MỘT TRƯỜNG HỢP CỦA ĐỊNH LÍ GIỚI HẠN TRUNG TÂM CHO DÃY BIẾN NGẪU NHIÊN PHỤ THUỘC

Phạm Thị Thu Hường và Phạm Thị Thu Hoa[1]

ABSTRACT

Central limit theorem plays an important role in probability theory and applied statistic. However, the findings of this theorem mainly focus on the sequences of independent random variables. Its results haven't been found so much in the case of the sequences of dependent radom variables. Although, the independence of the sequences of random variables is not easy to meet and satisfy. So we need to find conditions to limit the range of the sequences of dependent radom variables to get the results of the central limit theorem. In this paper, we find out a range of conditions for the sequences of dependent radom variables and prove that these conditions stronger than the results were outlined in the paper of Dvoretzky but this still satisfies the central limit theorem.

Keywords: probability theory, applied statistic, Central limit theorem, the sequences of independent random variables, the sequences of dependent radom variables

Title: A case of central limit theorem for the sequences of dependent random variables

TÓM TẮT

Định lý giới hạn trung tâm giữ một vai trò quan trọng trong lý thuyết xác suất và thống kê ứng dụng. Tuy nhiên, những kết quả nghiên cứu về định lý này chủ yếu tập trung vào dãy những biến ngẫu nhiên độc lập, còn trong trường hợp những biến ngẫu nhiên phụ thuộc kết quả nghiên cứu vẫn chưa được nhiều. Tuy nhiên, điều kiện độc lập của dãy các biến ngẫu nhiên không phải lúc nào cũng thỏa mãn và dễ thỏa mãn. Nên ta cần phải tìm điều kiện để hạn chế dãy những biến ngẫu nhiên phụ thuộc để có được kết quả của định lý giới hạn trung tâm. Trong bài báo này, chúng tôi nêu ra một điều kiện cho dãy biến ngẫu nhiên phụ thuộc và chứng minh điều kiện đưa ra chặt hơn kết quả đã nêu ra trong bài báo của Dvoretzky nhưng dãy biến ngẫu nhiên này vẫn thỏa mãn định lí giới hạn trung tâm.

Từ khóa: lí thuyết xác suất, thống kê ứng dụng, định lí giới hạn trung tâm, dãy biến ngẫu nhiên độc lập, dãy biến ngẫu nhiên phụ thuộc

1 CÁC ĐỊNH NGHĨA VÀ KẾT QUẢ CÓ LIÊN QUAN

Định lí 1: Định lí giới hạn trung tâm cho dãy biến ngẫu nhiên độc lập:

Xét dãy tam giác $(X_{1n}, X_{2n}, ..., X_{nn}), n = 1,2,...$ gồm các biến ngẫu nhiên sao cho đối với mỗi n, các biến ngẫu nhiên $X_{1n}, X_{2n}, ..., X_{nn}$ độc lập, $\sum_{k=1}^{n} D(X_{kn}) = 1$, và

[1] Trường Đại học An Giang

$EX_{kn} = 0, (k = 1,...,n)$. Đặt $S_n = \sum_{k=1}^{n} X_{kn}, \delta_{kn}^2 = D(X_{kn}), k \le n$, khi đó nếu với $s > 2$ nào đó, $\sum_{k=1}^{n} Emin(|X_{kn}|^2, |X_{kn}|^s) \to 0$ thì $S_n \xrightarrow{D} N(0,1)$.

Từ đây ta thấy, khi có nhiều nhân tố ngẫu nhiên độc lập tác động sao cho không có nhân tố nào vượt trội lấn át các nhân tố khác thì kết quả của chúng có dạng phân phối tiệm cận chuẩn.

Định nghĩa 1:

Cho không gian xác suất (Ω, F, P), $\mathcal{F}, \mathcal{G}$ là hai đại số của Ω khi đó ta định nghĩa: $\alpha(\mathcal{F}, \mathcal{G}) = \sup |P(F \cap G) - p(F)p(G)|$, supremum được lấy trên tất cả những tập $F \in \mathcal{F}$ và $G \in \mathcal{G}$.

Định nghĩa 2:

Cho dãy tam giác $(X_{n,k})_{n=1,2,..k=1,2,...,k_n}$, $\mathcal{F}_{n,k} = \mathcal{B}(X_{n,1},...,X_{n,k})$

$$\mathcal{G}_{n,k+m+1} = \mathcal{B}(X_{n,k},...,X_{n,k_n}).$$

Ta định nghĩa : $\alpha_n(m) = \sup_{1 \le k \le k_n - m} \alpha(\mathcal{F}_{n,k}, \mathcal{G}_{n,k+m+1})$

Định lí 2:

Cho ξ là một biến ngẫu nhiên giá trị phức thỏa mãn $|\xi| \le 1$, đặt $\mathcal{F} = \mathcal{B}(\xi)$ và $\mathcal{G}$ là δ- đại số trong không gian xác suất. Khi đó:

$$E \mid E(\xi \mid \mathcal{G}) - E\xi \mid \le 2\pi.\alpha(\mathcal{F}, \mathcal{G}).$$

Định lí 3:

Theo Dvoretzky (1972) ta có kết quả sau:

Cho một dãy tam giác biến ngẫu nhiên $(X_{n,k}), n = 1,2,...,k = 1,2,...,k_n$. Đặt $S_{n,a,b} = \sum_{k=a+1}^{b} X_{n,k}, S_{n,b} = S_{n,0,b}, S_n = S_{n,k_n} = \sum_{k=1}^{k_n} X_{n,k}$. Sao cho $EX_{n,k} = 0, \forall n = 1,2,....k = 1,2,...,k_n$. Một dãy tổng riêng của $(X_{n,k})$ là $(Y_{n,i}), n = 1,2,..i = 1,2,...,r_n$ với $0 = j_n(0) < j_n(1) < ... < j_n(r_n) = k_n$ sao cho:

$Y_{n,i} = \sum_{k=j_n(i-1)+1}^{j_n(i)} X_{n,k}$ thỏa mãn:

$$\lim_{n\to\infty} \sum_{i \text{ chẵn}} EY^2_{n,i} = 0 \quad (1)$$

$$\lim_{n\to\infty} \sum_{i \, le} EY^2_{n,i} = 1 \quad (2)$$

$$\lim_{n\to\infty}\sum_{i=1}^{r_n} E\left[Y^2{}_{n,i} I\left(\left| Y_{n,i} \right| > \varepsilon \right) \right] = 0, \forall \varepsilon > 0 \quad (3)$$

và $\lim\limits_{n\to\infty} r_n \alpha_n (m_n) = 0$ với $m_n = \min\limits_{1<i<k_n} [j_n(i) - j_n(i-1)]$ (4)

Thì $S_n \xrightarrow{D} N(0,1)$.

2 KẾT QUẢ MỚI CỦA BÀI BÁO

Chúng tôi nêu ra một điều kiện cho dãy biến ngẫu nhiên phụ thuộc dựa trên ý tưởng của Đào Quang Tuyến (2003) và chứng minh điều kiện đưa ra chặc hơn kết quả đã nêu ra trong bài báo của Dvoretzky (1972), nhưng dãy biến ngẫu nhiên này vẫn thỏa mãn định lí giới hạn trung tâm.

Từ kết quả trong bài báo Asymp của Dvoretzky (1972), chúng tôi tổng quát kết quả trên như sau:

Định lí 4:
Cho một dãy tam giác biến ngẫu nhiên $(X_{n,k}), n = 1,2,..., k = 1,2,...,k_n$ thỏa mãn

$EX_{n,k} = 0, \forall n = 1,2,.... k = 1,2,...,k_n$, và một dãy tổng riêng của $(X_{n,k})$ là

$(Y_{n,i}), n = 1,2,... i = 1,2,...,r_n$ với $0 = j_n(0) < j_n(1) < ... < j_n(r_n) = k_n$ sao cho:

$$Y_{n,i} = \sum_{k=j_n(i-1)+1}^{j_n(i)} X_{n,k} \text{ thỏa mãn:}$$

$$\lim_{n\to\infty} \sum_{i\,\text{chẳn}} EY^2{}_{n,i} = 0 \qquad (1')$$

$$\lim_{n\to\infty} \sum_{i\,\text{lÎ}} EY^2{}_{n,i} = 1 \qquad (2')$$

$$\lim_{n\to\infty}\sum_{i=1}^{r_n} E\left[Y^2{}_{n,i} I\left(\left| Y_{n,i} \right| > \varepsilon \right) \right] = 0, \forall \varepsilon > 0 \ (3') \text{ và với } \forall t \text{ thỏa điều kiện sau:}$$

$$\sum_{k=1}^{r_n} \left| Cov\left\{ exp\left(it \sum_{j\in(0,k)_{2k}} Y_{n,j} \right), exp(-itY_{n,k}) \right\} \right| \xrightarrow{n\to\infty} 0 \qquad (4')$$

Ở đây , $I_{2k} =$ tập hợp những số nguyên chẳn (lẻ) trong tập *I*, nếu k là số chẳn hay lẻ Thì $S_n \xrightarrow{D} N(0,1)$.

Chứng minh: ta có thể tổng quát được như trên vì ở đây điều kiện (4') chặc hơn so với điều kiện (4), cụ thể là:

$$\sum_{k=1}^{r_n}\left|Cov\left\{exp\left(it\sum_{j\in(0,k)_{2k}}Y_{n,j}\right),exp(-itY_{n,k})\right\}\right|\le 2\pi.r_n\alpha_n(m_n)$$

Hay ta có: $\left|Cov\left\{exp\left(it\sum_{j\in(0,k)_{2k}}Y_{n,j}\right),exp(-itY_{n,k})\right\}\right|\le \alpha_n(m_n)$.

Thật vậy,

$$\left|Cov\left\{exp\left(it\sum_{j\in(0,k)_{2k}}Y_{n,j}\right),exp(-itY_{n,k})\right\}\right| =$$

$$\left|Ee^{it\left(\sum_{j\in(0,k)_{2k}}Y_{n,j}+Y_{n,k}\right)}-Ee^{it\sum_{j\in(0,k)_{2k}}Y_{n,j}}Ee^{itY_{n,k}}\right|$$

$$=\left|E\left(E\left\{Ee^{it\left(\sum_{j\in(0,k)_{2k}}Y_{n,j}+Y_{n,k}\right)}-Ee^{it\sum_{j\in(0,k)_{2k}}Y_{n,j}}Ee^{itY_{n,k}}\right\}\middle|\mathcal{F}_{n,j}\right)\right| \text{ với } j\in(0,k)_{2k}$$

$$\le E\left|E\left\{Ee^{it\left(\sum_{j\in(0,k)_{2k}}Y_{n,j}+Y_{n,k}\right)}-Ee^{it\sum_{j\in(0,k)_{2k}}Y_{n,j}}Ee^{itY_{n,k}}\right\}\middle|\mathcal{F}_{n,j}\right| \text{ với } j\in(0,k)_{2k}$$

$$=\left|E\left\{Ee^{it\left(\sum_{j\in(0,k)_{2k}}Y_{n,j}+Y_{n,k}\right)}\middle|\mathcal{F}_{n,j}-(Ee^{it\sum_{j\in(0,k)_{2k}}Y_{n,j}}Ee^{itY_{n,k}})\middle|\mathcal{F}_{n,j}\right\}\right| \text{ với } j\in(0,k)_{2k}$$

$$=\left|Ee^{it\sum_{j\in(0,k)_{2k}}Y_{n,j}}\right|E\left(\left|Ee^{itY_{n,k}}-Ee^{itY_{n,k}}\middle|\mathcal{F}_{n,j}\right|\right) \text{ với } j\in(0,k)_{2k}$$

$$\le 1.2\pi\alpha(\mathcal{F}_{n,j},\mathcal{F}_{n,k}) \text{ với } \mathcal{F}_{n,j}=\mathcal{B}(Y_{n,1},...,Y_{n,j})=\mathcal{B}(X_{n,1},...,X_{n,j_n(j)})$$

$$\mathcal{F}_{n,k}=\mathcal{B}(X_{n,j_n(k-1)+1},...,X_{n,j_n(k)})$$

$$\le 2\pi\alpha_n(m_n) \text{ với } m_n=\min_{1<i<k_n}[j_n(i)-j_n(i-1)],\ \alpha_n(m)=\sup_{1\le k\le k_n-m}\alpha(\mathcal{F}_{n,k},\mathcal{G}_{n,k+m+1}).$$

(theo bổ đề 5.3 của Dvoretzky).

Vậy, $\sum_{k=1}^{r_n}\left|Cov\left\{exp\left(it\sum_{j\in(0,k)_{2k}}Y_{n,j}\right),exp(-itY_{n,k})\right\}\right|\le 2\pi.r_n\alpha_n(m_n)$.

Ta đi chứng minh định lý 4 thỏa mãn định lí giới hạn trung tâm:

Do ta có $\lim_{n\to\infty}\sum_{i\,\text{chẳn}}EY^2{}_{n,i}=0$ (1') và từ điều kiện

$$\sum_{k=1}^{r_n}\left|Cov\left\{exp\left(it\sum_{j\in(0,k)_{2k}}Y_{n,j}\right),exp(-itY_{n,k})\right\}\right|\xrightarrow{n\to\infty}0 \quad (4')$$

Ở đây , $I_{2k}=$ tập hợp những số nguyên chẳn (lẻ) trong tập *I*, nếu *k* là số chẳn hay lẻ. Nên ta có: $\sum_{k\ chẳn}Y_{n,2k+2}\xrightarrow{p}0$.

Mặc khác: $\lim_{n\to\infty}\sum_{i\text{lẻ}}EY^2{}_{n,i}=1$ (2'),

$EX_{n,k}=0,\forall n=1,2,....k=1,2,...,k_n$

Dựa theo phương pháp so sánh được trình bày trong [2] ta định nghĩa $(Y^*{}_{n,k}),k(lẻ)\le r_n$ là dãy những biến ngẫu nhiên độc lập, sao cho hàm phân bố của mỗi $Y^*{}_{n,k}$ trùng với $Y_{n,k}$ với mỗi n và mỗi $k(lẻ)\le r_n$.

Kết hợp với các điều kiện (2') và (3') ta có $\sum_{k(lẻ)}Y^*{}_{n,k}\xrightarrow{D}N(0,1)$.

Mặc khác:

$$\left|Ee^{it\sum_{k(lẻ)\in(0,r_n)}Y_{n,k}}-Ee^{it\sum_{k(lẻ)\in(0,r_n)}Y^*{}_{n,k}}\right|=\left|\sum_{k(lẻ)\in(0,r_n)}\left(Ee^{it\sum_{j(lẻ)\in(0,k)}Y_{n,j}}-Ee^{itY_{n,k-2}}Ee^{itY_{n,k}}\right)\prod_{k+2}^{r_n}Ee^{itY_{n,k}}\right|$$

$$\le\sum_{k(lẻ)\in(0,r_n)}\left|\left(Ee^{it\sum_{j(lẻ)\in(0,k)}Y_{n,j}}-Ee^{itY_{n,k-2}}Ee^{itY_{n,k}}\right)\right|\left|\prod_{k+2}^{r_n}Ee^{itY_{n,k}}\right|$$

$$=\sum_{k(lẻ)\in(0,r_n)}\left|Cov\left\{exp\left(it\sum_{j(lẻ)\in(0,k)}Y_{n,j}\right),exp(-itY_{n,k})\right\}\right|$$

$$\le \sum_{k=1}^{r_n}\left| Cov\left\{ exp\left(it \sum_{j\in(0,k)_{2k}} Y_{n,j}\right), exp(-itY_{n,k})\right\}\right| \xrightarrow{n} 0 \text{ (do 4')}$$

Nên $\sum_{k(lẻ)} Y_{n,k} \xrightarrow{D} N(0,1)$

Do đó, $S_n = \sum_{i=1}^{r_n} Yn,i \xrightarrow{D} N(0,1)$.

Ví dụ:

Cho (X_k) là dãy biến ngẫu nhiên bất kì có kỳ vọng và phương sai hữu hạn. Đặt

$X_{kn} = \dfrac{X_k - EX_k}{B_n}, k = 1,2,...,n$,

$S_n^* = \dfrac{\sum_{k=1}^{n}(X_k - EX_k)}{B_n}, B_n^2 = \sum_{k=1}^{n} D(X_k), S_{n,a,b}^* = \sum_{k=a+1}^{b} X_{kn}, S_{n,b}^* = S_{n,0,b}^*$. Khi đó với

$\varepsilon > 0$ bất kỳ $\dfrac{1}{B_n^2}\sum_{k=1}^{n} E[(X_k - EX_k)^2, |X_k - EX_k| > \varepsilon B_n] \to 0$ và điều kiện

$\sum_{k=1}^{n} |\operatorname{cov}\{exp(itS_{n,k-1}^*), exp(-itX_{kn})\}| \xrightarrow{n} 0$ được thỏa mãn thì $S_n^* \xrightarrow{D} N(0,1)$.

TÀI LIỆU THAM KHẢO

A. Dvoretzky, Asymptotic normality for sums of dependent random variables, Proc.Sixth Berkeley Symp. Math. Statist. Prob. Univ of California Press, 1972, 513-535.

Đào Quang Tuyến, Central limit theorems for Mixing Arrays, Vietnam journal of Mathematics 32 (2004), 277-292.

Nguyễn Duy Tiến - Vũ Viết Yên, Lý thuyết xác suất thống kê, nhà xuất bản giáo dục, 2003.

Y.S. Chow and H. Teicher, Probability theory, Springer, Newyork, Heidelberg, Berlin, 1978.