Filomat 31:14 (2017), 4369–4377
https://doi.org/10.2298/FIL1714369B

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

An Extension of Central Limit Theorem for Randomly Indexed m-Dependent Random Variables

Stefano Belloni[a,b]

[a]*Center for Modeling and Simulation in the Biosciences (BIOMS)*
[b]*Interdisciplinary Center for Scientific Computing (IWR)*

Abstract. In this note, we prove a conjecture of Shang about the sum of a random number N_n of m-dependent random variables. The random number N_n is supposed to converge in probability toward a positive random variable.

1. Introduction

In a number of applications we have to deal with random sums of dependent random variables. In order to study such random sums it is useful to have the possibility to rely on *Central Limit* results. Studies on random central limit theorems have a long tradition and the applicability of such results varies from random walk problems to Monte Carlo methods, passing through sequential analysis. The works of Rényi [10] and Blum et. al. [4] focused on central limit problems for the sum of a random number of independent random variables. More recent studies can be found in e.g. [6, 7, 9, 15], most of which, nevertheless, deal with independent cases. The case of dependent random variable has been the base of the work of Y. Shang [13]. In this note we prove the conjecture on the convergence of the random sum to a positive random variable instead of a constant.
In this article we will indicate with $\mathrm{Cov}(X, Y)$ the covariance between two jointly distributed real-valued random variables X and Y with finite second moments. $\mathrm{Var}(X)$ is the variance of a random variable X. Let $\{X_n\}$ be a sequence of random variable. We write

$$X_n \xrightarrow{P} X$$

to indicate that the sequence of random variable X_n converges in probability towards the random variable X, i.e. for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\Big(|X_n - X| \geq \varepsilon\Big) = 0.$$

Let F_n and F be the cumulative distribution functions of the random variables X_n and X, respectively. We write

$$X_n \xrightarrow{L} X$$

2010 *Mathematics Subject Classification*. Primary 60F05
Keywords. Central limit theorem, dependent random variables, stationary sequence
Received: 23 February 2016; Accepted: 08 June 2016
Communicated by Miljana Jovanović
Email address: stefano.belloni@bioquant.uni-heidelberg.de (Stefano Belloni)

if the sequence of random variable $\{X_n\}$ converges towards X in distribution, i.e.

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for every number $x \in \mathbb{R}$ at which F is continuous. In what follows we will indicate with $\Phi(x)$ the cumulative distribution of a standard normal random variable.

The result relies on the following hypothesis

- (A1) there exist some $k_0 \geq 0$ and $c > 0$ such that for any $\lambda > 0$ and $n > k_0$ we have

$$\mathbf{P}\Big(\max_{k_0<k_1\leq k_2\leq n} |S_{k_2} - S_{k_1} - (k_2 - k_1)\mu| \geq \epsilon\Big) \leq \frac{c \cdot \mathrm{Var}(S_n - S_{k_0})}{\epsilon^2},$$

  and

- (A2) $\mathrm{Cov}(X_1, X_i) \geq 0$ for $i = 2, \dots, m + 1$.

We report here the new central limit theorem for randomly indexed m-dependent random variables proved in [13]:

Theorem 1.1. *Let $\{X_j\}_{j\in\mathbb{N}}$ be a* stationary *m-dependent sequence of random variables. Let $\mathbb{E}[X_j] = \mu$, $0 < Var[X_i] = \sigma^2 < \infty$ and $S_n = \sum_{i=1}^n X_i$ be the partial sum. Let $\{N_n\}_{n\in\mathbb{N}}$ denote a sequence of positive integer-valued random variables such that*

$$\frac{N_n}{z_n} \xrightarrow{P} \lambda$$

as n goes to infinity, where $\{z_n\}_{n\in\mathbb{N}}$ is an arbitrary positive sequence tending to $+\infty$ and λ constant. If (A1) and (A2) hold, then

$$\frac{\sqrt{N_n}}{\tau}\Big(\frac{S_{N_n}}{N_n} - \mu\Big) \xrightarrow{L} \mathcal{N}(0, 1)$$

as $n \to \infty$, where $\tau^2 = \sigma^2 + \sum_{i=2}^{m+1} Cov(X_i, X_1)$.

The rest of the note is organized as follows: In Section 2 we present the main result and prove it. The section 3 is dedicated to an example while in the last one we outline some line of research.

2. Main Result

The main purpose of this section is to prove the following generalization of Theorem 1.1 [13].

Theorem 2.1. *Let $\{X_j\}_{j\in\mathbb{N}}$ be a* stationary *m-dependent sequence of random variables. Let $\mathbb{E}[X_j] = \mu$, $0 < Var[X_i] = \sigma^2 < \infty$ and $S_n = \sum_{i=1}^n X_i$ the partial sum. Let $\{N_n\}_{n\in\mathbb{N}}$ denote a positive sequence of positive integer-valued random variables such that*

$$\frac{N_n}{z_n} \xrightarrow{P} \lambda$$

as n goes to infinity, where $\{z_n\}_{n\in\mathbb{N}}$ is an increasing positive sequence tending to $+\infty$ and λ is a positive random variable. If (A1) and (A2) hold, then

$$\frac{\sqrt{N_n}}{\tau}\Big(\frac{S_{N_n}}{N_n} - \mu\Big) \xrightarrow{L} \mathcal{N}(0, 1)$$

as $n \to \infty$, where $\tau^2 = \sigma^2 + \sum_{i=2}^{m+1} Cov(X_i, X_1)$.

Remark 2.2. *As in [13] we can stress here that (A1) is related to the Anscomb condition [1] and may be regarded as a* relaxed *Kolmogorov inequality. The Condition (A2) says that each pair X_i, X_j is positively correlated.*

Remark 2.3. *In order to prove this theorem we follow the same steps as in [10] and [4]. We notice that while in their results the random variables are independent, here we suppose m-dependence. We further notice that the sequences N_n and X_n are not supposed to be independent, we in fact underline that the limit theorem used here is Theorem 1.1 [13], where such an hypotheses is not required. Last we point out that for clarity of the main proof we assume the sequence z_n is an increasing sequence.*

We begin by prooving the following result, which generalizes *Theorem 4* in [11]:

Lemma 2.4. *Let $(\xi_j)_{j\in\mathbb{N}}$ be a sequence of m-dependent and stationary random variables. We assume that there exist two divergent sequences of real numbers C_n, D_n and a distribution function $F(x)$ such that with*

$$\sigma_n = \frac{\sum_{j=1}^n \xi_j - C_n}{D_n}$$

we have

$$\lim_{n\to\infty} \mathbf{P}\Big(\sigma_n < x\Big) = F(x)$$

for every point of continuity x of $F(x)$.
Then for every probability measure $\mathbf{Q}$ on $(\Omega, \mathcal{F})$ absolutely continuous with respect to $\mathbf{P}$ we have

$$\lim_{n\to\infty} \mathbf{Q}\Big(\sigma_n < x\Big) = F(x)$$

for every point of continuity x of $F(x)$.

Proof. Let x be a continuity point for $F(x)$ such that $F(x) > 0$. For $n \geq n_0$, for n_0 big enough, we have $\mathbf{P}(\sigma_n < x) > 0$. Let us consider the sets $A_0 = \Omega$ and $A_n := \{\omega \in \Omega : \sigma_{n+n_0}(\omega) < x\}$, $n \in \mathbb{N}$. By following Theorem 1 and Theorem 2 in [11] we can simply prove that for all $k > n_0$

$$\lim_{n\to\infty} \mathbf{P}(\sigma_n < x \,|\, \sigma_k < x) = F(x).$$

The original proof uses the following

Lemma 2.5. *Let θ_n and ϵ_n be random variables such that $\lim_{n\to\infty} \mathbf{P}(\theta_n < x) = F(x)$ for every x, which is a point of continuity for $F(x)$ and $\lim_{n\to\infty} \mathbf{P}(|\epsilon_n| \geq \delta) = 0$ for every $\delta > 0$. Then $\lim_{n\to\infty} \mathbf{P}(\theta_n + \epsilon_n < x) = F(x)$ for every x, which is a point of continuity for $F(x)$.*

We apply this lemma with $\theta_n = \sigma_n$ and $\epsilon_n = -\frac{\sigma^*_{k+m}}{D_n}$, where $\sigma^*_n = \sum_{j=1}^n \xi_j$. Since $\lim_{n\to\infty} \mathbf{P}(\sigma_n < x) = F(x)$ and $D_n \to \infty$, it follows that

$$\lim_{n\to\infty} \mathbf{P}\Big(\sigma_n - \frac{\sigma^*_{k+m}}{D_n} < x\Big) = F(x).$$

For the hypothesis that the random variables are m-dependent, $\sigma_n - \frac{\sigma^*_{k+m}}{D_n} = \frac{\sigma^*_n - \sigma^*_{k+m} - C_n}{D_n}$ and σ_k are independent. It easily follows that

$$\mathbf{P}\Big(\sigma_n - \frac{\sigma^*_{k+m}}{D_n} < x \Big| \sigma_k < x\Big) = \mathbf{P}\Big(\sigma_n - \frac{\sigma^*_{k+m}}{D_n} < x\Big),$$

which in turn implies

$$\lim_{n\to\infty} \mathbf{P}\Big(\sigma_n - \frac{\sigma^*_{k+m}}{D_n} < x \Big| \sigma_k < x\Big) = F(x).$$

We can apply the lemma again with $\theta = \sigma_n - \frac{\sigma^*_{k+m}}{D_n}$ and $\epsilon = \frac{\sigma^*_{k+m}}{D_n}$ on the probability space $(\Omega, \mathcal{F}, \mu)$ where $\mu(A) = \mathbf{P}(A \,|\, \sigma_k < x)$. It follows that

$$\lim_{n\to\infty} \mu(\sigma_n < x) = \lim_{n\to\infty} \mathbf{P}(\sigma_n < x \,|\, \sigma_k < x) = F(x).$$

□

We will make use of the following lemma:

Lemma 2.6. *Under the same hypothesis of Theorem 2.1, let $\nu_n = [z_n\lambda]$ where λ is a positive random variable having a discrete distribution, then it holds*

$$\lim_{n\to\infty} \mathbf{P}\Big\{\frac{\sqrt{\nu_n}}{\tau}\Big(\frac{S_{\nu_n}}{\nu_n} - \mu\Big) < x\Big\} = \Phi(x).$$

Proof. For the proof we refer to [10], by using Lemma 2.4 instead of *Theorem 4* in [10]. □

We can now prove the theorem when λ is a positive and *discrete*-value random variable.

Proof. [Proof of the discrete version] It is possible to follow the proof in [10]. Without loosing generality let $\mu = 0$, $\frac{N_n}{z_n} \xrightarrow{P} \lambda : (\Omega, \mathcal{F}) \to \mathbb{R}$ and $\nu_n = [z_n\lambda]$. We can rearrange the random variable $\frac{\sqrt{N_n}}{\tau}\Big(\frac{S_{N_n}}{N_n}\Big)$ as follows:

$$\frac{S_{N_n}}{\tau\sqrt{N_n}} = \frac{S_{\nu_n}}{\tau\sqrt{\nu_n}} + \sqrt{\frac{\nu_n}{N_n}}\Big(\frac{S_{N_n} - S_{\nu_n}}{\tau\sqrt{\nu_n}}\Big) + \frac{S_{\nu_n}}{\tau\sqrt{\nu_n}}\Big(\sqrt{\frac{\nu_n}{N_n}} - 1\Big).$$

We have $\frac{N_n}{\nu_n} \xrightarrow{P} 1$, and then the convergence (in probability)

$$\frac{S_{\nu_n}}{\tau\sqrt{\nu_n}}\Big(\sqrt{\frac{\nu_n}{N_n}} - 1\Big) \xrightarrow{P} 0.$$

It is sufficient to prove that

$$\sqrt{\frac{\nu_n}{N_n}}\Big(\frac{S_{N_n} - S_{\nu_n}}{\tau\sqrt{\nu_n}}\Big) \xrightarrow{P} 0.$$

Let $B_n(\rho) := \{\omega \in \Omega : |N_n - \nu_n|(\omega) < \rho z_n\}$, with $\rho > 0$ and $n_k = [z_n l_k]$; Pick $\epsilon > 0$ (arbitrarily) and let us consider $C_{n_k} := \{\omega \in \Omega : \frac{|S_{N_n} - S_{n_k}|}{\tau\sqrt{n_k}} > \epsilon\}$; By the theorem of total probability we have

$$\mathbf{P}\Big(\Big|\frac{S_{N_n} - S_{\nu_k}}{\tau\sqrt{\nu_n}}\Big| > \epsilon\Big) \leq \sum_{k=1}^{\infty} \mathbf{P}(A_k \cap B_n(\rho) \cap C_{nk}) + \mathbf{P}(B_n(\rho)^C).$$

From Hypothesis (A1)

$$\mathbf{P}(A_k \cap B_n(\rho) \cap C_{nk}) \leq \mathbf{P}\Big(\max_{|l-n_k|\leq\rho n} \frac{|S_l - S_{n_k}|}{\sqrt{n_k}} > \epsilon\Big) \leq \frac{2c\rho\tau^2}{l_k\epsilon^2}.$$

We have now to control the tail of λ.

Let $D_M := \{\omega \in \Omega : \lambda(\omega) \geq l_M\}$. We have

$$\mathbf{P}\Big(\Big|\frac{S_{N_n} - S_{\nu_k}}{\tau\sqrt{\nu_n}}\Big| > \epsilon\Big) \leq \mathbf{P}(D_M) + \frac{2c\rho\tau^2}{\epsilon^2}\sum_{k=1}^{M-1}\frac{1}{l_k} + \mathbf{P}(B_n(\rho)^C).$$

Exactly as in [10] let $\delta > 0$ (arbitrary) and M big enough in order to have $\mathbf{P}(D_M) < \frac{\delta}{3}$. By fixing M we can find $\rho > 0$ such that

$$\frac{2\rho\tau}{\epsilon^2}\sum_{k=1}^{M-1}\frac{1}{l_k} < \frac{\delta}{3}.$$

By the hypothesis $\frac{N_n}{z_n} \xrightarrow{P} \lambda$, we can choose $n_0 = n_0(\epsilon, \delta)$ so that for $n \geq n_0$ it holds $\mathbf{P}(B_n(\rho)^C) < \frac{\delta}{3}$. From this follows the desired convergence for the sequence of stationary and m-dependent random variables. □

The proof of Theorem 2.1 is based on the following result in [4]:

Lemma 2.7. *[4] Let W_n, $X_{m,n}$, $Y^{(j)}_{m,n}$ and $Z^{(j)}_{m,n}$ be random variables. Suppose that*

$$W_n = X_{m,n} + \sum_{j=1}^{k} Y^{(j)}_{m,n} Z^{(j)}_{m,n},$$

and

1. *For every $\epsilon > 0$ and $j = 1, \dots, k$;*
$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbf{P}(|Y^{(j)}_{m,n}| > \epsilon) = 0$$
2. *$j = 1, \dots, k$;*
$$\lim_{M\to\infty} \lim_{m\to\infty} \limsup_{n\to\infty} \mathbf{P}(|Z^{(j)}_{m,n}| > M) = 0$$
3. *For every fixed m*
$$\lim_{n\to\infty} \mathbf{P}(X_{m,n} < x) = F(x).$$

It follows that

$$\lim_{n\to\infty} \mathbf{P}(W_n < x) = F(x).$$

Proof. [Sketch of the proof] For a complete proof of the Lemma we refer to [4], Lemma 2.
The idea is to control the quantity

$$\mathcal{R}(x) := \Big|\mathbf{P}(W_n \le x) - F(x)\Big| \tag{1}$$

for every continuity point x of $F(x)$. Let m, n be positive natural numbers and $\delta > 0$. We can control $\mathcal{R}(x)$ as follows:

$$\mathcal{R}(x) \le \max_{i\pm 1} \Big|\mathbf{P}(X_{m,n} \le x + i\cdot\delta) - F(x)\Big| + \mathbf{P}\Big(\sum_{j=1}^{k} Y^{(j)}_{m,n} Z^{(j)}_{m,n} > \delta\Big).$$

This follows from

Lemma 2.8 (Lemma 1 in [4]). *If X and Y are random variables and $\delta \ge 0$, then*

$$\Big|\mathbf{P}(X \le a) - \mathbf{P}(X + Y \le a + \delta)\Big| \le \mathbf{P}(|Y| > \delta).$$

For the first addend holds:

$$\begin{aligned} |\mathbf{P}(X_{m,n} \le x + i\delta) - F(x)\Big| &\le \max_{i\pm 1} |\mathbf{P}(X_{m,n} \le x + i\delta) - F(x + i\delta)\Big| \\ &\quad + \max_{i\pm 1} |F(x) - F(x + i\cdot\delta)\Big|. \end{aligned} \tag{2}$$

The second term can be controlled as follows

$$\mathbf{P}\Big(\Big|\sum_{j=1}^{k} Y^{(j)}_{m,n} Z^{(j)}_{m,n}\Big| > \delta\Big) \ge \sum_{j=1}^{k} \mathbf{P}\Big(|Y^{(j)}_{m,n}| > M\Big) + \sum_{j=1}^{k} \mathbf{P}\Big(|Z^{(j)}_{m,n}| > \frac{\delta}{k\cdot M}\Big). \tag{3}$$

Let $\epsilon > 0$. We can find $\delta > 0$ such that

$$\max_{i\pm 1} |\mathbf{P}(F(x) - F(x + i\delta)\Big| < \frac{\epsilon}{4},$$

and use the hypotheses (1), (2) and (3) to identify an n_0 such that for all $n > n_0$ holds

$$\mathcal{R}(x) < \epsilon.$$

This concludes the proof. □

We can now proof the main result of this article.

Proof. [Proof of Theorem 2.1] We follow the proof in [4].
Let $\mu_m = \frac{k}{2^m}$ when $\lambda \in [\frac{k-1}{2^m}, \frac{k}{2^m}[$

$$\mu_{m,n} = N_n + [z_n(\mu_n - \lambda)].$$

We notice that μ_m is discrete for every m, $0 < \mu_m - \lambda \leq \frac{1}{2^m}$ and that $\frac{\mu_{m,n}}{z_n} \xrightarrow{P} \mu_m > \lambda$.
We can rewrite the quantity $\frac{S_{N_n}}{\sqrt{N_n}}$ as

$$\begin{aligned}\frac{S_{N_n}}{\sqrt{N_n}} &= X_{m,n} + Y^{(1)}_{m,n}Z^{(2)}_{m,n} + Y^{(2)}_{m,n}Z^{(2)}_{m,n}\\ &= \frac{S_{\mu_{m,n}}}{\sqrt{\mu_{m,n}}} + \Big(\frac{S_{N_n} - S_{\mu_{m,n}}}{\sqrt{z_n\mu_m}}\Big)\sqrt{\frac{z_n\mu_m}{N_n}} + \Big(\frac{\sqrt{\mu_{m,n}} - \sqrt{N_n}}{\sqrt{N_n}}\Big)\frac{S_{\mu_{m,n}}}{\sqrt{\mu_{m,n}}}.\end{aligned}$$

From the proof of the discrete version we have the convergence of the law of the random variables

$$\frac{S_{\mu_{m,n}}}{\sqrt{\mu_{m,n}}} = X_{m,n} = Z^{(2)}_{m,n}$$

towards Φ for every m. If follows that

$$\limsup_{n\to\infty} \mathbf{P}\Big(\Big|\frac{S_{\mu_{m,n}}}{\sqrt{\mu_{m,n}}}\Big| > M\Big) = 1 - \Phi(M) + \Phi(-M).$$

For every fixed m we have that

$$\lim_{M\to\infty}\lim_{m\to\infty}\limsup_{n\to\infty} \mathbf{P}(|Z^{(2)}_{m,n}| > M) = 0.$$

For the quantity $\frac{m}{2^m} \leq N$, it holds

$$\lim_{n\to\infty}\Big|\frac{\sqrt{\mu_{m,n}} - \sqrt{N_n}}{\sqrt{N_n}}\Big| = \sqrt{\frac{\mu_m}{N_n}} - 1 \leq \sqrt{1 + \frac{1}{m}} - 1$$

so that the random variable $Y^{(2)}_{m,n}$ satisfies the hypothesis of Lemma 2.7, i.e. for every $\epsilon > 0$ and $j = 1, \dots, k$;

$$\lim_{m\to\infty}\limsup_{n\to\infty} \mathbf{P}(|Y^{(j)}_{m,n}| > \epsilon) = 0.$$

We can follow the proof of the last step as in [4], i.e. we have to prove that for every $\epsilon > 0$ and $j = 1, \dots, k$;

$$\mathcal{L}^\epsilon = \lim_{m\to\infty}\limsup_{n\to\infty} \mathbf{P}\Big(\Big|\frac{S_{N_n} - S_{\mu_{m,n}}}{\sqrt{z_n\mu_m}}\Big| > \epsilon\Big) = 0.$$

In order to calculate the limit, we will divide the event

$$\Big\{\Big|\frac{S_{N_n} - S_{\mu_{m,n}}}{\sqrt{z_n\mu_m}}\Big| > \epsilon\Big\}$$

in the union of simpler sets. With the help of the theorem of total probability we have the following estimation:

$$\begin{aligned}\mathcal{L}^\epsilon \leq &\limsup_{m\to\infty}\limsup_{n\to\infty} \mathbf{P}\Big(\Big|\frac{S_{N_n} - S_{\mu_{m,n}}}{\sqrt{z_n\mu_m}}\Big| > \epsilon, \Big|\frac{N_n}{z_n} - \lambda\Big| < \frac{1}{2^m}, \Big|\frac{\mu_{m,n}}{z_n} - \mu_m\Big| < \frac{1}{2^m}\Big)\\ &+ \limsup_{m\to\infty}\limsup_{n\to\infty} \mathbf{P}\Big(\Big|\frac{N_n}{z_n} - \lambda\Big| \geq \frac{1}{2^m}\Big)\\ &+ \limsup_{m\to\infty}\limsup_{n\to\infty} \mathbf{P}\Big(\Big|\frac{\mu_{m,n}}{z_n} - \mu_m\Big| \geq \frac{1}{2^m}\Big).\end{aligned}$$

By the hypothesis and observation on $\frac{N_n}{z_n}$ e $\frac{\mu_{m,n}}{z_n}$, these last two lim sup are equal to zero.
We can now concentrate on the following limit

$$\mathcal{L}_1^\epsilon = \limsup_{m\to\infty}\limsup_{n\to\infty}\mathbf{P}\Big(\max_{|\frac{i}{n}-\lambda|<\frac{1}{2^m},|\frac{j}{n}-\mu_n|<\frac{1}{2^m}}\Big|\frac{S_i-S_j}{\sqrt{z_n\mu_m}}\Big|>\epsilon\Big).$$

An application of the theorem of the total probability leads us to the following inequality

$$\mathcal{L}_1^\epsilon \leq \limsup_{m\to\infty}\limsup_{n\to\infty}\sum_{k=m}^{m2^m}\mathbf{P}\Big(\frac{k-1}{2^m}\leq\lambda<\frac{m}{2^m},\max_{\substack{|\frac{i}{z_n}-\lambda|<\frac{1}{2^m}\\|\frac{j}{z_n}-\lambda|<\frac{1}{2^{m-1}}}}\Big|\frac{S_i-S_j}{\sqrt{z_n\mu_m}}\Big|>\epsilon\Big)$$
$$+\limsup_{m\to\infty}\limsup_{n\to\infty}\mathbf{P}(\lambda<\frac{m-1}{2^m}\text{ or } m\leq\lambda).$$

Let $t=[z_n(k-3)s^{-m}]$. We can estimate $\mathcal{L}_1^\epsilon$ by the quantity

$$\mathcal{L}_2^\epsilon = \limsup_{m\to\infty}\limsup_{n\to\infty}\sum_{k=m}^{m2^m}\mathbf{P}\Big(\frac{k-1}{2^m}\leq\lambda<\frac{m}{2^m},\max_{\substack{|\frac{i}{z_n}-\lambda|<\frac{1}{2^m}\\|\frac{j}{z_n}-\lambda|<\frac{1}{2^{m-1}}}}\frac{|S_i-S_t|+|S_t-S_j|}{\sqrt{n\mu_m}}>\epsilon\Big).$$

We write now $A_{k,m}^\lambda=\{\frac{k-1}{2^m}\leq\lambda<\frac{k}{2^m}\}$. We have $\mathcal{L}_2^\epsilon\leq\mathcal{L}_3^\epsilon$, where

$$\mathcal{L}_3^\epsilon = \limsup_{m\to\infty}\sum_{k=m}^{m2^m}\limsup_{n\to\infty}2\mathbf{P}\Big(\max_{z_n(k-3)2^{-m}<r<z_n(k+3)2^{-m}}|S_r-S_t|>\frac{\epsilon}{2}\sqrt{\frac{z_nk}{2^m}}\Big|A_{k,m}^\lambda\Big)\mathbf{P}(A_{k,m}^\lambda).$$

By an application of (A2) we obtain

$$\mathrm{Var}\Big(S_{z_n(k+3)2^{-m}}-S_{z_n(k-3)2^{-m}}\Big)\leq\Big(\frac{z_n(k+3)}{2^m}-\frac{z_n(k-3)}{2^m}+1\Big)\tau^2=(6z_n2^{-m}+1)\tau^2.$$

We can use this last expression to estimate $\mathcal{L}_3^\epsilon$:

$$\begin{aligned}\mathbf{P}\Big(\max_{z_n(k-3)2^{-m}<r<z_n(k+3)2^{-m}}|S_r-S_t|>\frac{\epsilon}{2}\sqrt{\frac{z_nk}{2^m}}\Big) &\leq c\frac{2^{m+2}}{\epsilon^2z_nk}\mathrm{Var}\Big(S_{n(k+3)2^{-m}}-S_{n(k-3)2^{-m}}\Big)\\ &\leq c\frac{2^{m+2}}{\epsilon^2z_nk}(6n2^{-m}+1)\tau^2\\ &=c\tau^2\frac{24n+2^{m+2}}{\epsilon^2z_nk}.\end{aligned}$$

For $k\geq m$ holds

$$\limsup_{n\to\infty}\frac{24n+2^{m+2}}{\epsilon^2z_nk}c\tau^2\leq\frac{24}{\epsilon^2m}c\tau^2.$$

The proof of the theorem is a consequence of the following Lemma:

Lemma 2.9. *Let $\{k_n\}$ and $\{m_n\}$ be two divergent sequences and $A_n\in\sigma(\xi_{k_m},\ldots,\xi_{m_n})$. For every set A we have*

$$\limsup_{n\to\infty}\mathbf{P}(A_n|A)=\limsup_{n\to\infty}\mathbf{P}(A_n).$$

We have in fact that

$$0\leq\mathcal{L}^\epsilon\leq\limsup_{m\to\infty}\sum_{k=m}^{m2^m}\frac{48c\tau}{\epsilon^2m}\mathbf{P}(A_{k,m}^N)=\limsup_{m\to\infty}\frac{48}{\epsilon^2m}c\tau^2=0.$$

□

3. Example

Let us consider the example in [13]. Let N_n be a geometric sum with parameter $\frac{1}{n}$ so that in probability $\frac{N_n}{n} \to \lambda$, where λ is an exponential random variable.

$$\lim_{n\to\infty} \mathbf{P}\Big(\Big|\frac{N_n}{n} - \lambda\Big| > \delta\Big) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \mathbf{P}\Big(\Big|\frac{k}{n} - \lambda\Big| > \delta \,\Big| N_n = k\Big)\mathbf{P}(N_n = k)$$
$$= \lim_{n\to\infty} \sum_{k=1}^{\infty} \Big[(1 - e^{\delta - \frac{k}{n}}) + e^{-\delta - \frac{k}{n}}\Big]\Big(1 - \frac{1}{n}\Big)^{n\frac{k}{n}} \frac{1}{n}.$$

Since for every $k \le 1$

$$e^{-\frac{k}{n}} e^{\pm\delta} \le e^{-\frac{1}{n}} e^{\pm\delta} = e^{-\frac{1}{n} \pm \delta}$$
$$\lim_{n\to\infty} \sum_{k\in\mathbb{N}} \Big(1 - \frac{1}{n}\Big)^{n\frac{k}{n}} \frac{1}{n} = 2\int_0^{\infty} e^{-x} dx \le 2.$$

By the relation $\mathbf{P}(\lambda < 0) = 0$, for $\delta^{-1} \le n$

$$\mathbf{P}\Big(\Big|\frac{N_n}{n} - \lambda\Big| > \delta\Big) \le 1 - e^{\delta - \frac{1}{n}} + e^{-\delta - \frac{1}{n}} \mathbf{1}_{\delta^{-1} > n}$$
$$\mathbf{P}\Big(\Big|\frac{N_n}{n} - \lambda\Big| > \delta\Big) \le 1 - e^{\delta - \frac{1}{n}}.$$

Let $\delta > 0$. For every $\epsilon > 0$ solving $1 - e^{\delta - \frac{1}{n}} < \epsilon$ with respect to n we obtain an $n_0(\epsilon)$ such that for every $n \ge n_0$, holds $\lim_{n\to\infty} \mathbf{P}\Big(\Big|\frac{N_n}{n} - \lambda\Big| > \delta\Big) = 0$.
Applying the result

$$\frac{\sqrt{N_n}}{\tau}\Big(\frac{S_{N_n}}{N_n} - \mu\Big) \xrightarrow{L} \mathcal{N}(0, 1)$$

as $n \to \infty$, where $\tau^2 = \sigma^2 + \sum_{i=2}^{m+1} \mathrm{Cov}(X_i, X_1)$.

4. Concluding Remarks

In this section we would like to point out a possible extension of the invariance principle for randomly indexed m-dependent random variables. It would be desirable, in the framework of the previous sections, to prove the following

Theorem 4.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a stationary m-dependent sequence of random variables with zero means and finite variance. Let $\{N_n\}_{n\in\mathbb{N}}$ denote a positive sequence of positive integer-valued random variables such that*

$$\frac{N_n}{z_n} \xrightarrow{P} \lambda$$

as n goes to infinity, where $\{z_n\}_{n\in\mathbb{N}}$ is an arbitrary positive sequence tending to $+\infty$ and λ is a positive random variable. If

- *(A1) there exist some $k_0 \ge 0$ and $c > 0$ such that, for any $\lambda > 0$ and $n > k_0$ we have*

$$\mathbf{P}\Big(\max_{k_0 < k_1 \le k_2 \le n} |S_{k_2} - S_{k_1} - (k_2 - k_1)\mu| \ge \epsilon\Big) \le \frac{c \cdot Var(S_n - S_{k_0})}{\epsilon^2}$$

and

- *(A2)* $Cov(X_1, X_i) \geq 0$ *for* $i = 2, \ldots, m+1$,

then

$$\frac{\sum_{i=1}^{[\frac{N_n}{t}]} X_i + (N_n \cdot t - [\frac{N_n}{t}])X_{[\frac{N_n}{t}]+1}}{\sqrt{N_n}\tau} \xrightarrow{\mathcal{L}} W(t)$$

as $n \to \infty$, *where* $\tau^2 = \sigma^2 + \sum_{i=2}^{m+1} Cov(X_i, X_1)$.

Proof. [Idea of the Proof] To prove this result one might follow Theorem 3.1. in [3]. Form previous proofs we know that with $\tau\sqrt{N_n} := a_n$ the condition (i) in the theorem holds. It has to be checked if:
For each $c \in \mathbb{N}$ and each set $(\alpha_i, \beta_i : i = 1, \ldots, c) \in \mathbb{R}^{2c}$ and each $\epsilon > 0$, set $n_j := [\frac{jn}{c}]$ and

$$E_{n,r} := \{\omega \in \Omega : \alpha_j \leq \frac{S_{N_i}}{a_n}(\omega) \leq \beta_j : n_{j-1} < i \leq n_j, i < r \text{ and } i \neq r\}$$

holds

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{r=1}^{n} \mathbf{P}(E_{n,r} \cap \{|S_{r'} - S_r| \geq \epsilon a_n\}) = 0$$

where $r \mapsto r'$ via $r' = [n(N(j-1)+u)c^{-1}N^{-1}] := n_{j,u}$ such that $n_{j,u} < r \leq n_{j,u+1}$. □

This result may be useful in many practical situations where we are interested in random sums of a random number of objects: in these situations this result guarantees that we can still approximate the sample distribution with a normal random variable, independently of the convergence (in probability) of the normalized counting process.

References

[1] D. J. Aldous, Weak convergence of randomly indexed sequences of random variables, Math. Proc. Cambridge Philos. Soc. 83 (1978) 117–126.
[2] P. Billingsley, Probability and measure. John Wiley & Sons, 2008.
[3] P. Billingsley, The invariance principle for dependent random variables, Transactions of the American Mathematical Society (1956) 250–268.
[4] J. R. Blum, L. H. David, J. I. Rosenblatt, On the central limit theorem for the sum of a random number of independent random variables, Probability Theory and Related Fields 1.4 (1963) 389 – 393.
[5] V. Capasso, D. Bakstein, An introduction to continuous-time stochastic processes. Theory, models, 2005.
[6] I. Fakhre-Zakeri, F. Jamshid, A central limit theorem with random indices for stationary linear processes, Statistics & Probability Letters 17.2 (1993) 91–95.
[7] T. L. Hung, T. T. Tran, Some results on asymptotic behaviors of random sums of independent identically distributed random variables, Commun. Korean Math. Soc 25.1 (2010) 119–128.
[8] S. A. Orey, Central limit theorems for m-dependent random variables, Duke Math. J. 25 (1958) 543–546.
[9] M. Przystalski, Asymptotics for products of a random number of partial sums, Bull. Polish Acad. Sci. Math. 57 (2009) 163–167.
[10] A. Rényi, On the central limit theorem for the sum of a random number of independent random variables, Acta Mathematica Hungarica 11.1-2 (1960) 97–102.
[11] A. Rényi, On mixing sequences of sets, Acta Mathematica Hungarica 9.1-2 (1958) 215–228.
[12] Y. Shang, Central limit theorem for the sum of a random number of dependent random variables, Asian Journal of Mathematics and Statistics 4.3 (2011) 168–173.
[13] Y. Shang, A central limit theorem for randomly indexed m-dependent random variables, Filomat 26.4 (2012) 713–717.
[14] Y. Shang, A Note on the Central Limit Theorems for Dependent Random Variables, ISRN Probability and Statistics, vol. 2012, Article ID 192427, 2012.
[15] I. Sugiman, A random CLT for dependent random variables. Journal of multivariate analysis 20.2 (1986) 321–326.