

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

CHẶN CHERNOFF VÀ CHẶN Hoeffding
(Phần 2)

ThS Nguyễn Thu Hằng

Hà nội, tháng 1 năm 2024

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

CHẶN CHERNOFF VÀ CHẶN Hoeffding
(Phần 2)

Xác nhận của bộ môn

Hà Nội, tháng 1 năm 2024

MỤC LỤC

Lời giới thiệu	1
1. Kiến thức chuẩn bị	2
2. Chặn Hoeffding	4
3. Một số ví dụ về ứng dụng của chặn Hoeffding	9

LỜI GIỚI THIỆU

Hàm phân phối của tổng các biến ngẫu nhiên độc lập $Z = X_1 + X_2 + \dots + X_n$ là một thông tin quan trọng. Đôi khi, phân phối của Z còn có thể cho chúng ta biết về phân phối của $X_i, i = 1, 2, \dots$. Trong lịch sử, đã có rất nhiều kết quả nghiên cứu trong các trường hợp khác nhau. Trong trường hợp X_i độc lập cùng phân phối thì câu hỏi về hàm phân phối của Z được trả lời khá trọn vẹn trong rất nhiều trường hợp khi X_i có phân phối quen thuộc thường gặp và có ứng dụng quan trọng trong lý thuyết thống kê hiện đại. Tuy nhiên, trong trường hợp X_i không cùng phân phối, vấn đề trở nên phức tạp hơn rất nhiều. Đa số các kết quả nghiên cứu đáng chú ý liên quan đến phân phối của Z . Một câu hỏi nữa là độ chính xác của đáng chú ý liên quan đến đó? Một loạt các bất đẳng thức như bất đẳng thức Markov, bất đẳng thức Chebyshev... cho phép ta đánh giá xác suất đuôi $P(Z \geq a)$ trong trường hợp biết kỳ vọng, phương sai của biến ngẫu nhiên. Bất đẳng thức mạnh hơn cho một chặn trên của phân phối đuôi của Z giảm theo [hàm mũ](#) là bất đẳng thức Chernoff. Trong một số báo cáo học thuật trước, tôi đã trình bày lại một số bất đẳng thức xác suất quen thuộc này cùng một số ví dụ thực tiễn có thể áp dụng trong giảng dạy. Theo luồng nghiên cứu đó, trong báo cáo này, tôi tiếp tục nghiên cứu, tổng hợp và trình bày lại về chặn Hoeffding - một bất đẳng thức cho chặn trên của xác suất một tổng các biến ngẫu nhiên sai lệch với giá trị kỳ vọng. Bất đẳng thức Hoeffding được chứng minh bởi Wassily Hoeffding (1962) và được coi là mở rộng của bất đẳng thức Chernoff.

Báo cáo học thuật chia làm ba phần

Phần 1: Trình bày về hàm sinh mô-men và nhắc lại bất đẳng thức Markov và Chebyshev và bất đẳng thức Chernoff

Phần 2: Trình bày về chặn Hoeffding.

Phần 3. Trình bày một số ví dụ cơ bản có thể ứng dụng chặn Hoeffding.

1. Kiến thức chuẩn bị

1.1. Hàm sinh mômen

Định nghĩa 1.1. Cho X là một biến ngẫu nhiên bất kì. Hàm sinh mô-men của X là

$$M_X(t) = E[e^{tX}].$$

Nếu X là biến ngẫu nhiên rời rạc

$$M_X(t) = \sum_{i=1}^{\infty} e^{tx_i} p_i.$$

Nếu X là biến ngẫu nhiên liên tục

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx.$$

Tính chất của hàm sinh mô-men

i) Khai triển $e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \dots + \frac{t^n X^n}{n!} + \dots$

Suy ra

$$M_X(t) = 1 + tEX + \frac{t^2 E(X^2)}{2!} + \dots + \frac{t^n E(X^n)}{n!} + \dots$$

Do đó, hàm sinh mô-men bao gồm tất cả các mô-men của X .

ii) Nếu X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập và $X = X_1 + X_2 + \dots + X_n$ thì

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t).$$

Thật vậy, vì X_1, X_2, \dots, X_n độc lập nên $e^{tX_1}, e^{tX_2}, \dots, e^{tX_n}$ cũng độc lập. Ta có

$$\begin{aligned} M_X(t) &= E[e^{tX}] = E[e^{t(X_1 + X_2 + \dots + X_n)}] \\ &= E\left[\prod_{i=1}^n e^{tX_i}\right] = \prod_{i=1}^n E[e^{tX_i}] = \prod_{i=1}^n M_{X_i}(t). \end{aligned}$$

Hàm sinh mô-men của một số phân phối thường gặp

Phân phối	Hàm sinh mô-men
Phân phối 0-1 tham số p	$1 - p + pe^t$
Phân phối nhị thức $B(n, p)$	$(1 - p + pe^t)^n$
Phân phối Poisson	$e^{\lambda(e^t - 1)}$
Phân phối đều liên tục $U(a, b)$	$\frac{e^{tb} - e^{ta}}{t(b - a)}$
Phân phối mũ $E(\lambda)$	$(1 - t\lambda^{-1})^{-1}, t < \lambda$
Phân phối chuẩn $N(\mu, \sigma^2)$	$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

1.2. Hàm lồi

Định nghĩa 1.2. Hàm liên tục $f(x)$ được gọi là lồi trên khoảng I nếu và chỉ nếu với mọi $0 < p < 1$ và với mọi $x, y \in I$ ta có

$$f(px + (1 - p)y) \leq pf(x) + (1 - p)f(y).$$

Mệnh đề 1.3. Một hàm $f(x)$ liên tục và khả vi cấp 2 là lồi trên I nếu

$$f''(x) \geq 0, \forall x \in I.$$

Mệnh đề 1.4. (Bất đẳng thức Jensen) Nếu hàm $f(x)$ liên tục và lồi trên I thì với mọi số dương p_1, p_2, \dots, p_N thỏa mãn $p_1 + p_2 + \dots + p_N = 1$ và mọi số $x_1, x_2, \dots, x_N \in I$ ta có

$$f\left(\sum_{i=1}^N p_i x_i\right) \leq \sum_{i=1}^N p_i f(x_i).$$

Ta thấy rằng, nếu X là biến ngẫu nhiên nhận giá trị trong đoạn $[a, b]$ thì hàm e^{hX} là hàm lồi nên đồ thị của nó nằm dưới đường thẳng đi qua hai điểm $M(a, e^{ha})$ và $N(b, e^{hb})$. Do đó ta có

$$e^{hX} \leq \frac{b - X}{b - a} e^{ha} + \frac{X - a}{b - a} e^{hb}, a \leq X \leq b.$$

Từ đó ta có hệ quả sau

Hệ quả 1.5. Cho X là biến ngẫu nhiên nhận giá trị trong đoạn $[a, b]$. Khi đó, với mọi số thực h ta có

$$Ee^{hX} \leq \frac{b - EX}{b - a} e^{ha} + \frac{EX - a}{b - a} e^{hb}, a \leq X \leq b. \quad (1)$$

1.3. Một số bất đẳng thức xác suất khác

Bất đẳng thức Markov. Cho X là một biến ngẫu nhiên nhận giá trị không âm. Khi đó, với mọi $a \geq 0$ ta có

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Bất đẳng thức Chebyshev. Với mọi $a > 0$ ta có

$$P(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$

Bất đẳng thức Chernoff. Với mọi $t > 0$ ta có

$$P(X \geq a) \leq \min_{t>0} \frac{E[e^{tX}]}{e^{ta}}$$

và

$$P(X \leq a) \leq \min_{t>0} e^{ta} E[e^{-tX}].$$

2. Chặn Hoeffding

Cho X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập với môment bậc nhất và bậc hai hữu hạn.

Đặt $S = X_1 + X_2 + \dots + X_n$ và $\bar{X} = \frac{S}{n}$

$$\mu = E\bar{X} = \frac{ES}{n}; \sigma^2 = nVar(\bar{X}) = \frac{VarS}{n}.$$

Theo bất đẳng thức Chebychev ta có

$$P(|\bar{X} - \mu| \geq t) \leq \frac{\sigma^2}{nt^2},$$

và

$$P(\bar{X} - \mu \geq t) \leq \frac{1}{1 + \frac{nt^2}{\sigma^2}}.$$

Từ chặn Chernoff ta thu được

$$P(\bar{X} - \mu \geq t) = P(E - ES \geq nt) \leq Ee^{h(S-ES-nt)}, \quad (2)$$

Trong đó h là số dương bất kì. Trong trường hợp các biến ngẫu nhiên độc lập thì

$$Ee^{h(S-ES-nt)} = e^{-hnt} \prod_{i=1}^n Ee^{h(X_i-EX_i)}. \quad (3)$$

Tiếp theo ta đánh giá giá trị kì vọng trong vế phải của (2). Một loạt các định lí sau đây được phát biểu và chứng minh bởi Wassily Hoeffding (còn được gọi là chặn Hoeffding).

Định 2.1. Cho $X_i, i = 1, 2, \dots, n$ là các biến ngẫu nhiên độc lập, nhận giá trị trong đoạn $[0, 1]$ thì với $0 < t < 1 - \mu$ ta có

$$P(\bar{X} - \mu \geq t) \leq \left(\left(\frac{\mu}{\mu+t} \right)^{\mu+t} \left(\frac{1-\mu}{1-\mu-t} \right)^{1-\mu-t} \right)^n \quad (4)$$

$$\leq e^{-nt^2 g(\mu)} \quad (5)$$

$$\leq e^{-2nt^2}, \quad (6)$$

Trong đó

$$g(\mu) = \frac{1}{1-2\mu} \ln \frac{1-\mu}{\mu}, \quad 0 < \mu < \frac{1}{2},$$

$$g(\mu) = \frac{1}{2\mu(1-\mu)}, \quad \frac{1}{2} \leq \mu < 1.$$

Chứng minh. Từ (2) và (3) suy ra với $h > 0$ thì

$$P(\bar{X} - \mu \geq t) \leq e^{-hnt-hn\mu} \prod_{i=1}^n Ee^{hX_i}.$$

Đặt $\mu_i = EX_i$ thì $n\mu = \mu_1 + \mu_2 + \dots + \mu_n$. Áp dụng Hệ quả 1.5 cho $X = X_i$ và $a = 0, b = 1$ ta được

$$\prod_{i=1}^n E e^{hX_i} \leq \prod_{i=1}^n (1 - \mu_i + \mu_i e^h).$$

Áp dụng bất đẳng thức Cauchy ta lại có

$$\left(\prod_{i=1}^n (1 - \mu_i + \mu_i e^h) \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n (1 - \mu_i + \mu_i e^h) = 1 - \mu + \mu e^h.$$

Suy ra,

$$P(\bar{X} - \mu \geq t) \leq \left(e^{-hnt - hn\mu} (1 - \mu + \mu e^h) \right)^n, \forall h > 0 \quad (7)$$

Vế phải của (7) đạt cực tiểu tại $h = h_0$, với

$$h_0 = \ln \frac{(1 - \mu)(\mu + t)}{\mu(1 - \mu - t)}.$$

Mặt khác, do $0 < t < 1 - \mu$ nên $h_0 > 0$. Trong (7) thay h bởi h_0 ta thu được (4).

Để chứng minh (5), ta viết lại vế phải của (4) dạng $\exp(-nt^2 G(t, \mu))$, trong đó

$$G(t, \mu) = \frac{\mu + t}{t^2} \ln \frac{\mu + t}{\mu} + \frac{1 - \mu - t}{t^2} \ln \frac{1 - \mu - t}{1 - \mu}.$$

Ta chỉ cần chứng minh $G(t, \mu)$ đạt giá trị cực tiểu chính là $g(\mu)$. Thật vậy,

Đạo hàm 2 vế của $G(t, \mu)$ theo t ta được

$$\begin{aligned} t^2 \frac{\partial G(t, \mu)}{\partial t} &= \left(1 - 2 \frac{1 - \mu}{t} \right) \ln \left(1 - \frac{t}{1 - \mu} \right) - \left(1 - 2 \frac{\mu + t}{t} \right) \ln \left(1 - \frac{t}{\mu + t} \right) \\ &= H \left(\frac{t}{1 - \mu} \right) - H \left(\frac{t}{\mu + t} \right), \end{aligned}$$

Trong đó, $H(x) = \left(1 - \frac{2}{x} \right) \ln(1 - x)$. Vì giả thiết $0 < t < 1 - \mu$ nên $0 \leq \frac{t}{1 - \mu} \leq 1$ và

$0 \leq \frac{t}{\mu + t} \leq 1$. Hơn nữa, với $|x| < 1$ ta có khai triển

$$H(x) = 2 + \left(\frac{2}{3} - \frac{1}{2} \right) x^2 + \left(\frac{2}{4} - \frac{1}{3} \right) x^3 + \left(\frac{2}{5} - \frac{1}{4} \right) x^4 + \dots$$

với các hệ số trong khai triển đều dương nên $H(x)$ là hàm tăng trên $(0, 1)$. Do đó,

$\frac{\partial G(t, \mu)}{\partial t} > 0$ khi và chỉ khi $\frac{t}{1 - \mu} > \frac{t}{\mu + t}$ tức là $t > 1 - 2\mu$.

Suy ra, với $t > 1 - 2\mu$, $G(t, \mu)$ đạt cực tiểu tại $t = 1 - 2\mu$, giá trị cực tiểu là

$$G(1 - 2\mu, \mu) = \frac{\ln \frac{1 - \mu}{\mu}}{1 - 2\mu} = g(\mu).$$

Với $1 - 2\mu \leq 0$ thì $G(t, \mu)$ đạt cực tiểu tại $t = 0$ và giá trị cực tiểu là $\frac{1}{2\mu(1-\mu)} = g(\mu)$.

Ta chứng minh xong (5).

Lại thấy rằng $g(\mu) \geq g\left(\frac{1}{2}\right) = 2$. Từ đó dễ dàng suy ra (6). Ta hoàn thành chứng minh Định lí 1.

Chú ý.

- Trong trường hợp $a \leq X_i \leq b, i = 1, 2, \dots, n$, thì μ và t trong (4), (5), (6) được thay bằng $\frac{\mu-a}{b-a}$ và $\frac{t}{b-a}$.
- Nếu $t > 1 - \mu$ thì với giả thiết của Định lí 1, xác suất trong (4) bằng 0.
- Với $t = 1 - \mu$, bất đẳng thức (4) vẫn đúng nếu vế phải thay bằng các giới hạn của nó khi t tiến tới $1 - \mu$, tức là μ^n . Trong trường hợp đặc biệt này, bất đẳng thức (4) đạt được dấu bằng. Cụ thể là, nếu $t = 1 - \mu$ thì $P(\bar{X} - \mu \geq t) = P(\bar{X} = 1) = P(S = n)$ và $P(S = n) = \mu^n$ nếu $P(X_i = 0) = 1 - \mu, P(X_i = 1) = \mu, i = 1, \dots, n$. Hay nói cách khác S có phân phối nhị thức với tham số n và μ .

Như vậy, có thể nói chặn trong (4) là chặn tốt nhất có thể thu được từ bất đẳng thức (3) dưới các giả thiết của Định lí. Nghĩa là, đó là giá trị nhỏ nhất theo h trong vế phải của (3) khi X_i có phân phối nhị thức với tham số n và μ .

Định lí sau đây là mở rộng của (6) trong trường hợp các biến ngẫu nhiên X_i không nhất thiết phải bị chặn trong các khoảng giống nhau.

Định lí 2.2. Cho $X_i, i = 1, 2, \dots, n$ là các biến ngẫu nhiên độc lập và $a_i \leq X_i \leq b_i, i = 1, 2, \dots, n$. Khi đó, với $t > 0$ ta có

$$P(|\bar{X} - \mu| \geq t) \leq 2e^{-\frac{2n^2 t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

Chứng minh.

Đặt $\mu_i = EX_i$, (3) trở thành

$$Ee^{h(S-ES-nt)} = e^{-hnt} \prod_{i=1}^n Ee^{h(X_i-\mu_i)}.$$

Từ Hệ quả 1.5, với $h > 0$

$$Ee^{h(X_i - \mu_i)} \leq e^{-h\mu_i} \left(\frac{b_i - \mu_i}{b_i - a_i} e^{ha_i} + \frac{\mu_i - a_i}{b_i - a_i} e^{hb_i} \right) = e^{L(h_i)},$$

trong đó

$$L(h_i) = -h_i p_i + \ln(1 - p_i + p_i e^{h_i}),$$

$$h_i = h(b_i - a_i), \quad p_i = \frac{\mu_i - a_i}{b_i - a_i}.$$

Đạo hàm và đạo hàm cấp 2 của $L(h_i)$ theo h_i là

$$L'(h_i) = -p_i + \frac{p_i}{(1 - p_i)e^{-h_i} + p_i},$$

$$L''(h_i) = \frac{p_i(1 - p_i)e^{-h_i}}{((1 - p_i)e^{-h_i} + p_i)^2} = \frac{(1 - p_i)e^{-h_i}}{(1 - p_i)e^{-h_i} + p_i} \left(1 - \frac{(1 - p_i)e^{-h_i}}{(1 - p_i)e^{-h_i} + p_i} \right).$$

Ta thấy, $L''(h_i)$ có dạng $u(1 - u)$, $0 < u < 1$ nên $L''(h_i) \leq 1/4$. Đạo hàm cấp ba $L'''(h_i) < 0$. Áp dụng khai triển Taylor ta có

$$L(h_i) \leq L(0) + L'(0)h_i + \frac{1}{8}h_i^2 = \frac{1}{8}h_i^2 = \frac{1}{8}h^2(b_i - a_i)^2.$$

Do đó,

$$Ee^{h(X_i - \mu_i)} \leq e^{\frac{1}{8}h^2(b_i - a_i)^2}.$$

Kết hợp với (2) và (3) ta suy ra

$$P(\bar{X} - \mu \geq t) \leq e^{-hnt} \prod_{i=1}^n Ee^{h(X_i - EX_i)} \leq e^{-hnt + \frac{1}{8}h^2 \sum_{i=1}^n (b_i - a_i)^2} \quad (8)$$

Vế phải của (8) đạt cực tiểu tại $h_0 = \frac{4nt}{\sum_{i=1}^n (b_i - a_i)^2}$. Thay h_0 vào (8) ta được

$$P(\bar{X} - \mu \geq t) \leq e^{-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

Tương tự

$$P(\bar{X} - \mu \leq -t) = P(-(\bar{X} - \mu) \geq -t) \leq e^{-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

Suy ra

$$P(|\bar{X} - \mu| \geq t) \leq 2e^{-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

Định lí 2 được chứng minh.

Hệ quả 2.3. Nếu X_i là dãy biến ngẫu nhiên độc lập, cùng phân phối 0-1 tham số p . Khi đó $a_i = 0, b_i = 1$ và tổng $S_n = X_1 + \dots + X_n \sim B(n, p)$. Ta có

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - p\right| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Hệ quả 2.4. Nếu Y_1, \dots, Y_m và Z_1, \dots, Z_n là các biến ngẫu nhiên độc lập, cùng nhận giá trị trong $[a, b]$, $\bar{Y} = \frac{Y_1 + \dots + Y_m}{m}$, $\bar{Z} = \frac{Z_1 + \dots + Z_n}{n}$ thì với $t > 0$

$$P(\bar{Y} - \bar{Z} - (E\bar{Y} - E\bar{Z}) \geq t) \leq e^{-\frac{2t^2}{(m^{-1} + n^{-1})(b-a)^2}}.$$

Các ước lượng trong định lí sau đây phụ thuộc vào phương sai $\frac{\sigma^2}{n}$ của \bar{X} .

Định lí 2.5. Cho $X_i, i = 1, 2, \dots, n$ là các biến ngẫu nhiên độc lập, $EX_i = 0, X_i \leq b, i = 1, 2, \dots, n$. Khi đó, với $0 < t < b$ ta có

$$\begin{aligned} P(\bar{X} \geq t) &\leq \left(\left(1 + \frac{bt}{\sigma^2}\right)^{-\left(1 + \frac{bt}{\sigma^2}\right)\frac{\sigma^2}{b^2 + \sigma^2}} \left(1 - \frac{t}{b}\right)^{-\left(1 - \frac{t}{b}\right)\frac{b^2}{b^2 + \sigma^2}} \right)^n \\ &\leq e^{-\frac{nt}{b} \left[\left(1 + \frac{\sigma^4}{bt}\right) \ln\left(1 + \frac{bt}{\sigma^2}\right) - 1 \right]}. \end{aligned}$$

3. Một số ví dụ về ứng dụng của chặn Hoeffding

Ví dụ 1. Chặn Hoeffding có thể xác định độ tin cậy trong một phép thử. Xét một phép thử chỉ xảy ra hai khả năng (ví dụ như gieo đồng xu). Giả sử p là xác suất xảy ra biến cố thứ nhất A_1 . Sau n lần thực hiện phép thử thấy có k lần xuất hiện A_1 . Đặt $\hat{p} = \frac{k}{n}$ là tần suất xuất hiện A_1 . Đây cũng là một ước lượng cho p . Vậy \hat{p} với p có quan hệ với nhau như thế nào? Chúng ta có thể dùng chặn Hoeffding để ước lượng cho sai số giữa \hat{p} và p như sau:

Với $\varepsilon > 0$ ta có

$$P(\hat{p} - p \geq \varepsilon) \leq e^{-2n\varepsilon^2},$$

$$P(\hat{p} - p \leq -\varepsilon) \leq e^{-2n\varepsilon^2}$$

Suy ra

$$P(|\hat{p} - p| \leq \varepsilon) \geq 1 - e^{-2n\varepsilon^2}.$$

Như vậy, với độ tin cậy $1 - e^{-2n\varepsilon^2}$, ta có thể nói rằng p thuộc khoảng $(\hat{p} - \varepsilon, \hat{p} + \varepsilon)$. Nếu ε tăng thì $e^{-2n\varepsilon^2}$ giảm. Khi đó ta có thể chắc chắn hơn về kết quả tìm được.

Ta lấy một trường hợp cụ thể hơn. Gieo một đồng xu 1000 lần thấy có 552 lần xuất hiện mặt ngửa. Hỏi kết luận đồng xu cân đối tin tưởng được bao nhiêu phần trăm?

Theo như tính toán phía trên, $n = 1000, k = 552, \hat{p} = 0,552, p = 0,5$.

Đặt $\varepsilon = \hat{p} - p = 0,052$. Ta có $\alpha = e^{-2n\varepsilon^2} = 0,0045$. Vậy với độ tin cậy $1 - \alpha = 0,9955$ ta có thể nói đồng xu cân đối.

Ví dụ 2. Tìm vị trí của chất điểm.

Xét một chất điểm thực hiện bước đi ngẫu nhiên trên một đường thẳng. Nó bắt đầu tại vị trí ban đầu 0 và mỗi bước đi nó di chuyển một bước về phía trước hoặc về phía sau với xác suất như nhau. Tổng quát, từ vị trí i nó có thể di chuyển đến vị trí $i + 1$ hoặc $i - 1$ với xác suất như nhau và bằng $\frac{1}{2}$. Chúng ta muốn tìm hiểu sau n lần di chuyển, chất điểm sẽ ở vị trí nào?

Có thể kí hiệu sự di chuyển của chất điểm như một dãy biến ngẫu nhiên X_1, X_2, \dots, X_n thỏa mãn X_i nhận giá trị 1 hoặc -1 với cùng xác suất $\frac{1}{2}$.

Đặt $X = X_1 + X_2 + \dots + X_n$ chính là vị trí của chất điểm sau n lần di chuyển. Khi đó bài toán quy về tìm một giới hạn trên cho $|X|$.

Vì $EX = 0$ nên theo bất đẳng thức Hoeffding ta có

$$P\left(\left|\frac{1}{n}X - 0\right| \geq \varepsilon\right) \leq e^{-\frac{2n^2\varepsilon^2}{2n}}.$$

Do đó nếu chọn $\varepsilon = \sqrt{\frac{\log n}{n}}$ thì với xác suất ít nhất là $1 - \frac{1}{n}$ ta sẽ có $\left|\frac{X}{n}\right| \leq \sqrt{\frac{\log n}{n}}$ hay $|X| \leq \sqrt{n \log n}$. Tức là với khả năng cao, chất điểm sẽ không ở quá vị trí cách vị trí ban đầu không quá $\sqrt{n \log n}$.

Ví dụ 3. Gieo 1 quân xúc xắc cân đối n lần. Gọi X_i là số chấm trên mặt xuất hiện ở lần gieo thứ i . Dễ thấy $EX_i = \frac{7}{2}, i = 1, 2, \dots, n$. Đặt $X = X_1 + \dots + X_n$ là biến ngẫu nhiên chỉ tổng số

chấm trên các lần gieo. Theo bất đẳng thức Hoeffding, ta có thể ước lượng tổng số chấm trong n lần gieo theo chặn trong Định lí 2 với $t = 5\sqrt{n \ln n}$.

$$P\left(\left|X - \frac{7}{2}n\right| \geq 5\sqrt{n \ln n}\right) \leq \frac{2}{n^2}.$$

Nếu chúng ta muốn đảm bảo với xác suất lệch là không quá 0,01 thì chỉ cần thực hiện đủ n lần gieo trong đó

$$\frac{2}{n^2} < 0,01$$

Tương đương với $n > \sqrt{200} \approx 14$ lần gieo. Nói cách khác, với 15 lần gieo, chúng ta có thể tự tin rằng tổng số nốt nằm trong khoảng 20 đến 85.

KẾT LUẬN

Báo cáo đã trình bày về chặn Hoeffding với các dạng thể hiện khác nhau và một số ví dụ thực tế có thể sử dụng chặn Hoeffding. Tài liệu có thể dùng làm tài liệu tham khảo cho giảng viên, sinh viên, làm sinh động thêm những ứng dụng của môn xác suất thống kê đối với các vấn đề thực tiễn trong cuộc sống. Tuy nhiên, đây chỉ là một nội dung nhỏ trong vô vàn kết quả về bất đẳng thức xác suất. Trong thời gian tới, tác giả sẽ tiếp tục tìm hiểu về phân phối đuôi của biến ngẫu nhiên trong các trường hợp khác nhau của biến ngẫu nhiên cũng như nhiều bất đẳng thức khác nhằm hoàn thiện hơn về nội dung nghiên cứu và nâng cao chất lượng bài giảng cũng như làm tốt mục tiêu gắn lí thuyết với thực tiễn.

TÀI LIỆU THAM KHẢO

1. N.D.Tiến, V.V.Yên, *Lý thuyết xác suất*, NXB Giáo dục 2006.
2. N.C.Văn, T.T.Ninh, *Giáo trình Lý thuyết xác suất và thống kê toán*, NXB Đại học Kinh tế quốc dân, 2004
3. N.T.T.Hà, *Một số bất đẳng thức xác suất và ứng dụng*, Khóa luận tốt nghiệp Đại học.
4. Wassily Hoeffding, Probability Inequalities for Sums of Bounded Random Variables, *Journal of the American Statistical Association*, Vol. 58, No. 301 (Mar., 1963), pp. 13-30.