BIẾN NGẪU NHIÊN

Nguyễn Tiến Đạt

Bộ môn Xác suất-Thống kê

Khoa Toán-Tin học, trường ĐH Khoa học Tự nhiên, ĐHQG TPHCM

ndat@hcmus.edu.vn

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Outline

Giới thiệu

Thông thường, ta tóm tắt các kết quả từ một phép thử ngẫu nhiên bằng những giá trị số (chẳng hạn: số lượng khách hàng, số linh kiện bị lỗi trong một lô hàng, lượng mưa trong tháng, ...).

Như đã giới thiệu, không gian mẫu mô tả tất cả các kết quả có thể của phép thử. Hơn nữa, trong nhiều trường hợp, sẽ rất hữu ích khi ta "gán" một giá trị số cho mỗi kết quả trong Ω.

Vì tính ngẫu nhiên, ta không biết trước một kết quả cụ thể của phép thử có xảy ra hay không, do đó, giá trị số tương ứng được gán cũng không biết trước.

Vì vậy, "biến" (*variable*) gán giá trị số cho kết quả của phép thử, được xem là *biến ngẫu nhiên*.

Biến ngẫu nhiên - Định nghĩa

Định nghĩa 1.1

Biến ngẫu nhiên (*random variable*) X là một ánh xạ (*map*) từ không gian mẫu Ω vào $\mathbb{R}$,

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega). \end{aligned}$$

Thông thường, ta dùng

- các chữ cái in hoa $X, Y, Z, \ldots$ để ký hiệu các biến ngẫu nhiên;
- các chữ in thường $x, y, z, \ldots$ chỉ các giá trị của biến ngẫu nhiên.

Ví dụ 1.2

Xét phép thử tung hai đồng tiền xu. Khi đó, không gian mẫu tương ứng của phép thử này là

$$\Omega = \{SS, SN, NS, NN\}.$$

Gọi X là số đồng xu xuất hiện mặt "Ngửa". Khi đó, X là một ánh xạ từ không gian mẫu Ω vào $\mathbb{R}$ được xác định như sau:

ω	SS	SN	NS	NN
$X(\omega)$	0	1	1	2

Outline

Biến ngẫu nhiên rời rạc

Dựa theo miền giá trị của biến ngẫu nhiên (tức là $X(\Omega)$) mà ta có thể phân thành hai loại chính như sau:

Định nghĩa 2.1 (Biến ngẫu nhiên rời rạc - Discrete random variables)

Biến ngẫu nhiên X được gọi là ***rời rạc***, nếu miền giá trị tương ứng $X(\Omega)$ là một tập hữu hạn (*finite*) hoặc vô hạn đếm được (*countably infinite*).

Ví dụ 2.2

(a) Gọi X là biến ngẫu nhiên thể hiện số chấm xuất hiện của phép thử tung 1 hạt xúc sắc. Khi đó, X là một biến ngẫu nhiên rời rạc với miền giá trị của X là $X(\Omega) = \{1; 2; 3; 4; 5; 6\}$.

(b) Gọi Y là b.n.n. thể hiện số khách hàng đến một cửa hàng mỗi tháng. Khi đó, Y là b.n.n. rời rạc với miền giá trị của Y là $Y(\Omega) = \mathbb{N} = \{0; 1; 2; 3; 4; 5; ...\}$.

Biến ngẫu nhiên liên tục

Định nghĩa 2.3 (Biến ngẫu nhiên liên tục - Continuous random variables)

Biến ngẫu nhiên X được gọi là ***liên tục***, nếu miền giá trị tương ứng $X(\Omega)$ là một khoảng dạng $(a, b) \subset \mathbb{R}$ hoặc toàn bộ $\mathbb{R}$.

Nhắc lại: với $a, b \in \mathbb{R}, a < b$, một khoảng là tập con của $\mathbb{R}$ có thể có một trong các dạng sau: (a, b); $[a, b]$; $[a, b)$; $(a, b]$; $(-\infty; b)$; $(-\infty; b]$; $(a; +\infty)$ hoặc $[a; +\infty)$.

Ví dụ 2.4

Trong các thí dụ sau đây, X là biến ngẫu nhiên liên tục:

(a) Chọn ngẫu nhiên một thời điểm trong ngày và xét b.n.n. X thể hiện nhiệt độ không khí tại thời điểm đó.

(b) Chọn ngẫu nhiên một bóng đèn điện và xét b.n.n. X đo thời gian hoạt động bình thường của nó.

Outline

Phân phối xác suất của biến ngẫu nhiên

Biến cố ngẫu nhiên (*event*, xem lại bài giảng chương 1 "Khái niệm cơ bản của Xác suất") tương ứng với biến ngẫu nhiên.

Ký hiệu

Cho $D \subset \mathbb{R}$, ta ký hiệu biến cố:

$$(X \in D) := \{\omega \in \Omega : X(\omega) \in D\} \subset \Omega.$$

Chẳng hạn, với $a \in \mathbb{R}$, ta viết các biến cố:

$$(X = a) = \{\omega \in \Omega : X(\omega) = a\} \subset \Omega,$$

$$(X \leq a) = \{\omega \in \Omega : X(\omega) \leq a\} \subset \Omega.$$

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Phân phối xác suất của biến ngẫu nhiên rời rạc

Bảng phân phối xác suất

Định nghĩa 3.1

Để mô tả biến ngẫu nhiên rời rạc X nhận giá trị nào đó với xác suất tương ứng là bao nhiêu thì ta dùng bảng phân phối xác suất.

Bảng phân phối xác suất này có hai dòng như sau:

- dòng thứ nhất là các giá trị mà biến ngẫu nhiên X nhận.
- dòng thứ hai là xác suất để biến ngẫu nhiên X nhận các giá trị tương ứng đó.

X	x_1	x_2	...	x_n	...
$\mathbb{P}$	$\mathbb{P}(X = x_1)$	$\mathbb{P}(X = x_2)$	...	$\mathbb{P}(X = x_n)$	...

Phân phối xác suất của b.n.n. rời rạc (tt)

Hàm trọng lượng xác suất (probability mass function)

Giả sử biến ngẫu nhiên rời rạc X nhận các giá trị $x_1, x_2, ..., x_n, ...$, hay ta có $X(\Omega) = \{x_1; x_2; ...; x_n; ...\}$.

Định nghĩa 3.2

Hàm số f gọi là hàm trọng lượng xác suất (*probability mass function*), được định nghĩa như sau, với mỗi $x_j \in X(\Omega)$:

$$f(x_j) := \mathbb{P}(X = x_j).$$

Có thể dùng ký hiệu f hoặc f_X cho hàm trọng lượng xác suất.
Ta có một số tính chất:

- $0 \leq f(x_j) \leq 1, \quad \forall x_j \in X(\Omega)$;
- $\sum_{x_j \in X(\Omega)} f(x_j) = 1$;
- $f(x) := 0$ nếu $x \notin X(\Omega)$.

Bảng phân phối xác suất của b.n.n. rời rạc - thí dụ

Ví dụ:

Gọi X là số mặt Sấp xuất hiện khi thực hiện phép thử tung hai đồng xu cân đối, đồng chất.

Ta có, không gian mẫu: $\Omega = \{SS; SN; NS; NN\}$.

Khi đó, miền giá trị của X là: $X(\Omega) = \{0; 1; 2\}$.

Xác suất tương với các giá trị X là: $\mathbb{P}(X = 0) = \mathbb{P}(\{NN\}) = \dfrac{1}{4}$;

$$\mathbb{P}(X = 1) = \mathbb{P}(\{SN; NS\}) = \frac{2}{4}; \quad \text{và} \quad \mathbb{P}(X = 2) = \mathbb{P}(\{SS\}) = \frac{1}{4}.$$

Do đó, bảng phân phối xác suất của b.n.n. X là

X	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$
$\mathbb{P}(X = x_j),\ j \in \{1; 2; 3\}$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Phân phối xác suất biến ngẫu nhiên liên tục

Hàm mật độ xác suất: Định nghĩa

Định nghĩa 3.3

Cho biến ngẫu nhiên liên tục X. Hàm số $f : \mathbb{R} \to [0, +\infty)$, thỏa:

$$(i) \qquad \mathbb{P}(X \in I) = \int_I f(x)dx, \quad \forall I \subset \mathbb{R},$$

$$(ii) \qquad \int_{\mathbb{R}} f(x)dx = \int_{-\infty}^{+\infty} f(x)dx = 1,$$

được gọi là **hàm mật độ xác suất** (*probability density function*) của biến ngẫu nhiên X.
Có thể dùng ký hiệu f hoặc f_X cho hàm mật độ xác suất của X.

Chẳng hạn, ở tính chất (i), nếu $I = (a, b) \subset \mathbb{R}$, với $a < b$ thì ta có

$$\mathbb{P}(X \in (a, b)) = \mathbb{P}(a < X < b) = \int_a^b f(x)dx.$$

Hàm mật độ xác suất (tt)

Nhận xét 3.4

1) Mọi hàm số f không âm và thỏa điều kiện $\int_{\mathbb{R}} f(x)dx = 1$, đều là hàm mật độ xác suất của một biến ngẫu nhiên X nào đó, [1].

2) Trong trường hợp X là biến ngẫu nhiên liên tục, ta có $\mathbb{P}(X = a) = \int_a^a f(u)du = 0$, với $a \in \mathbb{R}$.

[1] Reference: e.g., Theorem 11.3, *Probability Essentials*, J.Jacod and P.E. Protter, Springer, 2nd Ed. 2004.

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Hàm phân phối xác suất: định nghĩa

Định nghĩa 3.5

Cho biến ngẫu nhiên X (rời rạc hoặc liên tục), hàm số thực

$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F(x) := \mathbb{P}(X \leq x) \end{aligned}$$

được gọi là **hàm phân phối xác suất** (*distribution function*) của X.

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

Ngoài ra, về mặt ký hiệu, ta có thể sử dụng: F_X hoặc F.

Hàm phân phối xác suất: tính chất

Mệnh đề 3.6

Hàm phân phối xác suất F của b.n.n. X có các tính chất sau:

(i) không giảm (*non-decreasing*): $x, y \in \mathbb{R}$ với $x \leq y \;\Rightarrow\; F(x) \leq F(y)$;

(ii) liên tục phải (*right-continuous*): $\lim_{x \to x_0^+} F(x) = F(x_0)$, $\forall x_0 \in \mathbb{R}$;

(iii) $\lim_{x \to -\infty} F(x) = 0$ và $\lim_{x \to +\infty} F(x) = 1$;

(iv) $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a)$,
với $a < b$ bất kỳ trong $\mathbb{R}$.

Hàm phân phối xác suất biến ngẫu nhiên rời rạc

Giả sử biến ngẫu nhiên rời rạc X có bảng phân phối xác suất

X	x_1	x_2	x_3	...	x_k	x_{k+1}	...
$\mathbb{P}$	p_1	p_2	p_3	...	p_k	p_{k+1}	...

trong đó $x_1 < x_2 < ... < x_k < ...$ và $p_j := \mathbb{P}(X = x_j)$. Ta có:

- Nếu $x < x_1$ thì $\{X \leq x\} = \emptyset$ và do đó

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\emptyset) = 0;$$

- Nếu $x_k \leq x < x_{k+1}$ thì,

$$\begin{aligned}\{X \leq x\} &= \{X \in \{x_1; x_2; ...; x_k\}\} \\ &= \{X = x_1\} \cup \{X = x_2\} \cup ... \cup \{X = x_k\};\end{aligned}$$

  Mặt khác, do các biến cố $\{X = x_1\}, \{X = x_2\}, ...$ và $\{X = x_k\}$ xung khắc nhau từng đôi một nên

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + ... + \mathbb{P}(X = x_k).$$

Hàm phân phối xác suất biến ngẫu nhiên rời rạc (tt)

Ta có, nếu $x_k \leq x < x_{k+1}$, thì:

$$\begin{aligned} F(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + ... + \mathbb{P}(X = x_k) \\ &= p_1 + p_2 + ... + p_k. \end{aligned}$$

Vậy, với $x \in \mathbb{R}$, hàm phân phối xác suất của b.n.n. rời rạc X là:

$$F(x) = \begin{cases} 0 & , \text{ nếu } x < x_1, \\ p_1 & , \text{ nếu } x_1 \leq x < x_2, \\ p_1 + p_2 & , \text{ nếu } x_2 \leq x < x_3, \\ ... & \\ p_1 + p_2 + ... + p_k & , \text{ nếu } x_k \leq x < x_{k+1}, \\ ... & \end{cases}$$

Khi X là b.n.n. rời rạc, cần lưu ý sự khác biệt có thể có giữa các biến cố

$$(X \leq x) \quad \text{và} \quad (X < x), \quad \text{với } x \in \mathbb{R}.$$

Hàm phân phối xác suất b.n.n. rời rạc - thí dụ 1

Ví dụ 3.7

Tung một đồng tiền xu cân đối và đồng chất. Gọi X là số lần mặt "Sấp" xuất hiện. Hãy lập bảng phân phối xác suất của X và xác định hàm phân phối tương ứng của X.

Gợi ý: Bảng phân phối xác suất của X là

X	0	1
$\mathbb{P}$	0.5	0.5

Hàm phân phối xác suất của X là:

$$F(x) = \begin{cases} 0 & , \text{ nếu } x < 0, \\ 0.5 & , \text{ nếu } 0 \leq x < 1, \\ 1 & , \text{ nếu } x \geq 1. \end{cases}$$

Hàm phân phối xác suất b.n.n. rời rạc - một vài thí dụ khác

Ví dụ 3.8

Gọi X là số chấm xuất hiện khi tung một hạt xúc sắc cân đối và đồng chất. Hãy lập bảng phân phối và xác định hàm phân phối xác suất của X.

Ví dụ 3.9

Tung đồng thời hai đồng tiền xu cân đối và đồng chất. Gọi Y là số mặt Sấp xuất hiện khi thực hiện phép thử. Hãy lập bảng phân phối xác suất và xác định hàm phân phối xác suất của Y.

Ví dụ 3.10

Một lô linh kiện điện tử có 450 sản phẩm, trong đó có 15 sản phẩm kém chất lượng. Chọn ngẫu nhiên lần lượt 2 sản phẩm không hoàn lại. Gọi Z là số sản phẩm chất lượng kém trong 2 sản phẩm được chọn. Hãy lập bảng phân phối xác suất và xác định hàm phân phối của Z.

Hàm phân phối xác suất biến ngẫu nhiên liên tục

Xét f là hàm mật độ xác suất của biến ngẫu nhiên liên tục X.

1) Từ định nghĩa về hàm mật độ xác suất, ta có với $x \in \mathbb{R}$ hàm phân phối xác suất của X là:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

2) Nếu F khả vi tại $y \in \mathbb{R}$:

$$F'(y) = \frac{d}{dx}F(y) = f(y);$$

3) Với $a, b \in \mathbb{R}$ bất kỳ, $a < b$, ta có

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b).$$

Hàm phân phối xác suất b.n.n. liên tục - thí dụ 1

Ví dụ 3.11

Xét hàm số:

$$f(x) = \begin{cases} 2x, & \text{nếu } x \in [0,1], \\ 0, & \text{nếu } x \in \mathbb{R}\backslash[0,1]. \end{cases}$$

a) Chứng tỏ rằng $f(x)$ là hàm mật độ xác suất của b.n.n. X nào đó.

b) Tìm hàm phân phối xác suất F của X.

c) Tính xác suất $\mathbb{P}\left(0 < X \leq \frac{1}{2}\right)$.

Hàm phân phối xác suất b.n.n. liên tục - thí dụ 2

Ví dụ 3.12

Giả sử tuổi thọ của một thiết bị (đơn vị: giờ) được thể hiện bởi biến ngẫu nhiên Y có hàm mật độ xác suất có dạng , với $a \in \mathbb{R}$ (giả sử a thỏa mãn f là hàm mật độ xác suất):

$$f(x) = \begin{cases} \frac{a}{x^2}, & \text{nếu } x \geq 100, \\ 0, & \text{nếu } x < 100. \end{cases}$$

a) Với giá trị nào của $a \in \mathbb{R}$ để f là hàm mật độ xác suất?
b) Hãy xác định hàm phân phối của Y.
c) Thiết bị được gọi là loại tốt nếu tuổi thọ của nó kéo dài ít nhất 400 giờ. Tính tỉ lệ (hay xác suất) để thiết bị là loại tốt.

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Outline

Các đặc trưng của biến ngẫu nhiên: Kỳ vọng (trung bình)

Định nghĩa 4.1

Kỳ vọng (hay trung bình) của biến ngẫu nhiên X (*the Expectation of* X), ký hiệu $\mathbb{E}(X)$ (hoặc μ), là một số thực được định nghĩa

- nếu X là biến ngẫu nhiên **rời rạc**:

$$\mathbb{E}(X) \equiv \mu := \sum_{x \in X(\Omega)} x.\mathbb{P}(X = x); \qquad (1)$$

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

$$\mathbb{E}(X) \equiv \mu := \int_{-\infty}^{+\infty} x.f(x)dx, \qquad (2)$$

trong đó f là hàm mật độ xác suất của X nếu X là b.n.n liên tục.

Remark: $\mathbb{E}(X)$ còn được gọi là trung bình (*mean*) của b.n.n. X.

Kỳ vọng của biến ngẫu nhiên rời rạc - Thí dụ

Ví dụ 4.2

Một hộp chứa 10 viên bi, trong đó có 3 viên bi nặng 10g, 5 viên bi nặng 50g và 2 viên bi nặng 20g. Chọn ngẫu nhiên ra 1 viên bi và gọi X là khối lượng (theo gram) của viên bi đó. Hãy tính $\mathbb{E}(X)$.

Ví dụ 4.3

Một chùm chìa khóa có 6 chìa, trong đó có 2 chìa mở được cửa. Thử lần lượt từng chìa (thử xong bỏ ra ngoài) cho đến khi mở được cửa. Hãy tính số lần thử trung bình để mở được cửa.

Kỳ vọng của biến ngẫu nhiên rời rạc - Thí dụ

Ví dụ 4.2

Một hộp chứa 10 viên bi, trong đó có 3 viên bi nặng 10g, 5 viên bi nặng 50g và 2 viên bi nặng 20g. Chọn ngẫu nhiên ra 1 viên bi và gọi X là khối lượng (theo gram) của viên bi đó. Hãy tính $\mathbb{E}(X)$.

Ví dụ 4.3

Một chùm chìa khóa có 6 chìa, trong đó có 2 chìa mở được cửa. Thử lần lượt từng chìa (thử xong bỏ ra ngoài) cho đến khi mở được cửa. Hãy tính số lần thử trung bình để mở được cửa.

Gợi ý: Gọi Y là số lần thử chìa khóa để mở được cửa. Trước tiên, ta tính được

$$\mathbb{P}(Y=1)=\frac{2}{6};\quad \mathbb{P}(Y=2)=\frac{4}{6}.\frac{2}{5};\quad \mathbb{P}(Y=3)=\frac{4}{6}.\frac{3}{5}.\frac{2}{4};\quad \mathbb{P}(Y=4)=\frac{4}{6}.\frac{3}{5}.\frac{2}{4}.\frac{2}{3};$$
$$\mathbb{P}(Y=5)=\frac{4}{6}.\frac{3}{5}.\frac{2}{4}.\frac{1}{3};$$

Kỳ vọng của biến ngẫu nhiên liên tục - Thí dụ

Ví dụ 4.4

Cho X là một biến ngẫu nhiên có hàm mật độ xác suất

$$f(x) = \begin{cases} 2x, & \text{nếu } x \in [0,1], \\ 0, & \text{nếu } x \in \mathbb{R}\backslash[0,1]. \end{cases}$$

Hãy tính kỳ vọng $\mathbb{E}(X)$.

Ví dụ 4.5

Cho Y là một biến ngẫu nhiên có hàm mật độ xác suất

$$g(y) = \begin{cases} \frac{2}{y^2}, & \text{nếu } y \in [1,2], \\ 0, & \text{nếu } y \in \mathbb{R}\backslash[1,2]. \end{cases}$$

Hãy tính kỳ vọng $\mathbb{E}(Y)$.

Tính chất của kỳ vọng

Một vài tính chất của kỳ vọng

Cho X, Y là hai biến ngẫu nhiên và các hằng số $c_1, c_2 \in \mathbb{R}$ bất kỳ, thì kỳ vọng của biến ngẫu nhiên có các tính chất sau:

(i) $\mathbb{E}(c_1) = c_1$;

(ii) $\mathbb{E}(c_1.X) = c_1.\mathbb{E}(X)$;

(iii) $\mathbb{E}(c_1.X + c_2.Y) = c_1.\mathbb{E}(X) + c_2.\mathbb{E}(Y)$;

(iv) Nếu hai b.n.n. X và Y **"độc lập"**, thì $\mathbb{E}(X.Y) = \mathbb{E}(X).\mathbb{E}(Y)$.

(Định nghĩa thế nào là hai b.n.n. độc lập với nhau sẽ được giới thiệu sau.)

Ý nghĩa của kỳ vọng:

- Kỳ vọng là giá trị **trung bình theo xác suất** của tất cả các giá trị có thể có của biến ngẫu nhiên.
- Kỳ vọng phản ánh giá trị trung tâm của phân phối xác suất.

Tính chất của kỳ vọng (tt)

Tính kỳ vọng của $h(X)$

- Nếu X là biến ngẫu nhiên **rời rạc**, với hàm số h bất kỳ, ta có

$$\mathbb{E}\big(h(X)\big) := \sum_{x \in X(\Omega)} h(x).\mathbb{P}(X = x).$$

- Nếu X là biến ngẫu nhiên **liên tục** có hàm mật độ xác suất f. Khi đó, nếu h là hàm số thỏa $\mathbb{E}\big(|h(X)|\big) < \infty$ hoặc $h(y) \geq 0, \forall y \in \mathbb{R}$, thì ta có, [2]

$$\mathbb{E}\big(h(X)\big) = \int_{\mathbb{R}} h(x).f(x)dx.$$

[2] Reference: e.g., Corollary 9.1, *Probability Essentials*, J.Jacod and P.E. Protter, Springer, 2nd Ed. 2004.

Tính chất của kỳ vòng - thí dụ

Ví dụ 4.6

Tung đồng thời hai cặp đồng xu cân đối, đồng chất. Gọi X là số mặt Sấp xuất hiện ở cặp thứ nhất và Y là số mặt Sấp xuất hiện ở cặp thứ hai. Tính $\mathbb{E}(X), \mathbb{E}(Y)$ và $\mathbb{E}(XY)$.

Gợi ý: Kỳ vọng của X được tính như sau

$$\mathbb{E}(X) = 0.\mathbb{P}(X = 0) + 1.\mathbb{P}(X = 1) + 2.\mathbb{P}(X = 2) = 0.\frac{1}{4} + 1.\frac{2}{4} + 2.\frac{1}{4} = 1;$$

Tương tự, $\mathbb{E}(Y) = 1$.
Đặt biến ngẫu nhiên $W = X.Y$, ta có bảng phân phối xác suất của W

W	0	1	2	4
$\mathbb{P}$	$\frac{1}{4} + \frac{1}{4}$	$\frac{2}{4}.\frac{2}{4}$	$\frac{1}{4}.\frac{2}{4} + \frac{2}{4}.\frac{1}{4}$	$\frac{1}{4}.\frac{1}{4}$

Do đó, ta có

$$\mathbb{E}(XY) = 0.(\frac{1}{4} + \frac{1}{4}) + 1.(\frac{2}{4}.\frac{2}{4}) + 2.(\frac{1}{4}.\frac{2}{4} + \frac{2}{4}.\frac{1}{4}) + 4.(\frac{1}{4}.\frac{1}{4}) = 1 = \mathbb{E}(X).\mathbb{E}(Y).$$

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Các đặc trưng của biến ngẫu nhiên: Phương sai

Định nghĩa 4.7

Nếu biến ngẫu nhiên X có kỳ vọng $\mathbb{E}(X)$ thì **phương sai của X** (*Variance of* X), ký hiệu $\mathrm{Var}(X)$ (hoặc σ^2), được định nghĩa:

$$\mathrm{Var}(X) \equiv \sigma^2 := \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^2\right]. \tag{3}$$

Remark

- Trong tính toán, để tính phương sai của b.n.n. X, ta thường sử dụng công thức:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2.$$

- Đôi khi, trong một số giáo trình, người ta còn dùng ký hiệu $D(X)$ để chỉ phương sai của X.

Độ lệch chuẩn và một số tính chất của phương sai

Định nghĩa 4.8 (Độ lệch chuẩn - Standard Deviation)

Độ lệch chuẩn của biến ngẫu nhiên X, ký hiệu $\sigma(X)$, là căn bậc hai của $\mathrm{Var}(X)$, nghĩa là:

$$\sigma(X) := \sqrt{\mathrm{Var}(X)}.$$

Một vài tính chất của phương sai

Cho hai biến ngẫu nhiên X, Y và hằng số $c \in \mathbb{R}$, ta có:

(i) $\mathrm{Var}(c) = 0$;

(ii) $\mathrm{Var}(c.X) = c^2.\mathrm{Var}(X)$;

(iii) Nếu b.n.n. X và Y **độc lập** với nhau, thì

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

(Định nghĩa thế nào là hai b.n.n. độc lập với nhau sẽ được giới thiệu sau.)

Phương sai của b.n.n. - Thí dụ 1

Ví dụ 4.2 (tt)

Một hộp chứa 10 viên bi, trong đó có 3 viên bi nặng 10g, 5 viên bi nặng 50g và 2 viên bi nặng 20g. Từ hộp bi đó, chọn ngẫu nhiên 1 viên bi và gọi X là khối lượng của viên bi đó. Hãy tính $\mathbb{E}(X)$ và $\mathrm{Var}(X)$.

Ví dụ 4.5 (tt)

Cho biến ngẫu nhiên Y có hàm mật độ xác suất

$$g(y) = \begin{cases} \frac{2}{y^2} & , \text{ nếu } y \in [1,2], \\ 0 & , \text{ nếu } y \in \mathbb{R} \backslash [1,2], \end{cases}$$

Hãy tính $\mathbb{E}(Y)$ và $\mathrm{Var}(Y)$.

Phương sai của b.n.n. - Thí dụ 2

Ví dụ 4.9

Cho hàm số f được xác định như sau:

$$f(u) = \begin{cases} 24.e^{-24.(u-10)} & , \text{ nếu } u \geq 10, \\ 0 & , \text{ nơi khác .} \end{cases}$$

(a) Chứng minh f là hàm mật độ xác suất của biến ngẫu nhiên W nào đó.

(b) Tính $\mathbb{P}(-5 \leq W < 16)$.

(c) Tính $\mathbb{E}(W)$ và $\mathrm{Var}(W)$.

Ý nghĩa của Phương sai của biến ngẫu nhiên

- Phương sai (Variance) là kỳ vọng của bình phương các sai lệch giữa X và $\mathbb{E}(X)$.

  Nói cách khác, phương sai là trung bình bình phương sai lệch.

  Nó phản ánh mức độ phân tán các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình.

- Trong công nghiệp, phương sai biểu thị độ chính xác trong sản xuất. Trong canh tác, phương sai biểu thị mức độ ổn định của năng suất...

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Đặc trưng của biến ngẫu nhiên: Mode

Định nghĩa 4.10 (Mode của biến ngẫu nhiên)

Mode của biến ngẫu nhiên X, ký hiệu $\mathrm{Mode}(X)$, là giá trị mà X nhận được với xác suất lớn nhất.

- Nếu X là b.n.n. rời rạc có bảng phân phối xác suất

X	x_1	x_2	...	x_n	...
$\mathbb{P}$	p_1	p_2	...	p_n	...

thì

$$\mathrm{Mode}(X) = x_i \quad \Leftrightarrow \quad p_i = \mathbb{P}(X = x_i) = \max\{p_1; p_2; \ldots\}.$$

- Nếu X có phân phối liên tục với hàm mật độ xác suất f, thì

$$\mathrm{Mode}(X) = x_0 \quad \Leftrightarrow \quad f(x_0) = \sup_{x \in \mathbb{R}} f(x).$$

Mode của biến ngẫu nhiên - Thí dụ

Ví dụ 4.11 (Trường hợp b.n.n. rời rạc)

Tìm Mode của b.n.n. X có phân phối rời rạc với bảng phân phối xác suất sau:

X	1	2	3	4	5
$\mathbb{P}$	0.3	0.25	0.18	0.14	0.13

Ví dụ 4.12 (Trường hợp b.n.n. liên tục)

Tìm Mode của b.n.n. X có hàm mật độ xác suất

$$f(x) = \begin{cases} \frac{3}{4}.x.(2-x) & , \text{ nếu } x \in [0,2], \\ 0 & , \text{ nếu } x \in \mathbb{R} \backslash [0,2]. \end{cases}$$

Outline

1. Định nghĩa biến ngẫu nhiên
2. Phân loại biến ngẫu nhiên
3. Phân phối xác suất
   - Bảng phân phối xác suất
   - Hàm mật độ xác suất (probability density functions)
   - Hàm phân phối xác suất (distribution functions)
4. Các tham số đặc trưng
   - Kỳ vọng (trung bình) - (Expectation)
   - Phương sai (Variance)
     - Độ lệch chuẩn (Standard deviation) và một vài tính chất của Phương sai
   - Mode
   - Trung vị (Median)

Đặc trưng của biến ngẫu nhiên: Trung vị (Median)

Định nghĩa 4.13

Cho biến ngẫu nhiên X bất kỳ, **trung vị** của X (*median of* X), ký hiệu $\mathrm{Med}(X)$, là giá trị m của biến ngẫu nhiên X sao cho

$$\mathbb{P}(X \leq m) \geq \frac{1}{2},$$

và

$$\mathbb{P}(X \geq m) \geq \frac{1}{2}.$$

Ta viết $\mathrm{Med}(X) = m$.

Đặc trưng của biến ngẫu nhiên: Trung vị

Nhận xét 4.14

Khi X là biến ngẫu nhiên liên tục thì trung vị của X chính là điểm chia phân phối xác suất thành hai phần bằng nhau. Nghĩa là,

$$\mathbb{P}(X \geq m) = \mathbb{P}(X \leq m) = 1/2,$$

tương đương với $\mathbb{P}(X \geq m) = 1/2$ hay $\mathbb{P}(X \leq m) = 1/2$.

Đặc trưng của biến ngẫu nhiên: Trung vị - thí dụ 1

Ví dụ 4.15 (Trường hợp b.n.n. rời rạc)

Xét b.n.n. X có phân phối rời rạc với bảng phân phối xác suất sau:

X	1	2	3	4
$\mathbb{P}$	0.1	0.2	0.3	0.4

Tìm $\text{Med}(X)$.

Ví dụ 4.16 (Trường hợp b.n.n. rời rạc - Trung vị không duy nhất)

Xét b.n.n. Y có phân phối rời rạc với bảng phân phối xác suất sau:

Y	1	2	3	4
$\mathbb{P}$	0.1	0.4	0.3	0.2

Tìm $\text{Med}(Y)$.

Đặc trưng của biến ngẫu nhiên: Trung vị - thí dụ 2

Ví dụ 4.17 (Trường hợp b.n.n. liên tục)

Xét biến ngẫu nhiên liên tục X có hàm mật độ xác suất cho bởi

$$f(x) = \begin{cases} 4x^3, & \text{nếu } x \in (0,1), \\ 0, & \text{nếu } x \in \mathbb{R}\backslash(0,1). \end{cases}$$

Tìm $\mathrm{Med}(X)$.

Ví dụ 4.18 (Trường hợp b.n.n. liên tục - Trung vị không duy nhất)

Xét biến ngẫu nhiên liên tục Y có hàm mật độ xác suất cho bởi

$$g(y) = \begin{cases} \frac{1}{2}, & \text{nếu } y \in [0,1], \\ 1, & \text{nếu } y \in [\frac{5}{2}, 3], \\ 0, & \text{nếu } y \in \mathbb{R}\backslash\big([0,1] \cup [\frac{5}{2}, 3]\big). \end{cases}$$

Tìm $\mathrm{Med}(Y)$.