Tạp chí Phát triển Khoa học và Công nghệ – Khoa học Tự nhiên, 3(4):294-298
Open Access Full Text Article

Bài Nghiên cứu

Luật yếu số lớn với dãy được đánh số ngẫu nhiên của các biến ngẫu nhiên *m* phụ thuộc

Trần Lộc Hùng[1], Nguyễn Tấn Nhựt[2,*]

TÓM TẮT

Trước tiên, chúng tôi thiết lập các bất đẳng thức liên quan đến chặn trên cho xác suất của tổng một số lượng ngẫu nhiên các biến ngẫu nhiên thỏa mãn những điều kiện nhất định. Cụ thể hơn, ở Định lí 1, các biến này được giả định phải nhận giá trị trên một khoảng bị chặn và đặc biệt là chúng được đặt dưới giả thiết *m* phụ thuộc thay vì độc lập theo thường lệ, trong đó độc lập chỉ là trường hợp riêng của m phụ thuộc khi m bằng 0. Đối với một số chỉ số có phân phối quen thuộc, có thể tiếp tục thực hiện những ước tính hợp lí cho số hạng kì vọng ở vế phải của hai bất đẳng thức trong Định lí 1 để nhận được các chặn kiểu Chernoff-Hoeffding. Với mỗi trường hợp đáp ứng như thế của biến ngẫu nhiên chỉ số, các chặn đó sẽ được sử dụng vào việc chứng minh rằng có luật yếu số lớn trên dãy biến ngẫu nhiên m phụ thuộc tương ứng và tốc độ hội tụ là mũ. Tiếp theo, ở Định lí 2, chỉ số có phân phối Poisson được chọn làm điển hình để trình bày. Cuối cùng, định lí này được minh họa thông qua một hình ảnh xây dựng từ những giá trị mô phỏng dành cho một dãy 1 phụ thuộc. Ở đây, cách thức tạo ra dãy 1 phụ thuộc từ một dãy độc lập đã thực hiện sẽ phần nào giúp độc giả hiểu rõ hơn về cấu trúc m phụ thuộc.

Từ khoá: luật yếu số lớn, tổng ngẫu nhiên các biến ngẫu nhiên, m phụ thuộc, bất đẳng thức Chernoff-Hoeffding

GIỚI THIỆU

Cho $Y_1, Y_2, \ldots$ là các biến ngẫu nhiên độc lập và đồng nhất phân phối với biến ngẫu nhiên có giá trị nguyên không âm Y. Cho $X_1, X_2, \ldots$ là các biến ngẫu nhiên đồng nhất phân phối với biến ngẫu nhiên X nhận giá trị trên khoảng đóng [0,1], độc lập với các biến ngẫu nhiên $Y_1, Y_2, \ldots$. Nội dung bài viết xoay quanh các tổng sau đây:

$N_n = \sum_{j=1}^{n} Y_j$, $S_{N_n} = \sum_{j=1}^{N_n} X_j$ và $\bar{X}_{N_n} = N_n^{-1} S_{N_n}$.

Khi $N_n = 0$, qui ước $S_0 = \bar{X}_0 = 0$. Định lí 1 ở Phần kết quả sẽ chỉ ra một chặn trên cho xác suất $P(S_{N_n} > N_n x)$ mà trong đó các biến ngẫu nhiên $X_1, X_2, \ldots$ được giả định là m phụ thuộc và x là một số thực thích hợp (xem Bổ đề 1, Định lí 1).

Định nghĩa 1 (*m phụ thuộc*). Cho m là một số nguyên không âm. Các biến ngẫu nhiên $X_1, X_2, \ldots$ được gọi là m phụ thuộc nếu với mọi số nguyên dương n, các tập biến ngẫu nhiên $\{X_1, X_2, \ldots, X_n\}$ và $\{X_{n+m+1}, X_{n+m+2}, \ldots\}$ độc lập[1].

Từ công trình của Hoeffding[2] những gì cần thiết để chứng minh Định lí 1, kết quả then chốt của bài viết này, được tóm tắt lại trong Bổ đề 1 dưới đây.

Với $\mu \in (0,1)$, định nghĩa hàm $I_\mu : [0,1] \to [0,\infty]$ bởi công thức

$$I_\mu(x) = x \log\left(\frac{x}{\mu}\right) + (1-x)\log\left(\frac{1-x}{1-\mu}\right).$$

Bổ đề 1. Nếu $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên m phụ thuộc và đồng nhất phân phối với biến ngẫu nhiên X nhận giá trị trên khoảng [0,1] và $\mathrm{E}X = \mu \in (0,1)$, thì

$$\mathrm{P}\left(\frac{1}{n}\sum_{j=1}^{n} X_j > x\right) \le \exp\left(-\left[\frac{n}{m+1}\right] I_\mu(x)\right)$$

với mọi $x \in [\mu, 1)$, và

$$\mathrm{P}\left(\frac{1}{n}\sum_{j=1}^{n} X_j < x\right) \le \exp\left(-\left[\frac{n}{m+1}\right] I_\mu(x)\right)$$

[1]*Trường Đại học Tài chính – Marketing TPHCM;*

[2]*Xã Bình Thành, Huyện Lấp Vò, Tỉnh Đồng Tháp*

Liên hệ

Nguyễn Tấn Nhựt, Xã Bình Thành, Huyện Lấp Vò, Tỉnh Đồng Tháp

Email: ntn.nhut@gmail.com

Lịch sử

- Ngày nhận: 24-11-2018
- Ngày chấp nhận: 22-7-2019
- Ngày đăng: 31-12-2019

DOI : 10.32508/stdjns.v3i4.528

Trích dẫn bài báo này: Hùng T L, Nhựt N T. **Luật yếu số lớn với dãy được đánh số ngẫu nhiên của các biến ngẫu nhiên *m* phụ thuộc**. *Sci. Tech. Dev. J. - Nat. Sci.;* 3(4):294-298.

với mọi $x \in (0,\mu]$.
Trong công thức trên, $\left[\frac{n}{m+1}\right]$ là số nguyên lớn nhất nhỏ hơn hay bằng $\frac{n}{m+1}$. Có thể chứng minh bổ đề này theo cách thức mà Hoeffding đã chứng minh định lí đầu tiên và một thảo luận mở rộng cho tổng các biến m phụ thuộc trong bài báo của ông ấy[2]. Một chặn trên giảm về 0 theo tốc độ hàm mũ đối với một xác suất có dạng như trong Bổ đề 1 thường gọi là chặn Chernoff-Hoeffding (xem[2,3]).
Từ phần Luật số lớn về sau bài viết sẽ xét những trường hợp cụ thể của các giả thiết, bắt đầu với Y có phân phối Poisson, Định lí 2 là một dạng luật yếu số lớn được xác định dựa trên việc áp dụng Hệ quả 1 ở Phần kết quả. Sau đó, trong Phần kết luận, X được lấy là trung bình cộng của hai biến ngẫu nhiên độc lập có cùng phân phối đều trên $[a,b]$ để cung cấp chất liệu cho mô phỏng ở **Hình 1**, đây là một thí dụ cho kết quả mà Định lí 2 xác định. Hơn nữa, những thảo luận trong phần này gợi ý rằng có thể thay giả thiết X nhận giá trị trên khoảng [0,1] thành một khoảng bất kì $[a,b]$, mà trong đó a, b hữu hạn.

KẾT QUẢ

Với những gì đã thiết lập ở Phần giới thiệu, định lí sau đây là kết quả chính.
Định lí 1. Với mọi $x \in [\mu,1)$,

$$P(S_{N_n} > xN_n) \leq E\exp\left(-\left[(m+1)^{-1}N_n\right] I_\mu(x)\right); \quad (1)$$

với mọi $x \in (0,\mu]$,

$$P(S_{N_n} < xN_n) \leq E\exp\left(-\left[(m+1)^{-1}N_n\right] I_\mu(x)\right). \quad (2)$$

Chứng minh. Chỉ cần chứng minh bất đẳng thức (1), bất đẳng thức (2) được chứng minh tương tự.
Trước tiên, áp dụng luật xác suất toàn phần và giả thiết N_n độc lập với các X_j để có

$$\begin{aligned} P(S_{N_n} > xN_n) &= \textstyle\sum_{k=0}^{\infty} P(N_n = k)\, P(S_{N_n} > kx | N_n = k) \\ &\leq P(N_n = 0) + \textstyle\sum_{k=1}^{\infty} P(N_n = k)\, P(S_k > kx) \end{aligned}$$

Sau đó, theo Bổ đề 1,

$$P(S_k > kx) \leq \exp\left(-\left[\frac{k}{m+1}\right] I_\mu(x)\right)$$

nên thu được đánh giá tiếp theo như sau:

$$\begin{aligned} P(S_{N_n} > xN_n) &\leq \textstyle\sum_{k=0}^{\infty} P(N_n = k) \exp\left(-\left[\frac{k}{m+1}\right] I_\mu(x)\right) \\ &= E\exp\left(-\left[\frac{N_n}{m+1}\right] I_\mu(x)\right). \end{aligned}$$

Chứng minh xong.
Trong một số trường hợp nhất định, sử dụng hệ quả bên dưới sẽ tiện lợi hơn để chỉ ra $\bar{X}_{N_n}$ hội tụ theo xác suất về μ với tốc độ mũ, mà Định lí 2 ở Phần Luật số lớn là một trường hợp như thế.
Hệ quả 1. Với mọi số dương $\varepsilon < \min(\mu, 1-\mu)$,

$$P(|S_{N_n} - \mu N_n| > \varepsilon N_n) \leq 2E\exp\left(-2\left[\frac{N_n}{m+1}\right]\varepsilon^2\right).$$

Chứng minh. Trước tiên, có bất đẳng thức

$$\begin{aligned} &P(|s_{N_n} - \mu N_n| > \varepsilon N_n) \\ &\leq P(S_{N_n} > (\mu+\varepsilon)N_n) + P(S_{N_n} < (\mu-\varepsilon)N_n). \end{aligned}$$

Theo Định lí 1,

$$P(S_{N_n} > (\mu+\varepsilon)N_n) \leq E\exp\left(-\left[\frac{N_n}{m+1}\right] I(\mu+\varepsilon)\right).$$

và

$$P(S_{N_n} < (\mu-\varepsilon)N_n) \leq E\exp\left(-\left[\frac{N_n}{m+1}\right] I(\mu-\varepsilon)\right).$$

Cuối cùng, vì $I_\mu(x) \geq 2(x-\mu)^2$ với mọi $x \in (0,1)$, nên hệ quả được chứng minh.

Tuy phần này chỉ xem xét một trường hợp cụ thể là Y có phân phối Poisson, một số phân phối quen thuộc khác như Bernoulli hay hình học vẫn chung một kết luận như thế. Định lí dưới đây có dạng của luật yếu số lớn.

Định lí 2. Nếu Y là biến Poisson với tham số $\lambda > 0$ thì $\bar{X}_{N_n}$ hội tụ theo xác suất về μ, nghĩa là với mọi $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(|\bar{X}_{N_n} - \mu| > \varepsilon\right) = 0$$

Chứng minh. Với hai số dương ε và ε' mà $\varepsilon < \min(\mu, 1-\mu) \le \varepsilon'$, vì

$$\mathrm{P}\left(|\bar{X}_{N_n} - \mu| > \varepsilon'\right) \le \mathrm{P}\left(|\bar{X}_{N_n} - \mu| > \varepsilon\right)$$

nên chỉ cần tiến hành chứng minh này với giả định $\varepsilon < \min(\mu, 1-\mu)$.

Trước tiên, bởi luật xác suất toàn phần và tính độc lập của N_n đối với các X_j từ giả thiết, dễ thấy

$$\begin{aligned}
P\left(\left|\bar{X}_{N_n} - \mu\right| > \varepsilon\right) &= \textstyle\sum_{k=0}^{\infty} P(N_n = k) P\left(\left|\bar{X}_{N_n} - \mu\right| > \varepsilon | N_n = k\right)\\
&\le P(N_n = 0) + \textstyle\sum_{k=1}^{\infty} P(N_n = k) P\left(\left|\bar{X}_{N_n} - \mu\right| > \varepsilon | N_n = k\right)\\
&= P(N_n = 0) + \textstyle\sum_{k=1}^{\infty} P(N_n = k) P\left(|S_{N_n} - \mu N_n| > \varepsilon N_n | N_n = k\right)\\
&\le P(N_n = 0) + \textstyle\sum_{k=0}^{\infty} P(N_n = k) P\left(|S_{N_n} - \mu N_n| > \varepsilon N_n | N_n = k\right)\\
&= P(N_n = 0) + P\left(|S_{N_n} - \mu N_n| > \varepsilon N_n\right).
\end{aligned}$$

Sử dụng Hệ quả 1 kết hợp với giả thiết N_n là biến Poisson tham số $n\lambda$ và $\left[\frac{N_n}{m+1}\right] \ge \frac{N_n}{m+1} - 1$, suy ra

$$\begin{aligned}
&\mathrm{P}\left(|S_{N_n} - \mu N_n| > \varepsilon N_n\right) \le 2\,\mathrm{E}\exp\left(-2\left[\tfrac{N_n}{m+1}\right]\varepsilon^2\right)\\
&\le 2\exp\left(2\varepsilon^2\right) \textstyle\sum_{k=0}^{\infty} \exp(-n\lambda) \frac{(n\lambda)^k}{k!} \exp\left(-\frac{2k\varepsilon^2}{m+1}\right)\\
&= 2\exp\left(2\varepsilon^2 - n\lambda\right) \textstyle\sum_{k=0}^{\infty} \frac{1}{k!} \left((n\lambda)\exp\left(-\frac{2\varepsilon^2}{m+1}\right)\right)^k\\
&= 2\exp\left(-n\lambda\left(1 - \exp\left(-\frac{2\varepsilon^2}{m+1}\right)\right) + 2\varepsilon^2\right).
\end{aligned}$$

Số hạng cuối cùng tiến về 0 khi n tiến ra vô cực, và $\mathrm{P}(N_n = 0) = \exp(-n\lambda)$ cũng là một số hạng tiến về 0 khi n tiến ra vô cực, do đó định lí được chứng minh.

Đáng tiếc rằng, mặc dù chứng minh trên đã đồng thời chỉ ra tốc độ hội tụ của luật yếu số lớn này là mũ, nhưng chặn được sử dụng trong phép chứng minh vẫn chưa phản ánh sát với thực tế khi n hoặc $n\lambda$ chưa đủ lớn. Một phép tính chi tiết sẽ chỉ rõ hạn chế này trong phần tiếp theo.

KẾT LUẬN

Để tạo ra một mô phỏng cho **Định lí 2** và đồng thời minh họa việc có thể áp dụng các kết quả đã đề cập ở những phần trước khi X nhận giá trị trên một khoảng bị chặn bất kì như thế nào, ở phần này, X được cụ thể là trung bình cộng của hai biến ngẫu nhiên độc lập cùng có phân phối đều, nhưng là trên khoảng $[a,b]$ thay vì khoảng $[0,1]$ như trước. Nói cách khác,

$X = \frac{U+V}{2}$, trong đó U và V độc lập và cùng có phân phối đều trên $[a,b]$.

Giá trị kì vọng của X là $\frac{a+b}{2}$ và của $\frac{X-a}{b-a}$ là $\frac{1}{2}$. Vì thế,

$$\mathrm{P}\left(\left|S_{N_n} - \tfrac{a+b}{2}N_n\right| > \varepsilon N_n\right) = \mathrm{P}\left(\left|\textstyle\sum_{j=1}^{N_n} \frac{X_j - a}{b-a} - \frac{1}{2}N_n\right| > \frac{\varepsilon}{b-a}N_n\right).$$

Nếu $\varepsilon < \frac{b-a}{2}$ thì $\frac{\varepsilon}{b-a} < \frac{1}{2}$. Khi đó, theo Hệ quả 1,

$$\mathrm{P}\left(\left|\textstyle\sum_{j=1}^{N_n} \frac{X_j - a}{b-a} - \frac{1}{2}N_n\right| > \frac{\varepsilon}{b-a}N_n\right) \le 2\,\mathrm{E}\exp\left(-2\left(\tfrac{\varepsilon}{b-a}\right)^2\left[\tfrac{N_n}{m+1}\right]\right).$$

Lưu ý, trong bất đẳng thức trên, $X_1, X_2, \ldots$ là dãy m phụ thuộc và đồng nhất phân phối với X vừa mô tả ở phần đầu của mục này. Giả sử $U_1, U_2, \ldots$ là các biến ngẫu nhiên độc lập lẫn nhau và cùng có phân phối đều trên $[a,b]$, khi đó với $n \ge 1$, nếu $X_n = \frac{U_n + U_{n+1}}{2}$ thì $X_1, X_2, \ldots$ là ví dụ cho dãy một phụ thuộc và các thành phần đồng nhất phân phối với X.

Tiếp tục ước lượng kì vọng ở vế phải theo các bước tương tự khi chứng minh Định lí 2 để thu được đánh giá

$$\begin{aligned}&\mathrm{E}\exp\left(-2\left(\tfrac{\varepsilon}{b-a}\right)^2\left[\tfrac{N_n}{m+1}\right]\right)\\ &\leq \exp\left(-n\lambda\left(1-\exp\left(-\tfrac{2}{m+1}\left(\tfrac{\varepsilon}{b-a}\right)^2\right)\right)+2\left(\tfrac{\varepsilon}{b-a}\right)^2\right).\end{aligned}$$

Tóm lại,

$$\begin{aligned}&\mathrm{P}\left(\left|\overline{X}_{N_n}-\tfrac{a+b}{2}\right|>\varepsilon\right)\\ &\leq 4\exp\left(-n\lambda\left(1-\exp\left(-\tfrac{2}{m+1}\left(\tfrac{\varepsilon}{b-a}\right)^2\right)\right)+2\left(\tfrac{\varepsilon}{b-a}\right)^2\right).\end{aligned}$$

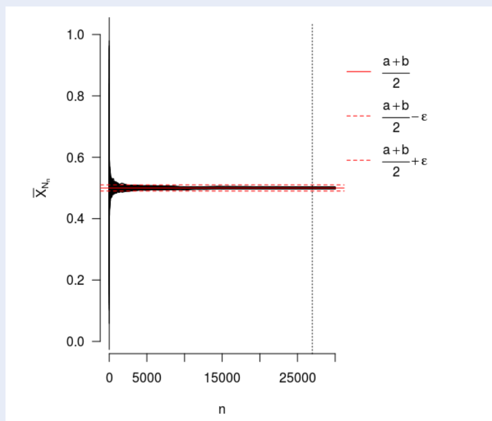
Nếu sử dụng bất đẳng thức này xác định n sao cho

$$\mathrm{P}\left(\left|\bar{X}_{N_n}-\frac{a+b}{2}\right|>\varepsilon\right)\leq\alpha$$

mà trong đó ε và α là các giá trị mong muốn cho trước, thì có thể sử dụng

$$n\geq\lambda^{-1}\left(1-\exp\left(-\tfrac{2}{m+1}\left(\tfrac{\varepsilon}{b-a}\right)^2\right)\right)^{-1}\left(\log\left(\tfrac{4}{\alpha}\right)+2\left(\tfrac{\varepsilon}{b-a}\right)^2\right).$$

Giả sử $\lambda=20$, $m=1, a=-1, b=2, \varepsilon=\frac{1}{100}$, $\alpha=\frac{1}{100}$, theo đó n được khuyên là không bé hơn 26962. Tuy nhiên, nếu quan sát **Hình 1** có thể sẽ thấy rằng đây là một ước lượng thô. Vài sửa đổi phù hợp trên các giả định ban đầu kết hợp với các bước đánh giá chặt chẽ hơn đối với các bất đẳng thức có lẽ sẽ mang lại một chặn trên hiệu quả hơn khi cần đến những tính toán số như thế này.



Hình 1: Minh họa luật số lớn trong trường hợp Y có phân phối Poisson tham số $\lambda=20$ và $X_1, X_2, \ldots$ là dãy một phụ thuộc, nghĩa là $m=1$, cụ thể $X_n=\frac{1}{2}(U_n+U_{n+1})$ mà trong đó $U_1, U_2, \ldots$ là các biến ngẫu nhiên độc lập lẫn nhau và cùng có phân phối đều trên khoảng [-1,2]. Ở đây có 200 quĩ đạo dừng tại $n=30000$ và $\varepsilon=\frac{1}{100}$.

XUNG ĐỘT LỢI ÍCH

Các tác giả không cạnh tranh lợi ích.

ĐÓNG GÓP CỦA CÁC TÁC GIẢ

Các tác giả có đóng góp như nhau cho bài viết này. Tất cả tác giả đã soạn bản thảo, đọc và duyệt phiên bản cuối cùng của bản thảo.

TÀI LIỆU THAM KHẢO

1. Ferguson TS. A course in large sample theory. Chapman & Hall texts in statistical science. In: Chapman and Hall/CRC; 1996.
2. Hoeffding W. Probability inequalities for sums of bounded random variables. Journal of the American statistical association. 1963;58(301):13–30.
3. Chernoff H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics. 1952;23(4):493–507.

Science & Technology Development Journal – Natural Sciences, 3(4):294-298

Open Access Full Text Article

Research Article

Weak law of large numbers for randomly indexed sequences of m-dependent random variables

Tran Loc Hung[1], **Nguyen Tan Nhut**[2,*]

ABSTRACT

First, we establish the inequalities related to the upper bound for the probability of the sum of a random number of random variables satisfying certain conditions. More specifically, in Theorem 1, these variables are assumed that get values on a bounded interval and, in particular, are setting under m-dependence assumption instead of the usual independence, where independence is merely the specific case of m-dependence when m equal to 0. For a random index with a familiar distribution, it is possible to proceed to make reasonable estimates for the expected terms on the right-hand side of the two inequalities in Theorem 1 to obtain Chernoff-Hoeffding-style bounds. Those bounds will be employed to prove that there is a weak law of large numbers for the sequence of m-dependent random variables correspondingly, and the convergence rate is exponential. Next, in Theorem 2, we had chosen the Poisson distributed index as a typical for presentation. Finally, this theorem is illustrated through an image which is constructed by simulated values of 1-dependent variables. Here, the way that we have applied to create a 1-dependent sequence from an independent sequence that it is likely will help readers understand more about m-dependence structure.

Key words: weak law of large numbers, random sums of random variables, m-dependence, Chernoff-Hoeffding inequality

[1]*University of Finance-Marketing*

[2]*Binh Thanh Commune, Lap Vo District, Dong Thap Province*

Correspondence

Nguyen Tan Nhut, Binh Thanh Commune, Lap Vo District, Dong Thap Province

Email: ntn.nhut@gmail.com

History

- Received: 24-11-2018
- Accepted: 22-7-2019
- Published: 31-12-2019

DOI : 10.32508/stdjns.v3i4.528

Cite this article : Loc Hung T, Tan Nhut N. **Weak law of large numbers for randomly indexed sequences of m-dependent random variables**. *Sci. Tech. Dev. J. - Nat. Sci.;* 3(4):294-298.