Độ lệch

Toán Chuyên Đề

HUST

Ngày 11 tháng 11 năm 2016

Tài liệu tham khảo

- Eric Lehman, F Thomson Leighton & Albert R Meyer, *Mathematics for Computer Science*, 2013 (Miễn phí)
- Michael Mitzenmacher và Eli Upfal, *Probability and Computing*, 2005
- Nguyễn Tiến Dũng và Đỗ Đức Thái, *Nhập Môn Hiện Đại Xác Suất & Thống Kê*.

Nội dung

Ví dụ

Trò chơi A

Bạn sẽ thắng $\$2$ với xác suất $2/3$ và thua $\$1$ với xác suất $1/3$.

Trò chơi B

Bạn sẽ thắng $\$1002$ với xác suất $2/3$ và thua $\$2001$ với xác suất $1/3$.

Ví dụ

Trò chơi A

Bạn sẽ thắng $\$2$ với xác suất $2/3$ và thua $\$1$ với xác suất $1/3$.

Trò chơi B

Bạn sẽ thắng $\$1002$ với xác suất $2/3$ và thua $\$2001$ với xác suất $1/3$.

Bạn nên chơi trò chơi nào?

Ví dụ

Trò chơi A

Bạn sẽ thắng $\$2$ với xác suất $2/3$ và thua $\$1$ với xác suất $1/3$.

Trò chơi B

Bạn sẽ thắng $\$1002$ với xác suất $2/3$ và thua $\$2001$ với xác suất $1/3$.

Bạn nên chơi trò chơi nào? Kỳ vọng lãi thu được từ mỗi trò chơi là bao nhiêu?

Kỳ vọng lãi

$$\mathrm{Ex}[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1$$

Kỳ vọng lãi

$$\mathrm{Ex}[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1$$
$$\mathrm{Ex}[B] = 1002 \cdot \frac{2}{3} + (-2001) \cdot \frac{1}{3} = 1$$

Kỳ vọng lãi

$$\mathrm{Ex}[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1$$
$$\mathrm{Ex}[B] = 1002 \cdot \frac{2}{3} + (-2001) \cdot \frac{1}{3} = 1$$

Câu hỏi

Trò chơi nào rủi ro hơn?

Định nghĩa

Phương sai của biến ngẫu nhiên R là

$$\mathrm{Var}[R] = \mathrm{Ex}[(R - \mathrm{Ex}[R])^2]$$

Nói cách khác, phương sai là trung bình của bình phương độ lệch so với trung bình.

Trò chơi A

$$A - \mathrm{Ex}[A] = \begin{cases} 1 & \text{với xác suất } 2/3 \\ -2 & \text{với xác suất } 1/3 \end{cases}$$

Trò chơi A

$$A - \mathrm{Ex}[A] = \begin{cases} 1 & \text{với xác suất } 2/3 \\ -2 & \text{với xác suất } 1/3 \end{cases}$$

$$(A - \mathrm{Ex}[A])^2 = \begin{cases} 1 & \text{với xác suất } 2/3 \\ 4 & \text{với xác suất } 1/3 \end{cases}$$

Trò chơi A

$$A - \mathrm{Ex}[A] = \begin{cases} 1 & \text{với xác suất } 2/3 \\ -2 & \text{với xác suất } 1/3 \end{cases}$$

$$(A - \mathrm{Ex}[A])^2 = \begin{cases} 1 & \text{với xác suất } 2/3 \\ 4 & \text{với xác suất } 1/3 \end{cases}$$

$$\mathrm{Ex}[(A - \mathrm{Ex}[A])^2] = 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3}$$

Trò chơi A

$$A - \mathrm{Ex}[A] = \begin{cases} 1 & \text{với xác suất } 2/3 \\ -2 & \text{với xác suất } 1/3 \end{cases}$$

$$(A - \mathrm{Ex}[A])^2 = \begin{cases} 1 & \text{với xác suất } 2/3 \\ 4 & \text{với xác suất } 1/3 \end{cases}$$

$$\mathrm{Ex}[(A - \mathrm{Ex}[A])^2] = 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3}$$

$$\mathrm{Var}[A] = 2.$$

Trò chơi B

$$B - \mathrm{Ex}[B] = \begin{cases} 1001 & \text{với xác suất } 2/3 \\ -2002 & \text{với xác suất } 1/3 \end{cases}$$

Trò chơi B

$$B - \mathrm{Ex}[B] = \begin{cases} 1001 & \text{với xác suất } 2/3 \\ -2002 & \text{với xác suất } 1/3 \end{cases}$$

$$(B - \mathrm{Ex}[B])^2 = \begin{cases} 1{,}002{,}001 & \text{với xác suất } 2/3 \\ 4{,}008{,}004 & \text{với xác suất } 1/3 \end{cases}$$

Trò chơi B

$$B - \text{Ex}[B] = \begin{cases} 1001 & \text{với xác suất } 2/3 \\ -2002 & \text{với xác suất } 1/3 \end{cases}$$

$$(B - \text{Ex}[B])^2 = \begin{cases} 1,002,001 & \text{với xác suất } 2/3 \\ 4,008,004 & \text{với xác suất } 1/3 \end{cases}$$

$$\text{Ex}[(B - \text{Ex}[B])^2] = 1,00,001 \cdot \frac{2}{3} + 4,008,004 \cdot \frac{1}{3}$$

Trò chơi B

$$B - \mathrm{Ex}[B] = \begin{cases} 1001 & \text{với xác suất } 2/3 \\ -2002 & \text{với xác suất } 1/3 \end{cases}$$

$$(B - \mathrm{Ex}[B])^2 = \begin{cases} 1,002,001 & \text{với xác suất } 2/3 \\ 4,008,004 & \text{với xác suất } 1/3 \end{cases}$$

$$\mathrm{Ex}[(B - \mathrm{Ex}[B])^2] = 1,00,001 \cdot \frac{2}{3} + 4,008,004 \cdot \frac{1}{3}$$

$$\mathrm{Var}[B] = 2,004,002.$$

Trò chơi nào rủi ro hơn?

Trò chơi A

$$\mathrm{Var}[A] = 2$$

Lãi suất thường gần với giá trị trung bình $\$1$.

Trò chơi nào rủi ro hơn?

Trò chơi A

$$\mathrm{Var}[A] = 2$$

Lãi suất thường gần với giá trị trung bình $\$1$.

Trò chơi B

$$\mathrm{Var}[B] = 2,0004,002$$

Lãi suất lệch rất xa so với giá trị trung bình là $\$1$.

Trò chơi nào rủi ro hơn?

Trò chơi A

$$\mathrm{Var}[A] = 2$$

Lãi suất thường gần với giá trị trung bình $\$1$.

Trò chơi B

$$\mathrm{Var}[B] = 2,0004,002$$

Lãi suất lệch rất xa so với giá trị trung bình là $\$1$.

Phương sai cao thường gắn với rủi ro nhiều. Ví dụ, trong 10 lần chơi trò chơi A, ta có lãi trung bình $\$10$ nhưng cũng có thể mất $\$10$. Còn với trò chơi B thì sao?

"Đơn vị" của phương sai

- Biến ngẫu nhiên và phương sai không cùng "đơn vị".

$$\mathrm{Var}[R] = \mathrm{Ex}[(R - \mathrm{Ex}[R])^2]$$

"Đơn vị" của phương sai

- Biến ngẫu nhiên và phương sai không cùng "đơn vị".

$$\mathrm{Var}[R] = \mathrm{Ex}[(R - \mathrm{Ex}[R])^2]$$

- Ví dụ, nếu đơn vị của biến ngẫu nhiên là $\$$, vậy thì đơn vị của phương sai là $\$^2$.

"Đơn vị" của phương sai

- Biến ngẫu nhiên và phương sai không cùng "đơn vị".

$$\mathrm{Var}[R] = \mathrm{Ex}[(R - \mathrm{Ex}[R])^2]$$

- Ví dụ, nếu đơn vị của biến ngẫu nhiên là \$, vậy thì đơn vị của phương sai là $\$^2$.
- Độ lệch chuẩn tương tự như phương sai nhưng cùng "đơn vị" với biến ngẫu nhiên.

Định nghĩa

Độ lệch chuẩn σ_R của biến ngẫu nhiên R là căn bậc hai của phương sai:

$$\sigma_R = \sqrt{\mathrm{Var}[R]} = \sqrt{\mathrm{Ex}[(R - \mathrm{Ex}[R])^2]}.$$

Định nghĩa

Độ lệch chuẩn σ_R của biến ngẫu nhiên R là căn bậc hai của phương sai:

$$\sigma_R = \sqrt{\mathrm{Var}[R]} = \sqrt{\mathrm{Ex}[(R - \mathrm{Ex}[R])^2]}.$$

Ví dụ

Độ lệch chuẩn của biến ngẫu nhiên lãi trong trò chơi A và B là

Định nghĩa

Độ lệch chuẩn σ_R của biến ngẫu nhiên R là căn bậc hai của phương sai:

$$\sigma_R = \sqrt{\mathrm{Var}[R]} = \sqrt{\mathrm{Ex}[(R - \mathrm{Ex}[R])^2]}.$$

Ví dụ

Độ lệch chuẩn của biến ngẫu nhiên lãi trong trò chơi A và B là

$$\sigma_A = \sqrt{\mathrm{Var}[A]} = \sqrt{2} \approx 1.14,$$

Định nghĩa

Độ lệch chuẩn σ_R của biến ngẫu nhiên R là căn bậc hai của phương sai:

$$\sigma_R = \sqrt{\mathrm{Var}[R]} = \sqrt{\mathrm{Ex}[(R - \mathrm{Ex}[R])^2]}.$$

Ví dụ

Độ lệch chuẩn của biến ngẫu nhiên lãi trong trò chơi A và B là

$$\sigma_A = \sqrt{\mathrm{Var}[A]} = \sqrt{2} \approx 1.14,$$
$$\sigma_B = \sqrt{\mathrm{Var}[B]} = \sqrt{2,004,002} \approx 1416.$$

Công thức khác cho phương sai

Bổ đề

Với mọi biến ngẫu nhiên R,

$$\mathrm{Var}[R] = \mathrm{Ex}[R^2] - (\mathrm{Ex}[R])^2.$$

Công thức khác cho phương sai

Bổ đề

Với mọi biến ngẫu nhiên R,

$$\mathrm{Var}[R] = \mathrm{Ex}[R^2] - (\mathrm{Ex}[R])^2.$$

Ví dụ

Với trò chơi A

Công thức khác cho phương sai

Bổ đề

Với mọi biến ngẫu nhiên R,

$$\mathrm{Var}[R] = \mathrm{Ex}[R^2] - (\mathrm{Ex}[R])^2.$$

Ví dụ

Với trò chơi A

$$\mathrm{Ex}[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1$$

Công thức khác cho phương sai

Bổ đề

Với mọi biến ngẫu nhiên R,

$$\mathrm{Var}[R] = \mathrm{Ex}[R^2] - (\mathrm{Ex}[R])^2.$$

Ví dụ

Với trò chơi A

$$\mathrm{Ex}[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1$$
$$\mathrm{Ex}[A^2] = 2^2 \cdot \frac{2}{3} + (-1)^2 \cdot \frac{1}{3} = 3$$

Công thức khác cho phương sai

Bổ đề

Với mọi biến ngẫu nhiên R,

$$\mathrm{Var}[R] = \mathrm{Ex}[R^2] - (\mathrm{Ex}[R])^2.$$

Ví dụ

Với trò chơi A

$$\begin{aligned}
\mathrm{Ex}[A] &= 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1 \\
\mathrm{Ex}[A^2] &= 2^2 \cdot \frac{2}{3} + (-1)^2 \cdot \frac{1}{3} = 3 \\
\mathrm{Var}[A] &= \mathrm{Ex}[A^2] - (\mathrm{Ex}[A])^2 = 3 - 1^2 = 2.
\end{aligned}$$

Bài tập

Hãy chứng minh bổ đề trước.

Phương sai của biến ngẫu nhiên chỉ báo

Bổ đề

Xét B là biến ngẫu nhiên chỉ báo với $\Pr[B = 1] = p$. Vậy thì

$$\mathrm{Var}[B] = p(1 - p).$$

Phương sai của biến ngẫu nhiên chỉ báo

Bổ đề

Xét B *là biến ngẫu nhiên chỉ báo với* $\Pr[B = 1] = p$. *Vậy thì*

$$\mathrm{Var}[B] = p(1 - p).$$

Bài tập

Hãy chứng minh bổ đề trên.

Số bước trung bình dẫn đến lỗi

- Hệ thống lỗi ở mỗi bước với xác suất p.

Số bước trung bình dẫn đến lỗi

- Hệ thống lỗi ở mỗi bước với xác suất p.
- Xét C là số bước để có lỗi đầu tiên xuất hiện (kể cả bước lỗi). Vậy

$$\mathrm{Ex}[C] =$$

Số bước trung bình dẫn đến lỗi

- Hệ thống lỗi ở mỗi bước với xác suất p.
- Xét C là số bước để có lỗi đầu tiên xuất hiện (kể cả bước lỗi). Vậy

$$\mathrm{Ex}[C] = 1/p.$$

Số bước trung bình dẫn đến lỗi

- Hệ thống lỗi ở mỗi bước với xác suất p.
- Xét C là số bước để có lỗi đầu tiên xuất hiện (kể cả bước lỗi). Vậy

$$\mathrm{Ex}[C] = 1/p.$$

- Phương sai của C bằng bao nhiêu?

$$\mathrm{Ex}[C^2] = \overbrace{1^2 \cdot p}^{\text{lỗi ngay bước đầu tiên}} + \overbrace{\mathrm{Ex}[(C+1)^2] \cdot (1-p)}^{\text{hoặc không}}$$

$$\begin{aligned}\mathrm{Ex}[C^2] &= \overbrace{1^2 \cdot p}^{\text{lỗi ngay bước đầu tiên}} \quad + \overbrace{\mathrm{Ex}[(C+1)^2] \cdot (1-p)}^{\text{hoặc không}} \\ &= p + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \mathrm{Ex}[C] \cdot (1-p) + (1-p)\end{aligned}$$

$$
\begin{aligned}
\mathrm{Ex}[C^2] &= \overbrace{1^2 \cdot p}^{\text{lỗi ngay bước đầu tiên}} + \overbrace{\mathrm{Ex}[(C+1)^2] \cdot (1-p)}^{\text{hoặc không}} \\
&= p + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \mathrm{Ex}[C] \cdot (1-p) + (1-p) \\
&= 1 + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \left(\frac{1-p}{p}\right)
\end{aligned}
$$

$$\begin{aligned}
\mathrm{Ex}[C^2] &= \overbrace{1^2 \cdot p}^{\text{lỗi ngay bước đầu tiên}} + \overbrace{\mathrm{Ex}[(C+1)^2] \cdot (1-p)}^{\text{hoặc không}} \\
&= p + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \mathrm{Ex}[C] \cdot (1-p) + (1-p) \\
&= 1 + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \left(\frac{1-p}{p}\right)
\end{aligned}$$

Ta được

$$p \cdot \mathrm{Ex}[C^2] = \frac{2-p}{p}.$$

$$\mathrm{Ex}[C^2] = \overbrace{1^2 \cdot p}^{\text{lỗi ngay bước đầu tiên}} + \overbrace{\mathrm{Ex}[(C+1)^2] \cdot (1-p)}^{\text{hoặc không}}$$
$$= p + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \mathrm{Ex}[C] \cdot (1-p) + (1-p)$$
$$= 1 + \mathrm{Ex}[C^2] \cdot (1-p) + 2 \cdot \left(\frac{1-p}{p}\right)$$

Ta được

$$p \cdot \mathrm{Ex}[C^2] = \frac{2-p}{p}.$$

và được

$$\mathrm{Ex}[C^2] = \frac{2-p}{p^2}.$$

Bài tập

Hãy tính tiếp $\mathrm{Var}[C]$.

Bài tập: Biến ngẫu nhiên đều

Với biến ngẫu nhiên đều R trên $\{1, 2, 3, \ldots, n\}$, phương sai của R bằng bao nhiêu?

Định lý

Nếu R_1, R_2 là hai biến ngẫu nhiên độc lập, vậy thì

$$\mathrm{Var}[R_1 + R_2] = \mathrm{Var}[R_1] + \mathrm{Var}[R_2].$$

Định lý

Nếu R_1, R_2 là hai biến ngẫu nhiên độc lập, vậy thì

$$\mathrm{Var}[R_1 + R_2] = \mathrm{Var}[R_1] + \mathrm{Var}[R_2].$$

Bài tập: Hãy chứng minh định lý trên.

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Chứng minh

Xem J như số mặt "ngửa" khi tung n đồng xu độc lập, mỗi đồng có xác suất xuất hiện mặt ngửa là p.

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Chứng minh

Xem J như số mặt "ngửa" khi tung n đồng xu độc lập, mỗi đồng có xác suất xuất hiện mặt ngửa là p. Đặt

$$J_i = 1 \quad \Leftrightarrow \quad \text{đồng thứ } i \text{ ngửa}$$

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Chứng minh

Xem J như số mặt "ngửa" khi tung n đồng xu độc lập, mỗi đồng có xác suất xuất hiện mặt ngửa là p. Đặt

$$J_i = 1 \quad \Leftrightarrow \quad \text{đồng thứ } i \text{ ngửa}$$

Vậy thì

$$\mathrm{Var}[J] =$$

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Chứng minh

Xem J như số mặt "ngửa" khi tung n đồng xu độc lập, mỗi đồng có xác suất xuất hiện mặt ngửa là p. Đặt

$$J_i = 1 \quad \Leftrightarrow \quad \text{đồng thứ } i \text{ ngửa}$$

Vậy thì

$$\mathrm{Var}[J] = \mathrm{Var}[J_1 + J_2 + \cdots + J_n]$$

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Chứng minh

Xem J như số mặt "ngửa" khi tung n đồng xu độc lập, mỗi đồng có xác suất xuất hiện mặt ngửa là p. Đặt

$$J_i = 1 \quad \Leftrightarrow \quad \text{đồng thứ } i \text{ ngửa}$$

Vậy thì

$$\begin{aligned}\mathrm{Var}[J] &= \mathrm{Var}[J_1 + J_2 + \cdots + J_n] \\ &= \mathrm{Var}[J_1] + \mathrm{Var}[J_2] + \cdots + \mathrm{Var}[J_n]\end{aligned}$$

Bổ đề

Xét J là biến ngẫu nhiên theo phân bố nhị thức với tham số n và p, vậy thì

$$\mathrm{Var}[J] = np(1-p).$$

Chứng minh

Xem J như số mặt "ngửa" khi tung n đồng xu độc lập, mỗi đồng có xác suất xuất hiện mặt ngửa là p. Đặt

$$J_i = 1 \quad \Leftrightarrow \quad \text{đồng thứ } i \text{ ngửa}$$

Vậy thì

$$\begin{aligned} \mathrm{Var}[J] &= \mathrm{Var}[J_1 + J_2 + \cdots + J_n] \\ &= \mathrm{Var}[J_1] + \mathrm{Var}[J_2] + \cdots + \mathrm{Var}[J_n] \\ &= np(1-p). \end{aligned}$$

Nội dung

Ví dụ

- Intelligent Quotients trung bình của mọi người là 100.

Ví dụ

- Intelligent Quotients trung bình của mọi người là 100.
- Vậy nhiều nhất chỉ $1/3$ dân số có IQ lớn hơn 300. Tại sao?

Ví dụ

- Intelligent Quotients trung bình của mọi người là 100.
- Vậy nhiều nhất chỉ $1/3$ dân số có IQ lớn hơn 300. Tại sao?
- Suy ra, xác suất một người ngẫu nhiên có IQ lớn hơn 300 là $\leq 1/3$.

Định lý (Markov)

Nếu R là biến ngẫu nhiên không âm, vậy thì với mọi $x > 0$,

$$\Pr[R \geq x] \leq \frac{\mathrm{Ex}[R]}{x}.$$

The Chinese Appetizer Problem

- Có n người ngồi ăn quanh một mâm tròn.
- Mỗi người có một món khai vị trước mặt. Giả sử các món khai vị này khác nhau.
- Lợi dụng lúc mọi người mải nói chuyện, ai đó đã quay mâm một cách ngẫu nhiên để mỗi người nhận được ngẫu nhiên một món khai vị.
- Hãy tính xác suất để cả n người đều nhận lại được đúng món khai vị của mình.

The Chinese Appetizer Problem

- Giả sử mỗi người nhận lại được món khai vị ban đầu của mình với xác suất $1/n$.

The Chinese Appetizer Problem

- Giả sử mỗi người nhận lại được món khai vị ban đầu của mình với xác suất $1/n$.
- Kỳ vọng của số người R nhận đúng món khai vị của mình là

$$\text{Ex}[R] = n \cdot \frac{1}{n}.$$

The Chinese Appetizer Problem

- Giả sử mỗi người nhận lại được món khai vị ban đầu của mình với xác suất $1/n$.
- Kỳ vọng của số người R nhận đúng món khai vị của mình là

$$\mathrm{Ex}[R] = n \cdot \frac{1}{n}.$$

- Theo định lý Markov,

$$\Pr[R = n] = \Pr[R \geq n] \leq \frac{\mathrm{Ex}[R]}{n} = \frac{1}{n}.$$

Bài tập

Hãy chứng minh định lý Markov.

Giả thiết R không âm là quan trọng

Xét biến ngẫu nhiên R với

$$\Pr[R = 1000] = 1/2 \quad \text{và} \quad \Pr[R = -1000] = 1/2.$$

Giả thiết R không âm là quan trọng

Xét biến ngẫu nhiên R với

$$\Pr[R = 1000] = 1/2 \quad \text{và} \quad \Pr[R = -1000] = 1/2.$$

Vậy thì

$$\mathrm{Ex}[R] = 0.$$
$$\Pr[R \geq 1000] = 1/2 \neq \frac{\mathrm{Ex}[R]}{1000} = 0.$$

Biến ngẫu nhiên bị chặn

- Giả sử IQ trung bình của sinh viên Bách Khoa là 150.

Biến ngẫu nhiên bị chặn

- Giả sử IQ trung bình của sinh viên Bách Khoa là 150.
- Xác suất một sinh viên Bách Khoa có IQ hơn 200 khoảng bao nhiêu?

Biến ngẫu nhiên bị chặn

- Giả sử IQ trung bình của sinh viên Bách Khoa là 150.
- Xác suất một sinh viên Bách Khoa có IQ hơn 200 khoảng bao nhiêu?

$$\Pr[B \geq 200] \leq \frac{\mathrm{Ex}[B]}{200} = \frac{150}{200} = \frac{3}{4}.$$

Biến ngẫu nhiên bị chặn

- Giả sử IQ trung bình của sinh viên Bách Khoa là 150.
- Xác suất một sinh viên Bách Khoa có IQ hơn 200 khoảng bao nhiêu?

$$\Pr[B \geq 200] \leq \frac{\mathrm{Ex}[B]}{200} = \frac{150}{200} = \frac{3}{4}.$$

- Biết thêm rằng không có sinh viên nào có IQ nhỏ hơn 100, vậy ước lượng trên có thể giảm xuống bằng bao nhiêu?

Xét $T = B + 100$, ta được

$$\Pr[B \geq 200] = \Pr[T \geq 100] \leq \frac{\mathrm{Ex}[T]}{100} = \frac{50}{100} = \frac{1}{2}.$$

Hệ quả

Nếu R là biến ngẫu nhiên không âm, vậy thì với mọi $c > 0$

$$\Pr\left[R \geq c \cdot \mathrm{Ex}[R]\right] \leq 1/c.$$

Chứng minh.

$\square$

Hệ quả

Nếu R là biến ngẫu nhiên không âm, vậy thì với mọi $c > 0$

$$\Pr\left[R \geq c \cdot \mathrm{Ex}[R]\right] \leq 1/c.$$

Chứng minh.

Thay $x = c \cdot \mathrm{Ex}[R]$ vào định lý Markov. $\square$

Định lý

Xét R *là biến ngẫu nhiên thỏa mãn* $R \leq u$*. Vậy thì với mọi* $x < u$*,*

$$\Pr[R \leq x] \leq \frac{u - \mathrm{Ex}[R]}{u - x}.$$

Định lý

Xét R là biến ngẫu nhiên thỏa mãn $R \leq u$. Vậy thì với mọi $x < u$,

$$\Pr[R \leq x] \leq \frac{u - \mathrm{Ex}[R]}{u - x}.$$

Bài tập: Hãy chứng minh định lý trên.

Bài tập

- Giả sử điểm thi giữa kỳ trung bình của lớp Toán Chuyên Đề là $7.5/10$.

Bài tập

- Giả sử điểm thi giữa kỳ trung bình của lớp Toán Chuyên Đề là $7.5/10$.
- Hãy ước lượng tỉ lệ sinh viên trong lớp có điểm nhỏ hơn hoặc bằng 5.

Giải

$R =$ điểm ngẫu nhiên của sinh viên

Giải

$$R = \text{ điểm ngẫu nhiên của sinh viên}$$

$$\max \text{ Điểm } = 10 = u$$

Giải

$$
\begin{aligned}
R = &\ \text{điểm ngẫu nhiên của sinh viên} \\
\max\ \text{Điểm} &= 10 = u \\
\mathrm{Ex}[R] &= 7.5
\end{aligned}
$$

Giải

$$\begin{aligned}
R &= \text{ điểm ngẫu nhiên của sinh viên} \\
\max\ \text{Điểm} &= 10 = u \\
\text{Ex}[R] &= 7.5 \\
\Pr[R \le 5.0] &\le \frac{10.0 - 7.5}{10.0 - 5.0} = \frac{2.5}{5.0} = \frac{1}{2}.
\end{aligned}$$

Giải

$$
\begin{aligned}
R &= \text{ điểm ngẫu nhiên của sinh viên} \\
\max \text{ Điểm } &= 10 = u \\
\mathrm{Ex}[R] &= 7.5 \\
\Pr[R \leq 5.0] &\leq \frac{10.0 - 7.5}{10.0 - 5.0} = \frac{2.5}{5.0} = \frac{1}{2}.
\end{aligned}
$$

Nói cách khác, chỉ nhiều nhất nửa lớp có điểm ≤ 5.

Nội dung

Bổ đề

Với mọi biến ngẫu nhiên R, $\alpha \in \mathbb{R}$, *và* $x > 0$,

$$\Pr[|R| \geq x] \leq \frac{\text{Ex}[|R|^\alpha]}{x^\alpha}$$

Bổ đề

Với mọi biến ngẫu nhiên R, $\alpha \in \mathbb{R}$, và $x > 0$,

$$\Pr[|R| \geq x] \leq \frac{\mathrm{Ex}[|R|^\alpha]}{x^\alpha}$$

Chứng minh.

Do

$$|R| \geq x \quad \Leftrightarrow \quad |R|^\alpha \geq x^\alpha$$

Áp dụng định lý Markov, ta suy ra bổ đề trên. □

Định lý (Chebyshev)

Xét R là một biến ngẫu nhiên và $x \in \mathbb{R}^+$. Vậy thì

$$\Pr[|R - \mathrm{Ex}[R]| \geq x] \leq \frac{\mathrm{Var}[R]}{x^2}.$$

Định lý (Chebyshev)

Xét R là một biến ngẫu nhiên và $x \in \mathbb{R}^+$. Vậy thì

$$\Pr[|R - \mathrm{Ex}[R]| \geq x] \leq \frac{\mathrm{Var}[R]}{x^2}.$$

Đây là một trường hợp riêng của bổ đề trước. Tại sao?

Hệ quả

Xét R là biến ngẫu nhiên, và xét c là một số thực dương

$$\Pr[|R - \mathrm{Ex}[R]| \geq c \cdot \sigma_R] \leq \frac{1}{c^2}.$$

Hệ quả

Xét R là biến ngẫu nhiên, và xét c là một số thực dương

$$\Pr[|R - \mathrm{Ex}[R]| \geq c \cdot \sigma_R] \leq \frac{1}{c^2}.$$

Bài tập: Hãy chứng minh hệ quả trên.

Ví dụ

$R =$ IQ của một người ngẫu nhiên. Giả sử

$$R \geq 0, \quad \text{Ex}[R] = 100, \quad \sigma_R = 15$$

Hãy ước lượng

$$\Pr[R \geq 250].$$

Ước lượng

- Bởi định lý Markov

Ước lượng

- Bởi định lý Markov

$$\Pr[R \geq 250] \leq \frac{\mathrm{Ex}[R]}{250} = \frac{100}{250} = 0.4$$

Ước lượng

- Bởi định lý Markov

$$\Pr[R \geq 250] \leq \frac{\mathrm{Ex}[R]}{250} = \frac{100}{250} = 0.4$$

- Bởi định lý Chebyshev

Ước lượng

- Bởi định lý Markov

$$\Pr[R \geq 250] \leq \frac{\text{Ex}[R]}{250} = \frac{100}{250} = 0.4$$

- Bởi định lý Chebyshev

$$\Pr[R \geq 250] = \Pr[R - 100 \geq 150]$$

Ước lượng

- Bởi định lý Markov

$$\Pr[R \geq 250] \leq \frac{\mathrm{Ex}[R]}{250} = \frac{100}{250} = 0.4$$

- Bởi định lý Chebyshev

$$\begin{aligned}\Pr[R \geq 250] &= \Pr[R - 100 \geq 150] \\ &= \Pr[R - \mathrm{Ex}[R] \geq 10 \cdot \sigma_R]\end{aligned}$$

Ước lượng

- Bởi định lý Markov

$$\Pr[R \geq 250] \leq \frac{\mathrm{Ex}[R]}{250} = \frac{100}{250} = 0.4$$

- Bởi định lý Chebyshev

$$\begin{aligned}
\Pr[R \geq 250] &= \Pr[R - 100 \geq 150] \\
&= \Pr[R - \mathrm{Ex}[R] \geq 10 \cdot \sigma_R] \\
&\leq \Pr[|R - \mathrm{Ex}[R]| \geq 10 \cdot \sigma_R]
\end{aligned}$$

Ước lượng

- Bởi định lý Markov

$$\Pr[R \geq 250] \leq \frac{\mathrm{Ex}[R]}{250} = \frac{100}{250} = 0.4$$

- Bởi định lý Chebyshev

$$\begin{aligned}
\Pr[R \geq 250] &= \Pr[R - 100 \geq 150] \\
&= \Pr[R - \mathrm{Ex}[R] \geq 10 \cdot \sigma_R] \\
&\leq \Pr[|R - \mathrm{Ex}[R]| \geq 10 \cdot \sigma_R] \\
&\leq \frac{1}{100}.
\end{aligned}$$

Định lý

Với mọi biến ngẫu nhiên R *và mọi* $c > 0$

$$\Pr[R - \mathrm{Ex}[R] \geq c \cdot \sigma_R] \leq \frac{1}{c^2 + 1}$$

và

$$\Pr[R - \mathrm{Ex}[R] \leq -c \cdot \sigma_R] \leq \frac{1}{c^2 + 1}.$$

Trở lại với IQ

Ví dụ

$R =$ IQ của một người ngẫu nhiên. Giả sử

$$R \geq 0, \quad \text{Ex}[R] = 100, \quad \sigma_R = 15$$

Hãy ước lượng

$$\Pr[R \geq 250].$$

Ước lượng

$$\Pr[R \geq 250] = \Pr[R - 100 \geq 150]$$

Ước lượng

$$
\begin{aligned}
\Pr[R \geq 250] &= \Pr[R - 100 \geq 150] \\
&= \Pr[R - \mathrm{Ex}[R] \geq 10 \cdot \sigma_R]
\end{aligned}
$$

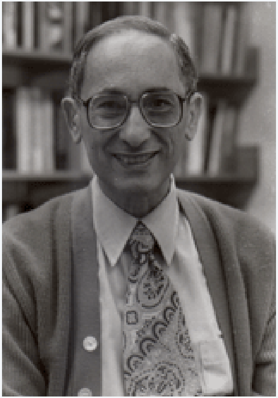
Ước lượng

$$\begin{aligned}
\Pr[R \geq 250] &= \Pr[R - 100 \geq 150] \\
&= \Pr[R - \mathrm{Ex}[R] \geq 10 \cdot \sigma_R] \\
&\leq \frac{1}{10^2 + 1} = \frac{1}{101}
\end{aligned}$$

Nội dung

Chặn Chernoff

Tổng của rất nhiều biến ngẫu nhiên có giá trị nhỏ và độc lập có nhiều khả năng sẽ không vượt quá trung bình của tổng.



Định lý (Chặn Chernoff)

Xét $T_1, T_2, \dots T_n$ *là các biến ngẫu nhiên độc lập thoả mãn* $0 \leq T_i \leq 1$ *với mọi* i*. Xét*

$$T = T_1 + T_2 + \cdots + T_n.$$

Vậy thì với mọi $c \geq 1$*,*

$$\Pr[T \geq c \operatorname{Ex}[T]] \leq e^{-k \operatorname{Ex}[T]}$$

trong đó $k = c \ln(c) - c + 1$*.*

Ví dụ

Tung 1000 đồng xu độc lập. Hãy tính xác suất của số mặt ngửa vượt quá kỳ vọng ít nhất 20%.

Lời giải

- Đặt T_i là biến ngẫu nhiên chỉ báo cho sự kiện đồng xu thứ i là ngửa.

Lời giải

- Đặt T_i là biến ngẫu nhiên chỉ báo cho sự kiện đồng xu thứ i là ngửa.
- Vậy thì tổng số mặt ngửa là

$$T = T_1 + T_2 + \cdots + T_{1000}.$$

Lời giải

- Đặt T_i là biến ngẫu nhiên chỉ báo cho sự kiện đồng xu thứ i là ngửa.
- Vậy thì tổng số mặt ngửa là

$$T = T_1 + T_2 + \cdots + T_{1000}.$$

- Cả hai điều kiện của Chernoff đều thoả mãn: Các biến T_i độc lập và $T_i \in [0, 1]$.

Lời giải

- Đặt T_i là biến ngẫu nhiên chỉ báo cho sự kiện đồng xu thứ i là ngửa.
- Vậy thì tổng số mặt ngửa là

$$T = T_1 + T_2 + \cdots + T_{1000}.$$

- Cả hai điều kiện của Chernoff đều thoả mãn: Các biến T_i độc lập và $T_i \in [0, 1]$.
- Theo định lý Chernoff

$$\Pr[T \geq c \operatorname{Ex}[T]] \leq e^{-k \operatorname{Ex}[T]}$$

với $c = 1.2$ và $k = c \ln(c) - c + 1 = 0.0187\ldots$.

$$
\begin{aligned}
\Pr[T \geq 1.2\ \mathrm{Ex}[T]] &\leq e^{-k\ \mathrm{Ex}[T]} \\
&= e^{-(0.0187\ldots)\cdot 500} \\
&< 0.0000834
\end{aligned}
$$

Ảnh hưởng của số biến và độ lệch

Xác suất sẽ nhỏ hơn rất nhiều nếu số đồng xu tăng lên.

Ví dụ

nếu tung 1 triệu đồng xu, xác suất để số mặt ngửa vượt quá kỳ vọng ít nhất 20% chỉ nhiều nhất là

$$e^{-(0.0187)\cdot 500000} < e^{-9392}.$$

Ảnh hưởng của số biến và độ lệch

Xác suất sẽ nhỏ hơn rất nhiều nếu số đồng xu tăng lên.

Ví dụ

nếu tung 1 triệu đồng xu, xác suất để số mặt ngửa vượt quá kỳ vọng ít nhất 20% chỉ nhiều nhất là

$$e^{-(0.0187)\cdot 500000} < e^{-9392}.$$

Xác suất cũng sẽ nhỏ hơn rất nhiều nếu độ lệch tăng lên.

Ví dụ

tung 1000 đồng xu, xác suất để số mặt ngửa vượt quá kỳ vọng ít nhất 30% chỉ nhiều nhất là

$$e^{-(0.0410)\cdot 500} < e^{-20.5} < 1/1,000,000,000.$$

Trò chơi Pick-4

- Bạn chọn một số bốn chữ số trong khoảng 0000 đến 9999.

Trò chơi Pick-4

- Bạn chọn một số bốn chữ số trong khoảng 0000 đến 9999.
- Nếu số bạn chọn là số ngẫu nhiên chương trình chọn, bạn sẽ được $\$5,000$.

Trò chơi Pick-4

- Bạn chọn một số bốn chữ số trong khoảng 0000 đến 9999.
- Nếu số bạn chọn là số ngẫu nhiên chương trình chọn, bạn sẽ được $\$5,000$.
- Xác suất thắng của bạn là $1/10,000$.

Trò chơi Pick-4

- Bạn chọn một số bốn chữ số trong khoảng 0000 đến 9999.
- Nếu số bạn chọn là số ngẫu nhiên chương trình chọn, bạn sẽ được $\$5,000$.
- Xác suất thắng của bạn là $1/10,000$.
- Nếu có 10 triệu người chơi, kỳ vọng số người thắng là 1000.

Trò chơi Pick-4

- Bạn chọn một số bốn chữ số trong khoảng 0000 đến 9999.
- Nếu số bạn chọn là số ngẫu nhiên chương trình chọn, bạn sẽ được $\$5,000$.
- Xác suất thắng của bạn là $1/10,000$.
- Nếu có 10 triệu người chơi, kỳ vọng số người thắng là 1000.
- Nỗi lo của công ty sổ xố: Số người thắng ít nhất là 2000.

Trò chơi Pick-4

- Bạn chọn một số bốn chữ số trong khoảng 0000 đến 9999.
- Nếu số bạn chọn là số ngẫu nhiên chương trình chọn, bạn sẽ được $\$5,000$.
- Xác suất thắng của bạn là $1/10,000$.
- Nếu có 10 triệu người chơi, kỳ vọng số người thắng là 1000.
- Nỗi lo của công ty sổ xố: Số người thắng ít nhất là 2000.
- Hãy tính xác suất để số người thắng ít nhất là 2000.

Trò chơi Pick-4 (lời giải)

- Đặt T_i là biến chỉ số cho sự kiện người thứ i thắng.

Trò chơi Pick-4 (lời giải)

- Đặt T_i là biến chỉ số cho sự kiện người thứ i thắng.
- Số người thắng là biến $T = T_1 + T_2 + \cdots + T_{10,000,000}$.

Trò chơi Pick-4 (lời giải)

- Đặt T_i là biến chỉ số cho sự kiện người thứ i thắng.
- Số người thắng là biến $T = T_1 + T_2 + \cdots + T_{10,000,000}$.
- Vì số người thắng gấp 2 lần kỳ vọng, ta chọn $c = 2$.

Trò chơi Pick-4 (lời giải)

- Đặt T_i là biến chỉ số cho sự kiện người thứ i thắng.
- Số người thắng là biến $T = T_1 + T_2 + \cdots + T_{10,000,000}$.
- Vì số người thắng gấp 2 lần kỳ vọng, ta chọn $c = 2$.
- Ta giả sử người chơi chọn số ngẫu nhiên đều và độc lập. Vậy thì Theo định lý Chernoff

$$
\begin{aligned}
k = c\ \ln(c) - c + 1 &= 0.386 \\
\Pr[T \geq 2000] &= \Pr[T \geq 2\ \mathrm{Ex}[T]] \\
&\leq e^{-k\ \mathrm{Ex}[T]} \\
&= e^{-(0.386\ldots)\cdot 1000} \\
&< e^{-386}
\end{aligned}
$$

Vậy hầu như không bao giờ công ty sổ xố phải trả gấp đôi kỳ vọng.

Trò chơi Pick-4 (tiếp)

Bài tập

Hãy tính xác suất để số người thắng cao hơn 10% so với kỳ vọng.

Trò chơi Pick-4 (tiếp)

Bài tập

Hãy tính xác suất để số người thắng cao hơn 10% so với kỳ vọng.

$$\begin{aligned} k = 1.1\ \ln(1.1) - 1.1 + 1 &= 0.00484 \\ \Pr[T \geq 1.1\ \mathrm{Ex}[T]] &\leq e^{-k\ \mathrm{Ex}[T]} \\ &= e^{-(0.00484)\cdot 1000} \\ &< 0.01 \end{aligned}$$

Nội dung

Cân bằng tải

- Hệ thống với n công việc $B_1, B_2, \ldots, B_n$ đến theo dòng.
- Công việc B_i cần L_i thời gian
- Các công việc cần xử lý ngay lập tức trên m máy $S_1, S_2, \ldots, S_n$.
- Hãy tìm cách gán mỗi công việc cho mỗi máy để hệ thống đảm bảo cân bằng tải.

Cân bằng tải

- Hệ thống với n công việc $B_1, B_2, \ldots, B_n$ đến theo dòng.
- Công việc B_i cần L_i thời gian
- Các công việc cần xử lý ngay lập tức trên m máy $S_1, S_2, \ldots, S_n$.
- Hãy tìm cách gán mỗi công việc cho mỗi máy để hệ thống đảm bảo cân bằng tải.

Phương pháp

Gán một cách ngẫu nhiên mỗi công việc đến cho một máy.

Dữ liệu thực tế

- Số công việc $n = 100,000$.
- Số lượng máy $m = 10$.
- Đặt
$$L = \sum_{j=1}^{n} L_j.$$
- Giả sử $L = 25,000$ giây.
- Vậy tải trung bình trên mỗi máy
$$\frac{L}{m} = \frac{25,000}{10} = 2500.$$

Phân tích

- Đặt R_{ij} là tải trên máy S_i từ công việc B_j. Tức là

$$R_{ij} = \begin{cases} L_j & \text{nếu máy } S_i \text{ được gán công việc } B_j \\ 0 & \text{ngược lại.} \end{cases}$$

- Vậy thì tải của máy S_i là

$$R_i = R_{i1} + R_{i2} + \cdots + R_{in}.$$

- Ta được

$$\begin{aligned} \mathrm{Ex}[R_i] &= \sum_{j=1}^{n} \mathrm{Ex}[R_{ij}] \\ &= \sum_{j=1}^{n} L_j/m \\ &= L/m. \end{aligned}$$

Đây là giá trị tối ưu cân bằng tải.

Phân tích 2: Tải của mỗi máy R_i

- Giả sử các $0 \leq R_{ij} \leq 1$, theo định lý Chernoff,

$$\Pr[R_i \geq c\ L/m] \leq e^{-k\ L/m}$$

  với $k = c\ln(c) - c + 1$.
- Với $c = 1.1$, ta được $k = 0.0048$,
- và với $L = 25,000$ ta được

$$\Pr[R_i \geq 1.1 \times L/m] \leq e^{-0.0048 \times 2500} \leq 1/160,000.$$

Phân tích 3: Máy phải chịu tải nhiều nhất

$$\begin{aligned}
&\Pr[\ \text{máy chịu tải nhiều nhất}\ \geq c\ L/m\ ]\\
&= \Pr[\ (R_1 \geq c\ L/m)\ \cup\ (R_2 \geq c\ L/m)\ \cup\ \cdots\ \cup\ (R_m \geq c\ L/m)\ ]\\
&\leq \sum_{i=1}^{m} \Pr[R_i \geq c\ L/m]\\
&\leq \frac{m}{160,000} = \frac{1}{16,000}.
\end{aligned}$$