THỐNG KÊ MÁY TÍNH & ỨNG DỤNG
Bài 4
KÌ VỌNG VÀ PHƯƠNG SAI

Vũ Quốc Hoàng
(vqhoang@fit.hcmus.edu.vn)
FIT-HCMUS, 2018

Nội dung

- Giới thiệu kì vọng
- Kì vọng của biến ngẫu nhiên rời rạc
- Kì vọng của biến ngẫu nhiên liên tục
- Biến ngẫu nhiên phái sinh và kì vọng
- Các tính chất của kì vọng
- Phương sai
- Các tính chất của phương sai

Giới thiệu kì vọng
Ngôn ngữ đời thường

- Một lớp học gồm 20 SV có điểm môn TKMT&UD như sau

Điểm	4	6	7	8	9	10
Số SV	4	5	5	3	2	1

- Hỏi: điểm trung bình môn TKMT&UD của lớp là bao nhiêu?
- Trả lời: điểm trung bình là

$$\frac{4 \times 4 + 6 \times 5 + 7 \times 5 + 8 \times 3 + 9 \times 2 + 10 \times 1}{20} = 6.65$$

Giới thiệu kì vọng
Ngôn ngữ xác suất

- Chọn ngẫu nhiên một SV trong lớp, khảo sát X là "điểm môn TKMT&UD". Ta có X là b.n.n rời rạc với hàm xác suất

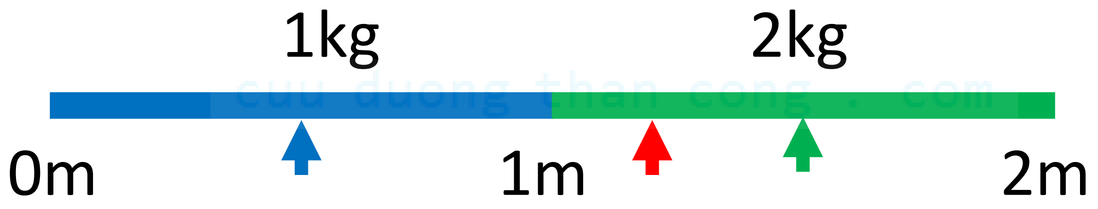
x	4	6	7	8	9	10
f(*x*)	4/20	5/20	5/20	3/20	2/20	1/20

- Hỏi: kì vọng của X là bao nhiêu?
- Trả lời: kì vọng của X là

$$4 \times \frac{4}{20} + 6 \times \frac{5}{20} + 7 \times \frac{5}{20} + 8 \times \frac{3}{20} + 9 \times \frac{2}{20} + 10 \times \frac{1}{20} = 6.65$$

Giới thiệu kì vọng
Ngôn ngữ đời thường

- Một hệ gồm 2 thanh đồng chất được hàn dính với nhau như hình sau. Thanh thứ nhất dài 1m, nặng 1kg. Thanh thứ hai dài 1m, nặng 2kg.



- Hỏi: điểm cân bằng của hệ là vị trí nào?
- Trả lời:
  - Điểm cân bằng của thanh thứ nhất ở vị trí 0.5m, của thanh thứ hai ở vị trí 1.5m.
  - Theo "qui tắc đòn bẩy" ta có: $1 \times l = 2 \times (1 - l) \Rightarrow \frac{1}{1-l} = \frac{2}{l} = \frac{3}{1} \Rightarrow l = 2/3$
  - Vậy điểm cân bằng của hệ ở vị trí $0.5 + 2/3 \approx 1.17\text{m}$

Giới thiệu kì vọng
Ngôn ngữ vật lý

- Ta có mật độ khối lượng của chất điểm tại vị trí x mét $(0 \leq x \leq 2)$ là:

$$f(x) = \begin{cases} \dfrac{1\text{kg}}{1\text{mét}} & 0 \leq x < 1 \\ \dfrac{2\text{kg}}{1\text{mét}} & 1 \leq x \leq 2 \end{cases}$$

- Trọng tâm của hệ là vị trí:

$$\frac{1}{3}\int_0^2 xf(x)dx = \frac{1}{3}\left(\int_0^1 x \times 1dx + \int_1^2 x \times 2dx\right) = \frac{1}{3}\left(\frac{1}{2} + 3\right) \approx 1.17\text{m}$$

Kì vọng của biến ngẫu nhiên rời rạc

- Cho b.n.n rời rạc X với hàm xác suất f, kì vọng (mean) của X, kí hiệu $E(X)$, là số thực được tính bởi (nếu tính được):

$$E(X) = \sum_{x} xP(X = x) = \sum_{x} xf(x)$$

- Kì vọng của X là giá trị trung bình của các giá trị mà X có thể nhận với trọng số là xác suất để X nhận các giá trị tương ứng đó
- Ví dụ: cho $X \sim \text{Bernoulli}(p)$, ta có:

$$E(X) = \sum_{x \in \{0,1\}} xP(X = x) = 0 \times P(X = 0) + 1 \times P(X = 1)$$
$$= 0 \times (1 - p) + 1 \times p = p$$

Kì vọng của biến ngẫu nhiên rời rạc
Ví dụ

- Xét thí nghiệm tung một đồng xu (đồng chất) 2 lần, đặt X là số lần được mặt ngửa. Khi đó X là b.n.n rời rạc có tập giá trị là $\{0, 1, 2\}$. X có hàm xác suất được cho bởi bảng sau:

x	0	1	2
$P(X = x)$	1/4	1/2	1/4

- Ta có kì vọng của X là:

$$E(X) = \sum_{x \in \{0,1,2\}} xP(X = x) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$

- Vậy: trung bình 2 lần tung thì được 1 lần ngửa

Kì vọng của biến ngẫu nhiên liên tục

- Cho b.n.n liên tục X với hàm mật độ xác suất f, kì vọng (mean) của X, kí hiệu $E(X)$, là số thực được tính bởi (nếu tính được):

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

- Kì vọng của X là trọng tâm (hay điểm cân bằng) của phân phối của X
- Ví dụ: cho $X \sim \text{Uniform}(0,1)$, ta có:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{1} (x \times 1) dx = \frac{x^2}{2} \Bigg|_{x=0}^{x=1} = \frac{1}{2}$$

Kì vọng của biến ngẫu nhiên liên tục
Ví dụ

- Cho X là b.n.n liên tục với hàm mật độ xác suất có dạng:

$$f(x) = \begin{cases} \dfrac{x}{8} & \text{với } 0 < x < 4 \\ 0 & \text{khác} \end{cases}$$

- Ta có kì vọng của X là:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{4} x \frac{x}{8} dx = \frac{x^3}{24} \Bigg|_{x=0}^{x=4} = \frac{8}{3}$$

- Vậy: trọng tâm (hay điểm cân bằng) của phân phối của X là $\frac{8}{3}$

Biến ngẫu nhiên phái sinh và kì vọng

- Cho b.n.n $X: \Omega \rightarrow \mathbb{R}$ và hàm số $r: \mathbb{R} \rightarrow \mathbb{R}$, ta nói $Y: \Omega \rightarrow \mathbb{R}$ là b.n.n phái sinh từ b.n.n X qua hàm số r, kí hiệu $Y = r(X)$ nếu Y được xác định bởi:

$$Y(\omega) = r\big(X(\omega)\big), \omega \in \Omega$$

- Ví dụ:
  - Gọi l là chiều dài của hình vuông và s là diện tích của hình vuông thì s là đại lượng phái sinh từ đại lượng l qua hàm số r, với $r(x) = x^2$. Ta kí hiệu: $s = r(l) = l^2$
  - Bấy chừ, nếu chọn ngẫu nhiên một hình vuông, gọi L, S là chiều dài và diện tích của hình vuông đó thì S là b.n.n phái sinh từ b.n.n L qua hàm số r. Ta kí hiệu: $S = r(L) = L^2$

Biến ngẫu nhiên phái sinh và kì vọng
Ví dụ

- Nhà cái rung hai đồng xu (đồng chất). Người chơi sẽ được 1$ nếu không ra ngửa, mất 1$ nếu ra hai ngửa và không được/mất gì nếu ra một ngửa:
  - Đặt X là số mặt ngửa
  - Đặt Y là số tiền người chơi kiếm được thì Y là b.n.n phái sinh từ b.n.n X qua hàm số r được xác định bởi:

$$r(x) = \begin{cases} 1 & x = 0 \\ 0 & x = 1 \\ -1 & x = 2 \end{cases}$$

  - Ta kí hiệu: $Y = r(X)$
- Câu hỏi: trung bình mỗi lần chơi thì người chơi được/mất bao nhiêu?

Biến ngẫu nhiên phái sinh và kì vọng
Ví dụ

- Trả lời: X là b.n.n rời rạc có tập giá trị là $\{0, 1, 2\}$ với hàm xác suất:

x	0	1	2
$P(X = x)$	1/4	1/2	1/4

- Y là b.n.n rời rạc có tập giá trị là $\{-1, 0, 1\}$ với hàm xác suất:

$$P(Y = -1) = P(X = 2) = 1/4$$
$$P(Y = 0) = P(X = 1) = 1/2$$
$$P(Y = 1) = P(X = 0) = 1/4$$

- Kì vọng của Y là:

$$E(Y) = \sum_{y \in \{-1,0,1\}} yP(Y = y) = -1 \times \frac{1}{4} + 0 \times \frac{1}{2} + 1 \times \frac{1}{4} = 0\$$$

Biến ngẫu nhiên phái sinh và kì vọng

- Công thức tính kì vọng cho b.n.n phái sinh: Cho X là b.n.n và $Y = r(X)$ thì kì vọng của Y có thể được tính từ phân phối của X bằng công thức:

$$E(Y) = E\big(r(X)\big) = \begin{cases} \sum_x r(x)f(x) & \text{nếu } X \text{ là b.n.n rời rạc} \\ \int_{-\infty}^{\infty} r(x)f(x)dx & \text{nếu } X \text{ là b.n.n liên tục} \end{cases}$$

- Ở ví dụ trên ta có thể tính trung bình số tiền được/mất như sau:

$$E(Y) = E\big(r(X)\big) = \sum_{x\in\{0,1,2\}} r(x)P(X = x) = 1 \times \frac{1}{4} + 0 \times \frac{1}{2} + (-1) \times \frac{1}{4}$$

$$= 0\$$$

Biến ngẫu nhiên phái sinh và kì vọng
Ví dụ

- Chọn ngẫu nhiên một hình vuông có chiều dài trong khoảng $[0, 1]\mathrm{m}$. Tính trung bình diện tích hình vuông được chọn?
- Gọi L là chiều dài của hình vuông được chọn thì $L \sim \mathrm{Uniform}(0, 1)$ nên L có hàm mật độ xác suất là:

$$f(l) = \begin{cases} 1 & \text{khi } 0 \leq l \leq 1 \\ 0 & \text{khác} \end{cases}$$

- Gọi S là diện tích của hình vuông được chọn thì S là b.n.n phái sinh từ b.n.n L qua hàm số r, với $r(x) = x^2$. Từ đó ta có kì vọng của S là:

$$E(S) = E\big(r(L)\big) = \int_{-\infty}^{\infty} r(l) f(l) dl = \int_0^1 (l^2 \times 1) dl = \frac{1}{3}\mathrm{m}^2$$

- Lưu ý: dễ bị trực giác lừa là $\frac{1}{2}\mathrm{m}^2$ (S là b.n.n liên tục trên $[0, 1]\mathrm{m}^2$ nhưng không có phân phối đều)

Các tính chất của kì vọng

- Với a, b là hai hằng số thực, X là b.n.n và $Y = aX + b$ (nghĩa là Y là b.n.n phái sinh từ X qua hàm số r, với $r(x) = ax + b$) thì:

$$E(Y) = E(aX + b) = aE(X) + b$$

- Hệ quả: nếu $X = c$ (với c là hằng số) thì $E(X) = E(c) = 0$
- Ví dụ: chọn ngẫu nhiên một thanh có chiều dài trong khoảng $[0, 1]\mathrm{m}$. Kéo dãn thanh dài gấp đôi và nối thêm $0.5\mathrm{m}$ thì được thanh có chiều dài trung bình là:

$$E(Y) = E(2X + 0.5) = 2E(X) + 0.5 = 2 \times 0.5 + 0.5 = 1.5\mathrm{m}$$

(với $X \sim \mathrm{Uniform}(0, 1)$ là chiều dài ban đầu của thanh và Y là chiều dài sau khi biến đổi của thanh)

Các tính chất của kì vọng

- Với X, Y là hai b.n.n và $Z = X + Y$ (nghĩa là Z là b.n.n phái sinh từ X, Y qua hàm số 2 biến r, với $r(x, y) = x + y$) thì:
$$E(Z) = E(X + Y) = E(X) + E(Y)$$
- Hệ quả: nếu $E(aX + bY + c) = aE(X) + bE(Y) + c$
- Ví dụ: chọn ngẫu nhiên hai thanh có chiều dài trong khoảng $[0, 1]\mathrm{m}$. Kéo dãn thanh thứ nhất dài gấp đôi, cắt bỏ một nửa thanh thứ hai, nối lại thì được thanh có chiều dài trung bình là:
$$E(Z) = E(2X + Y/2) = 2E(X) + E(Y)/2 = 2 \times 0.5 + 0.5/2 = 1.25\mathrm{m}$$

Các tính chất của kì vọng

- Hai b.n.n X, Y được gọi là độc lập (nhau) nếu với mọi $A, B \subset \mathbb{R}$ ta có:
$$P\big((X \in A) \cap (Y \in B)\big) = P(X \in A)P(Y \in B)$$
  - Nghĩa là việc X nhận giá trị nào cũng không ảnh hưởng đến khả năng nhận giá trị nào đó của Y (và ngược lại)
- Với X, Y là hai b.n.n độc lập và $Z = XY$ thì:
$$E(Z) = E(XY) = E(X)E(Y)$$
- Ví dụ: chọn ngẫu nhiên một thanh có chiều dài trong khoảng $[0, 1]\mathrm{m}$, chọn (độc lập) ngẫu nhiên một hệ số trong khoảng $[1, 2]$, kéo dãn thanh theo hệ số đã chọn thì được thanh có chiều dài trung bình là:
$$E(Y) = E(RX) = E(R)E(X) = 1.5 \times 0.5 = 0.75\mathrm{m}$$

(với $X \sim \mathrm{Uniform}(0, 1)$ là chiều dài ban đầu của thanh, Y là chiều dài sau khi biến đổi của thanh, $R \sim \mathrm{Uniform}(1, 2)$ là hệ số chọn được)

Phương sai

- Một lớp học gồm 20 SV có “phổ điểm” các môn Toán, Lý, Hóa như sau:

Điểm	1	2	3	4	5	6	7	8	9	10
Toán	0	0	0	0	10	10	0	0	0	0
Lý	10	0	0	0	0	0	0	0	0	10
Hóa	2	2	2	2	2	2	2	2	2	2

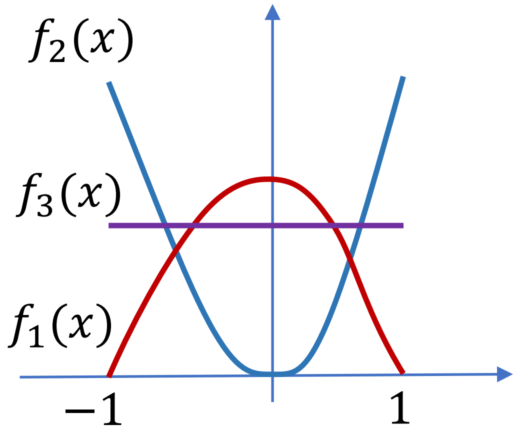
- Điểm trung bình các môn là:
  - Toán: $(5 \times 10 + 6 \times 10)/20 = 5.5$
  - Lý: $(1 \times 10 + 10 \times 10)/20 = 5.5$
  - Hóa: $(1 + 2 + \cdots + 10) \times 2/20 = 5.5$
- Điểm trung bình không thể hiện được sự phân tán của phổ điểm

Phương sai

- Xét các b.n.n liên tục X_1, X_2, X_3 có cùng tập hỗ trợ $[-1, 1]$ với các hàm mật độ xác suất lần lượt là f_1, f_2, f_3 được cho bởi:

$$f_1(x) = -\frac{3x^2}{4} + \frac{3}{4}, \quad f_2(x) = \frac{3x^2}{2}, \quad f_3(x) = \frac{1}{2}$$

- Kì vọng của các b.n.n là:
  - $E(X_1) = \int_{-\infty}^{\infty} x f_1(x) dx = \int_{-1}^{1} x \left(-\frac{3x^2}{4} + \frac{3}{4}\right) dx = 0$
  - $E(X_2) = \int_{-\infty}^{\infty} x f_2(x) dx = \int_{-1}^{1} x \left(\frac{3x^2}{2}\right) dx = 0$
  - $E(X_3) = \int_{-\infty}^{\infty} x f_3(x) dx = \int_{-1}^{1} x \left(\frac{1}{2}\right) dx = 0$



- Kì vọng không thể hiện được sự phân tán của phân phối

Phương sai

- Cho X là b.n.n có kì vọng $\mu = E(X)$, phương sai (variance) của X, kí hiệu $Var(X)$, là số thực được tính bởi (nếu tính được):

$$Var(X) = E\left((X-\mu)^2\right) = \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{nếu } X \text{ là b. n. n rời rạc} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & \text{nếu } X \text{ là b. n. n liên tục} \end{cases}$$

- Khi đó ta cũng nói $\sqrt{Var(X)}$ là độ lệch chuẩn (standard deviation) của X. Lưu ý: độ lệch chuẩn có cùng đơn vị với X nhưng phương sai thì không
- Ta còn kí hiệu phương sai là σ^2 và độ lệch chuẩn là σ
- Phương sai (hay độ lệch chuẩn) phản ánh sự phân tán của phân phối

Phương sai
Ví dụ

- Một lớp học gồm 20 SV có "phổ điểm" các môn Toán, Lý, Hóa như sau:

Điểm	1	2	3	4	5	6	7	8	9	10
Toán	0	0	0	0	10	10	0	0	0	0
Lý	10	0	0	0	0	0	0	0	0	10
Hóa	2	2	2	2	2	2	2	2	2	2

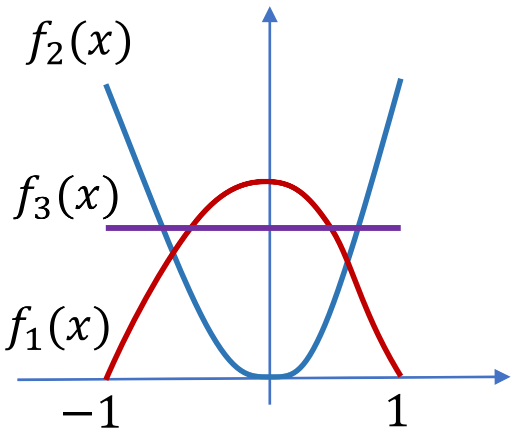
- Điểm trung bình các môn đều là 5.5. Nhưng độ lệch chuẩn khác nhau: Toán là 0.51, Lý là 4.62, Hóa là 2.95
- Ta mô tả điều này bằng cách nói: điểm Toán là 5.5 ± 0.51, điểm Lý là 5.5 ± 4.62, điểm Hóa là 5.5 ± 2.95
- Như vậy: phổ điểm môn Lý rộng nhất và phổ điểm môn Toán hẹp nhất

Phương sai
Ví dụ

- Xét các b.n.n liên tục X_1, X_2, X_3 có cùng tập hỗ trợ $[-1, 1]$ với các hàm mật độ xác suất lần lượt là f_1, f_2, f_3 được cho bởi:

$$f_1(x) = -\frac{3x^2}{4} + \frac{3}{4}, \quad f_2(x) = \frac{3x^2}{2}, \quad f_3(x) = \frac{1}{2}$$

- Các biến có cùng kì vọng là 0 nhưng khác phương sai:
  - $Var(X_3) = \int_{-\infty}^{\infty} (x-0)^2 f_3(x)dx = \int_{-1}^{1} \frac{x^2}{2} dx = \frac{1}{3}$
  - $Var(X_2) = ?$ (tự tính nha)
  - $Var(X_1) = ?$
- Như vậy: phân phối của X_2 là rộng nhất và phân phối của X_1 là hẹp nhất



Các tính chất của phương sai

- Cho X là b.n.n có phương sai. Ta có:
  - $Var(X) = E(X^2) - (E(X))^2$
  - $Var(X) \geq 0$
  - $Var(X) = 0$ khi và chỉ khi $P(X = c) = 1$ với c là hằng số nào đó
  - Cho a, b là hai hằng số thì:
  $$Var(aX + b) = a^2 Var(X)$$
- Ví dụ: chọn ngẫu nhiên một thanh có chiều dài trong khoảng $[0, 1]\mathrm{m}$. Kéo dãn thanh dài gấp đôi và nối thêm $0.5\mathrm{m}$ thì được thanh có chiều dài với độ lệch chuẩn là:
$$\sigma = \sqrt{Var(Y)} = \sqrt{Var(2X + 0.5)} = \sqrt{2^2 Var(X)} = \sqrt{4 \times 1/12} = 0.58\mathrm{m}$$
  - Lưu ý: $X \sim \mathrm{Uniform}(0, 1)$ thì
$$Var(X) = E(X^2) - \big(E(X)\big)^2 = \int_0^1 (x^2 \times 1)dx - \left(\int_0^1 (x \times 1)dx\right)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

Các tính chất của phương sai

- Nếu X, Y là hai b.n.n độc lập có phương sai thì:
$$Var(X+Y) = Var(X) + Var(Y)$$
- Ví dụ: chọn ngẫu nhiên (độc lập) hai thanh có chiều dài trong khoảng $[0,1]\mathrm{m}$. Kéo dãn thanh thứ nhất dài gấp đôi, cắt bỏ một nửa thanh thứ hai, nối lại thì được thanh có chiều dài với độ lệch chuẩn là:
$$\sigma = \sqrt{Var(Z)} = \sqrt{Var(2X + Y/2)} = \sqrt{4Var(X) + Var(Y)/4}$$
$$= \sqrt{4 \times \frac{1}{12} + \frac{1}{4} \times \frac{1}{12}} = 0.60\mathrm{m}$$