

A WEAK LAW OF LARGE NUMBERS FOR DEPENDENT RANDOM VARIABLES *

IOANNIS KARATZAS †

WALTER SCHACHERMAYER ‡

April 20, 2022

Abstract

Every sequence f_1, f_2, \dots of random variables with $\lim_{M \rightarrow \infty} (M \sup_{k \in \mathbb{N}} \mathbb{P}(|f_k| > M)) = 0$ contains a subsequence f_{k_1}, f_{k_2}, \dots that satisfies, along with all its subsequences, the weak law of large numbers: $\lim_{N \rightarrow \infty} ((1/N) \sum_{n=1}^N f_{k_n} - D_N) = 0$, in probability. Here D_N is a “corrector” random variable with values in $[-N, N]$, for each $N \in \mathbb{N}$; these correctors are all equal to zero if, in addition, $\liminf_{k \rightarrow \infty} \mathbb{E}(f_k^2 \mathbf{1}_{\{|f_k| \leq M\}}) = 0$ holds for every $M \in (0, \infty)$.

AMS 2020 Subject Classification: Primary 60A10; Secondary 60F99.

Keywords: Weak law of large numbers, hereditary convergence, weak convergence, truncation, generalized mathematical expectation, nonlinear expectation

1 Introduction

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider real-valued measurable functions f_1, f_2, \dots . If these are independent and have the same distribution with $\mathbb{E}(|f_1|) < \infty$, the celebrated KOLMOGOROV strong law of large numbers (SLLN: [9]; [8]; [3], section 2.4) states that the “sample average” $(f_1 + \dots + f_N)/N$ converges \mathbb{P} -a.e. to the “ensemble average” $\mathbb{E}(f_1) = \int_{\Omega} f_1 d\mathbb{P}$, as $N \rightarrow \infty$.

A bit more generally, if the functions $f_k(\omega) = f(T^{k-1}(\omega))$, $k \geq 1$, $\omega \in \Omega$ are the images of an integrable function $f_1 : \Omega \rightarrow \mathbb{R}$ along the orbit of successive actions of a measure-preserving transformation $T : \Omega \rightarrow \Omega$, then the above sample average converges \mathbb{P} -a.e. as $N \rightarrow \infty$ to the conditional expectation $f_* = \mathbb{E}(f_1 | \mathcal{I})$ of f_1 , given the σ -algebra \mathcal{I} of T -invariant sets, by BIRKHOFF’s pointwise ergodic theorem ([1]; [3], p. 333).

A deep result of KOMLÓS [10], already 55 years old but always very striking, says that such “stabilization via averaging” occurs within *any* sequence f_1, f_2, \dots of measurable, real-valued functions with $\sup_{n \in \mathbb{N}} \mathbb{E}(|f_n|) < \infty$. More precisely, there exist then an integrable function f_* and a

* We are indebted to Albert SHIRYAEV for pointing out to us the concept of generalized mathematical expectations in KOLMOGOROV [9], which prompted this investigation. We are grateful to Nathan SOEDJAK for his careful reading of the manuscript and his incisive comments.

† Departments of Mathematics and Statistics, Columbia University, New York, NY 10027 (e-mail: ik1@columbia.edu). Support from National Science Foundation Grant DMS-20-04977 is gratefully acknowledged.

‡ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria (email: walter.schachermayer@univie.ac.at). Support from the Austrian Science Fund (FWF) under grant P-28861 and grant P-35197 is gratefully acknowledged.

subsequence $\{f_{k_n}\}_{n \in \mathbb{N}}$ such that $(f_{k_1} + \dots + f_{k_N})/N$ converges to f_* , \mathbb{P} -a.e. as $N \rightarrow \infty$; and the same is true “hereditarily”, that is, for any further subsequence of $\{f_{k_n}\}_{n \in \mathbb{N}}$.

We have also another celebrated result of KOLMOGOROV, the weak law of large numbers (WLLN: [9]; [2], section 5.2; [3], § 2.2.3) for a sequence f_1, f_2, \dots of real-valued, measurable functions which are independent. If these are identically distributed and satisfy the weak- \mathbb{L}^1 -type condition

$$\lim_{M \rightarrow \infty} \left(M \cdot \mathbb{P}(|f_1| > M) \right) = 0 \quad (1.1)$$

(rather than $\mathbb{E}(|f_1|) < \infty$), then we have the WLLN

$$\lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N f_n - D_N \right) = 0, \quad \text{in probability} \quad (1.2)$$

for the sequence of “correctors”

$$D_N := \mathbb{E} \left(f_1 \mathbf{1}_{\{|f_1| \leq N\}} \right), \quad N \in \mathbb{N}; \quad (1.3)$$

whereas, if the independent functions f_1, f_2, \dots do not have the same distribution but satisfy

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N \mathbb{P}(|f_n| > N) = 0, \quad \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{n=1}^N \mathbb{E} \left(f_n^2 \mathbf{1}_{\{|f_n| \leq N\}} \right) = 0, \quad (1.4)$$

then again the convergence in probability (WLLN) in (1.2) holds, though now with correctors

$$D_N := \frac{1}{N} \sum_{n=1}^N \mathbb{E} \left(f_n \mathbf{1}_{\{|f_n| \leq N\}} \right), \quad N \in \mathbb{N}. \quad (1.5)$$

It was shown in [5], [4], ([2], Theorem 5.2.3) that, for independent f_1, f_2, \dots , the conditions in (1.4) are not only sufficient but also necessary for the existence of a sequence D_1, D_2, \dots of real numbers with the property (1.2).

Let us also note, that the correctors in both (1.3), (1.5) satisfy $|D_N| \leq N$; and that they are all equal to zero, if each of the f_1, f_2, \dots has distribution symmetric around the origin.

The purpose of this Note is to present a KOMLÓS-type version of the weak law of large numbers. This is formulated in the next section, and proved in section 3. The proof, considerably simpler than its counterpart for the strong law in [10], is based on truncation and on weak- \mathbb{L}^2 convergence arguments, which give also sufficient conditions for the resulting correctors to be equal to zero. Examples and ramifications are taken up in section 4.

2 Result

We consider real-valued measurable functions f_1, f_2, \dots on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and introduce for every $M \in (0, \infty)$ the quantities

$$\tau_n(M) := M \cdot \mathbb{P}(|f_n| > M), \quad \tau(M) := \sup_{n \in \mathbb{N}} \tau_n(M). \quad (2.1)$$

Theorem 2.1. A General, Hereditary WLLN. *In the above context, suppose that the weak- \mathbb{L}^1 -type condition*

$$\lim_{M \rightarrow \infty} \tau(M) = 0 \quad (2.2)$$

holds. There exist then a sequence of “corrector” random variables D_1, D_2, \dots with

$$\mathbb{P}(|D_N| \leq N) = 1 \quad \text{for every } N \in \mathbb{N}, \quad (2.3)$$

and a subsequence $\{f_{k_n}\}_{n \in \mathbb{N}}$, such that the WLLN

$$\lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N f_{k_n} - D_N \right) = 0, \quad \text{in probability} \quad (2.4)$$

is satisfied “hereditarily”; i.e., not just by $\{f_{k_n}\}_{n \in \mathbb{N}}$ but also by all its subsequences. If, in addition,

$$\liminf_{k \rightarrow \infty} \mathbb{E} \left(f_k^2 \mathbf{1}_{\{|f_k| \leq M\}} \right) = 0 \quad (2.5)$$

holds for every $M \in (0, \infty)$, the correctors in (2.4), (2.3) can be chosen as $D_N = 0$ for every $N \in \mathbb{N}$.

The correctors D_1, D_2, \dots correspond to the generalized mathematical expectations in KOLMOGOROV [9], §6.4; they are also related to the nonlinear expectations developed by PENG in [11].

3 Proof

We start with the simple but crucial idea of “truncation”. This goes back to the work of KHINTCHINE and KOLMOGOROV ([5], [7]), where it plays a major role in the proofs of laws of large numbers and of convergence results for series of random variables.

Lemma 3.1. *Under the condition (2.2), we have*

$$\lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N f_n - \frac{1}{N} \sum_{n=1}^N f_n \mathbf{1}_{\{|f_n| \leq N\}} \right) = 0, \quad \text{in probability.} \quad (3.1)$$

Proof: For every $\varepsilon > 0$, the expression

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_{n=1}^N f_n \mathbf{1}_{\{|f_n| > N\}} \right| > \varepsilon \right) \leq \mathbb{P} \left(\bigcup_{n=1}^N \{|f_n| > N\} \right) \leq \sum_{n=1}^N \mathbb{P}(|f_n| > N) \leq N \cdot \max_{1 \leq n \leq N} \mathbb{P}(|f_n| > N)$$

is dominated by $N \cdot \sup_{n \in \mathbb{N}} \mathbb{P}(|f_n| > N)$, which converges to zero as $N \uparrow \infty$ on account of (2.2). \square

It follows that, in order to establish (2.4), it is enough to prove

$$\lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N f_n \mathbf{1}_{\{|f_n| \leq N\}} - D_N \right) = 0, \quad \text{in probability} \quad (3.2)$$

for a suitable sequence D_1, D_2, \dots of correctors, and along an appropriate subsequence of $\{f_n\}_{n \in \mathbb{N}}$ denoted by the same symbols for economy of exposition—as well as along all further subsequences of this subsequence.

PROOF OF THEOREM 2.1: For each integer $N \in \mathbb{N}$ we consider the truncated functions

$$f_n^{[-N,N]} := f_n \mathbf{1}_{\{|f_n| \leq N\}}, \quad n \in \mathbb{N} \quad (3.3)$$

that appear in (3.1), (3.2). These are bounded in \mathbb{L}^∞ (as they take values in $[-N, N]$), thus bounded in \mathbb{L}^2 as well. As a result we can extract, for each $N \in \mathbb{N}$, a subsequence of $\{f_n\}_{n \in \mathbb{N}}$ denoted by the same symbols for economy of exposition, such that the sequence in (3.3) converges weakly in \mathbb{L}^2 to some $D_N \in \mathbb{L}^2$:

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(f_n^{[-N,N]} \cdot \xi \right) = \mathbb{E} (D_N \cdot \xi), \quad \forall \xi \in \mathbb{L}^2. \quad (3.4)$$

And by standard diagonalization arguments, we can extract then a further subsequence of $\{f_n\}_{n \in \mathbb{N}}$, denoted again by the same symbols, and such that (3.4) holds for *every* $N \in \mathbb{N}$.

It is fairly straightforward to check that these weak- \mathbb{L}^2 limits in (3.4) satisfy (2.3). On the other hand, the lower-semicontinuity of the \mathbb{L}^2 -norm under weak- \mathbb{L}^2 convergence, in this case

$$\|D_N\|_{\mathbb{L}^2} \leq \liminf_{n \rightarrow \infty} \|f_n^{[-N,N]}\|_{\mathbb{L}^2},$$

gives $\mathbb{P}(D_N = 0) = 1$ for every $N \in \mathbb{N}$, under the condition (2.5).

We introduce now, for each $M \in (0, \infty)$, the quantities

$$\sigma_n(M) := \frac{1}{M} \mathbb{E} \left(f_n^2 \mathbf{1}_{\{|f_n| \leq M\}} \right), \quad \sigma(M) := \sup_{n \in \mathbb{N}} \sigma_n(M). \quad (3.5)$$

As shown by FELLER ([4], p. 235; see also [3], § 2.3.3), these quantities are related to those in (2.1) via

$$0 \leq \sigma_n(M) = \frac{2}{M} \int_0^M \tau_n(t) dt - \tau_n(M) \leq \frac{2}{M} \int_0^M \tau(t) dt \quad (3.6)$$

for every $n \in \mathbb{N}$, $M \in (0, \infty)$,¹ thus

$$0 \leq \sigma(M) \leq \frac{2}{M} \int_0^M \tau(t) dt, \quad M \in (0, \infty). \quad (3.7)$$

From this bound (3.7) and the assumption (2.2), it follows that we have

$$\lim_{M \rightarrow \infty} \sigma(M) = 0. \quad (3.8)$$

We note also

$$\mathbb{E} \left(f_n^{[-M,M]} \right)^2 = \mathbb{E} \left(f_n^2 \mathbf{1}_{\{|f_n| \leq M\}} \right) \leq M \cdot \sigma_n(M) \leq M \cdot \sigma(M) \quad (3.9)$$

for all $n \in \mathbb{N}$, $M \in (0, \infty)$, therefore

$$\mathbb{E} (D_M^2) \leq \sup_{n \in \mathbb{N}} \mathbb{E} \left(f_n^{[-M,M]} \right)^2 \leq M \cdot \sigma(M) = o(M), \quad \text{as } M \rightarrow \infty. \quad (3.10)$$

We observe at this point that, in order to prove (3.2) and thus (2.4) as well, along a suitable subsequence, it is enough to show convergence along this subsequence in \mathbb{L}^2 , namely

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \cdot \mathbb{E} \left(\sum_{n=1}^N \left(f_n^{[-N,N]} - D_N \right) \right)^2 = 0. \quad (3.11)$$

¹ In the integrand of this expression, as it appears on page 235 of [4], there is a typo; this is here corrected.

And developing the square, we need to show that the expectations of both the sum of squares

$$\sum_{n=1}^N \mathbb{E} \left(f_n^{[-N,N]} - D_N \right)^2 \leq 2 \sum_{n=1}^N \mathbb{E} \left(f_n^{[-N,N]} \right)^2 + 2N \cdot \mathbb{E} (D_N)^2 \quad (3.12)$$

and the sum of cross-products

$$2 \sum_{n=1}^N \sum_{1 \leq j < n} \mathbb{E} \left[\left(f_j^{[-N,N]} - D_N \right) \left(f_n^{[-N,N]} - D_N \right) \right] \quad (3.13)$$

are of order $o(N^2)$, as $N \rightarrow \infty$, for the subsequence in question and for all its subsequences. Now, from (3.9), (3.10), the expression in (3.12) is already dominated by $4N^2 \cdot \sigma(N) = o(N^2)$, as $N \rightarrow \infty$, on account of (3.8).

Let us recall what happens at this juncture in the case of independent f_1, f_2, \dots : the correctors D_N are real constants, given as in (1.5), so the differences $f_n^{[-N,N]} - D_N$, $n = 1, \dots, N$ are independent with zero mean, thus uncorrelated. The expectations of their cross-product in (3.13) vanish, and the argument ends here.

In the general case, when *nothing* is assumed about the finite-dimensional distributions of the f_1, f_2, \dots (in particular, when these functions are not independent) we need to guarantee, by passing to a further subsequence if necessary, that the expression in (3.13) is also of order $o(N^2)$, as $N \rightarrow \infty$. One way to accomplish this, is to ensure that the differences $f_n^{[-N,N]} - D_N$, $n = 1, \dots, N$ are *very close to being uncorrelated*.

We do this by induction, in the following manner: Suppose the terms f_1, \dots, f_{n-1} of the subsequence have been chosen. We select the next term f_n in such a way, that the difference $f_n^{[-N,N]} - D_N$, with $N \leq e^{n^2}$, is “almost orthogonal” to all of the $f_1^{[-N,N]} - D_N, \dots, f_{n-1}^{[-N,N]} - D_N$; to wit,

$$\left| \mathbb{E} \left[\left(f_j^{[-N,N]} - D_N \right) \left(f_n^{[-N,N]} - D_N \right) \right] \right| \leq e^{-n^2} \leq \frac{1}{N} \quad (3.14)$$

for every $j = 1, \dots, n-1$. Such a choice of f_n is certainly possible on account of (3.4), and completes the induction step.

Returning to (3.13), we note that the summation

$$2 \sum_{n=1}^{\sqrt{\log N}} \sum_{1 \leq j < n} \left| \mathbb{E} \left[\left(f_j^{[-N,N]} - D_N \right) \left(f_n^{[-N,N]} - D_N \right) \right] \right|$$

is straightforward to control: each summand is bounded by $N \cdot \sigma(N)$ on account of (3.9), (3.10), so the entire summation is of the order

$$N\sigma(N) \sum_{n=1}^{\sqrt{\log N}} 2n \sim N\sigma(N) \cdot \log N = o(N^2),$$

as $N \rightarrow \infty$. On the other hand, the validity of (3.14) for every $j = 1, \dots, n-1$ implies that the summation

$$2 \sum_{n=1+\sqrt{\log N}}^N \sum_{1 \leq j < n} \left| \mathbb{E} \left[\left(f_j^{[-N,N]} - D_N \right) \left(f_n^{[-N,N]} - D_N \right) \right] \right|$$

is of the order

$$2 \sum_{n=1+\sqrt{\log N}}^N n e^{-n^2/2} \sim \int_{\sqrt{\log N}}^N 2x e^{-x^2/2} dx = \frac{1}{N} - e^{-N},$$

as $N \rightarrow \infty$, thus certainly of order $o(N^2)$. And it follows that the expression of (3.13) is of order $o(N^2)$ as well.

The argument is now complete. It is also straightforward to check that it works just as well for an arbitrary subsequence, of the subsequence just constructed. \square

4 Ramifications and Examples

The condition (3.2), which reads

$$\lim_{M \rightarrow \infty} \left(\sup_{n \in \mathbb{N}} \tau_n(M) \right) = 0,$$

can be weakened to

$$\lim_{M \rightarrow \infty} \left(\liminf_{n \in \mathbb{N}} \tau_n(M) \right) = 0 \tag{4.1}$$

Indeed, by passing to a subsequence, this becomes

$$\lim_{M \rightarrow \infty} \left(\limsup_{n \in \mathbb{N}} \tau_n(M) \right) = 0, \tag{4.2}$$

and one checks relatively easily that (4.2) can replace (3.2) in the inductive construction of the subsequence (of) $\{f_n\}_{n \in \mathbb{N}}$. We note also that the condition (4.2) can be satisfied, while (3.2) fails.

To see this, take $g \in \mathbb{L}^0$ with

$$\limsup_{M \rightarrow \infty} \left(M \cdot \mathbb{P}(|g| > M) \right) > 0 \tag{4.3}$$

and define the functions

$$f_n := g \cdot \mathbf{1}_{\{|g| > n\}}, \quad n \in \mathbb{N}. \tag{4.4}$$

We have then $\tau_n(M) = M \cdot \mathbb{P}(|g| > M \vee n)$, $\tau(M) = M \cdot \mathbb{P}(|g| > M)$, so (4.3) means that (3.2) fails. However,

$$\lim_{n \rightarrow \infty} \tau_n(M) = M \cdot \lim_{n \rightarrow \infty} \mathbb{P}(|g| > n) = 0$$

holds for every $M \in (0, \infty)$, so (4.2) is satisfied. We obtain this way the WLLN (2.4) for a suitable sequence of correctors D_1, D_2, \dots .

It is also checked readily that the condition (2.5) is satisfied here, so all these correctors can actually be chosen equal to zero.

Example 4.1. To provide another illustration of Theorem 2.1 that highlights the role of both conditions (2.1) and (2.5) in a more substantial way, let us revisit an old example from [7] (see also section 5.2 of [2]) in slightly modified form. Suppose that the functions f_1, f_2, \dots satisfy

$$\mathbb{P}(f_n = 0) = \varrho_n; \quad \mathbb{P}(f_n = k) = \frac{(1 - \varrho_n)c}{k^2 \log k}, \quad k = 2, 3, \dots \tag{4.5}$$

with constants $0 < \varrho_n < 1$ and $2c = \left(\sum_{k \geq 2} k^{-2} (1/\log k) \right)^{-1}$, for every $n \in \mathbb{N}$.

We do *not* impose any condition on the finite-dimensional joint distributions of the f_1, f_2, \dots ; in particular, we do *not* require the f_1, f_2, \dots to be independent.

In this setting,

$$\tau_n(M) = 2cM(1 - \varrho_n) \sum_{k > M} \frac{1}{k^2 \log k} \sim \frac{2c}{\log M} (1 - \varrho_n)$$

holds for integers $M \geq 2$ in the notation of (2.1). Thus, $\tau(M) = \sup_{n \in \mathbb{N}} \tau_n(M) \leq (2c)/\log M$, and the condition (2.2) is satisfied. On the other hand, we have also

$$\mathbb{E}(f_n^2 \mathbf{1}_{\{|f_n| \leq M\}}) = 2c(1 - \varrho_n) \sum_{2 \leq k \leq M} k^2 \cdot \frac{1}{k^2 \log k} \leq \frac{2cM}{\log 2} (1 - \varrho_n)$$

and this shows that the condition (2.5) is also satisfied when

$$\limsup_{n \rightarrow \infty} \varrho_n = 1. \tag{4.6}$$

We conclude that, under the condition (4.6), there is a subsequence of f_1, f_2, \dots , denoted again by the same symbols, for which the WLLN holds with $D_N \equiv 0$, and hereditarily: that is,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f_n = 0 \quad \text{holds in probability}$$

for f_1, f_2, \dots and for every one of its subsequences.

Remark 4.2. Theorem 2.1 has a direct extension, with only very obvious notational changes, to the case where f_1, f_2, \dots take values in some Euclidean space \mathbb{R}^d , rather than the real line. In such an extension it does not matter whether balls or cubes are considered in the truncation scheme (3.3).

References

- [1] BIRKHOFF, G.D. (1931) Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences* **17**, 656-660.
- [2] CHUNG, K.L. (1974) *A Course in Probability Theory*. Second Edition. Probability and Mathematical Statistics: A Series of Monographs and Textbooks, Vol. 21. Academic Press, New York.
- [3] DURRETT, R. (2010) *Probability: Theory and Examples*. Cambridge University Press.
- [4] FELLER, W. (1970) *An Introduction to Probability Theory and Its Applications: Volume II*. J. Wiley & Sons, New York.
- [5] GNEDENKO, B. & KOLMOGOROV, A.N. (1954) *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley Publishing Co., Cambridge, Mass.
- [6] KHINTCHINE, A. & KOLMOGOROV, A.N. (1925) Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. *Mat. Sbornik* **32**, 668-677.

- [7] KOLMOGOROV, A.N. (1928/29) Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Mathematische Annalen* **99**, 309-319; **102**, 484-488.
- [8] KOLMOGOROV, A.N. (1930) Sur la loi forte des grandes nombres. *Comptes Rendus Acad. Sci. Paris* **191**, 910-912.
- [9] KOLMOGOROV, A.N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik **2** (3). Springer-Verlag, Berlin. English Translation: *Foundations of Probability Theory*, Chelsea Publishing Co., New York, 1950.
- [10] KOMLÓS, J. (1967) A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hung.* **18**, 217-229.
- [11] PENG, SH. (2019) *Nonlinear Expectations and Stochastic Calculus under Uncertainty*. Springer-Verlag, NY.