ON THE STRONG LAW OF LARGE NUMBERS FOR $\varphi$-SUB-GAUSSIAN RANDOM VARIABLES

K. Zajkowski

UDC 519.21

For $p \geq 1$, let $\varphi_p(x) = x^2/2$ if $|x| \leq 1$ and $\varphi_p(x) = 1/p|x|^{p} - 1/p + 1/2$ if $|x| > 1$. For a random variable $\xi$, let $\tau_{\varphi_p}(\xi)$ denote $\inf\{a \geq 0 : \forall \lambda \in \mathbb{R} \ln E \exp(\lambda \xi) \leq \varphi_p(a \lambda)\}$; $\tau_{\varphi_p}$ is a norm in a space $\text{Sub}_{\varphi_p} = \{\xi : \tau_{\varphi_p}(\xi) < \infty\}$ of $\varphi_p$-sub-Gaussian random variables. We prove that if, for a sequence $(\xi_n) \subset \text{Sub}_{\varphi_p}$, $p > 1$, there exist positive constants $c$ and $\alpha$ such that, for every natural number $n$, the inequality $\tau_{\varphi_p}\left(\sum_{i=1}^{n} \xi_i\right) \leq cn^{1-\alpha}$ holds, then $n^{-1} \sum_{i=1}^{n} \xi_i$ converges almost surely to zero as $n \to \infty$. This result is a generalization of the strong law of large numbers for independent sub-Gaussian random variables [see R. L. Taylor and T.-C. Hu, Amer. Math. Monthly, 94, 295 (1987)] to the case of dependent $\varphi_p$-sub-Gaussian random variables.

1. Introduction

The classical Kolmogorov strong laws of large numbers deal with independent variables. The investigations of limit theorems for dependent random variables are extensive and episodic. The strong law of large numbers (SLLN) for various types of associated random variables one can find, e.g., in Bulinski and Shashkin [2] (Ch. 4). Most of these variables are considered in the spaces of integrable functions. It is also of interest to describe general conditions under which the SLLN is true in spaces of random variables other than the $L_p$-spaces. In the present paper, we investigate almost sure convergence of the arithmetic mean (but not only) sequences of $\varphi$-sub-Gaussian random variables.

The notion of sub-Gaussian random variables was introduced by Kahane in [8]. A random variable $\xi$ is called sub-Gaussian if its moment generating function is majorized by the moment generating function of a centered Gaussian random variable with variance $\sigma^2$, i.e.,

$$E \exp(\lambda \xi) \leq E \exp(\lambda g) = \exp(\sigma^2 \lambda^2/2),$$

where $g \sim \mathcal{N}(0, \sigma^2)$ (see [4] or [3], Chap. 1). In terms of cumulant generating functions, this condition takes the form $\ln E \exp(\lambda \xi) \leq \sigma^2 \lambda^2/2$.

We can generalize the notion of sub-Gaussian random variables to the classes of $\varphi$-sub-Gaussian random variables (see [3], Chap. 2). A continuous even convex function $\varphi(x)$ ($x \in \mathbb{R}$) is called an $N$-function if the following conditions are satisfied:

(a) $\varphi(0) = 0$ and $\varphi(x)$ is monotone increasing for $x > 0$,

(b) $\lim_{x \to 0} \varphi(x)/x = 0$ and $\lim_{x \to \infty} \varphi(x)/x = \infty$.

This function is called a quadratic $N$-function if, in addition, $\varphi(x) = cx^2$ for all $|x| \leq x_0$ with $c > 0$ and $x_0 > 0$.

Institute of Mathematics, University of Bialystok, Bialystok, Poland; e-mail: kryza@math.uwb.edu.pl.

Published in Ukraïns’kyi Matematychnyi Zhurnal, Vol. 73, No. 3, pp. 431–436, March, 2021. Ukrainian DOI: 10.37863/umzh.v73i3.197. Original article submitted July 11, 2018.

506 0041-5995/21/7303–0506 © 2021 Springer Science+Business Media, LLC
The quadratic condition is needed to ensure the nontriviality of the classes of \( \varphi \)-sub-Gaussian random variables (see [3, p. 67]).

**Example 1.1.** For \( p \geq 1 \), let

\[
\varphi_p(x) = \begin{cases} 
\frac{x^2}{2}, & \text{if } |x| \leq 1, \\
\frac{1}{p} |x|^p - \frac{1}{p} + \frac{1}{2}, & \text{if } |x| > 1.
\end{cases}
\]

The function \( \varphi_p \) is an example of quadratic \( N \)-function, which is a standardization of the function \( |x|^p \) (see Lemma 2.5 in [10]). We emphasize that, for \( p = 2 \), we have the case of sub-Gaussian random variables.

Let \( \varphi \) be a quadratic \( N \)-function. A random variable \( \xi \) is said to be \( \varphi \)-sub-Gaussian if there is a constant \( a > 0 \) such that \( \ln \mathbb{E} \exp(\lambda \xi) \leq \varphi(a \lambda) \). The \( \varphi \)-sub-Gaussian standard (norm) \( \tau_\varphi(\xi) \) is defined as follows:

\[
\tau_\varphi(\xi) = \inf \{ a \geq 0 : \forall \lambda \in \mathbb{R} \quad \ln \mathbb{E} \exp(\lambda \xi) \leq \varphi(a \lambda) \}.
\]

Note that a space \( \text{Sub}_\varphi = \{ \xi : \tau_\varphi(\xi) < \infty \} \) with the norm \( \tau_\varphi \) is a Banach space (see [3], Ch. 2, Theorem 4.1).

Let \( \varphi(x), x \in \mathbb{R}, \) be a real-valued function. The function \( \varphi^*(y) (y \in \mathbb{R}) \) defined by

\[
\varphi^*(y) = \sup_{x \in \mathbb{R}} \{ xy - \varphi(x) \}
\]

is called the Young–Fenchel transform or the convex conjugate of \( \varphi \) (in general, \( \varphi^* \) may take the value \( \infty \)). It is known that if \( \varphi \) is a quadratic \( N \)-function, then \( \varphi^* \) is also a quadratic \( N \)-function. Indeed, since our \( \varphi_p \) is a differentiable function (even for \( \pm 1 \)), one can easily check (see [10], Lemma 2.6) that \( \varphi_p^* = \varphi_q \) for \( p, q > 1 \) if \( 1/p + 1/q = 1 \).

**Remark 1.1.** We can define the space \( \text{Sub}_{\varphi_p} \) by using the Luxemburg norm of the form

\[
\| \xi \|_{\psi_q} = \inf \{ K > 0 : \mathbb{E} \exp |\xi|/K |^q \leq 2 \}, \quad q = p/(p - 1).
\]

Then (cf. [10])

\[
\text{Sub}_{\varphi_p} = \{ \xi : \| \xi \|_{\psi_q} < \infty \text{ and } \mathbb{E} \xi = 0 \},
\]

the space \( L^0_{\psi_q} = \text{Sub}_{\varphi_p} \). Note that \( \| E \xi \|_{\psi_q} = \| \xi \|_{\psi_q} \) and we conclude that if \( \| \xi \|_{\psi_q} < \infty \), then \( \xi - E \xi \in \text{Sub}_{\varphi_p} \).

**Example 1.2.** The standard normal random variable \( g \) belongs to \( \text{Sub}_{\varphi_2} \) and \( \tau_{\varphi_2}(g) = 1 \) because

\[
E \exp(tg) = \exp(t^2/2) = \exp(\varphi_2(t)).
\]

Since \( g^2 \) has the \( \chi^2_1 \)-distribution with one degree of freedom whose moment generating function is

\[
E \exp(tg) = (1 - 2t)^{-1/2} \quad \text{for} \quad t < 1/2,
\]
we conclude that
\[ \mathbb{E} \exp \left( \frac{g^2}{K^2} \right) = \left( 1 - 2/K^2 \right)^{-1/2}, \]
which is less or equal 2 if \( K \geq \sqrt{8/3} \). This gives \( \|g\|_{\psi_2} = \sqrt{8/3} \). We observe that the \( \psi_2 \)-norm of \( |g| \) is equal to the \( \psi_2 \)-norm of \( g \). This yields \( |g| - \mathbb{E}|g| \in \text{Sub}_{\psi_2} \). As above, we can show that
\[ \|g\|^{2/q}_{\psi_q} = \inf \{ K > 0 ; \ \mathbb{E} \exp(g^2/K^q) \leq 2 \} = (8/3)^{1/q} < \infty. \]
Thus, we get
\[ |g|^{2/q} - \mathbb{E}|g|^{2/q} \in \text{Sub}_{\varphi_p}, \]
where \( 1/p + 1/q = 1 \).

Recall that convex conjugation is order-reversing and possesses a certain scaling property. If \( \varphi_1 \geq \varphi_2 \), then \( \varphi_1^* \leq \varphi_2^* \). For \( a > 0 \) and \( b \neq 0 \), let \( \psi(x) = a \varphi(bx) \). Then \( \psi^*(y) = a \varphi^*(y/(ab)) \) (see, e.g., [6], Chap. X, Proposition 1.3.1).

The convex conjugate of the cumulant generating function can be used to estimate the “tails” of distribution of a centered random variable. By \( \mathbb{E}\xi = 0 \) and \( \psi_\xi \) we denote the cumulant generating function of \( \xi \), i.e.,
\[ \psi_\xi(\lambda) = \ln \mathbb{E} \exp(\lambda \xi). \]
Thus, for \( \varepsilon > 0 \),
\[ P(\xi \geq \varepsilon) \leq \exp(-\psi_\xi(\varepsilon)). \]

Note that, for \( \xi \in \text{Sub}_\varphi \), by the definition of \( \tau_\varphi(\xi) \), we get the inequality
\[ \psi_\xi(\lambda) \leq \varphi(\tau_\varphi(\xi) \lambda) \]
and, by the order-reversing and scaling properties, we obtain
\[ \psi_\xi^*(\varepsilon) \geq \varphi^*(\varepsilon/\tau_\varphi(\xi)). \]

We can now establish a weaker form of the above estimate but with the help of the general function \( \varphi \):
\[ P(|\xi| \geq \varepsilon) \leq 2 \exp\left( -\varphi^* \left( \frac{\varepsilon}{\tau_\varphi(\xi)} \right) \right) \quad (1) \]
(see [3], Chap. 2, Lemma 4.3).

2. Results

First, we show that if we have an upper estimate for \( \tau_\varphi \), then, in inequality (1) we can substitute this estimate for \( \tau_\varphi \).

**Lemma 2.1.** If \( \tau_\varphi(\xi) \leq C(\xi) \) for every \( \xi \in \text{Sub}_\varphi \), then
\[ P(|\xi| \geq \varepsilon) \leq 2 \exp\left( -\varphi^* \left( \frac{\varepsilon}{C(\xi)} \right) \right). \]
Proof. Since $\varphi$ is even and monotonically increasing for $x > 0$, we get

$$\varphi(\tau_\varphi(\xi)x) = \varphi(\tau_\varphi(\xi)|x|) \leq \varphi(C(\xi)|x|) = \varphi(C(\xi)x).$$

Thus, by using the order-reversing and scaling properties once again, we obtain

$$\varphi^*(\frac{y}{\tau_\varphi(\xi)}) \geq \varphi^*(\frac{y}{C(\xi)}).$$

In combination with (1), this establishes the required inequality.

With regard for the presented preliminaries, we can prove the main result of the paper.

Theorem 2.1. Let $(\xi_n) \subset \text{Sub}_{p}$ for some $p > 1$. If there exist positive constants $c$ and $\alpha$ such that, for every natural number $n$, the condition $\tau_{\varphi_p}(\sum_{i=1}^{n} \xi_i) \leq cn^{1-\alpha}$ is satisfied, then the term $n^{-1} \sum_{i=1}^{n} \xi_i$ converges almost surely to zero as $n \to \infty$.

Proof. Since $\varphi_p^* = \varphi_{q}$, by Lemma 2.1 and the condition of the theorem, we have

$$P\left(\left|\sum_{i=1}^{n} \xi_i\right| \geq n\varepsilon\right) \leq 2 \exp\left(-\varphi_{q}\left(\frac{n^{\alpha}\varepsilon}{c}\right)\right).$$

For sufficiently large $n$ ($n > (c/\varepsilon)^{1/\alpha}$), we find $n^{\alpha}\varepsilon/c > 1$ and, as a consequence,

$$\varphi_{q}\left(\frac{n^{\alpha}\varepsilon}{c}\right) = n^{q\alpha} \left(\frac{\varepsilon}{c}\right)^{q} - \frac{1}{q} + \frac{1}{2}.$$

Thus, we get the estimate

$$P\left(\left|\sum_{i=1}^{n} \xi_i\right| \geq n\varepsilon\right) \leq 2 \exp\left(\frac{1}{q} - \frac{1}{2}\right) \exp\left(-n^{q\alpha} \left(\frac{\varepsilon}{c}\right)^{q}\right)$$

for every $\varepsilon$ and $n > (c/\varepsilon)^{1/\alpha}$. Hence, by the integral test, we conclude that the series $\sum_{n=1}^{\infty} P\left(\left|\sum_{i=1}^{n} \xi_i\right| \geq n\varepsilon\right)$ is convergent. This implies the complete and, as a consequence, almost sure convergence of $n^{-1} \sum_{i=1}^{n} \xi_i$ to zero.

Remark 2.1. We emphasize that the above theorem is a generalization of the theorem (SLLN) (see [9, p. 297]) to the case of $\varphi_p$-sub-Gaussian random variables but not only to sub-Gaussian variables. Moreover, we do not assume their independence. For this reason we used a modified condition for the behavior of a norm $\tau_p$ rather than the Taylor and Hu condition. It is described below.

Since $\tau_\varphi$ is a norm, we obtain

$$\tau_\varphi\left(\sum_{i=1}^{n} \xi_i\right) \leq \sum_{i=1}^{n} \tau_\varphi(\xi_i).$$
If, e.g., \( \xi_i, i = 1, \ldots, n \), are copies of the same variable \( \xi \), then, in the above, the equality holds and
\[
\tau_{\varphi} \left( \sum_{i=1}^{n} \xi_i \right) = n \tau_{\varphi}(\xi).
\]
Note that, in this case, the assumption of Theorem 2.1 is not satisfied. Additional information about the form of dependence (or independence) sometimes allows us to improve this estimate. Hence, for an independent sequence \((\xi_n)\), if there is some \( r \in (0, 2] \) such that \( \varphi(|x|^{1/r}) \) is convex, then
\[
\tau_{\varphi} \left( \sum_{i=1}^{n} \xi_i \right)^r \leq \sum_{i=1}^{n} \tau_{\varphi}(\xi_i)^r
\]
(see [3], Sec. 2, Theorem 5.2). If \( r \) is larger, then the estimate is better. For the function \( \varphi_p \), we can always take \( r = \min\{p, 2\} \). In Taylor’s and Hu’s SLLN, the variables \( \xi_n \) are sub-Gaussian and independent. Moreover, it is assumed that \( p = 2 \). We emphasize that if, in addition, \( \xi_1, \ldots, \xi_n \) have the same distribution in this case as \( \xi \), then
\[
\tau_{\varphi} \left( \sum_{i=1}^{n} \xi_i \right) \leq \sqrt{n} \tau_{\varphi}(\xi)
\]
and the condition of Theorem 2.1 is satisfied \((c = \tau_{\varphi}(\xi) \text{ and } \alpha = 1/2)\).

Note that other assumptions about the dependence of \( \xi_1, \ldots, \xi_n \) may give the same estimate for the norm of \( \tau_{\varphi} \left( \sum_{i=1}^{n} \xi_i \right) \). In the paper by G. Antonini, et al. [5] (Lemma 3), it was proved that, for \( \varphi \)-sub-Gaussian acceptable random variables, the inequality (2) holds if \( \varphi(|x|^{1/r}) \) is convex. The definition of acceptability of a sequence of random variables can be found therein. For our presentation, it is especially important that these estimates are identical. In this paper, there is a version of the Marcinkiewicz–Zygmund law of large numbers for \( \varphi \)-sub-Gaussian random variables obtained as a corollary of a much more general theorem. We present an independent proof of this corollary but under modified assumptions.

Proposition 2.1. Let \((\xi_n), p > 1\), be a bounded sequence of \( \varphi_p \)-sub-Gaussian random variables and let \( r = \min\{p, 2\} \). If, in addition,
\[
\tau_{\varphi_p} \left( \sum_{i=1}^{n} \xi_i \right)^r \leq \sum_{i=1}^{n} \tau_{\varphi_p}(\xi_i)^r,
\]
then \( n^{-1/s} \sum_{i=1}^{n} \xi_i \to 0 \) almost surely for any \( 0 < s < r \).

Remark 2.2. Since \( \varphi_p(|x|^{1/r}) \) is convex, estimate (3) is satisfied, e.g., by sequences of independent or acceptable random variables.

Proof. Let \( b = \sup_{n \geq 1} \tau_{\varphi_p}(\xi_n) \). Then
\[
\sum_{i=1}^{n} \tau_{\varphi_p}(\xi_i)^r \leq nb^r
\]
and, as a consequence,
\[
\tau_{\varphi_p} \left( \sum_{i=1}^{n} \xi_i \right) \leq n^{1/r}b.
\]
For a positive number \( s \) smaller than \( r \), by Lemma 2.1, we obtain
\[
\mathbb{P}\left(\sum_{i=1}^{n} \xi_i \geq n^{1/s} \varepsilon\right) \leq 2 \exp\left(-\varphi_q\left(\frac{n^{1/s} \varepsilon}{n^{1/r} b}\right)\right) = 2 \exp\left(-\varphi_q\left(n^{(1/s-1/r)} \varepsilon\right)\right).
\]

For \( n > (b/\varepsilon)^{(1/s-1/r)^{-1}} \), we get
\[
\varphi_q\left(n^{(1/s-1/r)} \varepsilon\right) = n q^{(1/s-1/r)} \frac{1}{q} \left(\frac{\varepsilon}{b}\right)^q - \frac{1}{q} + \frac{1}{2}
\]
and, consequently,
\[
\sum_{n=1}^{\infty} \exp\left(-\varphi_q\left(n^{(1/s-1/r)} \varepsilon\right)\right) < \infty.
\]

In view of the Borel–Cantelli lemma, this completes the proof.

**Remark 2.3.** Since we apply the function \( \varphi_p(x) \) instead of \( |x|^p \), we should not restrict ourselves to \( p \) not greater than 2 in order to guarantee the validity of the quadratic condition for the function \( |x|^p \). Moreover, we use the metric property (3) instead of assumptions about the form of dependence of the random variables (cf. [5], Corollary 7).

**Example 2.1.** The proof of the Hoeffding–Azuma inequality for a sequence \((\xi_n)\) of bounded random variables such that \( |\xi_n| \leq d_n \) a.s. and \( \mathbb{E}\xi_n = 0 \) is based on an estimate of the moment generating function of the partial sum \( \sum_{i=1}^{n} \xi_i \). Under the assumption that \( \xi_n \) are independent (Hoeffding) or \( \xi_n \) are martingales increments (Azuma), the following inequality holds:
\[
\mathbb{E}\exp\left(\lambda \sum_{i=1}^{n} \xi_i \right) \leq \exp\left(\frac{\lambda^2 \sum_{i=1}^{n} d_i^2}{2}\right)
\]
(see [7, 1]). We emphasize that, in [1], Azuma proved the above estimate under more general assumptions about \((\xi_n)\), which satisfy centered bounded martingales increments. Inequality (4) means that
\[
\tau_{\varphi_2}\left(\sum_{i=1}^{n} \xi_i \right) \leq \left(\sum_{i=1}^{n} d_i^2\right)^{1/2}.
\]
If we take \( d_n = 1 \) for \( n = 1, 2, \ldots \), then we get the condition
\[
\tau_{\varphi_2}\left(\sum_{i=1}^{n} \xi_i \right) \leq \sqrt{n},
\]
which means that the sequence \((\xi_n)\) satisfies the assumptions of Proposition 2.1 with \( p = r = 2 \) and the norm \( \tau_{\varphi_2}(\xi_n) \leq 1 \). Thus, we get the almost sure convergence of \( n^{-1/s} \sum_{i=1}^{n} \xi_i \) to 0 for any \( 0 < s < 2 \). Note that, for \( s = 1 \), we obtain the SLLN for this sequence.
REFERENCES

1. K. Azuma, “Weighted sums of certain dependent random variables,” *Tokohu Math. J.* (2), 19, 357–367 (1967).
2. A. Bulinski and A. Shashkin, “Limit theorems for associated random fields and related systems,” *Advanced Series on Statistical Science & Applied Probability*, 10, World Scientific, Hackensack (2007).
3. V. Buldygin and Yu. Kozachenko, “Metric characterization of random variables and random processes,” *Amer. Math. Soc.*, Providence, RI (2000).
4. V. Buldygin and Yu. Kozachenko, “Sub-Gaussian random variables,” *Ukr. Mat. Zhit.*, 32, No. 6, 723–730 (1980); *English translation*: *Ukr. Math. J.*, 32, No. 6, 483–489 (1980).
5. R. G. Antonini, Yu. Kozachenko, and A. Volodin, “Convergence of series of dependent $\varphi$-sub-Gaussian random variables,” *J. Math. Anal. Appl.*, 338, No. 2, 1188–1203 (2008).
6. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms. II*, Springer-Verlag, Berlin, Heidelberg (1993).
7. W. Hoeffding, “Probability for sums of bounded random variables,” *J. Amer. Statist. Assoc.*, 58, 13–30 (1963).
8. J. P. Kahane, “Local properties of functions in terms of random Fourier series [in French],” *Studia Math.*, 19, No. 1, 1–25 (1960).
9. R. L. Taylor and T.-C. Hu, “Sub-Gaussian techniques in proving strong laws of large numbers,” *Amer. Math. Monthly*, 94, 295–299 (1987).
10. K. Zajkowski, “On norms in some class of exponential type Orlicz spaces of random variables,” *Positivity*, 24, 1231–1240 (2020).