến cố sơ cấp trong S là đồng khả năng nên khả năng xuất hiện của mỗi biến cố là 1/8:

$$P(\text{NNS}) = P(\text{N})P(\text{N})P(\text{S}) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

+ Gọi Y là số tiền mà người chơi sẽ đạt được. Khi đó các giá trị mà biến ngẫu nhiên Y là:

5 USD nếu biến cố E_1 = {NNN, SSS} xuất hiện: $P(E_1) = 1/4$

-3USD nếu biến cố E_2 = { NNS, NSN, SNN, NSS, SNS, SSN } xuất hiện: $P(E_2) = 3/4$.

+ Từ đó người chơi sẽ hy vọng (trung bình) sẽ kiếm được số tiền là:

$$\mu = E(Y) = (5)\left(\frac{1}{4}\right) + (-3)\left(\frac{3}{4}\right) = -1.$$

Ví dụ 3. Đặt X là tuổi thọ tính theo giờ của một thiết bị điện tử nào đó, X là một biến ngẫu nhiên. Giả sử hàm mật độ xác suất của X là:

$$f(x) = \begin{cases} \dfrac{20000}{x^3}, x > 100 \\ 0, \quad x \le 100 \end{cases}$$

Hãy tính tuổi thọ trung bình của thiết bị điện tử loại này.

Giải: + Tuổi thọ trung bình của thiết bị chính bằng kỳ vọng của biến ngẫu nhiên X và bằng:

$$\mu = E(X) = \int_{100}^{+\infty} x\frac{20000}{x^3}dx = \int_{100}^{+\infty}\frac{20000}{x^2}dx = 200$$

+ Do đó, ta có thể hy vọng rằng loại thiết bị điện tử này có tuổi thọ trung bình là 200 giờ.

2.Kỳ vọng của hàm các biến ngẫu nhiên.

Như ta đã biết, khi X là một biến ngẫu nhiên và $g = g(t)$ là một hàm nào đó thì $g(X)$ cũng là một biến ngẫu nhiên. Khi đó $g(X)$ có kỳ vọng là bao nhiêu?

Định lí 1. Cho X là biến ngẫu nhiên với phân phối xác suất là $f(x)$ và $g = g(t)$ là hàm số xác định trên miền giá trị của biến ngẫu nhiên X. Khi đó **kỳ vọng** của biến ngẫu nhiên $g(X)$ được xác định bởi:

$$\mu_{g(X)} = E[g(X)] = \sum g(x)f(x)$$, nếu X là ***biến ngẫu nhiên rời rạc***,

và

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$, nếu X là ***biến ngẫu nhiên liên tục***.

Ví dụ 4. Giả sử số lượng xe ôtô X đến cửa hàng rửa xe vào khoảng thời gian từ 4 giờ chiều đến 5 giờ chiều của một ngày thứ sáu khô ráo, là một biến ngẫu nhiên rời rạc có phân phối xác suất như sau:

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Đặt $g(X) = 2X - 1$ là số tiền (tính theo USD), mà người chủ cửa hàng phải trả cho nhân công rửa xe. Người công nhân rửa xe hy vọng sẽ kiếm được bao nhiêu tiền trong khoảng thời gian nói trên?

Giải: Số tiền người công nhân rửa xe có thể hy vọng nhận được(trung bình) là:

$$\begin{aligned} E[g(X)] &= E(2X - 1) \\ &= \sum_{x=4}^{9} (2x-1)f(x) \\ &= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) \\ &= 12,67 \quad \text{USD} \end{aligned}$$

Ví dụ 5. Cho X là biến ngẫu nhiên liên tục có hàm mật độ là:

$$f(x) = \begin{cases} \dfrac{x^2}{3}, \text{khi} \quad x \in (-1;2), \\ 0, \quad \text{khi} \quad x \notin (-1;2) \,. \end{cases}$$

Hãy tìm kỳ vọng của $g(X) = 4X + 3$.

Giải: Theo Định lí 1, ta có

$$E(g(X)) = E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3} dx = \frac{1}{3}\int_{-1}^{2} (4x^3 + 3x^2) dx = 8.$$

Chú ý: Định lý 1 cho phép ta tính kỳ vọng của biến ngẫu nhiên $g(X)$ mà không cần biết đến phân phối xác suất của biến ngẫu nhiên đó, ta chỉ cần biết thông tin về biến ngẫu nhiên X mà thôi.

Sau đây, ta sẽ mở rộng cho trường hợp các biến ngẫu nhiên X và Y có phân phối xác suất đồng thời $f(x,y)$.

Định lý 2. Cho X và Y là các biến ngẫu nhiên với phân phối xác suất đồng thời là $f(x,y)$, hàm số $g = g(t,u)$ xác định trên miền giá trị của biến ngẫu nhiên hai chiều (X,Y). Khi đó kỳ vọng của biến ngẫu nhiên $g(X,Y)$ được xác định bởi:

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_{x}\sum_{y} g(x,y)f(x,y), \text{ nếu } X \text{ và } Y \text{ là các BNN rời rạc,}$$

và

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy, \text{nếu } X \text{ và } Y \text{ là các BNN liên tục.}$$

Ta cũng có thể tính kỳ vọng của hàm n biến ngẫu nhiên với $n \geq 2$.

Ví dụ 6. Hãy tính $E\left(\frac{Y}{X}\right)$ biết rằng hàm mật độ đồng thời là:

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4} & \text{khi } (x;y) \in (0;2)\times(0;1) \\ 0 & \text{khi } (x;y) \notin (0;2)\times(0;1) \end{cases}$$

Giải: Ta có

$$E\left(\frac{Y}{X}\right) = \int_0^1\int_0^2 \frac{y(1+3y^2)}{4} dxdy = \int_0^1 \frac{y+3y^3}{2} dy = \frac{5}{8}.$$

Chú ý:

a.Nếu $g(X,Y) = X$ thì ta có: $E(X) = \begin{cases} \sum_{x}\sum_{y} xf(x,y) = \sum_{x} xg(x) \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy = \int_{-\infty}^{+\infty} xg(x)dx \end{cases}$, trong đó $g(x)$ là phân phối biên duyên của X. Do đó, khi tính *E*(*X*) trong không gian hai chiều, ta có thể sử dụng phân phối xác suất đồng thời của *X* và *Y* hoặc phân phối biên duyên của *X*.

b.Tương tự, nếu $g(X,Y) = Y$ ta xác định được: $E(Y) = \begin{cases} \sum_{x}\sum_{y} yf(x,y) = \sum_{y} yh(y) \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy = \int_{-\infty}^{+\infty} yh(y)dy \end{cases}$, trong đó $h(y)$ là phân phối biên duyên của Y.

3. Một số tính chất

i) Với C là hằng: $E(C) = C$

ii) Tính chất tuyến tính:

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

+ Với số thực k ta có: $E(kX) = k.E(X)$

iii) Nếu X và Y là hai biến ngẫu nhiên độc lập thì ta có: $E(XY) = E(X \cap Y) = E(X).E(Y)$.

iv) Tính đơn điệu: Nếu $X \geq Y$ thì $E(X) \geq E(Y)$.

v) Nếu f là một hàm lồi và X là một biến ngẫu nhiên thì ta có:

$$E(f(X)) \geq f(E(X)).$$

Áp dụng trong Ví dụ 3: Cho X là biến ngẫu nhiên liên tục có hàm mật độ là

$$f(x) = \begin{cases} \dfrac{x^2}{3}, \text{khi} & x \in (-1;2), \\ 0, \quad \text{khi} & x \notin (-1;2) \end{cases}.$$

Hãy tìm kỳ vọng của $g(X) = 4X + 3$.

+ Ta có: $E(4X+3) = 4E(X) + E(3) = 4E(X) + 3$

+ Mà $E(X) = \int\limits_{-\infty}^{+\infty} xf(x)dx = \int\limits_{-1}^{2} \dfrac{x^3}{3}dx = \dfrac{5}{4}$

+ Do đó: $E(4X+3) = 4E(X) + 3 = 8$.

II. PHƯƠNG SAI.

Giá trị trung bình hay kỳ vọng của biến ngẫu nhiên X là một giá trị đặc biệt quan trọng trong thống kê, nó chính là giá trị trung bình (theo xác suất) của biến ngẫu nhiên, nó phản ánh giá trị trung tâm của phân phối xác suất. Tuy nhiên, với chỉ với con số đó ta chưa thể biết được đầy đủ thông tin về phân phối xác suất của biến ngẫu nhiên. Nhiều khi ta quan tâm tới mức độ phân tán các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình.

Để đánh giá mức độ phân tán của biến ngẫu nhiên X nào đó xung quanh giá trị trung bình $\mu = E(X)$ của nó ta cần có cách đánh giá "độ lệch" giữa X và kỳ vọng $\mu = E(X)$. Để ý rằng:

$$E(X-\mu) = E(X) - \mu = 0 \text{ với mọi biến ngẫu nhiên } X,$$

nên ta sẽ không xét giá trị trung bình (kỳ vọng) của sai số mà sẽ quan tâm đến trung bình (kỳ vọng) của bình phương sai số, và đại lượng đó được gọi là **phương sai:** đó là số thực để đo sự phân tán của biến ngẫu nhiên X và kí hiệu là $\mathrm{Var}(X)$ hoặc là σ_X^2, hoặc là σ^2 khi biến ngẫu nhiên đã rõ ràng.

1. Phương sai và độ lệch chuẩn của biến ngẫu nhiên.

Định nghĩa: Cho X là biến ngẫu nhiên với phân phối xác suất $f(x)$ và kỳ vọng là μ. **Phương sai** của X là một số thực được xác định bởi:

$$\sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x) \text{ ,nếu } X \text{ là rời rạc,}$$

và

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \text{ ,nếu } X \text{ liên tục.}$$

Căn bậc hai của phương sai là σ và được gọi là **độ lệch chuẩn** của biến ngẫu nhiên *X*.

Ví dụ 7. Gọi X là biến ngẫu nhiên biểu thị số xe ôtô được sử dụng cho mục đích kinh doanh chính thức trong một ngày làm việc nào đó. Phân phối xác suất của *X* tại công ty A là:

x	**1**	**2**	**3**
***f*(*x*)**	**0,3**	**0,4**	**0,3**

và tại công ty B là:

x	**0**	**1**	**2**	**3**	**4**
***f*(*x*)**	**0,2**	**0,1**	**0,3**	**0,3**	**0,1**

Hãy chỉ ra rằng phương sai của phân phối xác suất tại công ty B là lớn hơn so với tại công ty A.

Giải: + Tại công ty A, ta tính được

$$\mu = E(X) = (1)(0,3) + (2)(0,4) + (3)(0,3) = 2,0$$

+ Khi đó

$$\sigma^2 = \sum_{x=1}^{3} (x-2)^2 f(x) = \left(1-2\right)^2 (0,3) + \left(2-2\right)^2 (0,4) + \left(3-2\right)^2 (0,3)$$
$$= 0,6.$$

+ Tại công ty B, ta có

$$\mu = E(X) = \left(0\right)(0,2) + (1)(0,1) + (2)(0,3) + (3)(0,3) + (4)(0,1)$$
$$= 2,0$$

+ Khi đó

$$\sigma^2 = \sum_{x=0}^{4} (x-2)^2 f(x)$$
$$= \left(0-2\right)^2 (0,2) + \left(1-2\right)^2 (0,1) + \left(2-2\right)^2 (0,3) + \left(3-2\right)^2 (0,3) + \left(4-2\right)^2 (0,1) = 1,6$$

Rõ ràng, phương sai của số ôtô được sử dụng cho những mục đích kinh doanh chính thức tại công ty B là lớn hơn công ty A.

Sau đây là một công thức thường được sử dụng nhiều hơn trong việc tính toán để tìm số σ^2.

Định lí 3. Phương sai của biến ngẫu nhiên X có thể được xác định bởi công thức:

$$\sigma^2 = E(X^2) - \mu^2.$$

Ví dụ 8. Gọi X là biến ngẫu nhiên biểu thị số thiết bị hỏng trong một hệ thống gồm 3 thiết bị được kiểm tra của một chiếc máy. Phân phối xác suất của *X* như sau:

x	**0**	**1**	**2**	**3**
***f*(*x*)**	**0,51**	**0,38**	**0,10**	**0,01**

Hãy tính σ^2.

Giải: + Trước tiên, ta tính:

$$\mu = (0)(0,51) + (1)(0,38) + (2)(0,10) + (3)(0,01) = 0,61.$$

+ Ta lại có:

$$E(X^2) = (0)^2(0,51) + (1)^2(0,38) + (2)^2(0,10) + (3)^2(0,01) = 0,87.$$

+ Vậy nên:

$$\sigma^2 = E(X^2) - \mu^2 = 0,87 - (0,61)^2 = 0,4979.$$

Ví dụ 9. Nhu cầu hàng tuần đối với Pepsi, theo đơn vị 1000 lít, tại một chuỗi các cửa hàng ở một địa phương nào đó, là một biến ngẫu nhiên liên tục *X* với hàm mật độ xác suất như sau

$$f(x) = \begin{cases} 2(x-1), & x \in (0;2) \\ 0, & x \notin (0;2) \end{cases}$$

Hãy tìm kỳ vọng và phương sai của *X*.

Giải: + Kỳ vọng của biến ngẫu nhiên cần tìm là:

$$\mu = E(X) = 2\int_1^2 x(x-1)dx = \frac{5}{3}.$$

+ Ta có

$$E(X^2) = 2\int_1^2 x^2(x-1)dx = \frac{17}{6}$$

+ Vậy nên:

$$\sigma^2 = E(X^2) - \mu^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

Chú ý. Phương sai hay độ lệch chuẩn chỉ có ý nghĩa khi ta so sánh hai hay nhiều phân phối có cùng đơn vị đo. Do đó, ta chỉ có thể so sánh phương sai của các biến ngẫu nhiên cùng loại, cùng có đơn vị đo.

2. Phương sai của hàm các biến ngẫu nhiên

Ta sẽ mở rộng khái niệm phương sai của một biến ngẫu nhiên cho hàm của biến ngẫu nhiên X. Giả sử $g = g(t)$ là hàm số xác định trên miền giá trị của biến ngẫu nhiên X, khi đó $g(X)$ cũng là một biến ngẫu nhiên và phương sai của nó sẽ được kí hiệu là $\sigma^2_{g(X)}$.

Định lý 4. Cho X là biến ngẫu nhiên với phân phối xác suất là $f(x)$. Phương sai của biến ngẫu nhiên $g(X)$ là:

$$\sigma^2_{g(X)} = E\left\{[g(X) - \mu_{g(X)}]^2\right\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$, nếu X là ***biến ngẫu nhiên rời rạc***,

và

$$\sigma^2_{g(X)} = E\left\{[g(X) - \mu_{g(X)}]^2\right\} = \int_{-\infty}^{+\infty} [g(x) - \mu_{g(X)}]^2 f(x)dx$$, nếu X là ***biến ngẫu nhiên liên tục***.

Ví dụ 10. Tính phương sai của biến ngẫu nhiên $g(X) = 2X + 3$, trong đó X là biến ngẫu nhiên với phân phối xác suất như sau

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Giải: + Tính kỳ vọng của biến ngẫu nhiên $g(X) = 2X + 3$ theo Định lý 1:

$$\mu_{2X+3} = E(2X+3) = \sum_{x=0}^{3} (2x+3)f(x) = 6.$$

+ Sử dụng Định lí 4, ta có:

$$\sigma^2_{2X+3} = E\left\{[(2X+3) - \mu_{2X+3}]^2\right\} = E\left\{[(2X+3) - 6]^2\right\} = E\left\{[2X-3]^2\right\}$$

$$= E(4X^2 - 12X + 9) = \sum_{x=0}^{3} (4x^2 - 12x + 9)f(x) = 4.$$

Về nhà:

Tự đọc các Mục **4.4**

Bài tập: Tr. **107, 117**

Đọc trước các Mục 4.2(phần lien quan Covariance), 4,3 chuẩn bị cho **Bài số 6:**

Covariance và hệ số tương quan