TẠP CHÍ KHOA HỌC & CÔNG NGHỆ ĐẠI HỌC DUY TÂN
DTU Journal of Science and Technology 5(60) (2023) 74-79

TRƯỜNG ĐẠI HỌC DUY TÂN
DUYTAN UNIVERSITY

# Laws of large number for random weighted sums of random variables with regularly varying tailed and application

## Luật số lớn đối với tổng có trọng số ngẫu nhiên các biến ngẫu nhiên có xác suất đuôi biến đổi đều và ứng dụng

Tran Van Tam, Van Nguyen Yen Ngoc, Ho Duy Nguyen*
Trần Văn Tâm, Văn Nguyễn Yến Ngọc, Hồ Duy Nguyên*

*Faculty of Mathematics, the University of Da Nang - University of Science and Education, Vietnam*
*Khoa Toán, Trường Đại học Sư phạm - Đại học Đà Nẵng, Việt Nam*

*(Ngày nhận bài: 11/03/2023, ngày phản biện xong: 27/03/2023, ngày chấp nhận đăng: 04/9/2023)*

## Abstract

In this paper, we study laws of large number for random weighted sums of random variables with regularly varying tailed and application. Firstly, we use the theory of slowly varying functions to establish the law of large. Then, we apply this result to estimate non parametric regression models by the k-nearest neighbors.

*Keywords:* Laws of large number; random weighted sums; regularly varying tailed; the k-nearest neighbors.

## Tóm tắt

Trong bài báo này, chúng tôi nghiên cứu luật số lớn đối với tổng có trọng số ngẫu nhiên các biến ngẫu nhiên có xác suất đuôi biến đổi đều và ứng dụng. Đầu tiên, chúng tôi sử dụng lý thuyết hàm biến đổi chậm thiết lập luật số. Sau đó chúng tôi áp dụng kết quả thu được vào ước lượng mô hình hồi quy phi tham số bằng phương pháp ước lượng k-láng giềng gần nhất.

*Từ khóa:* Luật số lớn; tổng có trọng số ngẫu nhiên; xác suất đuôi biến đổi đều; k-láng giềng gần nhất.

## 1. Introduction

Let $\{X_n; n \geq 1\}$ be a sequence of random variables defined on a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\{a_{nj}; 1 \leq j \leq n, n \geq 1\}$ be an array of real numbers, many authors studied the weak laws of larger numbers for weighted sum type

$$S_n = \sum_{i=1}^{n} a_{ni} X_i.$$

Recently, Xuan et al. (T. D. Xuan, 2021) investigate laws of large numbers for this weighted sum of pairwise independent with heavy tails and study convergence in probability for the estimator of nonparametric regression model based on pairwise independent errors with heavy tails.

**Corresponding Author:* Ho Duy Nguyen
*Email:* nguyenlqddn@gmail.com

In this paper, we study weak laws of larger numbers for random weighted sum $S_n = \sum_{i=1}^{n} w_{ni} X_i$, where $\{w_{ni};\ 1 \le i \le n,\ n \ge 1\}$ be a triangle array of random variables, $\{X_n;\ n \ge 1\}$ be a sequence of pairwise asymptotically independent and identically distributed random variables. This result is applied to the nonparametric regression models with random design.

## 2. Preliminaries

Let $\{a_n;\ n \ge 1\}$ and $\{b_n; n \ge 1\}$ be sequences of positive real numbers. We use notation $a_n \asymp b_n$ instead of $0 < \liminf a_n/b_n \le \limsup a_n/b_n < \infty$; $a_n = o(b_n)$ means that $\lim_{n\to\infty} a_n/b_n = 0$; notation $a_n \sim b_n$ is used for $\lim_{n\to\infty} a_n/b_n = 1$. These notations are also used for positive real functions $f(x)$ and $g(x)$. The indicator function of $A$ is denoted by $I(A)$. Throughout this paper, the symbol $C$ will denote a generic constant $(0 < C < \infty)$ which is not necessarily the same one in each appearance.

**Definition 2.1.** (Y. Yang, 2017) A distribution $F$ on $\mathbb{R}$ is said to be *regularly varying tailed*, denoted by $F \in \Re_\alpha$ for some $\alpha > 0$, if $\limsup_{x\to\infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = y^{-\alpha}$ for any $y \ge 1$, where $\bar{F}(x) = 1 - F(x)$.

**Definition 2.2.** (Y. Yang, 2017) A sequence of random variables $\{X_n; n \ge 1\}$ is said to be *pairwise asymptotically independent* (PAI) if

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_i > x, X_j > x)}{\mathbb{P}(X_k > x)} = 0, k = i, j$$

holds for each $i \ne j$.

For a distribution $F$ on $\mathbb{R}$, denote its upper Matuszewska index by

$$\bar{F}_*(y) = \lim_{x\to\infty} \frac{\bar{F}(xy)}{\bar{F}(y)}\ for\ y \ge 1.$$

Additionally, denote $L_F = \lim_{y\to 1} inf\ \bar{F}_*(y)$.

**Lemma 2.3.** Let $\alpha > 0$, $\{X_n;\ n \ge 1\}$ *be a sequence of pairwise asymptotically independent and identically distributed random variables with distribution* $F \in \Re_\alpha$ *and* $L_F > 0$. *Let* $\{w_{ni};\ 1 \le i \le n,\ n \ge 1\}$ *be a triangle array of random variables. Then,*

$$\lim sup \frac{P(S_n^+ > x)}{\sum_{i=1}^{n} P(w_{ni} X_i > x)} \le \frac{1}{L_F},$$

*where* $S_n = \sum_{i=1}^{n} w_{ni} X_i$.

*Proof.* We firstly consider the upper bound. For any $0 < v < 1$ and $x > 0$,

$$\begin{aligned}
\mathbb{P}(S_n^+ > x) &\le \mathbb{P}\left(\bigcup_{j=1}^{n} \{w_{nj} X_j > (1-v)x\}\right) \\
&+ \mathbb{P}\left(\sum_{i=1}^{n} w_{ni} X_i^+ > x, \bigcap_{j=1}^{n} \{w_{nj} X_j^+ \le (1-v)x\}\right) \\
&\le \sum_{i=1}^{n} \mathbb{P}(w_{ni} X_i > (1-v)x) \\
&+ \mathbb{P}\left(\sum_{i=1}^{n} w_{ni} X_i^+ > x, \frac{x}{n} < \bigvee_{1\le j\le n} w_{nj} X_j^+ \le (1-v)x\right) \\
&:= K_1 + K_2.
\end{aligned}$$

Let $g(\cdot)$ be a positive function such that $g(x) \downarrow 0, xg(x) \uparrow \infty$ *and* $\mathbb{P}(w_{ni} > xg(x)) = o(\bar{F}(x)),\ 1 \le i \le n$. We have that

$$K_1 = \sum_{i=1}^{n} \mathbb{P}\left(w_{ni}X_i > (1-v)x, \bigcap_{j=1}^{n} \{w_{nj} \le xg(x)\}\right)$$

$$+\sum_{i=1}^{n} \mathbb{P}\left(w_{ni}X_i > (1-v)x, \bigcup_{j=1}^{n} \{w_{nj} > xg(x)\}\right)$$

$$\le \int_0^{xg(x)} \cdots \int_0^{xg(x)} \sum_{i=1}^{n} \bar{F}_i\left(\frac{(1-v)x}{\iota_i}\right) d\mathbb{P}(w_{n1} \le t_1, \cdots, w_{nn} \le \iota_n)$$

$$+n\sum_{i=1}^{n} \mathbb{P}(w_{ni} > xg(x))$$

$$= \int_0^{xg(x)} \cdots \int_0^{xg(x)} \frac{\bar{F}((1-v)x/L_F)}{\bar{F}(x/L_F)} \cdot \sum_{i=1}^{n} \bar{F}\left(\frac{x}{L_F}\right) d\mathbb{P}(w_{n1} \le t_1, \cdots, w_{nn} \le \iota_n)$$

$$+o(\bar{F}(x)) \lesssim \left(\bar{F}_*((1-v)^{-1})\right)^{-1} \sum_{i=1}^{n} \mathbb{P}(w_{ni}X_i > x).$$

As for $K_2$, we have that

$$K_2 \le \sum_{j=1}^{n} \mathbb{P}\left(\sum_{i=1, i\ne j}^{n} w_{ni}X_i^+ > vx, w_{nj}X_j > \frac{x}{n}\right) \le \sum_{j=1}^{n} \sum_{i=1, i\ne j}^{n} \mathbb{P}\left(w_{ni}X_i > \frac{vx}{n-1}, w_{nj}X_j > \frac{x}{n}\right)$$

$$\le \sum_{j=1}^{n} \sum_{i=1, i\ne j}^{n} \mathbb{P}\left(w_{ni}X_i > \frac{vx}{n}, w_{nj}X_j > \frac{vx}{n}\right) = o(1)\sum_{i=1}^{n} \mathbb{P}(w_{ni}X_i > x).$$

Hence

$$\frac{\mathbb{P}(S_n^+ > x)}{\sum_{i=1}^{n} \mathbb{P}(w_{ni}X_i > x)} \le \lim_{v\downarrow 0} \left(\bar{F}_*((1-v)^{-1})\right)^{-1} = L_F^{-1}.$$

We complete the proof.

**Definition 2.4. (K. Joag-Dev, 1983)** A collection $\{\boldsymbol{X_1}, \dots, \boldsymbol{X_n}\}$ of random variables is said to be negatively associated (NA) if for any disjoint subsets $\boldsymbol{A}, \boldsymbol{B}$ of $\{\boldsymbol{1}, \dots, \boldsymbol{n}\}$ and any real coordinatewise nondecreasing functions f on $\mathbb{R}^{|\boldsymbol{A}|}$ and g on $\mathbb{R}^{|\boldsymbol{B}|}$,

$$cov\left(f(X_k, k \in A), g(X_k, k \in B)\right) \le 0,$$

whenever the covariance exists, where $|A|$ denotes the cardinality of $A$. A sequence $\{X_n, n \ge 1\}$ of random variables is said to be negatively associated if every finite subfamily is negatively associated.

## 3. Results

In the first theorem, we establish the Marcinkiewicz laws of large numbers type for weighted sum of pairwise independent and identically distributed random variables with heavy tails.

**Theorem 3.1.** Let $\alpha > 0$, $\{X_n;\ n \geq 1\}$ *be a sequence of pairwise asymptotically independent and identically distributed random variables with distribution* $F \in \Re_\alpha$ *and* $L_F > 0$. *Let* $\{w_{ni};\ 1 \leq i \leq n,\ n \geq 1\}$ *be a triangle array of random variables such that*

$$\sum_{i=1}^{n} E(|w_{ni}|^\alpha) = o(1).$$

*Then,*

$$\sum_{i=1}^{n} w_{ni} X_i \xrightarrow{p} 0 \text{ as } n \to \infty.$$

*Proof.* We can write

$$\sum_{i=1}^{n} w_{ni} X_i = \sum_{i=1}^{n} \left(w_{ni}^+ X_i^+ - E(w_{ni}^+ X_i^+)\right) + \sum_{i=1}^{n} (w_{ni}^- X_i^- - E(w_{ni}^- X_i^-))$$
$$- \sum_{i=1}^{n} (w_{ni}^+ X_i^- - E(w_{ni}^+ X_i^-)) - \sum_{i=1}^{n} (w_{ni}^- X_i^+ - E(w_{ni}^- X_i^+)).$$

Thus, without loss of general, we may assume that $w_{ni} \geq 0$ a.s for all $1 \leq i \leq n$ and $X_n \geq 0$ a.s. for all $n \geq 1$. Put $S_n = \sum_{i=1}^{n} w_{ni} X_i$.

Let $\epsilon > 0$ be arbitrary, applying Lemma 3.3 we get

$$P(S_n > \epsilon) \leq L_F^{-1} \sum_{i=1}^{n} \mathbb{P}(w_{ni} X_i > \epsilon)$$
$$= L_F^{-1} \sum_{i=1}^{n} \int_0^\infty \mathbb{P}\left(X_i > \frac{\epsilon}{t_i}\right) dP(w_{ni} < t)$$
$$\leq L_F^{-1} \sum_{i=1}^{n} \int_0^\infty t_i^{\alpha \overline{F}}(\epsilon) d\mathbb{P}(w_{ni} < t)$$
$$= L_F^{-1} \overline{F}(\epsilon) \sum_{i=1}^{n} \int_0^\infty t_i^\alpha dP(w_{ni} < t)$$
$$= L_F^{-1} \overline{F}(\epsilon) \sum_{i=1}^{n} E(w_{ni}^\alpha) \to 0 \text{ as } n \to \infty.$$

We complete the proof.

Now, consider the following nonparametric regression model:

$$Y_{ni} = f(X_{ni}) + \varepsilon_{ni},\ 1 \leq i \leq n, \qquad (1)$$

where $X_{ni}$ are known random design points from a compact set $\boldsymbol{A} \subset \mathbb{R}$, $f(x)$ is an unknown regression function defined on $\boldsymbol{A}$, $\varepsilon_{ni}$ are random errors. As an estimator of $f(x)$, the following weighted regression estimator will be considered

$$\hat{f}_n(x) = \sum_{i=1}^{n} W_{ni}(x) Y_{ni}, \qquad (2)$$

where $W_{ni}(x) = W(x, x_{n1}, \dots x_{nn})$ are random weighted functions.

Assume that $X_{n1}, \dots, X_{nn}$ are independent and identically distributed continuous random variables on the interval $[0,1]$, $(X_{ni}; 1 \le i \le n)$ and $\{\varepsilon_{ni}; 1 \le i \le n\}$ are independent. For any $x \in (0,1)$, let $R_{ni}$ be the rank of $|X_{ni} - x|$ for $i = 1,2,\dots,n$. Then, it follows by Joag-Dev and Proschan [2] that $\{R_{n1}, R_{n2}, \dots, R_{nn}\}$ is NA. Let $0 < h_n < n$, we define the nearest neighbor weight functions $\{W_{ni}(x); 1 \le i \le n\}$ by $W_{ni}(x) = h_n^{-1} I(R_{ni} \le h_n)$. Then $\{W_{n1}(x), \dots, W_{nn}(x)\}$ is also NA. On the other hand, because

$$\mathbb{P}(R_{ni} = r) = \frac{1}{n} \text{ for } r = 1,2,\dots,n.$$

Then,

$$\sum_{i=1}^{n} E(|W_{ni}|^{\alpha}) = \sum_{i=1}^{n} \frac{1}{h_n^{\alpha}} \mathbb{P}(R_{ni} \le h_n) = \frac{1}{h_n^{\alpha-1}}.$$

**Theorem 3.2.** *In the model (2), assume that* $\{\varepsilon_i;\ 1 \le i \le n\}$ *is a sequence of PAI and identically distributed errors with zero mean and* regularly varying tailed for some $\alpha > 1$. Let $0 < h_n < n$ *and* $1/h_n^{\alpha-1} = o(1)$. *If* $f$ *is a Lipschitz function on [0;1], then*

$$\hat{f}_n(x) \xrightarrow{p} f(x) \text{ as } n \to \infty,$$

*Proof.* For any $x \in [0;1]$, it is obvious that

$$\hat{f}_n(x) - f(x) = \sum_{i=1}^{n} W_{ni}(x)\varepsilon_{nj} + [E(\hat{f}_n(x)) - f(x)].$$

Applying Theorem 3.1, we have that

$$\sum_{i=1}^{n} W_{ni}(x)\varepsilon_{ni} \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Thus, in order to complete the proof, we need to show that

$$|\sum_{i=1}^{n} W_{ni}(x) f(X_{ni}) - f(x)| \to 0 \text{ as } n \to \infty.$$

For any $a > 0$, we have

$$|\sum_{i=1}^{n} W_{ni}(x) f(X_{ni}) - f(x)|$$

$$\le \sum_{i=1}^{n} |W_{ni}(x)||f(X_{ni}) - f(x)| I(|X_{ni} - x| \le a)$$

$$+ \sum_{i=1}^{n} |W_{ni}(x)|\,|f(X_{ni}) - f(x)| I(|X_{ni} - x| > a)\}$$

$$+ \left|\sum_{i=1}^{n} W_{ni}(x) - 1\right| |f(x)|$$

$$\le \epsilon + \sum_{i=1}^{n} |W_{ni}(x)||f(X_{ni}) - f(x)| I(|X_{ni} - x| > a).$$

On the other hand, it follows by Lemma 6.1 in (L. Györfi, 2002) that for any $x \in (0,1)$, $|X_{n,R_{n,h_n}} - x| \to 0$ a.s. as $n \to \infty$, then

$$\sum_{i=1}^{n} |W_{ni}| \, |f(X_{ni}) - f(x)| I(|X_{ni} - x| > a)\}$$
$$\leq 6 \left| X_{n,R_{n,h_n}} - x \right| \to 0 \text{ as } n \to \infty.$$

## References

[1] K. Joag-Dev, F. P. (1983). Negative association of random variables with applications. *Ann. Stat., 11*, 286–295.

[2] L. Györfi, M. K. (2002). *A Distribution-Free Theory of Nonparametric Regression.* New York: Springer.

[3] T. D. Xuan, N. T. (2021). Convergence in probability for the estimator of nonparametric regression model based on pairwise independent errors with heavy tails. *DTU Journal of Science & Technology, 45*(2), 51-57.

[4] Y. Yang, T. Z. (2017). Approximations for finite-time ruin probability in a dependent discrete-time risk model with CMC simulations. *Journal of Computational and Applied Mathematics*, *321*, 143-159.