General Random Sum Limit Theorems for Martingales with large $O$ — Rates

PAUL L. BUTZER (and D. SCHULZ)'


Работа посвящена $O$-оценкам погрешности для сходимости в распределении случайных сумм из случайных величин, которые не обязательно независимы. Применением некоторой общей теоремы выводятся варианты центральной предельной теоремы и слабого закона больших чисел для случайного ряда мартингалов в случае случайного суммирования по последовательности частного выбора предельной случайной переменной. В обоих теоремах даются $O$-оценки для скорости сходимости.

This paper is concerned with large-$O$ error estimates for convergence in distribution of random sums of not necessarily independent random variables. As applications of a general theorem one obtains the random-sum versions of the central limit theorem and of the weak law of large numbers for martingale difference sequences by specializing the limiting random variable. Both theorems are equipped with $O$-rates.

Dedicated to the memory of WOLFGANG RICHTER (1932—1972), a scholar of the theory of randomly indexed random variables.

1. Introduction and History

The central limit theorem (CLT), perhaps the most important limit theorem of probability theory, may be formulated as follows: Let $(X_i)_{i \in \mathbb{N}} (\mathbb{N} = \{1, 2, \ldots\})$ be a sequence of real, independent, square-integrable random variables (r.v.s.) defined on a probability space $(\Omega, \mathcal{F}, P)$, let $S_n := \sum_{i=1}^{n} X_i$ denote its $n$th partial sum, $E[S_n]$ the expectation, and $\text{Var} [S_n]$ the variance of $S_n$. Then $(X_i)_{i \in \mathbb{N}}$ is said to satisfy the CLT provided the sequence $P_{S_n}$ of distributions of the normalized sums

$$
\bar{T}_n := (S_n - E[S_n])/(\text{Var} [S_n])^{1/2}
$$

converges weakly to the standard normal distribution, i.e.,

$$
\lim_{n \to \infty} E[f(\bar{T}_n)] = E[f(X^*)] \quad (f \in C_b),
$$

(1.1)

1) The research of the second named author was supported by DFG grant Bu 166/37.
where $X^*$ is the standard normally distributed r.v., and $C_b := C_b (\mathbb{R})$ the class of all bounded, uniformly continuous, real-valued functions $f$ defined on the axis $\mathbb{R}$, endowed with norm $\|f\|_{C_b} := \sup_{x \in \mathbb{R}} |f(x)|$.

In 1948 H. Robbins [37] gave sufficient conditions for the validity of the random-sum CLT. He generalized the classical CLT in the sense that he replaced the index "$n$" of $S_n$ by a positive, N-valued r.v. $N$, depending on a parameter $\lambda \in \mathbb{R}^+$, the family $(N_i)_{i \in \mathbb{N}}$ being defined on the same probability space $(\Omega, \mathcal{F}, P)$ as the sequence $(X_i)_{i \in \mathbb{N}}$. Relation (1.1) in the case of the CLT for randomly indexed sequences of r.v.s reads

$$\lim_{i \to \infty} \mathbb{E}[\bar{f}(T_{N_i})] = \mathbb{E}[f(X^*)] \quad (f \in C_b),$$

where $\bar{T}_{N_i} := (S_{N_i} - \mathbb{E}[S_{N_i}])/(\text{Var } S_{N_i})^{1/2}$, $S_{N_i} := \sum_{i=1}^{N_i} X_i$.

Whereas H. Robbins [37] assumed the r.v.s. $N_i, \lambda \in \mathbb{R}^+$, to be independent of the $X_i, i \in \mathbb{N}$. F. J. Anscombe [2] was the first to consider conditions for the validity of the random CLT without this restriction. These two classical papers were followed up by a series of papers on limit theorems for randomly indexed sequences of r.v.s., such as A. Rényi [34], J. R. Blum, D. L. Hanson, J. Rosenblatt [7], W. Richter [36], S. H. Sirazdinov—G. Orozov [43], Z. Rychlik [39] and D. J. Aldous [1]. An excellent survey on limit theorems in this connection is to be found in the Habilitation—thesis of Richter [35].

Another possibility to generalize (1.1) consists in dropping the independency assumption upon the r.v.s. $X_i, i \in \mathbb{N}$. Since it is generally difficult to find sufficient conditions for the convergence of arbitrary dependent r.v.s, one usually restricts oneself to particular types of dependency. In this respect martingale difference sequences (MDS) and martingale difference arrays have been examined to an especially great extent. The pioneering papers and books here are those of P. Lévy [29, 30, 31, p. 242] and J. L. Doob [18: p. 383], which were followed up by [6, 25, 19, 9, 42, 28, 21], for example. By a MDS is meant the following: Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of real r.v.s. defined on $(\Omega, \mathcal{F}, P)$, and let $(\mathcal{G}_i)_{i \in \mathbb{N}}$ be an increasing sequence of sub-$\sigma$-algebras of $\mathcal{F}$ such that $X_i$ is $\mathcal{G}_i$-measurable for each $i \in \mathbb{N}$. Then $(X_i, \mathcal{G}_i)_{i \in \mathbb{N}}, X_0 := 0$ is called a MDS if

$$\mathbb{E}[X_i | \mathcal{G}_{i-1}] = 0 \quad \text{a.s.} \quad (i \in \mathbb{N}).$$

For the martingale random CLT the reader is referred to [16, 32, 17, 40].

The paper by M. Csörgő [16] cited is, according to the best of our knowledge, the first ever concerned with the CLT for martingales in the case of randomly indexed r.v.s. The first result dealing with rates of convergence for the CLT for martingales is apparently due to I. A. Ibragimov (see [25]). The latter paper is the forerunner of a series of results in this field (see e.g. [24, 22, 3, 44; 15: p. 314, 26, 8, 4, 38, 23: Sect. 3.6, 14], the rates in [4] being established for $\mathbb{R}^n$-valued r.v.s., and in [38, 14] for Banach-space valued r.v.s.).

The purpose of this paper is to study rates of convergence for martingales in the instance of randomly indexed sequences of r.v.s, a topic that has so far been considered only by B. L. S. Prakasa Rao [33] and J. Strobel in his dissertation [44]. More concretely, the aim is to deduce the random CLT as well as the random weak law of large numbers (WLLN), both taken with rates, as applications of one general theorem. To carry out this unified approach, conditions are given which lead to large-$\mathcal{O}$ estimates for the difference

$$|\mathbb{E}[\bar{f}(T_{N_i})] - \mathbb{E}[\bar{f}(Z)]|$$

(1.4)
for all \( f \in C_b^0(\mathbb{R}) \). Here

\[
T_{N_j} := \phi(N_j) S_{N_j},
\]

(1.5)

\((X_i)_{i\in\mathbb{P}}\) are the first components of the MDS \((X_i, \mathfrak{F}_i)_{i\in\mathbb{P}}, \phi: \mathbb{N} \rightarrow \mathbb{R}^+\) is a positive, normalizing function, and \( Z \) is a limiting r.v. that is assumed to be \( \phi \)-decomposable. This means that for each \( n \in \mathbb{N} \) there exist independent r.vs. \( Z, Z_i \equiv Z_{i,n}, 1 \leq i \leq n \), such that the distribution \( P_Z \) of \( Z \) can be represented as

\[
P_Z = P_{\frac{1}{\phi(n)\sum_{i=1}^n Z_i}}.
\]

Furthermore, for \( r \in \mathbb{P} \) we have set \( C_b^0(\mathbb{R}) = C_b(\mathbb{R}) \),

\[
C_b^0(\mathbb{R}) := \{ f \in C_b(\mathbb{R}) : f^{(j)} \in C_b(\mathbb{R}), 1 \leq j \leq r \}.
\]

(1.7)

For details concerning the relationship between the concepts of \( \phi \)-decomposability and infinite divisibility see [14].

In the sequel it will always be assumed that the r.vs. \( N_i, \lambda \in \mathbb{R}^+ \), and \( X_i, i \in \mathbb{N} \), are independent, and that \( N_i \rightarrow \infty \) in probability for \( \lambda \rightarrow \infty \).

Since the limiting r.v. \( Z \) can be chosen rather generally, in particular as the Gaussian r.v. \( X^* \) and the degenerate r.v. \( X_0 \) (which takes on the value 0 with probability 1); respectively, both being \( \phi \)-decomposable, the random — sum CLT and WLLN with rates will be deduced as particular cases. The latter does not seem to have been considered at all in this frame.

Inspite of our general approach for MDS, our convergence rates are even better than those for sums of independent r.vs. due to Z. Rychlik and D. Szynal [41] on account of our use of \( K \)-functional methods. They are indeed just as sharp as those of P. L. Butzer and L. H"{a}ns [11, 12] in the case of non-random summation of independent r.vs. Returning to the proofs again; our main theorem is based upon a modification of the Trotter operator-theoretic method to the situation of not necessarily independent r.vs. as already applied in [3, 14]. This time it is tailored to the situation of randomly indexed r.vs. \( X_i \) which are independent of the index variable \( N_i, \lambda \in \mathbb{R}^+ \).

As an illustration of our results let us formulate a particular case of our random CLT with \( \theta \)-rates:

Let \((X_i, \mathfrak{F}_i)_{i\in\mathbb{P}}\) be a MDS, and let \( 0 < \alpha \leq 3 \). If \( f \in \text{Lip} (\alpha; 3, C_b) \) (see (2.3)) and

\[
E[|X_i|^3] < \infty \quad (i \in \mathbb{N},)
\]

(1.8)

as well as

\[
E[X_i | \mathfrak{F}_{i-1}] = E[X^*] \quad \text{a.s.} \quad (1 \leq j \leq 2, i \in \mathbb{N}),
\]

(1.9)

then

\[
\left| E[f(S_{N_j} l/\sqrt{N_j})) - E[f(X^*)] \right|
\leq C_f \left\{ E \left[ N_j^{-3/2} \sum_{i=1}^{N_j} (E[|X_i|^3] + E[|X^*|^3]) \right] \right\}^{1/3} \quad (\lambda \rightarrow \infty).
\]

(1.10)

In particular, if the r.vs. \( X_i \) are identically distributed, then the order of approximation in (1.10) is \( \mathcal{O}((E[N_i^{-1/2}])^{1/3}) \).

In the case of independent, identically distributed r.vs. \( X_i, i \in \mathbb{N}, Z. Rychlik \) and D. Szynal [41] deduced the rate \( \mathcal{O}(E[N_i^{-\beta/2}]) \) with \( 0 < \beta \leq 1 \) under corresponding assumptions upon the moments of \( X_i \).

By applying a result of V. M. Zolotarev [47] on the Kolmogorov distance between the distribution functions \( F_X \) and \( F_Y \) of two r.vs. \( X \) and \( Y \) (see (4.2) below), the follow-
ing estimate concerning the strong convergence of distributions can be derived from (1.10):

If \((X_i, \tilde{\gamma}_i)_{i \in \mathbb{Z}}\) is a MDS and \(f \in \text{Lip}(3; 3; \mathcal{N})\), then one has under the assumptions (1.8), (1.9)

\[
\sup_{t \in \mathbb{R}} \left| F_{S_{N_i}}(t) - F_{X^*}(t) \right| \leq M \left( E \left[ \frac{N_i^{-3/2}}{N_i} \sum_{i=1}^{N_i} \left( E[|X_i|^3] + E[|X^*|^3] \right) \right] \right)^{1/4},
\]

and in the case of identically distributed r.v.s. \((X_i)_{i \in \mathbb{Z}}\) this estimate is of order \(O(\lambda^{-1/3})\), \(\lambda \to \infty\), provided \(E[N_i^{-1/2}] = O(\lambda^{-1/2})\).

The best possible convergence rate reached by D. LANDERS and L. ROGGE [27] in the case of independent, identically distributed r.v.s., namely \(O(\lambda^{-1/2})\), cannot be achieved by our methods. This is due to the ZOLOTAREV estimate (see [47, 48]) used (see also Sec. 4).

Section 2 is concerned with questions of notation, the \(K\)-functional, moduli of continuity and Lipschitz classes. Section 3 is devoted to our general theorem on the convergence in distribution of the r.v.s. \(T_{N_i}\) towards a \(q\)-decomposable r.v. \(Z\). In Section 4 this theorem is applied to the strong convergence in distribution, and in Sections 5 and 6 to the random CLT and WLLN, respectively.

2. Notations and Preliminaries

The \(K\)-funktional and modulus of continuity, defined in terms of the spaces \(C_{B} \) and \(C_{B}^r, r \in \mathbb{N}\) (cf. (1.1) and (1.7)), need to be recalled. For any \(f \in C_{B}\) and \(t \geq 0\) the former is defined by

\[
K(t; f; C_{B}, C_{B}^r) := \inf_{g \in C_{B}^r} \{ \|f - g\|_{C_{B}} + t \|g\|_{C_{B}^r} \},
\]

\(|\cdot|\) being the semi-norm on \(C_{B}^r\), given by \(\|g\|_{C_{B}^r} := \|g^{(r)}\|_{C_{B}}\), and the \(r\)th modulus of continuity by

\[
\omega_r(t; f; C_{B}) := \sup_{|u| \leq t} \left\| \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} f(u + k\delta) \right\|.
\]

For each \(f \in C_{B}\) and each \(t \geq 0\) the \(K\)-functional is equivalent to this modulus (see [10: pp. 492, 258]), i.e., there are positive constants \(c_{1,r}\) and \(c_{2,r}\), independent of \(f\) and \(t \geq 0\), such that

\[
c_{1,r} \omega_r(t^{1/r}; f; C_{B}) \leq K(t; f; C_{B}, C_{B}^r) \leq c_{2,r} \omega_r(t^{1/r}; f; C_{B}).
\]

This enables one to define a Lipschitz class of index \(r \in \mathbb{N}\) and order \(\alpha, 0 < \alpha \leq r\), by

\[
\text{Lip} (\alpha; r; C_{B}) := \{ f \in C_{B} : \omega_r(t; f; C_{B}) \leq L_f^\alpha, t > 0 \},
\]

\(L_f^\alpha\) being the Lipschitz constant.

The concept of \(q\)-decomposability, defined in (1.6), can be extended to randomly indexed r.v.s. since the range of the index variable \(N_i\) is a subset of \(\mathbb{N}\). In fact, if \(Z\) is a \(q\)-decomposable r.v., then (1.6) implies

\[
P_{Z} = P_{q(N_i) \sum_{i=1}^{N_i} Z_i} (\lambda \in \mathbb{R}^+).
\]
If $p_n = p_n(n)$ denotes the probability with which the index variable $N$ takes on the value $n \in N$, then $\sum_{n=1}^{\infty} p_n = 1$, and so (2.4) yields that

$$P_{Z} = \sum_{n=1}^{\infty} p_n p(n) \sum_{i=1}^{n} Z_i$$

provided the r.vs. $Z_i$, $i \in N$, are independent of $N$ for each $n \in R^+$. Likewise one has for the expectations $E[Z]$ and $E[T_{N}]$ of $Z$ and $T_{N}$ (recall (1.5)), respectively,

$$E[Z] = \sum_{n=1}^{\infty} p_n E\left[\varphi(n) \sum_{i=1}^{n} Z_i\right], \quad (2.5)$$

$$E[T_{N}] = \sum_{n=1}^{\infty} p_n E[T_n]. \quad (2.6)$$

3. General Convergence Theorem for MDS with Rates

This section is concerned with the general limit theorem with rates described in the introduction. The proof is based upon the Trotter operator approach, first applied to the CLT in [45]. However, it has to be generalized so as to be applicable to MDS in the case of random summation instead of just sequences of independent r.vs. Although the proof may appear rather long and technical, it is nevertheless elementary. It uses Taylor series expansions, the operational rules for conditional expectations, and standard K-functional arguments of approximation theory.

**Theorem 1**: Let $(X_i, \mathfrak{S}_i) \in \mathfrak{P}$ be a MDS and $Z$ be a $\varphi$-decomposable r.v. with $E[Z] = 0$, such that

$$\xi_{r, i} := E[X_i | \mathfrak{S}_i] < \infty, \quad \xi_{r, i} := E[Z_i | \mathfrak{S}_i] < \infty \quad (i \in N) \quad (3.1; 3.2)$$

for some $r \in N$, $r \geq 2$, as well as

$$E[X_i | \mathfrak{S}_{i-1}] = E[Z_i] \quad a.s. \quad (i \in N, 1 \leq j \leq r - 1). \quad (3.3)$$

a) Under these hypotheses one has for $f \in C_B$

$$|E[f(T_{N})] - E[f(Z)]| \leq c_{2, \omega_r} \left( E \left[ \left( \varphi(N) \sum_{i=1}^{N} (\xi_{r, i} + \xi_{r, i}) \right) \right]^{1/r} \right) \quad (3.4)$$

$c_{2, \omega}$ being the constant of (2.2). In particular, if $f \in \text{Lip} (\alpha; r; C_B)$, $\alpha \in (0, r]$, then

$$|E[f(T_{N})] - E[f(Z)]| \leq 2c_{2, \omega_r} L_f \left( E \left[ \left( \varphi(N) \sum_{i=1}^{N} (\xi_{r, i} + \xi_{r, i}) \right) \right] \right)^{1/r}. \quad (3.5)$$

b) If the r.vs. $X_i$, $i \in N$, as well as the decomposition components $Z_i$, $i \in N$, are identically distributed, then

$$|E[f(T_{N})] - E[f(Z)]| \leq 2c_{2, \omega_r} \left\{ (\xi_{r, 1} + \xi_{r, 1}) E[\varphi(N)]^{1/r} \right\}. \quad (3.6)$$

If $f \in \text{Lip} (\alpha; r; C_B), 0 < \alpha \leq r$, then the left side of (3.6) has the bound

$$2c_{2, \omega_r} \left\{ (\xi_{r, 1} + \xi_{r, 1}) E[\varphi(N)]^{1/r} \right\}. \quad (3.7)$$
Proof: a) Let \( f \in C_B, g \in C_{B^*} \) be arbitrary. Making use of the fact that 
\[ |E[f(Z)]| \leq \|f\|_{C_B} \text{ for all } f \in C_B \text{ and any r.v. } Z, \]
\[ |E[f(T_{N_k})] - E[f(Z)]| \leq |E[f(T_{N_k})] - E[g(T_{N_k})]| + |E[g(T_{N_k})] - E[g(Z)]| + |E[g(Z)] - E[f(Z)]| \]
\[ \leq 2 \|f - g\| + |E[g(T_{N_k})] - E[g(Z)]|. \quad (3.8) \]

In order to estimate the second term, if \((Z_i)_{i \in \mathbb{N}}\) is the sequence of independent r.v.s. of (1.6), not only can the \((Z_i)_{i \in \mathbb{N}}\) be chosen to be independent of the r.v.s. \(N_1, \lambda \in \mathbb{R}^+\) by a proper choice of the underlying probability space (c.f. [20: p. 79/80]), but also the \(\sigma\)-algebras generated by the \(Z_i\) can be assumed to be independent of the sub-\(\sigma\)-
algebras \(\mathcal{F}_{i,i}, i \in \mathbb{N}\).

Setting \(R_{n,i} := \sum_{k=1}^{n-1} X_k + \sum_{k=i+1}^n Z_k, 1 \leq k \leq n, n \in \mathbb{N}\), a double application of

Taylor's formula for \(g \in C_{B^*}\) yields

\[
g(T_n) - g \left( \varphi(n) \sum_{i=1}^n Z_i \right) = \sum_{i=1}^n \left\{ g \left( \varphi(n) R_{n,i} + \varphi(n) X_i \right) - g \left( \varphi(n) R_{n,i} + \varphi(n) Z_i \right) \right\} \]

\[
= \sum_{i=1}^n \sum_{j=1}^{r-1} \frac{1}{(r-2)!} \int_0^1 \left( 1 - ty^{r-2} \left\{ g^{(r-1)}(\varphi(n) R_{n,i} + t\varphi(n) X_i) - g^{(r-1)}(\varphi(n) R_{n,i}) \right\}(\varphi(n) X_i)^{r-1} dt \right)
- g^{(r-1)}(\varphi(n) R_{n,i}) \left( \varphi(n) X_i \right)^{r-1} dt \]
\]

Since \(g \in C_{B^*}\) one has \(g^{(r-1)} \in \text{Lip}(1; 1; C_B)\) with \(L_g := \|g^{(r)}\|\), and so for \(0 < t \leq 1, \)
\[
\left\{ g^{(r-1)}(\varphi(n) R_{n,i} + t\varphi(n) X_i) - g^{(r-1)}(\varphi(n) R_{n,i}) \right\}(\varphi(n) X_i)^{r-1} \]
\[ \leq \|g^{(r)}\| \left\{ \varphi(n) \right\}^r \left| X_i \right|^r \text{ a.s., and analogously for the r.v.s. } Z_i. \]

In view of (2.5) and (2.6) this leads to

\[
|E[g(T_{N_k})] - E[g(Z)]| \leq \sum_{n=1}^\infty \sum_{i=1}^n \left\| g^{(r)}(\varphi(n) R_{n,i}) \left( X_i(\varphi(n))^i \right) \right\| \frac{1}{(r-1)!} \left\{ \varphi(n) \right\}^r \sum_{i=1}^n \left( \xi_{r,i} + \xi_{r,i} \right). \quad (3.9) \]
Let us now show that

\[ E[X, i g^{(i)}(\varphi(n) R_{i, n})] = E[Z, i g^{(i)}(\varphi(n) R_{i, n})] (1 \leq i \leq n, n \in N; 1 \leq j \leq r). \tag{3.10} \]

Setting \( \mathcal{F}_{i, n} := \mathcal{F}(\mathcal{F}_{i-1} \cup \mathcal{F}(Z_{i+1}, \ldots, Z_n)), \mathcal{F}(\mathcal{G}) \) and \( \mathcal{F}(X) \) being the \( \sigma \)-algebras generated by \( \mathcal{G} \) and \( X \), respectively, where \( \mathcal{G} \subset \mathcal{F}(\Omega) \), one has by standard arguments for the conditional expectation of real r.v.s,

\[ E[X, i g^{(i)}(\varphi(n) R_{i, n})] = E[Z, i g^{(i)}(\varphi(n) R_{i, n})] \]

since \( g^{(i)}(\varphi(n) R_{i, n}) \) is measurable with respect to \( \mathcal{F}_{i, n} \). Since moreover \( \mathcal{F}(Z_{i+1}, \ldots, Z_n) \) is independent of \( \mathcal{F}(\mathcal{F}_{i-1} \cup \mathcal{F}(X, i)) \), one has (cf. [5: p. 295]) \( E[X, i | \mathcal{F}_{i, n}] = E[X, i | \mathcal{F}_{i-1}] \) a.s. As \( \mathcal{F}(Z_i) \) is independent of \( \mathcal{F}_{i, n} \), one finally has \( E[Z, i | \mathcal{F}_{i, n}] = E[Z, i] \) a.s., and so \( E[X, i | \mathcal{F}_{i, n}] = E[Z, i | \mathcal{F}_{i, n}] \) a.s. on account of assumption (3.3). This establishes the validity of (3.10). Since the double sum in (3.9) vanishes, (3.8) yields

\[ \left| E[f(T_{N_1})] - E[f(Z)] \right| \leq 2 \| f - g \| + \frac{\| g^{(r)} \|_1}{(r - 1)!} E \left[ (\varphi(N_1))^{\frac{N_j}{2}} \sum_{i=1}^{N_j} (\xi_{i, i} + x_{i, i}) \right] \]

on account of (2.5) and (2.6). But the left side of this inequality is independent of \( g \), so that taking the infimum over all \( g \in C_{b, r} \) yields by (2.1)

\[ \left| E[f(T_{N_1})] - E[f(Z)] \right| \leq 2K \left( E \left[ (\varphi(N_1))^{\frac{N_j}{2}} \sum_{i=1}^{N_j} (\xi_{i, i} + x_{i, i}) \right] \right)^{1/r}; C_{b, r} \]

The first assertion, namely (3.4), of part a) now follows immediately by (2.2), and the second, namely (3.5), by (2.3). Part b) is a particular case of a). \[ \]

Remark 1: Note that Theorem 1 covers the situation that the r.v.s. \( X_i \) are independent (since such r.v.s. with \( E[X_i] = 0 \) form a MDS). In this instance condition (3.3) reduces to \( E[X, i] = E[Z, i], i \in N, 1 \leq j \leq r - 1 \), an assumption already used in [12], for example.

Remark 2: The estimate in (3.4) is only of practical interest provided the modulus \( \omega_r \) tends to zero for \( r \to \infty \), that is, if

\[ E \left[ (\varphi(N_1))^{\frac{N_j}{2}} \left( E[|X_i|^r] + E[|Z_i|^r] \right) \right] \]

tends to 0 for \( r \to \infty \). If the r.v.s. \( X_i, i \in N \), and \( Z_i, i \in N \), are in particular identically distributed, and one sets \( \varphi(N_1) = N_1^{-1/2}, r \geq 3 \) (see Theorem 3 b)), then the special bounds in (3.6) and (3.7) tend to zero since \( E[N_1^{1/2-r/2}] \to 0 \) for \( r \to \infty \) in view of the hypothesis that \( N_1 \to \infty \) in probability.

4. Approximation Theorem for Distribution Functions with Rates

In this section we shall examine the rate of approximation for the strong convergence in distribution of the r.v.s. \( T_{N_1} \) towards \( Z \). For this purpose, it suffices to apply the following result of V. M. Zolotarev, contained implicitly in [47]; it permits one to pass from weak convergence to strong convergence in distribution;
Lemma: Let $Y$ be a real r.v. with distribution function $F_Y$ for which there exists a constant $M_Y = M_Y > 0$ such that

$$|F_Y(t) - F_Y(s)| \leq M_Y |t - s| \quad (s, t \in \mathbb{R}, s < t),$$

(4.1)

and let $r \in \mathbb{N}$. For each r.v. $X$ and each constant $M_2 > 0$ there exists a constant $M = M(M_1, M_2) > 0$ such that the so-called Kolmogorov distance between the distribution functions $F_X$ and $F_Y$, namely

$$\sup_{t \in \mathbb{R}} |F_X(t) - F_Y(t)| \leq M \left( \sup_{f \in D} |E[f(X)] - E[f(Y)]| \right)^{1/(r + 1)},$$

(4.2)

the function class $D = D(M_2, r)$ being defined by

$$D := \{ f \in C^r_{-1}; f^r \in \text{Lip}_{M}(1; 1; C_0) \},$$

(4.3)

the Lipschitz constant $L_{jfr}$ being uniformly bounded by $M_2$.

As an immediate consequence of Theorem 1 and (4.2) we have the following theorem.

Theorem 2: Let the assumptions of Theorem 1 be satisfied, and let (4.1) hold for the limiting r.v. $Z$. Then

a. $\sup_{t \in \mathbb{R}} |F_{T_{N_2}}(t) - F_Z(t)| \leq M \left( \sup_{f \in D} |E[f(N)]f \sum_i (\xi_{t, i} + \xi_{r, i})| \right)^{1/(r + 1)}$

b. If the r.v.s $X_i$ and $Z_i$, $i \in \mathbb{N}$, are additionally identically distributed, and $\varphi(N_i) := N_i^{-1/2}$, then

$$\sup_{t \in \mathbb{R}} |F_{T_{N_2}}(t) - F_Z(t)| = \tilde{c} \left( \left( E \left[ N_i^{2/(r+1)} \right] \right)^{1/(r+1)} \right) \quad (\lambda \to \infty).$$

The order of approximation deduced for the weaker convergence can in general not be transferred to the associated strong convergence in distribution although it is known that both types of convergence are equivalent when considering convergence per se (without rates). (Further details to this and to other assertions equivalent to the convergence in distribution without as well as with rates may be found in Butzer-Hahn [13]).

Since the supremum of the right side of (4.2) over all $f \in D$ is smaller than the same supremum taken over all $f \in C_{r+1}(\mathbb{R})$, it is to be expected that the power $1/(r + 1)$ cannot be dropped in estimates of type (4.2) for general r.v.s $Y$ that need not satisfy condition (4.1). This will now be shown by means of a simple example taken from [46].

If one considers the distribution function $F_{X_0}$ of the degenerate r.v. $X_0$ (cf. Section 1), and if $F_{\tilde{X}_t}$ is a further distribution function of the form $F_{\tilde{X}_t}(t) = (1 - \varepsilon) \times F_{X_0}(t) + \varepsilon F_{X_0}(t - \varepsilon)$, $t \in \mathbb{R}$, for some $\varepsilon \in (0, 1)$, then one has immediately

$$\sup_{t \in \mathbb{R}} |F_{X_0}(t) - F_{\tilde{X}_t}(t)| = \varepsilon.$$

(4.4)

Now with the help of the metric $\kappa_\varepsilon$ defined in [46] by

$$\kappa_\varepsilon(X, Y) := s \int |t|^s-1 |F_X(t) - F_Y(t)| dt \quad (s \geq 1)$$

for any real r.v. $X, Y$, Zolotarev established the inequality

$$\sup_{f \in D} |E[f(X)] - E[f(Y)]| \leq \frac{1}{I(r)} \kappa_r(X, Y),$$

(4.5)

where $r \in \mathbb{N}$ is taken as in the definition of the class $D$ of (4.3).
In the case of our example, the r.vs. $X_0$ and $X_0$ are at a distance $\kappa_r(X, X_0) = \left(\frac{r}{r+1}\right)^{r+1}$ apart. So one has by (4.4)

$$\sup_{t \in \mathbb{R}} |F_{X_n}(t) - F_{\overline{X}_n}(t)| \leq \left(2\kappa_r(X_0, \overline{X}_0)\right)^{1/(r+1)} \quad (r \in \mathbb{N}).$$

(4.6)

A comparison of (4.5) with (4.6) therefore shows that when passing from the distance $\sup_{t \in \mathbb{R}} |E[f(X_0)] - E[f(\overline{X}_0)]|$ to the Kolmogorov-metric $\sup_{t \in \mathbb{R}} |F_{X_n}(t) - F_{\overline{X}_n}(t)|$, the rate of convergence becomes poorer, at least for those r.vs. $Z$ which do not necessarily satisfy a Lipschitz condition of type (4.1). Indeed, the r.v. $X_0$ does not satisfy (4.1). However, whether the estimate (4.2) and so the convergence rate in Theorem 2 could possibly be improved (in the sense that the power $1/(r+1)$ in the estimate of Theorem 2 could be dropped or increased) for those limiting r.vs. $Z$ satisfying (4.1) is a fact that is unknown to the authors.

5. The Random Sum CLT for MDS with Rates

Let us apply our two theorems to a concrete limiting r.v. $Z$, namely to the Gaussian distributed r.v. $X^*$ with mean zero and variance 1. Theorem 1 yields the following random CLT with large-0 rates for martingales.

**Theorem 3:** Let $(X_i, \mathcal{F}_i)_{i \in \mathbb{N}}$ be a MDS, $r \in \mathbb{N}$ and $(a_i)_{i \in \mathbb{N}}$ any sequence of positive real numbers. Assume that (3.1) holds, i.e.,

$$\zeta_{r,i} := E[|X_i|^r] < \infty \quad (i \in \mathbb{N}),$$

as well as

$$E[X_i^i | \mathcal{F}_{i-1}] := a_i E[X^* | i] \quad \text{a.s.} \quad (i \in N, 1 \leq j \leq r - 1).$$

(5.1)

a) Under these hypotheses one has for $f \in C_b$,

$$E[f(A_{X_1}^i S_{X_1})] - E[f(X^*)] \leq 2c_{r,r} \omega_r \left( E \left[ A_{X_1}^r \sum_{i=1}^{N_1} (\zeta_{r,i} + a_i E[|X^*|^r]) \right]^{1/r} ; f; C_B \right).$$

(5.2)

c. being the constant of (2.2), and $A_{N_1} := \left( \sum_{i=1}^{N_1} a_i^2 \right)^{1/2}$. In particular, if $f \in \text{Lip}(\alpha; r; C_B)$, $\alpha \in (0, r]$, then the bound in (5.2) takes on the form

$$2c_{r,r} L_f \left( E \left[ A_{X_1}^r \sum_{i=1}^{N_1} (\zeta_{r,i} + a_i E[|X^*|^r]) \right] \right).$$

b) If, in addition, the r.vs. $X_i, i \in \mathbb{N}$, are identically distributed, and (5.1) holds for $a_i = 1, i \in \mathbb{N}$, then

$$E[f(S_{X_1}/N_1^{1/2})] - E[f(X^*)] \leq 2c_{r,r} \omega_r \left( (\zeta_{r,1} + E[|X^*|^r]) E[N_1^{-1/2}] \right)^{1/r} ; f; C_B \right).$$

(5.3)

If $f \in \text{Lip}(\alpha; r; C_B)$, $\alpha \in (0, r]$, then the bound in (5.3) reads

$$2c_{r,r} L_f (\zeta_{r,1} + E[|X^*|^r]) E[N_1^{-(r-1)/2}] r^{1/r}.$$  

(5.4)

**Proof:** The r.v. $X^*$ is $\varphi$-decomposable for each $n \in \mathbb{N}$ into $n$ independent, normally distributed r.vs. $Z_i, 1 \leq i \leq n$, namely $Z_i = a_i X^*$. Moreover, one can ensure
as in the proof of Theorem 1 that the $Z_i, i \in \mathbb{N}, N_i, \lambda \in \mathbb{R}^+$, as well as the sub-$\sigma$-algebras $\mathcal{F}_i, i \in \mathbb{N}$, are all independent. So $X^*$ can be decomposed according to (1.6) with $\varphi(n) := \left( \sum_{i=1}^n a_i^2 \right)^{-1/2}$. Since $E[Z_i] = E[a_i X^*] = a_i E[X^*]$, for $i \in \mathbb{N}$, assumption (3.2) is satisfied on account of (5.1). So Theorem 1 may be applied since the moments of (4.1) exist here, too.

Concerning the previous literature in the matter, the only paper known to the authors is that by Basu [4], who in the classical case of non-random summation deduced the estimate

$$\left| E[f(S_n/n^{1/2})] - E[f(X^*)] \right| = \mathcal{O}(n^{-5/2}) \quad (n \to \infty)$$

provided that $(X_i)_{i \in \mathbb{N}}$ defines a MDS of stationary r.v.s satisfying the assumptions of Theorem 3 b) in the particular case $r = 3$ and $f \in \text{Lip} (\alpha; C_d), \alpha \in (2, 3)$. Comparing this estimate with ours, (5.4) for $r = 3$ and non-random summation gives the order $\mathcal{O}(n^{-3/2}), \alpha \in (2, 3)$, even in the case of non-stationary sequences $(X_i)_{i \in \mathbb{N}}$.

As an application of Theorem 2 to $X^*$ we have:

**Theorem 4**: a) Under the assumptions of Theorem 3 one has

$$\sup_{t \in \mathbb{R}} \left| F_{S_{N_1}/N_1}(t) - F_{X^*}(t) \right| \leq M \left\{ E \left[ A_{N_1}^{-1} \sum_{i=1}^{N_1} (a_i + a_i E[X^*]) \right] \right\}^{1/(r+1)}$$

b) If the r.vs. $X_i$ are identically distributed, and (5.1) holds for $a_i = 1, i \in \mathbb{N}$, then

$$\sup_{t \in \mathbb{R}} \left| F_{S_{N_1}/N_1}(t) - F_{X^*}(t) \right| = \mathcal{O}(n^{-1/2}) \quad (n \to \infty)$$

If $r = 3$, and $E[N_i^{-1/2}] = \mathcal{O}(\lambda^{-1/2}), \lambda \to \infty$, then the latter estimate is of order $\mathcal{O}(\lambda^{-1/8})$.

**Remark 3**: In his dissertation [44] Strobel obtains under additional assumptions as a corollary of results in a more general setting the estimate

$$\sup_{t \in \mathbb{R}} \left| F_{S_{N_1}/N_1}(t) - F_{X^*}(t) \right| = \mathcal{O}(n^{-1/8} + P(\nu_n > n)) \quad (n \to \infty)$$

for a sequence of independent, identically distributed, normalized r.vs. $(X_i)_{i \in \mathbb{N}}$, and a sequence of stopping rules $(\nu_n)_{n \in \mathbb{N}}$ for $\mathcal{F}_n := \mathcal{F}((X_1, \ldots, X_n))$.

6. The Random - Sum WLLN for MDS with Rates

The final application of Theorem 1 will be the WLLN of the title with $\varphi$-error estimates, a result that does not seem to have been considered before. When examining the WLLN in connection with r.vs. $T_{N_1}$ defined in (1.5), one normally thinks of stochastic convergence of $T_{N_1}$ towards zero, namely

$$\lim_{k \to \infty} P(|T_{N_1}| \geq \varepsilon) = 0 \quad (\varepsilon > 0). \quad (6.1)$$

Instead, we consider a formulation of the random WLLN which can be shown to be equivalent to (6.1), just as in the classical situation (see e.g. [5: p. 220]), namely

$$\lim_{k \to \infty} \left| E[f(T_{N_1})] - f(0) \right| = 0 \quad (f \in C^*) \quad (6.2)$$
for any \( r \in \mathbb{N} \). Version (6.2) has the advantage that Theorem 1 can be applied directly; one just needs to choose the limiting r.v. \( Z \) as the degenerate r.v. \( X_0 \) defined in Section 1. In this frame the random WLLN for MDS with \( \theta \)-rates reads

**Theorem 5:** Let \((X_i, \mathcal{F}_i)_{i \in \mathbb{N}}\) be a MDS such that \( E(|X_i|^2) < \infty, i \in \mathbb{N} \).

a) If \( f \in \text{Lip} (\alpha; 2; C_B) \), then

\[
|E[f(T_{N_i})] - f(0)| \leq 2c_2 \sqrt{\sum_{i=1}^{N_i} E(|X_i|^2)}^{\alpha/2}.
\]

b) Under the assumption

\[
\sum_{i=1}^{N_i} E(|X_i|^2) = o((\varphi(N_i)^{-2}) \text{ a.s. } \varphi(\to \infty),
\]

one has

\[
\lim_{i \to \infty} P(||T_{N_i}|| \geq \varepsilon) = 0 \quad (\varepsilon > 0).
\]

In particular, if \( \varphi(N_i) = N_i^{-1} \) in (6.3), then \( S_{N_i}/N_i \to 0 \) in probability for \( \lambda \to \infty \), i.e., \((X_i)_{i \in \mathbb{N}}\) satisfies the random WLLN.

**Proof:** First note that \( E[f(X_0)] = \int f(x) dP_{X_0}(x) = f(0) \). The distribution \( P_{X_0} \) can be decomposed as \( P_{X_0} = \sum_{n=1}^{\infty} p_n P_{\varphi(n)} \mathcal{Z}_n \) with \( P_{\mathcal{Z}} = P_{X_0} \) for all \( i \in \mathbb{N} \). Since

\[
\int |x|^s dP_{X_0}(x) = 0 \text{ for any } s > 0 \text{, condition (3.2) of Theorem 1 is satisfied. Then assumption (3.3) follows immediately from (1.3) and the distribution of the r.v.s. } \mathcal{Z}_n.
\]

Part b) follows from a), noting the equivalence of (6.1) and (6.2). 

The authors would like to thank Dr. L. Hahn for-kindly supervising the early stages of the work of the second named author, as well as Prof. M. Csööö, Ottawa, Canada, for his generous help in connection with the literature.

**REFERENCES**


Manuskripteingang: 2. 4. 1982

VERFASSER:

Prof. Dr. PAUL L. BÜTZER und D. SCHULZ
Lehrstuhl A für Mathematik der Rheinisch-Westfälischen Technischen Hochschule
D-5100 Aachen, Templergraben 55