§3 CÁC SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

3.1 Kỳ vọng (*Expectation value*)

*** Định nghĩa.** *Kỳ vọng của* X, với $f(x)$ là hàm mật độ XS đã cho của X liên tục hoặc $p(x)$ là hàm xác suất của X rời rạc, được tính như sau:

$$EX = \sum_{\forall x} xp(x), X \text{ rời rạc;} \tag{1a}$$

hoặc
$$EX = \int_{-\infty}^{+\infty} xf(x)dx, X \text{ liên tục.} \tag{1b}$$

*** Tính chất**

(i) $E(c) = c, c = const$;

(ii) $E(cX) = cEX$;

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

(iv) Nếu X và Y độc lập $\Longrightarrow E(XY) = EX.EY$.

Chú ý: tính chất (ii) và (iii) được gọi là *tính tuyến tính* của phép toán, vốn quen thuộc trong phép tính đạo hàm hay tích phân.

Ý nghĩa thực tế: kỳ vọng là *giá trị trung bình* (*mean value*) của biến ngẫu nhiên và đóng vai trò *định vị* biến.

*** <u>Thí dụ 1</u>**. Cho biến ngẫu nhiên $X \sim \mathcal{B}(n, p)$. Tính kỳ vọng của X.

Giải. Sử dụng công thức Bernoulli ta có

$$EX = \sum_{\forall x} xp(x) = \sum_{x=0}^{n} xC_n^x p^x q^{n-x} = np.$$

Việc tính tổng ở trên không đơn giản. Ta có cách làm khác dễ hơn như sau: Do $X = X_1 + X_2 + \ldots\ X_n$, trong đó $X_i \sim \mathcal{B}(1, p)$ (*PPXS Bernoulli*) với $p(0) = P(X = 0) = q$ và $p(1) = p$ và bảng PPXS

$X = x$	0	1
$p(x)$	q	p

Dễ thấy $EX_i = p, \forall i \quad \Rightarrow \quad EX = \sum_{i=1}^{n} EX_i = np$ (dùng tính chất (iii) ở trên).

* **<u>Thí dụ 2</u>**. Cho biến ngẫu nhiên $X \sim \mathcal{E}(\lambda)$. Tính kỳ vọng của X.

Giải. $$EX = \int_0^{+\infty} x\,\lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

* **<u>Thí dụ 3</u>**. Một người mua 10000 đồng một số đề. Tính số tiền thắng trung bình trong lần chơi đó.

Giải. Gọi X là số tiền thắng trong lần chơi, rõ ràng

$X = x$	0	$700000đ$
$p(x)$	99%	1%

và số tiền thắng trung bình $EX = 7000$ đồng.

*** <u>Thí dụ 4</u>**. Một người hàng ngày đi bộ từ nhà đến nơi làm việc trên quãng đường dài 600 m với vận tốc đều $V\, m/s$. Biết thời gian đi bộ của người đó là một biến ngẫu nhiên có phân bố đều trong khoảng từ 6 phút đến 10 phút. Tìm kỳ vọng của X.

Giải. Gọi T là thời gian đi bộ ở trên, rõ ràng $T \sim \mathcal{U}(6; 10)$ và

$V = \dfrac{600}{60T} = \dfrac{10}{T}\,(m/s)$. Để ý $f_T(x) = \frac{1}{4}$, $6 < t < 10$, từ đó

$$EV = \frac{1}{4}\int_6^{10} \frac{10dt}{t} = \frac{5}{2}\ln\frac{10}{6} \approx 1{,}277\ (m/s).$$

3.2 Phương sai (*Variance value*)

*** Định nghĩa 1.** Phương sai của biến ngẫu nhiên X, ký hiệu là VX, được xác định như sau

$$VX = \mathrm{var}(X) = E[(X - EX)^2]. \qquad (2)$$

Nhận xét: $X - EX$ là độ lệch của X so với trung bình của nó

⇒ Phương sai – trung bình của bình phương độ lệch

⇒ Đặc trưng cho độ phân tán của X quanh trung bình

độ bất định; dung sai; độ rủi ro …

Chú ý: Công thức (2) có dạng tương đương

$$VX = E(X^2) - (EX)^2.$$

Cách tính:

- Với X rời rạc

$$VX = \sum_{\forall x}(x - EX)^2\, p(x) = \sum_{\forall x} x^2\, p(x) - (EX)^2;$$

- Với X liên tục

$$VX = \int_{-\infty}^{+\infty}(x - EX)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - (EX)^2.$$

*** Tính chất**

(i) $V(c) = 0$, $c = const$;

(ii) $V(cX) = c^2 VX$;

(iii) Nếu X và Y độc lập $\Longrightarrow V(X + Y) = VX + VY$.

*** <u>Thí dụ 5</u>**. Tính phương sai của biến ngẫu nhiên $X \sim \mathcal{B}(n, p)$ (xem thí dụ 1).

Giải: Ta đã có trong thí dụ 3 $EX = np$. Có thể tính trực tiếp

$$VX = \sum_{\forall x}(x - EX)^2\, p(x) = \sum_{x=0}^{n}(x - EX)^2\, C_n^x p^x q^{n-x}.$$

Tuy nhiên ta có thể dùng cách tiếp cận ở thí dụ 1. Từ ý nghĩa thực tế ta có các X_i độc lập và dễ thấy $E(X_i^{\,2}) = p$, $VX_i = pq$, nên

$$\Rightarrow VX = \sum_{i=1}^{n} VX_i = npq.$$

*** <u>Thí dụ 6</u>**. Tìm kỳ vọng của biến ngẫu nhiên X trong thí dụ 4.

Giải. Từ thí dụ 4 ta có

$$EV = \frac{1}{4} \int_6^{10} \frac{100\, dt}{t^2} = \frac{5}{3}$$

$$\Rightarrow \operatorname{var}(V) = E(V^2) - (EV)^2 \approx 0{,}0358.$$

Để ý VX là một số không âm, tuy nhiên về mặt vật lý nó không cùng thứ nguyên với X, vì vậy ta đưa ra khái niệm sau:

*** Định nghĩa 2.** *Độ lệch chuẩn* (*standard deviation*) của biến ngẫu nhiên X, ký hiệu là $\sigma(X)$, được xác định bằng

$$\sigma(X) = \sqrt{VX} \tag{3}$$

Từ (3) ta có thể ký hiệu phương sai là $\sigma^2(X)$ hoặc đơn giản σ^2.

Chú ý: $\sigma(cX) = |c|\, \sigma(X)$

và nếu X và Y độc lập thì $\sigma(X+Y) = \sqrt{\sigma^2(X) + \sigma^2(Y)}$.

Một hệ quả quan trọng của tính chất (iii) của VX: Nếu ta có n biến ngẫu nhiên độc lập $X_i, i = \overline{1;n}$ và $VX_i = \sigma^2\ \forall i = \overline{1;n}$, thì

$$V\bar{X} = V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

3.2 Một số đặc số khác

1. Mốt (*mode*)

Mốt là giá trị của biến ngẫu nhiên có khả năng xuất hiện lớn nhất (trong một lân cận nào đó).

- Biến rời rạc: mốt là giá trị có XS lớn nhất.
- Biến liên tục: giá trị làm hàm mật độ đạt *max*.

$\Rightarrow$ Biến ngẫu nhiên có thể có nhiều mốt.

*** <u>Thí dụ 7</u>**. Tìm mốt của biến $X \sim \mathcal{B}(n, p)$

Giải. Mốt của X là phần nguyên của số thực $(n+1)p$. Nếu số này là số nguyên thì ta có hai mốt là $(n+1)p$ và $(n+1)p - 1$. Trong ứng dụng mốt còn có tên gọi là *số lần xuất hiện chắc nhất* (của A trong lược đồ Bernoulli tương ứng).

*** <u>Thí dụ 8</u>**. Tìm mốt của biến X có *phân phối Weibull* với hàm mật độ

$$f(x) = \begin{cases} \frac{x}{2}\, e^{-x^2/4}; & x \geq 0 \\ 0; & x < 0 \end{cases}.$$

Giải. Mốt của X là nghiệm của phương trình

$$f'(x) = \frac{1}{2}\ e^{-x^2/4} - \frac{x^2}{4}\ e^{-x^2/4} = 0.$$

Từ đó mốt là nghiệm của $1 - x^2/2 = 0$, nhưng do $x > 0$ suy ra

$\mathrm{mod}(X) = \sqrt{2} \approx 1{,}414$.

2. Trung vị (*median*)

Nếu ký hiệu trung vị là med$X$, thì

$$P(X <\ medX) = P(X \geq\ medX) = 1/2.$$

Có nghĩa là để tìm trung vị ta phải phải giải phương trình

$$F(x) = 1/2.$$

Chú ý cả mốt và trung vị đều có ý nghĩa định vị biến, trong một số trường hợp còn hay hơn kỳ vọng. Chẳng hạn xét tập điểm số của bốn lần kiểm tra giữa kỳ {2, 2, 2, 10}. Trung bình ở đây bằng 4, có vẻ mốt cho đánh giá chính xác hơn?

*** <u>Thí dụ 9</u>**. Tìm trung vị của biến X có *phân phối Weibull*.

Giải. Rõ ràng trung vị là nghiệm của phương trình

$$\int_0^{\mathrm{med}X} f(x)dx = 0{,}5 \text{ hay } 1 - e^{-(\mathrm{med}X)^2/4} = 0{,}5;$$

$$\Rightarrow \quad \mathrm{med}X = 1{,}665.$$

Nói chung 3 đặc trưng kỳ vọng, mốt và trung vị không bằng nhau; ở đây với biến X có phân phối Weibull ta thấy $EX = 1{,}772$; mod$X$ $= 1{,}414$; med$X = 1{,}665$. Trong trường hợp phân phối đối xứng (và đồ thị mật độ chỉ có 1 đỉnh) thì cả ba đặc trưng trùng nhau.

3. Phân vị (percentile)

Ta gọi x_α là *phân vị* α của X, nếu $P(X < x_\alpha) = \alpha \Leftrightarrow F(x_\alpha) = \alpha$. Số thực α được gọi là bậc của phân vị. Trong xác suất người ta quan tâm nhiều đến các bậc phân vị 25%, 50% (trung vị) và 75% (còn gọi là các tứ phân vị); còn trong thống kê hay sử dụng các phân vị có bậc lớn hoặc bé (chẳng hạn 90%, 95%, 99%, ... hoặc 0,5%, 1%, 5%, ...).

*** Thí dụ 10**. Tìm các phân vị 25%, 50%, 95%, 99% của biến ngẫu nhiên $X \sim \mathcal{N}(0; 1)$.

Giải. Sử dụng bảng hàm Laplace ta sẽ có:

$F(x_{0,25}) = 0{,}25 \Rightarrow \phi(-x_{0,25}) = 0{,}25 \Rightarrow x_{0,25} = -0{,}675$;

$F(x_{0,5}) = 0{,}5 \Rightarrow x_{0,5} = 0 = \mathrm{mod}(X)$;

$F(x_{0,95}) = 0{,}95 \Rightarrow \phi(x_{0,95}) = 0{,}45 \Rightarrow x_{0,95} = 1{,}645$;

$F(x_{0,99}) = 0{,}99 \Rightarrow \phi(x_{0,99}) = 0{,}49 \Rightarrow x_{0,95} = 2{,}325.$

4. Mô men (*moment*)

Mô men cấp k đối với a của biến ngẫu nhiên X, ký hiệu $\nu_k(a)$ là một số thực được xác định như sau

$$\nu_k(a) = E[(X-a)^k].$$

Nếu $a = 0$, ký hiệu $\nu_k = \nu_k(0) = E(X^k)$, là *mô men gốc cấp k*.

Nếu $a = EX$, ký hiệu $\mu_k = \mu_k(EX) = E[(X-EX)^k]$, là *mô men trung tâm cấp k*.

$$\Rightarrow \quad EX = \nu_1; VX = \mu_2.$$

Có thể thấy quan hệ giữa hai loại mô men:

$$\mu_2 = \nu_2 - \nu_1^2 = \sigma^2,$$

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3,$$

$$\mu_4 = \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\,\nu_2 - 3\nu_1^4, \ldots$$

Một số mô men cho ta đặc trưng về hình dạng phân phối XS:

- *Hệ số bất đối xứng* (*skewness*) là tỷ số $\beta_1 = \mu_3/\sigma^3$; nếu $\beta_1 = 0$ ta có đường cong mật độ đối xứng, còn nếu nó khác không thì phân phối bất đối xứng.
- *Hệ số nhọn* (*kurtosis* hay *peakedness*) là tỷ số $\beta_2 = \mu_4/\sigma^4 - 3$;

nếu β_2 càng lớn thì đường cong mật độ có đỉnh càng nhọn;

đường cong mật độ chuẩn có $\beta_2 = 0$.

*** Thí dụ 11**. Cho biến ngẫu nhiên $X \sim \mathcal{E}(0{,}5)$. Tìm các hệ số bất đối xứng và hệ số nhọn.

Giải. Dễ dàng tìm được $EX = \nu_1 = 2$; $\sigma^2 = \mu_2 = 4$. Từ đó suy ra:

$$\beta_1 = \mu_3/\sigma^3 = 16/8 = 2;$$

$$\beta_2 = \mu_4/\sigma^4 - 3 = 208/16 - 3 = 10.$$

Chú ý khi tính các tích phân ta sử dụng công thức

$$\int_0^{\infty} x^n e^{-ax}\, dx = \frac{\Gamma(n+1)}{a^{n+1}} = \frac{n!}{a^{n+1}}.$$

*** Thí dụ 12**. Cho biến ngẫu nhiên X có hàm mật độ

$$f(x) = A e^{\frac{1}{4} + x - x^2}, x \in \mathbb{R}.$$

Tìm A, sau đó tính các mô men gốc cấp k, $k = 1, 2, 3, 4$.

Giải. Dùng tính chất (iv) của hàm mật độ, ta có

$$1 = A \int_{-\infty}^{+\infty} e^{\frac{1}{4} + x - x^2} dx = A \int_{-\infty}^{+\infty} e^{\frac{1}{2} - (x - \frac{1}{2})^2}\, dx$$

$$= A\sqrt{e\pi}\, \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2.(\frac{1}{2})}(x - \frac{1}{2})^2}\, dx = A\sqrt{e\pi}$$

$$\Rightarrow \qquad A = \frac{1}{\sqrt{e\pi}}.$$

Mặt khác, có thể thấy rằng $X = \sqrt{e}Z$, với $Z \sim \mathcal{N}\left(\frac{1}{2}, \frac{1}{2}\right)$, từ đó:

$$\nu_1 = \sqrt{e}/2; \nu_2 = \mu_2 + \nu_1^2 = 3e/4;$$

$$\nu_3 = 3\nu_2\nu_1 - 2\nu_1^3 = 7e\sqrt{e}/8;$$

$$\nu_4 = \mu_4 + 4\nu_1\nu_3 - 6\nu_1^2\,\nu_2 + 3\nu_1^4 = \frac{1}{8}(8 - 11e^2).$$

Chú ý: Đối với biến $X \sim \mathcal{N}(\mu, \sigma^2)$ ta có:

$$EX = \mu;\ VX = \sigma^2; \mathrm{mod}X = \mathrm{med}X = EX; \mu_3 = 0; \mu_4 = 3\sigma^4.$$

BÀI TẬP

1. Tìm kỳ vọng và phương sai của biến ngẫu nhiên X có:
   a) phân phối Poisson $\mathcal{P}(\lambda)$ với $\lambda = 4$;
   b) phân phối mũ $\mathcal{E}(\lambda)$ với với $\lambda = 0{,}4$.

2. Cho X và Y là hai biến ngẫu nhiên X độc lập.
   a) Giả sử $X \sim \mathcal{B}(1; 0{,}2)$ và $Y \sim \mathcal{B}(2; 0{,}2)$, lập bảng PPXS của $Z = X + Y$, sau đó tính EZ và VZ.
   b) Giả sử $X \sim \mathcal{B}(1; 0{,}5)$ và $Y \sim \mathcal{B}(2; 0{,}2)$, lập bảng PPXS của $Z = X + Y$; Z có phân phối nhị thức không?

3. Một người cho thuê 3 xe con. Anh ta hàng ngày phải nộp thuế 8$ cho 1 xe (dù xe có được thuê hay không), mỗi xe được cho thuê với giá 20$. Giả sử số yêu cầu thuê xe trong một ngày tuân theo luật Poisson với $\lambda = 2{,}8$.
   a) Tìm PPXS của số tiền anh ta thu được trong 1 ngày, sau đó tính số tiền trung bình thu được trong ngày.
   b) Giải bài toán trong trường hợp có 4 xe. So sánh nên có 3 hay 4 xe.

4. Gieo một đồng tiền cho đến khi xuất hiện mặt ngửa thì dừng. XS xuất hiện mặt ngửa là p. Gọi X là số lần gieo cần thiết.
   a) Tính trị trung bình X.
   b) Tìm PPXS của X với điều kiện trong n lần gieo đầu tiên chỉ có đúng 1 lần xuất hiện mặt ngửa.

5. Cho hàm mật độ của X
   $$f(x) = \frac{1}{\sqrt{2\pi}} x^{0,5} e^{-0,5x}, x > 0.$$
   a) Tìm trị trung bình, phương sai và mốt của X.
   b) Tính các hệ số β_1 và β_2.
   Gợi ý: Dùng $\int_0^{+\infty} x^{s-1}\, e^{-px} dx = \Gamma(s)/p^s$, $p>0$.

6. Một tổ có 5 công nhân với XS mắc bệnh nghề nghiệp A sau 10 năm làm việc là 0,3.
   a) Lập bảng PPXS của số người mắc bệnh nghề nghiệp A sau 10 năm công tác.
   b) Tìm số người chắc nhất mắc bệnh.

7. Một người đi làm từ nhà đến cơ quan phải qua 3 ngã tư, XS gặp đèn đỏ ở các ngã tư tương ứng là 0,2; 0,4 và 0,5. Gọi X là số đèn đỏ mà người đó gặp trong một lần đi làm
   a) Lập bảng PPXS và hàm phân phối của X. Tính EX, VX.
   b) Hỏi thời gian trung bình người đó phải dừng vì đèn đỏ là bao nhiêu, biết rằng mỗi khi gặp đèn đỏ phải đợi khoảng 3 phút?

8. Tuổi thọ của một loại côn trùng là biến ngẫu nhiên X (đơn vị tháng) với hàm mật độ
$$f(x) = \begin{cases} kx^2(4-x); & x \in [0;4]; \\ 0; & x \notin [0;4]. \end{cases}$$
   a) Tìm k và mod$X$.
   b) Tính XS để côn trùng chết trước khi nó tròn 1 tháng tuổi.

9. Cho biến ngẫu nhiên X có hàm mật độ
$$f(x) = \begin{cases} kx^2e^{-2x}; & x \geq 0; \\ 0; & x < 0. \end{cases}$$
   a) Tìm k và hàm phân phối của X.
   b) Tìm kỳ vọng, phương sai và mốt của $\boldsymbol{X}$.

10. Cho X là biến ngẫu nhiên có phân phối mũ với tham số $\lambda = 2$, tính kỳ vọng và độ lệch chuẩn của $Y = e^{-X}$.

11. Một nhà máy bán một loại sản phẩm với giá 1$/1 sản phẩm. Trọng lượng một sản phẩm là biến ngẫu nhiên có phân phối chuẩn với kỳ vọng μ *kg* và độ lệch chuẩn 1 *kg*. Giá thành một sản phẩm là $c = 0{,}05\mu + 0{,}3$. Nếu trọng lượng bé hơn 8 *kg* thì phải loại bỏ vì không bán được. Hãy xác định μ để lợi nhuận của nhà máy là lớn nhất.

12. Đối với luật phân phối đối xứng có thể lấy độ lệch trung bình θ xác định từ điều kiện
$$P(|X - EX| < \theta) = 1/2$$
làm thước đo độ phân tán của biến ngẫu nhiên. Hãy tìm phương sai và độ lệch trung bình của biến ngẫu nhiên X có hàm mật độ
$$f(x) = \begin{cases} \frac{1}{\pi\sqrt{a^2-x^2}}, & |x| \leq a; \\ 0, & |x| > a. \end{cases}$$