

On the discrepancy of random subsequences of $\{n\alpha\}$ II

István Berkes* and Bence Borda†

Abstract

Let α be an irrational number, let X_1, X_2, \dots be independent, identically distributed, integer-valued random variables, and put $S_k = \sum_{j=1}^k X_j$. Assuming that X_1 has finite variance or heavy tails $\mathbb{P}(|X_1| > t) \sim ct^{-\beta}$, $0 < \beta < 2$, in [4] we proved that, up to logarithmic factors, the order of magnitude of the discrepancy $D_N(S_k\alpha)$ of the first N terms of the sequence $\{S_k\alpha\}$ is $O(N^{-\tau})$, where $\tau = \min(1/(\beta\gamma), 1/2)$ (with $\beta = 2$ in the case of finite variances) and γ is the strong Diophantine type of α . This shows a change of behavior of the discrepancy at $\beta\gamma = 2$. In this paper we determine the exact order of magnitude of $D_N(S_k\alpha)$ for $\beta\gamma < 1$, and determine the limit distribution of $N^{-1/2}D_N(S_k\alpha)$. We also prove a functional version of these results describing the asymptotic behavior of a wide class of functionals of the sequence $\{S_k\alpha\}$. Finally, we extend our results to the discrepancy of $\{S_k\}$ for general random walks S_k without arithmetic conditions on X_1 , assuming only a mild polynomial rate on the weak convergence of $\{S_k\}$ to the uniform distribution.

MSC2020: 11K38, 11K60, 60G50, 60F17

Keywords: i.i.d. sums, random walk, discrepancy, Diophantine approximation, functional central limit theorem, functional law of the iterated logarithm

*Alfréd Rényi Institute of Mathematics, 1053 Budapest, Reáltanoda u. 13-15, Hungary. e-mail: berkes.istvan@renyi.hu. Research supported by NKFIH grant K 125569.

†Alfréd Rényi Institute of Mathematics, 1053 Budapest, Reáltanoda u. 13-15, Hungary and Graz University of Technology, 8010 Graz, Steyrergasse 30, Austria. e-mail: borda.bence@renyi.hu. Research supported by the Austrian Science Fund (FWF), project Y-901.

1 Introduction

Let α be an irrational number with strong Diophantine type γ ; that is,

$$0 < \liminf_{q \rightarrow \infty} \|q\alpha\| q^\gamma < \infty,$$

where $\|\cdot\|$ denotes distance from the nearest integer. Let X_1, X_2, \dots be independent, identically distributed (i.i.d.) integer-valued random variables, and put $S_k = \sum_{j=1}^k X_j$. Assuming that X_1 has finite variance or heavy tails $\mathbb{P}(|X_1| > t) \sim ct^{-\beta}$, $0 < \beta < 2$, in our previous paper [4] we investigated the order of magnitude of the discrepancy $D_N(S_k\alpha)$ of the first N terms of the sequence $\{S_k\alpha\}$, where $\{\cdot\}$ denotes fractional part. Under some additional regularity conditions (including $\mathbb{E}X_1 = 0$ in the case $1 < \beta < 2$) we proved that

$$D_N(S_k\alpha) = O\left(\sqrt{\frac{\log \log N}{N}} \log N\right), \quad D_N(S_k\alpha) = \Omega\left(\sqrt{\frac{\log \log N}{N}}\right) \quad \text{a.s.} \quad (1)$$

if $\beta\gamma \leq 2$, and

$$D_N(S_k\alpha) = O\left(\left(\frac{\log \log N}{N}\right)^{1/(\beta\gamma)}\right), \quad D_N(S_k\alpha) = \Omega\left(\frac{1}{N^{1/(\beta\gamma)}}\right) \quad \text{a.s.} \quad (2)$$

if $\beta\gamma > 2$. Here a.s. (almost surely) means that a given relation holds with probability 1. These relations show the surprising fact that the asymptotic behavior of $D_N(S_k\alpha)$ changes substantially at the critical value $\gamma = 2/\beta$. A similar result holds in the case $\mathbb{E}X_1^2 < \infty$, $\mathbb{E}X_1 = 0$ (which corresponds formally to $\beta = 2$) when the critical value is $\gamma = 1$, and in the case $\mathbb{E}|X_1| < \infty$, $\mathbb{E}X_1 \neq 0$, when the critical value is $\gamma = 2$. (In the latter case the result was proved only for some special distributions.) Note that there is a gap between the upper and lower bounds in (1) and (2), and one purpose of the present paper is to remove the factor $\log N$ in the upper bound of (1), yielding a precise law of the iterated logarithm (LIL) for $D_N(S_k\alpha)$ in the case $\beta\gamma < 1$. Under the same condition we will also prove that $N^{-1/2}D_N(S_k\alpha)$ has a nondegenerate limit distribution. In fact, we will prove the functional version of these results describing the asymptotic behavior of a large class of functionals of the sequence $\{S_k\alpha\}$, including the L^p -discrepancy $D_N^{(p)}(S_k\alpha)$, $1 \leq p \leq \infty$. Recall that $D_N^{(p)}(S_k\alpha)$ is defined as the $L^p([0, 1])$ -norm of $F_N(t) - t$, where

$$F_N(t) = \frac{1}{N} \sum_{k=1}^N I_{[0,t)}(\{S_k\alpha\}), \quad 0 \leq t \leq 1$$

is the empirical distribution function, and that the discrepancy can also be expressed in terms of F_N as $D_N(S_k\alpha) = \sup_{0 \leq s < t \leq 1} |(F_N(t) - t) - (F_N(s) - s)|$. We define F_N , $D_N^{(p)}(a_k)$ and $D_N(a_k)$ for an arbitrary sequence of reals a_k analogously.

From a purely probabilistic point of view $\{S_k\alpha\}$ is a Markov chain with countable state space, but we can also view it on the state space $[0, 1]$; in the latter case its stationary distribution is the uniform distribution on $[0, 1]$. According to a classical result of Lévy [17], the distribution of $\{S_k\alpha\}$ converges weakly to the uniform distribution as $k \rightarrow \infty$ if and only if X_1 is nondegenerate (i.e. attains

more than one integers with positive probability). Since the limit distribution is continuous, we then have

$$\psi_\alpha(k) := \sup_{0 \leq t \leq 1} |\mathbb{P}(\{S_k \alpha\} \leq t) - t| \longrightarrow 0 \text{ as } k \rightarrow \infty.$$

Note that $D_N^{(\infty)}(S_k \alpha)$ is the empirical analogue of $\psi_\alpha(k)$. The order of magnitude of $\psi_\alpha(k)$ was investigated by Schatte [20], Diaconis [12], Su [24], Hensley and Su [15]; improving their results, in [3] we showed that

$$\psi_\alpha(k) = O(k^{-1/(\beta\gamma)}).$$

Moreover, under mild additional regularity conditions on the distribution of X_1 this estimate is sharp. This shows that, in contrast to $D_N(S_k \alpha)$, there is no change of behavior of $\psi_\alpha(k)$ at $\beta\gamma = 2$. Sharp results for the remainder term in the central limit theorem (CLT) and the corresponding Edgeworth expansion for i.i.d. sums (without taking the fractional part) under Diophantine conditions were given by Bobkov [9], [10].

By the previous result, condition $\beta\gamma < 1$ amounts to the same as

$$\psi_\alpha(k) \ll k^{-(1+\delta)} \quad (\delta > 0),$$

and in this paper we will extend our results on $D_N(S_k \alpha)$ to the discrepancy $D_N(S_k)$ with any i.i.d. sequence X_1, X_2, \dots satisfying $\psi(k) \ll k^{-(1+\delta)}$ with some $\delta > 0$, where

$$\psi(k) := \sup_{t \in [0,1]} |\mathbb{P}(\{S_k\} \leq t) - t|. \quad (3)$$

As in the arithmetic case, we will also prove the functional version of our results, describing the asymptotics of general functionals of the sequence $\{S_k\}$. The LIL for $D_N(S_k)$ was proved by Schatte [21] under the assumption $\psi(k) \ll k^{-(5/2+\delta)}$, $\delta > 0$; he also claimed, without proof, that the result remains valid under $\psi(k) \ll k^{-(2+\delta)}$, $\delta > 0$. We believe that our condition $\psi(k) \ll k^{-(1+\delta)}$, $\delta > 0$ is sharp, but its optimality remains open. For earlier results on the discrepancy of $D_N(S_k)$ in the case of absolutely continuous X_1 (when $\psi(k) \rightarrow 0$ exponentially fast), see Bazarova et al. [2], Berkes and Raseta [5].

To formulate our results, let X_1, X_2, \dots be i.i.d. real-valued random variables, let $S_k = \sum_{j=1}^k X_j$, and let $\psi(k)$ be as in (3). Let \mathcal{F} denote the set of all 1-periodic functions $f: \mathbb{R} \rightarrow \mathbb{R}$ such that f is of bounded variation on $[0, 1]$ and $\int_0^1 f(x) dx = 0$. For any $f, g \in \mathcal{F}$ set

$$C(f, g) = \mathbb{E}f(U)g(U) + \sum_{k=1}^{\infty} \mathbb{E}f(U)g(U + S_k) + \sum_{k=1}^{\infty} \mathbb{E}g(U)f(U + S_k), \quad (4)$$

where U is a random variable uniformly distributed on $[0, 1)$, independent of X_1, X_2, \dots .

Theorem 1. *Assume that $\psi(k) \ll k^{-(1+\delta)}$ for some $\delta > 0$. Then for any $f, g \in \mathcal{F}$ the series in (4) are absolutely convergent, and $C(f, g)$ is bilinear, symmetric and positive semidefinite in f, g ; moreover, $C(f, f) = 0$ if and only if $f = 0$ a.e. For any $f \in \mathcal{F}$ the sum $\sum_{k=1}^N f(S_k)$ satisfies the central limit theorem*

$$\frac{\sum_{k=1}^N f(S_k)}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad (5)$$

and the law of the iterated logarithm

$$\limsup_{N \rightarrow \infty} \frac{\sum_{k=1}^N f(S_k)}{\sqrt{2N \log \log N}} = \sigma \quad \text{a.s.}$$

with $\sigma = \sqrt{C(f, f)}$.

Relation (5) means convergence in distribution to the mean zero normal distribution of variance σ^2 if $\sigma > 0$, and to the constant zero if $\sigma = 0$.

Next, let

$$F_N(t) = \frac{1}{N} \sum_{k=1}^N I_{[0,t)}(\{S_k\}), \quad 0 \leq t \leq 1$$

denote the empirical distribution function of the sequence $\{S_k\}$. Further, for any interval $J \subseteq [0, 1]$ of length $\lambda(J)$ let $f_J(x) = I_J(\{x\}) - \lambda(J)$ denote the centered indicator of J extended with period 1, and put $\Gamma(s, t) = C(f_{[0,s)}, f_{[0,t)})$, $s, t \in [0, 1]$. The following two results describe the asymptotic behavior of $F_N(t)$ in the distributional and a.s. sense. For basic facts about functional limit theorems we refer to Billingsley [8].

Theorem 2. *Assume that $\psi(k) \ll k^{-(1+\delta)}$ for some $\delta > 0$. Then the function $\Gamma(s, t)$ is symmetric, positive semidefinite and continuous on the unit square, and*

$$\sqrt{N}(F_N(t) - t) \xrightarrow{D[0,1]} K(t) \quad \text{as } N \rightarrow \infty, \quad (6)$$

where $K(t)$ is a mean zero Gaussian process on $[0, 1]$ with covariance function $\Gamma(s, t)$.

Relation (6) expresses weak convergence in the Skorohod space $D[0, 1]$, and implies that the distribution of any functional Ψ of the process $\sqrt{N}(F_N(t) - t)$, $0 \leq t \leq 1$ converges weakly to the distribution of $\Psi(K(t))$, provided that Ψ is continuous in the $D[0, 1]$ topology. This gives the limit distribution of a large class of functionals of the sequence $\{S_k\}$. Since the discrepancy and the L^p -discrepancy are continuous functionals, we immediately obtain

$$\begin{aligned} \sqrt{N}D_N(S_k) &\xrightarrow{d} \sup_{0 \leq s < t \leq 1} |K(t) - K(s)|, \\ \sqrt{N}D_N^{(p)}(S_k) &\xrightarrow{d} \|K\|_{L^p([0,1])} \quad (1 \leq p \leq \infty). \end{aligned}$$

Note that $\Gamma(0, 0) = \Gamma(1, 1) = 0$. From Theorem 1 it follows easily that Γ is (strictly) positive definite on the open interval $(0, 1)$; in particular, the limit distributions in the previous formula are nondegenerate.

The next result is the LIL analogue of Theorem 2.

Theorem 3. *Assume that $\psi(k) \ll k^{-(1+\delta)}$ for some $\delta > 0$. Then with probability 1, the sequence of functions*

$$\frac{N(F_N(t) - t)}{\sqrt{2N \log \log N}}, \quad 0 \leq t \leq 1$$

is relatively compact in the Skorohod space $D[0, 1]$, and its class of limit functions in the $D[0, 1]$ topology is identical with the closed unit ball $B(\Gamma)$ of the reproducing kernel Hilbert space determined by the covariance function $\Gamma(s, t)$.

As a consequence, with probability 1 we have

$$\limsup_{N \rightarrow \infty} \frac{ND_N(S_k)}{\sqrt{2N \log \log N}} = \sup_{y \in B(\Gamma)} \sup_{0 \leq s < t \leq 1} |y(t) - y(s)|,$$

$$\limsup_{N \rightarrow \infty} \frac{ND_N^{(p)}(S_k)}{\sqrt{2N \log \log N}} = \sup_{y \in B(\Gamma)} \|y\|_{L^p([0,1])} \quad (1 \leq p \leq \infty).$$

Note that the value of these limsups are positive and finite. Indeed, finiteness is a consequence of the fact that $B(\Gamma)$ is compact in the supremum norm (see the proof of Theorem 3), whereas positivity follows from Theorem 1 with $f = f_{[0,t]}$ for some $0 < t < 1$. Again, Theorem 3 yields laws of the iterated logarithm for a large class of functionals of the process $\sqrt{N}(F_N(t) - t)$, $0 \leq t \leq 1$; see Strassen [23].

The connection of reproducing kernel Hilbert spaces with the functional LIL for Gaussian processes was first observed by Oodaira [18]; analogous results for mixing processes and lacunary series were obtained by Philipp [19]. The latter paper provides a good introduction to the probabilistic aspects of reproducing kernel Hilbert spaces for the reader not familiar with this concept. The same paper also provides the background for the chaining techniques required in the proof of our theorems. For the general theory of reproducing kernel Hilbert spaces we refer to Aronszajn [1].

If X_1 is uniformly distributed on $[0, 1]$, then it is not difficult to see that $\{S_k\}$ is an i.i.d. sequence of uniformly distributed random variables. In this simplest case we have $C(f, g) = \mathbb{E}f(U)g(U)$, and consequently $\Gamma(s, t) = \min\{s, t\} - st$ is the covariance function of the Brownian bridge. The corresponding functional CLT and functional LIL are due to Donsker [13] and Finkelstein [14].

In general, $\{S_k\}$ is a sequence of weakly dependent random variables, and there exist in the literature functional CLT's and LIL's for the empirical distribution function for such sequences, but their assumptions are considerably more restrictive than those for the central limit theorem. For example, for φ -mixing sequences with mixing rate φ_n , a functional CLT for the empirical distribution function holds under $\sum_{n=1}^{\infty} n^2 \varphi_n^{1/2} < \infty$ (see Billingsley [8, p. 197, Theorem 22.1]), while the (ordinary) CLT for partial sums holds under $\sum_{n=1}^{\infty} \varphi_n^{1/2} < \infty$. Also, most results in the literature deals with stationary processes, while $\{S_k\}$ is nonstationary.

2 Couplings and the Kolmogorov metric

We start with two useful observations on the relationship between the Kolmogorov metric, functions of bounded variation and couplings. The total variation of a function f on $[0, 1]$ is denoted by $V(f)$, and the $L^p([0, 1])$ -norm by $\|f\|_p$.

Lemma 1. *For any $[0, 1]$ -valued random variables X and Y , and any function $f : [0, 1] \rightarrow \mathbb{R}$ of bounded variation,*

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \leq V(f) \sup_{0 \leq t \leq 1} |\mathbb{P}(X \leq t) - \mathbb{P}(Y \leq t)|.$$

In the special case when the distribution of X is finitely supported with equal weights and Y is uniformly distributed on $[0, 1]$, this is known as Koksma's inequality [16,

Chapter 2, Theorem 5.1]. In particular, for any real-valued random variables X and Y , and any $f \in \mathcal{F}$ we have

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \leq V(f) \sup_{0 \leq t \leq 1} |\mathbb{P}(\{X\} \leq t) - \mathbb{P}(\{Y\} \leq t)|. \quad (7)$$

The claim formally follows from integration by parts; however, as one-sided continuity of f is not assumed, integration by parts strictly speaking cannot be applied.

Proof of Lemma 1. First, let $g : [0, 1] \rightarrow \mathbb{R}$ be a nondecreasing function with $g(0) = 0$. Note that for any $x \in [0, g(1)]$ the set $L(x) = \{y \in [0, 1] : g(y) > x\}$ is an interval of the form $[t, 1]$ or $(t, 1]$ for some $0 \leq t \leq 1$. Therefore

$$\begin{aligned} |\mathbb{E}g(X) - \mathbb{E}g(Y)| &= \left| \int_0^{g(1)} (\mathbb{P}(g(X) > x) - \mathbb{P}(g(Y) > x)) \, dx \right| \\ &= \left| \int_0^{g(1)} (\mathbb{P}(X \in L(x)) - \mathbb{P}(Y \in L(x))) \, dx \right| \\ &\leq g(1) \sup_{0 \leq t \leq 1} |\mathbb{P}(X \leq t) - \mathbb{P}(Y \leq t)|. \end{aligned}$$

Here $g(1) = V(g)$. Since the left hand side and $V(g)$ are both invariant under adding a constant to g , we have

$$|\mathbb{E}g(X) - \mathbb{E}g(Y)| \leq V(g) \sup_{0 \leq t \leq 1} |\mathbb{P}(X \leq t) - \mathbb{P}(Y \leq t)| \quad (8)$$

for any nondecreasing function $g : [0, 1] \rightarrow \mathbb{R}$. Finally, consider the Jordan decomposition $f = g_1 - g_2$, where g_1, g_2 are nondecreasing functions and $V(f) = V(g_1) + V(g_2)$. Applying (8) to g_1 and g_2 , the claim follows. \square

Lemma 2. Let X and ξ be independent $[0, 1]$ -valued random variables, and assume that ξ is uniformly distributed on $[0, 1]$. Then there exists a uniformly distributed $[0, 1]$ -valued random variable U such that U is a function of X and ξ , and $|U - X| \leq \sup_{0 \leq t \leq 1} |\mathbb{P}(X \leq t) - t|$.

Proof. Let $F(x) = \mathbb{P}(X \leq x)$ and $F^-(x) = \mathbb{P}(X < x)$, and define

$$U = \xi F(X) + (1 - \xi)F^-(X).$$

We have $|U - X| \leq \sup_{0 \leq t \leq 1} |F(t) - t|$ by construction. To show that U is uniformly distributed on $[0, 1]$, let D denote the set of points at which F is not continuous. If $x \in [0, 1] \setminus \bigcup_{d \in D} [F^-(d), F(d))$, then $t = \sup\{u \in [0, 1] : F(u) \leq x\}$ is seen to satisfy $x = F(t)$. Thus $\{U \leq x\} = \{F(X) \leq F(t)\} = \{X \leq t\}$ as random events, and so $\mathbb{P}(U \leq x) = \mathbb{P}(X \leq t) = F(t) = x$. If $x \in [F^-(d), F(d))$ for some $d \in D$, then

$$\{U \leq x\} = \{X < d\} \cup \left\{ X = d \text{ and } \xi \leq \frac{x - F^-(d)}{F(d) - F^-(d)} \right\}$$

as random events, hence using the independence of X and ξ we get

$$\begin{aligned} \mathbb{P}(U \leq x) &= \mathbb{P}(X < d) + \mathbb{P}(X = d) \mathbb{P}\left(\xi \leq \frac{x - F^-(d)}{F(d) - F^-(d)}\right) \\ &= F^-(d) + (F(d) - F^-(d)) \cdot \frac{x - F^-(d)}{F(d) - F^-(d)} = x. \end{aligned}$$

We note that if X has a continuous distribution, then $F = F^-$ and so $U = F(X)$ has the same property; a classical observation of probability theory. \square

Given two $[0, 1]$ -valued random variables X and Y with distribution functions $F(t) = \mathbb{P}(X \leq t)$ and $G(t) = \mathbb{P}(Y \leq t)$, their distance in the Kolmogorov metric (or uniform metric) is defined as

$$d_{\text{Kol}}(X, Y) = \sup_{0 \leq t \leq 1} |\mathbb{E}I_{[0,t]}(X) - \mathbb{E}I_{[0,t]}(Y)| = \sup_{0 \leq t \leq 1} |F(t) - G(t)|.$$

Lemma 1 states that here the family of test functions $I_{[0,t]}$, $0 \leq t \leq 1$ can be replaced by the larger class of functions of bounded variation; that is,

$$d_{\text{Kol}}(X, Y) = \sup_{V(f) \leq 1} |\mathbb{E}f(X) - \mathbb{E}f(Y)|.$$

Another classical distance for $[0, 1]$ -valued random variables is the p -Wasserstein metric, defined as

$$W_p(X, Y) = \inf \left\{ \|X' - Y'\|_p : X' \stackrel{d}{=} X \text{ and } Y' \stackrel{d}{=} Y \right\} \quad (1 \leq p \leq \infty).$$

Here $X' \stackrel{d}{=} X$ means that X' and X have the same distribution; the infimum is thus over all couplings (X', Y') of X and Y . The p -Wasserstein metric can be expressed in terms of the distribution functions [11, Theorem 2.10] as $W_p(X, Y) = \|F^{-1} - G^{-1}\|_p$, $1 \leq p \leq \infty$, where $F^{-1}(x) = \inf\{t \in [0, 1] : F(t) \geq x\}$, $0 < x < 1$ is the generalized inverse of F .

It is not difficult to see that for $G(t) = t$,

$$\|F^{-1} - G^{-1}\|_\infty = \sup_{0 < x < 1} |F^{-1}(x) - x| = \sup_{0 < t < 1} |F(t) - t|.$$

Indeed, by the left-continuity of F^{-1} the essential supremum and the actual supremum on $0 < x < 1$ are equal; to see the second equality note that $|F(t) - t| \leq c$ geometrically means that the graph of $F(t)$, $0 < t < 1$ stays between the lines $t \pm c$, and that this is invariant under taking the inverse function. Since in the definition of the Kolmogorov metric the supremum can also be taken over $0 < t < 1$, it follows that $W_\infty(X, U) = d_{\text{Kol}}(X, U)$ for a uniformly distributed U ; Lemma 2 is basically a constructive form of this fact. We summarize these observations in the following lemma.

Lemma 3. *Let X and U be $[0, 1]$ -valued random variables, and assume that U is uniformly distributed. Then*

$$W_\infty(X, U) = \sup_{V(f) \leq 1} |\mathbb{E}f(X) - \mathbb{E}f(U)|,$$

where the supremum is over all functions of bounded variation $f : [0, 1] \rightarrow \mathbb{R}$ of total variation $V(f) \leq 1$.

This duality between W_∞ and functions of bounded variation seems not to have been observed before. The classical Kantorovich duality theorem [11, Theorem 2.5]

$$W_1(X, Y) = \sup_{\text{Lip}(f) \leq 1} |\mathbb{E}f(X) - \mathbb{E}f(Y)|,$$

where the supremum is over all $f : [0, 1] \rightarrow \mathbb{R}$ such that $|f(x) - f(y)| \leq |x - y|$, establishes a duality between W_1 and Lipschitz functions. Amusingly, the general inequality $W_1(X, U) \leq W_\infty(X, U)$ thus reduces to the fact that every Lipschitz function is of bounded variation.

3 The variance

We now find the precise asymptotics of the variance.

Lemma 4. *Suppose that $\psi(k) \ll k^{-(1+\delta)}$ for some $\delta > 0$.*

- (i) *For any $f, g \in \mathcal{F}$ the series in (4) are absolutely convergent, and $|C(f, g)| \ll V(f)\|g\|_1 + V(g)\|f\|_1$. Further, $C(f, g)$ is bilinear, symmetric and positive semidefinite in f, g . In addition, $C(f, f) = 0$ if and only if $f = 0$ a.e.*
- (ii) *For any $f \in \mathcal{F}$ and any integers $M \geq 0$ and $N \geq 1$,*

$$\mathbb{E} \left(\sum_{k=M+1}^{M+N} f(S_k) \right)^2 = C(f, f)N + O \left(V(f) \left(\|f\|_1 + V(f)(M+1)^{-\delta} \right)^{\delta/(1+\delta)} \times \left(\|f\|_1 N + V(f)(M+1)^{-\delta} \right)^{1/(1+\delta)} \right).$$

The implied constants in (i) and (ii) depend only on the distribution of X_1 and δ .

Proof. First, we prove (i). In the definition (4) of $C(f, g)$ the variable U is independent of X_1, X_2, \dots , therefore we can write

$$\mathbb{E} f(U)g(U + S_k) = \mathbb{E} \int_0^1 f(u)g(u + S_k) du = \mathbb{E} G(S_k),$$

where $G(x) = \int_0^1 f(u)g(u + x) du$. It is not difficult to see e.g. directly from the definition of total variation that $V(G) \leq V(g)\|f\|_1$. Applying (7) to $X = S_k$ and a random variable Y uniformly distributed on $[0, 1]$ and noting that $\mathbb{E} G(Y) = 0$, we get $|\mathbb{E} f(U)g(U + S_k)| \leq V(g)\|f\|_1 \psi(k)$. Repeating the same argument for $\mathbb{E} g(U)f(U + S_k)$, the absolute convergence of the series in (4) follows. Since $|\mathbb{E} f(U)g(U)| \leq V(g)\|f\|_1$, we also get $|C(f, g)| \ll V(f)\|g\|_1 + V(g)\|f\|_1$. Clearly $C(f, g)$ is bilinear and symmetric in f, g .

Let $\hat{f}(h) = \int_0^1 f(x)e^{-2\pi i h x} dx$, $h \in \mathbb{Z}$ denote the Fourier coefficients of f . We claim that

$$C(f, f) = \sum_{h \neq 0} |\hat{f}(h)|^2 \frac{1 - |\varphi(2\pi h)|^2}{|1 - \varphi(2\pi h)|^2}, \quad (9)$$

where $\varphi(x) = \mathbb{E} \exp(ixX_1)$ is the characteristic function of X_1 . Here $|\varphi(2\pi h)| < 1$ for all integers $h \neq 0$; indeed, otherwise X_1 would have a lattice distribution with a rational span, which is impossible by the assumption $\psi(k) \rightarrow 0$. From (9) it will follow immediately that $C(f, f) \geq 0$, and that $C(f, f) = 0$ if and only if $\hat{f}(h) = 0$ for all integers $h \neq 0$; the latter condition is equivalent to $f = 0$ a.e. We will also have

$$\sum_{i,j=1}^r C(f_i, f_j) x_i x_j = C \left(\sum_{i=1}^r f_i x_i, \sum_{i=1}^r f_i x_i \right) \geq 0$$

for any $f_1, f_2, \dots, f_r \in \mathcal{F}$ and any $x_1, x_2, \dots, x_r \in \mathbb{R}$, proving that $C(f, g)$ is positive semidefinite.

To see (9), assume first that f is a trigonometric polynomial; that is, $\hat{f}(h) = 0$ for all but finitely many integers h . As before, $\mathbb{E} f(U)g(U + S_k) = \mathbb{E} G(S_k)$ with $G(x) =$

$\int_0^1 f(u)f(u+x)du$. Since $\hat{G}(h) = |\hat{f}(h)|^2$, we have $G(x) = \sum_{h \neq 0} |\hat{f}(h)|^2 e^{2\pi i h x}$. Therefore

$$\sum_{k=1}^{\infty} \mathbb{E}G(S_k) = \sum_{h \neq 0} \sum_{k=1}^{\infty} |\hat{f}(h)|^2 \varphi(2\pi h)^k = \sum_{h \neq 0} |\hat{f}(h)|^2 \frac{\varphi(2\pi h)}{1 - \varphi(2\pi h)}.$$

Combining the h and $-h$ terms, and using $\mathbb{E}f(U)^2 = \sum_{h \neq 0} |\hat{f}(h)|^2$, we obtain (9). For a general $f \in \mathcal{F}$, consider

$$f_H(x) = \int_0^1 f(x-y)F_H(y)dy = \sum_{0 < |h| < H} \left(1 - \frac{|h|}{H}\right) \hat{f}(h) e^{2\pi i h x},$$

where $F_H(x) = \sum_{|h| < H} (1 - |h|/H) e^{2\pi i h x}$ is the Fejér kernel of order H . Note that f_H is a trigonometric polynomial; in fact, it is a Cesàro mean of the formal Fourier series of f . By the special case proved above, we have

$$C(f_H, f_H) = \sum_{0 < |h| < H} \left(1 - \frac{|h|}{H}\right)^2 |\hat{f}(h)|^2 \frac{1 - |\varphi(2\pi h)|^2}{|1 - \varphi(2\pi h)|^2} \rightarrow \sum_{h \neq 0} |\hat{f}(h)|^2 \frac{1 - |\varphi(2\pi h)|^2}{|1 - \varphi(2\pi h)|^2}$$

as $H \rightarrow \infty$. On the other hand, letting $G(x) = \int_0^1 f(u)f(u+x)du$ and $G_H(x) = \int_0^1 f_H(u)f_H(u+x)du$, by Lemma 1 we have

$$\begin{aligned} |C(f, f) - C(f_H, f_H)| &= \left| \mathbb{E}f(U)^2 + 2 \sum_{k=1}^{\infty} \mathbb{E}G(S_k) - \mathbb{E}f_H(U)^2 - 2 \sum_{k=1}^{\infty} \mathbb{E}G_H(S_k) \right| \\ &\leq |\mathbb{E}f(U)^2 - \mathbb{E}f_H(U)^2| + 2 \sum_{k=1}^{\infty} |\mathbb{E}(G - G_H)(S_k)| \\ &\leq \left| \|f\|_2^2 - \|f_H\|_2^2 \right| + 2 \sum_{k=1}^{\infty} V(G - G_H)\psi(k) \\ &\ll \left| \|f\|_2^2 - \|f_H\|_2^2 \right| + V(G - G_H). \end{aligned}$$

Since $f_H \rightarrow f$ in $L^2([0, 1])$, it will be enough to prove $V(G - G_H) \rightarrow 0$ as $H \rightarrow \infty$. Writing $f = f_H + (f - f_H)$ in the definition of $G(x)$ and using the integral transformation $u \mapsto u - x \pmod{1}$ in one of the terms, we get

$$(G - G_H)(x) = \int_0^1 (f - f_H)(u) (f_H(u - x) + f(u + x)) du,$$

and consequently $V(G - G_H) \leq \|f - f_H\|_1 (V(f_H) + V(f))$. Using $F_H \geq 0$ it is readily seen that $V(f_H) \leq V(f)$. Hence $V(G - G_H) \leq 2V(f)\|f - f_H\|_1 \rightarrow 0$, finishing the proof of (9) for a general $f \in \mathcal{F}$.

Next, we prove (ii). We may assume $V(f) = 1$. Let U be as in the definition (4) of $C(f, f)$. Expanding the square in the claim we get

$$\mathbb{E} \left(\sum_{k=M+1}^{M+N} f(S_k) \right)^2 = \sum_{k=M+1}^{M+N} \mathbb{E}f(S_k)^2 + 2 \sum_{M+1 \leq k < \ell \leq M+N} \mathbb{E}f(S_k)f(S_\ell). \quad (10)$$

To estimate the diagonal terms, let us apply (7) to $X = S_k$ and $Y = U$. Noting that $V(f^2) \leq 2V(f)^2 = 2$, we obtain $\mathbb{E}f(S_k)^2 = \mathbb{E}f(U)^2 + O(\psi(k))$. The assumption $\psi(k) \ll k^{-(1+\delta)}$ thus ensures that $\sum_{k=M+1}^{M+N} \mathbb{E}f(S_k)^2 = N\mathbb{E}f(U)^2 + O((M+1)^{-\delta})$.

We now estimate the off-diagonal terms in two different ways. For given $k < \ell$ let us write $f(S_k)f(S_\ell) = f(S_k)f(S_k + (S_\ell - S_k))$. First, let \mathcal{F}_k denote the σ -algebra generated by X_1, X_2, \dots, X_k , and note that $S_\ell - S_k$ is independent of \mathcal{F}_k . Let us apply (7) to $X = S_k$, a random variable Y uniformly distributed on $[0, 1)$ independent of \mathcal{F}_k , and the function $g(x) = f(S_k)f(S_k + x)$ to estimate the conditional expectation with respect to \mathcal{F}_k . Since $\mathbb{E}(g(Y)|\mathcal{F}_k) = 0$ and $V(g) \leq |f(S_k)|V(f) = |f(S_k)|$, we have

$$|\mathbb{E}(f(S_k)f(S_\ell) | \mathcal{F}_k)| \leq |f(S_k)|\psi(\ell - k).$$

Taking the (total) expectation, we obtain $|\mathbb{E}f(S_k)f(S_\ell)| \leq \mathbb{E}|f(S_k)|\psi(\ell - k)$. We can apply (7) again, this time to $X = S_k$, a random variable Y uniformly distributed on $[0, 1)$ and $|f|$. Since $\mathbb{E}|f(Y)| = \|f\|_1$ and $V(|f|) \leq V(f) = 1$, we get

$$|\mathbb{E}f(S_k)f(S_\ell)| \leq (\|f\|_1 + \psi(k))\psi(\ell - k). \quad (11)$$

Second, let \mathcal{G}_k denote the σ -algebra generated by X_{k+1}, X_{k+2}, \dots . Note that S_k and U are independent of \mathcal{G}_k . Let us apply (7) to $X = S_k$, $Y = U$ and the function $g(x) = f(x)f(x + (S_\ell - S_k))$. Since $V(g) \leq 2V(f)^2 = 2$, we obtain

$$\mathbb{E}(f(S_k)f(S_\ell) | \mathcal{G}_k) = \mathbb{E}(f(U)f(U + (S_\ell - S_k)) | \mathcal{G}_k) + O(\psi(k)).$$

Note that $f(U)f(U + (S_\ell - S_k))$ has the same distribution as $f(U)f(U + S_{\ell-k})$. Taking the (total) expectation, we thus get

$$\mathbb{E}f(S_k)f(S_\ell) = \mathbb{E}f(U)f(U + S_{\ell-k}) + O(\psi(k)). \quad (12)$$

As observed in (i), here $|\mathbb{E}f(U)f(U + S_{\ell-k})| \leq \|f\|_1\psi(\ell - k)$.

Let $1 \leq K \leq N$ be an integer, to be chosen. We will estimate the off-diagonal terms for which $\ell - k > K$, and those for which $\ell - k \leq K$ separately. The assumption $\psi(k) \ll k^{-(1+\delta)}$ and (11) show that

$$\begin{aligned} \sum_{\substack{M+1 \leq k < \ell \leq M+N \\ \ell - k > K}} \mathbb{E}f(S_k)f(S_\ell) &\ll \left(\|f\|_1 N + (M+1)^{-\delta}\right) K^{-\delta}, \\ \sum_{M+N-K < k < \ell \leq M+N} \mathbb{E}f(S_k)f(S_\ell) &\ll \|f\|_1 K + (M+1)^{-\delta}. \end{aligned}$$

For a fixed $M+1 \leq k \leq M+N-K$, from (12) we obtain

$$\begin{aligned} \sum_{\ell=k+1}^{k+K} \mathbb{E}f(S_k)f(S_\ell) &= \sum_{d=1}^K \mathbb{E}f(U)f(U + S_d) + O(K\psi(k)) \\ &= \sum_{d=1}^{\infty} \mathbb{E}f(U)f(U + S_d) + O\left(\|f\|_1 \sum_{d=K+1}^{\infty} \psi(d) + K\psi(k)\right). \end{aligned}$$

Summing over $M+1 \leq k \leq M+N-K$ and using the assumption $\psi(k) \ll k^{-(1+\delta)}$, we altogether get

$$\sum_{M+1 \leq k < \ell \leq M+N} \mathbb{E} f(S_k) f(S_\ell) = N \sum_{d=1}^{\infty} \mathbb{E} f(U) f(U+S_d) + O\left(\left(\|f\|_1 N + (M+1)^{-\delta}\right) K^{-\delta} + \left(\|f\|_1 + (M+1)^{-\delta}\right) K\right).$$

Choosing $K = \lfloor (\|f\|_1 N + (M+1)^{-\delta})^{1/(1+\delta)} / (\|f\|_1 + (M+1)^{-\delta})^{1/(1+\delta)} \rfloor$, in (10) we thus have

$$\mathbb{E} \left(\sum_{k=M+1}^{M+N} f(S_k) \right)^2 = C(f, f)N + O\left(\left(\|f\|_1 + (M+1)^{-\delta}\right)^{\delta/(1+\delta)} \left(\|f\|_1 N + (M+1)^{-\delta}\right)^{1/(1+\delta)}\right).$$

□

4 Approximation by independent variables

In this section we approximate $\sum_{k=M+1}^{M+N} f(S_k)$ by a sum of independent random variables. A basic tool is the following result, implicit in Schatte [21], [22].

Lemma 5. *Let ξ_1, ξ_2, \dots be independent random variables uniformly distributed on $[0, 1)$, independent of the X_n 's, and let I_1, I_2, \dots be disjoint blocks of integers such that between I_n and I_{n+1} there are $\ell_n \geq 1$ integers. Then there exists a sequence $\delta_1, \delta_2, \dots$ of random variables such that δ_n is measurable with respect to ξ_n and the X_j 's between I_n and I_{n+1} , and $|\delta_n| \leq \psi(\ell_n)$, $n = 1, 2, \dots$; further, the random vectors*

$$(\{S_i\}, i \in I_1), (\{S_i - \delta_1\}, i \in I_2), \dots, (\{S_i - \delta_{n-1}\}, i \in I_n), \dots$$

are independent and they have, except for the first one, uniformly distributed components.

Proof. We begin with a simple remark. Let X be a random variable and let U be a uniformly distributed random variable independent of X . Then $\{X + U\}$ is uniformly distributed on $[0, 1)$ and is independent of X , as one can see immediately by conditioning $\{X + U\}$ on $\{X = c\}$.

Let now

$$I_n := \{p_n + 1, p_n + 2, \dots, p_n + b_n\}, \quad n = 1, 2, \dots$$

We claim that

$$\delta_n = \{S_{p_{n+1}} - S_{p_n+b_n}\} - U(\{S_{p_{n+1}} - S_{p_n+b_n}\}, \xi_n), \quad n = 1, 2, \dots$$

satisfies the lemma, where $U(\{S_{p_{n+1}} - S_{p_n+b_n}\}, \xi_n)$ is the uniform random variable constructed in Lemma 2 for the variables $\{S_{p_{n+1}} - S_{p_n+b_n}\}$ and ξ_n . Since $S_{p_{n+1}} - S_{p_n+b_n} \stackrel{d}{=} S_{\ell_n}$, by the definition of $\psi(k)$ and the claim of Lemma 2, $|\delta_n| \leq \psi(\ell_n)$.

Furthermore, we have

$$\begin{aligned} \{S_{p_2+1} - \delta_1\} &= \{(X_1 + \cdots + X_{p_1+b_1}) + X_{p_2+1} + U(\{S_{p_2} - S_{p_1+b_1}\}, \xi_1)\} \\ &\vdots \\ \{S_{p_2+b_2} - \delta_1\} &= \{(X_1 + \cdots + X_{p_1+b_1}) + (X_{p_2+1} + \cdots + X_{p_2+b_2}) \\ &\quad + U(\{S_{p_2} - S_{p_1+b_1}\}, \xi_1)\}. \end{aligned}$$

Put

$$\begin{aligned} V &= \{X_1 + \cdots + X_{p_1+b_1} + U(\{S_{p_2} - S_{p_1+b_1}\}, \xi_1)\}, \\ X &= (X_{p_2+1}, X_{p_2+1} + X_{p_2+2}, \dots, X_{p_2+1} + \cdots + X_{p_2+b_2}). \end{aligned}$$

Let ω be a point of the underlying probability space, and put

$$c = (X_1 + \cdots + X_{p_1+b_1})(\omega).$$

The conditional distribution of the vector $(\{S_{p_2+1} - \delta_1\}, \dots, \{S_{p_2+b_2} - \delta_1\})$ relative to $\sigma\{X_1, \dots, X_{p_1+b_1}\}$, evaluated at ω , equals the distribution of

$$\{X + c + U(\{S_{p_2} - S_{p_1+b_1}\}, \xi_1)\},$$

where adding the scalar $V^* = c + U(\{S_{p_2} - S_{p_1+b_1}\}, \xi_1)$ to X is meant coordinatewise. But $U(\{S_{p_2} - S_{p_1+b_1}\}, \xi_1)$ is a uniform random variable independent of X and so is $\{V^*\}$, showing that the distribution of $\{X + V^*\}$ does not depend on c and consequently, $(\{S_{p_2+1} - \delta_1\}, \dots, \{S_{p_2+b_2} - \delta_1\})$ is independent of $\sigma\{X_1, \dots, X_{p_1+b_1}\}$ and thus of $(\{S_i\}, i \in I_1)$. By our remark at the beginning of the proof, the coordinates of $(\{S_{p_2+1} - \delta_1\}, \dots, \{S_{p_2+b_2} - \delta_1\})$ are uniformly distributed.

Now let $n \geq 2$. As before, the vector

$$(\{S_{p_{n+1}+1} - \delta_n\}, \dots, \{S_{p_{n+1}+b_{n+1}} - \delta_n\}) \quad (13)$$

can be written as $\{X + V\}$, where

$$\begin{aligned} V &= \{X_1 + \cdots + X_{p_n+b_n} + U(\{S_{p_{n+1}} - S_{p_n+b_n}\}, \xi_n)\}, \\ X &= (X_{p_{n+1}+1}, X_{p_{n+1}+1} + X_{p_{n+1}+2}, \dots, X_{p_{n+1}+1} + \cdots + X_{p_{n+1}+b_{n+1}}). \end{aligned}$$

Repeating the argument used in the case $n = 1$, it follows that the vector (13) is independent of $\sigma\{X_1, \dots, X_{p_n+b_n}, \xi_1, \dots, \xi_{n-1}\}$ and thus of

$$(\{S_i\}, i \in I_1), (\{S_i - \delta_1\}, i \in I_2), \dots, (\{S_i - \delta_{n-1}\}, i \in I_n),$$

and has uniform coordinates. This completes the proof of the lemma. \square

Fix a function $f \in \mathcal{F}$, an integer $M \geq 0$ and a sequence of auxiliary random variables ξ_2, ξ_3, \dots uniformly distributed on $[0, 1)$, independent of X_1, X_2, \dots . Consider a decomposition of the infinite set of integers $\{M + 1, M + 2, \dots\}$ into nonempty, consecutive blocks $H_1, J_1, H_2, J_2, \dots$, and let $T_i = T_i^{(f)} = \sum_{k \in H_i} f(S_k)$ and $D_i = D_i^{(f)} = \sum_{k \in J_i} f(S_k)$ for all $i \geq 1$. If $M + N = \max J_R$ for some $R \geq 1$, we thus have

$$\sum_{k=M+1}^{M+N} f(S_k) = \sum_{i=1}^R T_i + \sum_{i=1}^R D_i.$$

Using Lemma 5 we can introduce two sequences of random variables $\delta_2, \delta_3, \dots$ and $\delta'_2, \delta'_3, \dots$ with the following properties. First, δ_i is measurable with respect to the σ -algebra generated by $X_k, k \in J_{i-1}$ and ξ_i , and $|\delta_i| \leq \psi(|J_{i-1}|)$ for all $i \geq 2$; similarly, δ'_i is measurable with respect to the σ -algebra generated by $X_k, k \in H_i$ and ξ_i , and $|\delta'_i| \leq \psi(|H_i|)$ for all $i \geq 2$. Letting $T_i^* = T_i^{(f)*} = \sum_{k \in H_i} f(S_k - \delta_i)$ and $D_i^* = D_i^{(f)*} = \sum_{k \in J_i} f(S_k - \delta'_i)$, we also have that T_2^*, T_3^*, \dots are independent and $\mathbb{E}T_i^* = 0$ for all $i \geq 2$; similarly, D_2^*, D_3^*, \dots are independent and $\mathbb{E}D_i^* = 0$ for all $i \geq 2$. Under the condition $\psi(k) \ll k^{-(1+\delta)}$ the variances are given by

$$\begin{aligned}\mathbb{E}T_i^{*2} &= C(f, f)|H_i| + O\left(V(f)^2|H_i|^{1/(1+\delta)}\right), \\ \mathbb{E}D_i^{*2} &= C(f, f)|J_i| + O\left(V(f)^2|J_i|^{1/(1+\delta)}\right)\end{aligned}\tag{14}$$

with implied constants depending only on the distribution of X_1 and δ . Indeed, conditioning on the σ -algebra \mathcal{A}_i generated by the variables $X_k, k \in [1, M] \cup H_1 \cup J_1 \cup \dots \cup H_{i-1} \cup J_{i-1}$ and ξ_i , we can write $T_i^* = \sum_{\ell=1}^{|H_i|} f(\tilde{S}_\ell - Z_i)$ with some \mathcal{A}_i -measurable random variable Z_i , where \tilde{S}_ℓ are the partial sums of the i.i.d. sequence $X_k, k \in H_i$. Applying Lemma 4 to $\tilde{f}(x) = f(x - Z_i)$ and noting that $C(\tilde{f}, \tilde{f}) = C(f, f)$ and $V(\tilde{f}) = V(f)$, we obtain $\mathbb{E}(T_i^{*2} | \mathcal{A}_i) = C(f, f)|H_i| + O(V(f)^2|H_i|^{1/(1+\delta)})$. Taking the (total) expectation, the first relation in (14) follows; the proof of the second relation is analogous.

Lemma 6. *Assume that $\psi(k) \ll k^{-(1+\delta)}$ with some $0 < \delta < 1$. If we choose the sizes of the blocks as $|H_i| = \lceil i^c \rceil$ and $|J_i| = \lceil i^{c'} \rceil$ for all $i \geq 1$, where $1/2 < c < 1$ is a fixed constant and $c' = (c + \delta/2)/(1 + \delta)$, then*

$$\mathbb{E}\left(\sum_{i=2}^R (T_i - T_i^*)\right)^2 \ll V(f)^2 R^{1-\delta/2}, \quad \mathbb{E}\left(\sum_{i=2}^R (D_i - D_i^*)\right)^2 \ll V(f)^2 R^{1-\delta/2}$$

for all $R \geq 2$ with implied constants depending only on the distribution of X_1 , δ and c .

Proof. We shall only give a proof for $\sum_{i=2}^R (T_i - T_i^*)$, as the proof for $\sum_{i=2}^R (D_i - D_i^*)$ is analogous. We may assume $V(f) = 1$. Note that $c' < c$. Expanding the square we get

$$\mathbb{E}\left(\sum_{i=2}^R (T_i - T_i^*)\right)^2 = \sum_{i=2}^R \mathbb{E}(T_i - T_i^*)^2 + 2 \sum_{2 \leq i < j \leq R} \mathbb{E}(T_i - T_i^*)(T_j - T_j^*). \tag{15}$$

Let us first estimate the diagonal term $\mathbb{E}(T_i - T_i^*)^2$. Let \mathcal{F}_i denote the σ -algebra generated by the variables $X_k, k \in J_{i-1}$ and ξ_i , and let $Y_i = \sum_{k \in J_{i-1}} X_k$. Note that in $T_i - T_i^* = \sum_{k \in H_i} (f((S_k - Y_i) + Y_i) - f((S_k - Y_i) + Y_i - \delta_i))$ the variables Y_i and δ_i are \mathcal{F}_i -measurable, and $(S_k - Y_i)$ is independent of \mathcal{F}_i . Hence we can apply Lemma 4 to the i.i.d. sequence obtained from X_1, X_2, \dots by deleting the terms $X_k, k \in J_{i-1}$ to estimate the conditional expectation with respect to \mathcal{F}_i . It is not difficult to see e.g. directly from the definition of total variation that the function $g(x) = f(x + Y_i) - f(x + Y_i - \delta_i)$ satisfies $V(g) \leq 2V(f) = 2$ and $\|g\|_1 \ll |\delta_i|$.

Therefore

$$\mathbb{E}((T_i - T_i^*)^2 \mid \mathcal{F}_i) \ll |\delta_i| \cdot |H_i| + \left(|\delta_i| + (\min H_i - |J_{i-1}|)^{-\delta} \right)^{\delta/(1+\delta)} \left(|\delta_i| \cdot |H_i| + (\min H_i - |J_{i-1}|)^{-\delta} \right)^{1/(1+\delta)}.$$

Here $\min H_i - |J_{i-1}| \gg i^{c+1}$, and $|\delta_i| \leq \psi(|J_{i-1}|) \ll i^{-c-\delta/2}$. Taking the (total) expectation, after simplifying the exponent we get $\mathbb{E}(T_i - T_i^*)^2 \ll i^{-\delta/2}$. Summing over $2 \leq i \leq R$ we get that the contribution of the diagonal terms in (15) is $\sum_{i=2}^R \mathbb{E}(T_i - T_i^*)^2 \ll R^{1-\delta/2}$.

We now estimate the off-diagonal terms in (15) similarly to the proof of Lemma 4. For given $i < j$ let us write

$$(T_i - T_i^*)(T_j - T_j^*) = \sum_{\substack{k \in H_i \\ \ell \in H_j}} (f(S_k) - f(S_k - \delta_i)) \times (f((S_\ell - W_{i,j}) + W_{i,j}) - f((S_\ell - W_{i,j}) + W_{i,j} - \delta_j)),$$

where $W_{i,j} = \sum_{k \in [1, M] \cup H_1 \cup J_1 \cup \dots \cup H_{i-1} \cup J_{i-1} \cup H_i \cup J_{j-1}} X_k$. Let $\mathcal{G}_{i,j}$ denote the σ -algebra generated by the variables X_k , $k \in [1, M] \cup H_1 \cup J_1 \cup \dots \cup H_{i-1} \cup J_{i-1} \cup H_i \cup J_{j-1}$ and ξ_i, ξ_j . Note that $W_{i,j}$, δ_i, δ_j and S_k , $k \in H_i$ are $\mathcal{G}_{i,j}$ -measurable. Moreover, $S_\ell - W_{i,j}$ is the sum of $\ell - M - |H_1| - |J_1| - \dots - |H_{i-1}| - |J_{i-1}| - |H_i| - |J_{j-1}|$ independent copies of X_1 , and it is independent of $\mathcal{G}_{i,j}$. We can thus apply (7) to $X = S_\ell - W_{i,j}$ and a random variable Y uniformly distributed on $[0, 1]$, independent of $\mathcal{G}_{i,j}$ to obtain

$$\begin{aligned} & |\mathbb{E}((T_i - T_i^*)(T_j - T_j^*) \mid \mathcal{G}_{i,j})| \ll \\ & \sum_{\substack{k \in H_i \\ \ell \in H_j}} |f(S_k) - f(S_k - \delta_i)| \psi(\ell - M - |H_1| - |J_1| - \dots - |H_{i-1}| - |J_{i-1}| - |H_i| - |J_{j-1}|). \end{aligned} \quad (16)$$

Next, we condition with respect to the σ -algebra \mathcal{F}_i generated by X_k , $k \in J_{i-1}$ and ξ_i . Let $Y_i = \sum_{k \in J_{i-1}} X_k$, and let us write $|f(S_k) - f(S_k - \delta_i)| = |f((S_k - Y_i) + Y_i) - f((S_k - Y_i) + Y_i - \delta_i)|$, as before. Note that Y_i and δ_i are \mathcal{F}_i -measurable, while $S_k - Y_i$ is independent of \mathcal{F}_i . Let us apply (7) to $X = S_k - Y_i$, a random variable Y uniformly distributed on $[0, 1]$, independent of \mathcal{F}_i and $g(x) = |f(x + Y_i) - f(x + Y_i - \delta_i)|$. As observed before, $\|g\|_1 \ll |\delta_i| \leq \psi(|J_{i-1}|)$, hence

$$\mathbb{E}(|f(S_k) - f(S_k - \delta_i)| \mid \mathcal{F}_i) \ll \psi(|J_{i-1}|) + \psi(k - |J_{i-1}|) \ll i^{-c-\delta/2}. \quad (17)$$

Taking the (total) expectation in (16) and (17), and summing over $k \in H_i$ we finally deduce

$$\begin{aligned} & |\mathbb{E}(T_i - T_i^*)(T_j - T_j^*)| \ll \\ & i^{-\delta/2} \sum_{\ell \in H_j} \psi(\ell - M - |H_1| - |J_1| - \dots - |H_{i-1}| - |J_{i-1}| - |H_i| - |J_{j-1}|). \end{aligned} \quad (18)$$

Now fix $2 \leq i < R$. In the case $j = i + 1$ the sum on the right hand side of (18) is $\ll 1$, thus $|\mathbb{E}(T_i - T_i^*)(T_{i+1} - T_{i+1}^*)| \ll i^{-\delta/2}$. In the case

$j - i \geq 2$ we have $(\ell - M - |H_1| - |J_1| - \dots - |H_{i-1}| - |J_{i-1}| - |H_i| - |J_{j-1}|) \gg (j - i - 1)j^c$, and hence $\left| \mathbb{E}(T_i - T_i^*)(T_j - T_j^*) \right| \ll (j - i - 1)^{-1-\delta} j^{-c\delta} i^{-\delta/2}$. Therefore $\sum_{j=i+1}^R \left| \mathbb{E}(T_i - T_i^*)(T_j - T_j^*) \right| \ll i^{-\delta/2}$ for any fixed $2 \leq i < R$. Summing over $2 \leq i < R$ we finally obtain that the contribution of the off-diagonal terms in (15) is also $\ll R^{1-\delta/2}$. \square

5 Exponential bounds

The crucial step in the proof of laws of the iterated logarithm is to prove sharp bounds for the tail probabilities of the partial sums of the considered random variables. In this section we prove such an inequality for shifted sums $\sum_{k=M+1}^{M+N} f(S_k)$.

Lemma 7. *Assume that $\psi(k) \ll k^{-(1+\delta)}$ for some $0 < \delta < 1$, and fix a constant $0 < \rho < \delta/16$. For any integers $M \geq 0$ and $N \gg 1$, any $f \in \mathcal{F}$ such that $V(f) \leq 2$ and $\|f\|_1 \gg N^{-1/2-\delta/8}$, and any $t \gg 1$ we have*

$$\mathbb{P} \left(\left| \sum_{k=M+1}^{M+N} f(S_k) \right| \geq t \|f\|_1^\rho \sqrt{N} \right) \ll \exp \left(-C \|f\|_1^{-\rho} t_0 t \right) + t^{-2} \|f\|_1^{-2\rho} N^{-1/2-\delta/8}$$

with some constant $C > 0$ and $t_0 = \min\{t, N^{\delta/16}\}$. The implied constants and C depend only on the distribution of X_1 , δ and ρ .

Proof. Let $c = 1 - \delta/2$ and $c' = (c + \delta/2)/(1 + \delta)$. Let us split the set $\{M + 1, M + 2, \dots\}$ into blocks $H_1, J_1, H_2, J_2, \dots$ such that $|H_i| = \lceil i^c \rceil$ and $|J_i| = \lceil i^{c'} \rceil$ for all $i \geq 1$. Let T_i, D_i , $i \geq 1$ and T_i^*, D_i^* , $i \geq 2$ be as in Section 4. Fix a large enough integer N , and let R be such that $M + N \in H_R \cup J_R$. Clearly $R = \Theta(N^{1/(c+1)})$, and $|T_i| \ll R^c \ll N^{(2-\delta)/(4-\delta)}$ for all $1 \leq i \leq R$, and the same holds for D_i . Since $N^{(2-\delta)/(4-\delta)} \ll \|f\|_1^\rho N^{1/2}$, the contribution of any individual block to $\sum_{k=M+1}^{M+N} f(S_k)$ is negligible. Therefore we may assume that $M + N = \max J_R$, and it will be enough to prove the claim for $\sum_{i=2}^R (T_i + D_i)$ instead of $\sum_{k=M+1}^{M+N} f(S_k)$. Recall that T_i^* , $i = 2, 3, \dots, R$ are independent, and $\mathbb{E}T_i^* = 0$; similarly, D_i^* , $i = 2, 3, \dots, R$ are independent, and $\mathbb{E}D_i^* = 0$. By Lemma 6 we have

$$\mathbb{E} \left(\sum_{i=2}^R (T_i - T_i^*) \right)^2 \ll R^{1-\delta/2} \ll N^{(1-\delta/2)/(2-\delta/2)} \leq N^{1/2-\delta/8},$$

and the same holds for $\sum_{i=2}^R (D_i - D_i^*)$. Hence by the Chebyshev inequality

$$\mathbb{P} \left(\left| \sum_{i=2}^R (T_i + D_i) - \sum_{i=2}^R (T_i^* + D_i^*) \right| \geq t \|f\|_1^\rho \sqrt{N} \right) \ll t^{-2} \|f\|_1^{-2\rho} N^{-1/2-\delta/8}. \quad (19)$$

From (14) and $C(f, f) \ll \|f\|_1$ we get

$$\sum_{i=2}^R \mathbb{E}T_i^{*2} \ll \|f\|_1 N + \sum_{i=2}^R i^{c/(1+\delta)} \ll \|f\|_1 N + N^{1-\delta/8} \ll \|f\|_1^{3\rho} N.$$

Assume $|\lambda| < (4\lceil R^c \rceil)^{-1}$, then $|\lambda T_i^*| \leq 2|\lambda||H_i| \leq 1/2$, and thus using the previous relations and $e^x \leq 1 + x + x^2$ for $|x| \leq 1/2$ we get

$$\begin{aligned} \mathbb{E} \exp \left(\lambda \sum_{i=2}^R T_i^* \right) &= \prod_{i=2}^R \mathbb{E} e^{\lambda T_i^*} \leq \prod_{i=2}^R \mathbb{E} (1 + \lambda T_i^* + \lambda^2 T_i^{*2}) \\ &= \prod_{i=2}^R (1 + \lambda^2 \mathbb{E} T_i^{*2}) \leq \exp \left(\lambda^2 \sum_{i=2}^R \mathbb{E} T_i^{*2} \right) \leq \exp(C_1 \lambda^2 \|f\|_1^{3\rho} N) \end{aligned}$$

with some constant $C_1 > 0$. Note that $\lambda = C_2 t_0 \|f\|_1^{-2\rho} N^{-1/2} \leq C_2 \|f\|_1^{-2\rho} N^{-1/2+\delta/16}$ with a small enough constant $C_2 > 0$ satisfies $|\lambda| < (4\lceil R^c \rceil)^{-1}$. Using the Markov inequality we thus get

$$\begin{aligned} \mathbb{P} \left(\sum_{i=2}^R T_i^* \geq t \|f\|_1^\rho \sqrt{N} \right) &\leq \exp \left(-\lambda t \|f\|_1^\rho \sqrt{N} \right) \mathbb{E} \exp \left(\lambda \sum_{i=2}^R T_i^* \right) \\ &\leq \exp \left(-\lambda t \|f\|_1^\rho \sqrt{N} + C_1 \lambda^2 \|f\|_1^{3\rho} N \right) \\ &\leq \exp \left(-C_2 \|f\|_1^{-\rho} t_0 t + C_1 C_2^2 \|f\|_1^{-\rho} t_0^2 \right) \\ &\leq \exp \left(-C \|f\|_1^{-\rho} t_0 t \right) \end{aligned}$$

with some constant $C > 0$. Using $-\lambda$ instead, we get that the same upper estimate holds for $\mathbb{P} \left(\sum_{i=2}^R T_i^* \leq -t \|f\|_1^\rho \sqrt{N} \right)$. Repeating the same arguments for $\sum_{i=2}^R D_i^*$, we altogether have

$$\mathbb{P} \left(\left| \sum_{i=2}^R (T_i^* + D_i^*) \right| \geq t \|f\|_1^\rho \sqrt{N} \right) \leq 4 \exp(-C \|f\|_1^{-\rho} t_0 t).$$

The last relation, together with (19) implies Lemma 7. \square

6 Chaining

Recall that for any interval $J \subseteq [0, 1)$ of length $\lambda(J)$, $f_J(x) = I_J(\{x\}) - \lambda(J)$ denotes the centered indicator of J , extended with period 1. Lemma 7 in the previous section gives a useful estimate for the tail probabilities

$$\mathbb{P} \left(\left| \sum_{k=M+1}^{M+N} f_J(S_k) \right| \geq t \lambda(J)^\rho \sqrt{N} \right)$$

which leads to the upper half of the LIL for the partial sums $\sum_{k=1}^N f_J(S_k)$ for a fixed interval $J \subseteq [0, 1)$ and along an exponentially growing subsequence of N 's. Proving the LIL for the discrepancy $D_N(S_k)$ requires, on the other hand, tail estimates for the variable

$$\max_{2^n \leq N < 2^{n+1}} \sup_{J \subseteq [0, 1)} \left| \sum_{k=1}^N f_J(S_k) \right|.$$

Such inequalities will be proved in this section, using the method of chaining. We follow closely Philipp [19].

Lemma 8. Assume that $\psi(k) \ll k^{-(1+\delta)}$ for some $0 < \delta < 1$. For any $a \in [0, 1)$, any $0 < \Delta \leq 1 - a$ and any integer $N \gg 1$ we have

$$\mathbb{P} \left(\sup_{0 < y \leq \Delta} \left| \sum_{k=1}^N f_{[a, a+y)}(S_k) \right| \gg \Delta^{\delta/30} \sqrt{N} + N^{1/2-\delta/30} \right) \ll \Delta^4 + N^{-\delta/30}.$$

The implied constants depend only on the distribution of X_1 and δ .

Proof. For any $h \geq 1$, $1 \leq j \leq 2^h$ let $\varphi_h^{(j)}$ denote the indicator function of the interval $[(j-1)2^{-h}, j2^{-h})$ extended with period 1. We observe that if $0 < y \leq \Delta$ has the dyadic expansion

$$y = \sum_{\ell=1}^{\infty} \varepsilon_{\ell} 2^{-\ell}, \quad \varepsilon_{\ell} = 0, 1, \quad (20)$$

and $H \geq 1$ is an arbitrary integer, then

$$\sum_{h=1}^H \varrho_h(x) \leq I_{[0, y)}(\{x\}) \leq \sum_{h=1}^H \varrho_h(x) + \sigma_H(x) \quad (21)$$

where ϱ_h is the indicator function of $\left[\sum_{\ell=1}^{h-1} \varepsilon_{\ell} 2^{-\ell}, \sum_{\ell=1}^h \varepsilon_{\ell} 2^{-\ell} \right)$ and σ_H is the indicator function of $\left[\sum_{\ell=1}^H \varepsilon_{\ell} 2^{-\ell}, \sum_{\ell=1}^H \varepsilon_{\ell} 2^{-\ell} + 2^{-H} \right)$, both extended with period 1. For $\varepsilon_h = 0$ clearly $\varrho_h \equiv 0$, and thus (21) remains valid if in the sums we keep only those terms where $\varepsilon_h = 1$. Also, for $\varepsilon_h = 1$, ϱ_h coincides with one of the $\varphi_h^{(j)}$, and σ_H also coincides with one of the $\varphi_H^{(j)}$. It follows that

$$\begin{aligned} f_{[0, y)}(x) &\leq \sum_{1 \leq h \leq H}^* (\varrho_h(x) - 2^{-h}) + (\sigma_H(x) - 2^{-H}) + 2^{-H}, \\ f_{[0, y)}(x) &\geq \sum_{1 \leq h \leq H}^* (\varrho_h(x) - 2^{-h}) - 2^{-H}, \end{aligned}$$

where \sum^* means that the summation is extended only for those h such that $\varepsilon_h = 1$. Setting $x = S_k - a$ and summing for $1 \leq k \leq N$ we get

$$\begin{aligned} \sum_{k=1}^N f_{[a, a+y)}(S_k) &\leq \sum_{k=1}^N \sum_{1 \leq h \leq H}^* (\varrho_h(S_k - a) - 2^{-h}) + \sum_{k=1}^N (\sigma_H(S_k - a) - 2^{-H}) + N2^{-H}, \\ \sum_{k=1}^N f_{[a, a+y)}(S_k) &\geq \sum_{k=1}^N \sum_{1 \leq h \leq H}^* (\varrho_h(S_k - a) - 2^{-h}) - N2^{-H}. \end{aligned}$$

Letting $A = \log_2 \Delta^{-1}$, observe that the dyadic expansion (20) of any $0 < y \leq \Delta$ starts with an index $\ell \geq A$. Thus it follows that for any integers $N \geq 1$ and $H \geq 1$

$$\sup_{0 < y \leq \Delta} \left| \sum_{k=1}^N f_{[a, a+y)}(S_k) \right| \leq 2 \sum_{A \leq h \leq H} \max_{1 \leq j \leq 2^h} \left| \sum_{k=1}^N \left(\varphi_h^{(j)}(S_k - a) - 2^{-h} \right) \right| + N2^{-H}. \quad (22)$$

Choose $H = \lfloor (1/2 + \delta/30) \log_2 N \rfloor$, and introduce the events

$$G(N, j, h) = \left\{ \left| \sum_{k=1}^N \left(\varphi_h^{(j)}(S_k - a) - 2^{-h} \right) \right| \geq D 2^{-(\delta/30)h} \sqrt{N} \right\},$$

$$G_N = \bigcup_{A \leq h \leq H} \bigcup_{1 \leq j \leq 2^h} G(N, j, h),$$

where $D > 0$ is a large constant, to be chosen. For any $h \leq H$ we have $2^{-h} \geq N^{-1/2-\delta/30}$, and thus applying Lemma 7 to the function $\varphi_h^{(j)}(x - a) - 2^{-h}$ with $t = t_0 = D$ and $\rho = \delta/30$ we get

$$\mathbb{P}(G(N, j, h)) \ll \exp(-CD^2 2^{(\delta/30)h}) + 2^{(\delta/15)h} N^{-1/2-\delta/8},$$

and consequently

$$\mathbb{P}(G_N) \ll \sum_{h \geq A} \exp(h - CD^2 2^{(\delta/30)h}) + N^{-1/2-\delta/8} \sum_{h \leq (1/2+\delta/30) \log_2 N} 2^{(1+\delta/15)h}. \quad (23)$$

Here $h - CD^2 2^{(\delta/30)h} \leq -4h$ provided D is large enough. The first sum in (23) is thus $\ll \exp(-4A) \ll \Delta^4$. The second term on the right hand side of (23) is $\ll N^{-\delta/30}$, hence $\mathbb{P}(G_N) \ll \Delta^4 + N^{-\delta/30}$. It follows that with the exception of a set with probability $\ll \Delta^4 + N^{-\delta/30}$, the right hand side of (22) is

$$\ll \sum_{A \leq h \leq H} 2^{-(\delta/30)h} \sqrt{N} + N 2^{-H} \ll 2^{-(\delta/30)A} \sqrt{N} + N^{1/2-\delta/30} \ll \Delta^{\delta/30} \sqrt{N} + N^{1/2-\delta/30}.$$

□

Lemma 9. Assume that $\psi(k) \ll k^{-(1+\delta)}$ for some $0 < \delta < 1$. For any $a \in [0, 1)$, any $0 < \Delta \leq 1 - a$ and any integer $n \gg 1$ we have

$$\mathbb{P} \left(\max_{2^n \leq N < 2^{n+1}} \sup_{0 < y \leq \Delta} \left| \sum_{k=1}^N f_{[a, a+y)}(S_k) \right| \gg \Delta^{\delta/30} \sqrt{2^n \log \log 2^n} + \sqrt{2^n} \right) \ll n^{-2}.$$

The implied constants depend only on the distribution of X_1 and δ .

Proof. As in the proof of Lemma 8, for any $h \geq 1$, $1 \leq j \leq 2^h$ let $\varphi_h^{(j)}$ denote the indicator function of the interval $[(j-1)2^{-h}, j2^{-h})$ extended with period 1. For any integers $M \geq 0$ and $N \geq 1$ let

$$F(M, N, j, h) = \left| \sum_{k=M+1}^{M+N} \left(\varphi_h^{(j)}(S_k - a) - 2^{-h} \right) \right|.$$

Following the proof of Lemma 8, we get that for any integers $M \geq 0$, $N \geq 1$ and $H \geq 1$

$$\sup_{0 < y \leq \Delta} \left| \sum_{k=M+1}^{M+N} f_{[a, a+y)}(S_k) \right| \leq 2 \sum_{A \leq h \leq H} \max_{1 \leq j \leq 2^h} F(M, N, j, h) + N 2^{-H}, \quad (24)$$

where $A = \log_2 \Delta^{-1}$. Let now $n \geq 1$, and let $2^n \leq N < 2^{n+1}$ have dyadic expansion $N = 2^n + \beta_{n-1}2^{n-1} + \dots + \beta_1 2 + \beta_0$, $\beta_i = 0, 1$. Letting $M_\ell = 2^n + \beta_{n-1}2^{n-1} + \dots + \beta_\ell 2^\ell$ we can decompose the interval of integers $[1, N]$ as

$$[1, N] = [1, 2^n] \cup \bigcup_{\substack{1 \leq \ell \leq n \\ \beta_{\ell-1}=1}} [M_\ell + 1, M_\ell + 2^{\ell-1}].$$

Note that the union of $[M_\ell + 1, M_\ell + 2^{\ell-1}]$ over $1 \leq \ell < n/2$ has total size $\ll 2^{n/2}$. Further, $M_\ell = 2^n + m_\ell 2^\ell$ for some integer $0 \leq m_\ell < 2^{n-\ell}$. Applying (24) on the interval $[1, 2^n]$ with $H = n/2$ and on the intervals $[M_\ell + 1, M_\ell + 2^{\ell-1}]$ for all $n/2 \leq \ell \leq n$ such that $\beta_{\ell-1} = 1$ with the choice $H = 2^{\ell/2}$, we get that

$$\begin{aligned} \sup_{0 < y \leq \Delta} \left| \sum_{k=1}^N f_{[a, a+y)}(S_k) \right| &\ll \sum_{A \leq h \leq n/2} \max_{1 \leq j \leq 2^h} F(0, 2^n, j, h) + 2^{n/2} \\ &+ \sum_{\substack{n/2 \leq \ell \leq n \\ \beta_{\ell-1}=1}} \left(\sum_{A \leq h \leq \ell/2} \max_{1 \leq j \leq 2^h} F(2^n + m_\ell 2^\ell, 2^{\ell-1}, j, h) + 2^{\ell/2} \right). \end{aligned}$$

Here $\sum_{n/2 \leq \ell \leq n} 2^{\ell/2} \ll 2^{n/2}$, therefore we altogether have

$$\begin{aligned} \max_{2^n \leq N < 2^{n+1}} \sup_{0 < y \leq \Delta} \left| \sum_{k=1}^N f_{[a, a+y)}(S_k) \right| &\ll \\ &\sum_{A \leq h \leq n/2} \max_{1 \leq j \leq 2^h} F(0, 2^n, j, h) + 2^{n/2} \\ &+ \sum_{n/2 \leq \ell \leq n} \sum_{A \leq h \leq \ell/2} \max_{0 \leq m < 2^{n-\ell}} \max_{1 \leq j \leq 2^h} F(2^n + m 2^\ell, 2^{\ell-1}, j, h). \end{aligned} \quad (25)$$

Let us introduce the events

$$\begin{aligned} G(n, j, h) &= \left\{ F(0, 2^n, j, h) \geq D 2^{-(\delta/30)h} \sqrt{2^n \log \log 2^n} \right\}, \\ G_n &= \bigcup_{A \leq h \leq n/2} \bigcup_{1 \leq j \leq 2^h} G(n, j, h), \\ G^*(n, j, h, m, \ell) &= \left\{ F(2^n + m 2^\ell, 2^{\ell-1}, j, h) \geq D 2^{-(\delta/30)(h+n-\ell)} \sqrt{2^n \log \log 2^n} \right\}, \\ G_n^* &= \bigcup_{n/2 \leq \ell \leq n} \bigcup_{A \leq h \leq \ell/2} \bigcup_{0 \leq m < 2^{n-\ell}} \bigcup_{1 \leq j \leq 2^h} G^*(n, j, h, m, \ell), \end{aligned}$$

where $D > 0$ is a large constant, to be chosen. Applying Lemma 7 to the function $\varphi_h^{(j)}(x - a) - 2^{-h}$ with $M = 0$, $N = 2^n$, $t = t_0 = D \sqrt{\log \log 2^n}$ and $\rho = \delta/30$ we get

$$\mathbb{P}(G(n, j, h)) \ll \exp(-CD^2 2^{(\delta/30)h} \log \log 2^n) + 2^{(\delta/15)h} 2^{(-1/2-\delta/8)n},$$

and consequently

$$\mathbb{P}(G_n) \ll \sum_{h=0}^{\infty} \exp(h - CD^2 2^{(\delta/30)h} \log \log 2^n) + 2^{(-1/2-\delta/8)n} \sum_{h \leq n/2} 2^{(1+\delta/15)h}. \quad (26)$$

Here $h - CD^2 2^{(\delta/30)h} \log \log 2^n \leq -h - 2 \log \log 2^n$ provided D is large enough, therefore the first sum in (26) is $\ll \exp(-2 \log \log 2^n) \ll n^{-2}$. The second term in (26) is $\ll 2^{-(\delta/30)n}$, hence $\mathbb{P}(G_n) \ll n^{-2}$.

Similarly, applying Lemma 7 to the function $\varphi_h^{(j)}(x-a) - 2^{-h}$ with $M = 2^n + m2^\ell$, $N = 2^{\ell-1}$, $t = D2^{(1/2-\delta/30)(n-\ell)+1/2} \sqrt{\log \log 2^n}$, $t_0 \geq D\sqrt{\log \log 2^n}$ and $\rho = \delta/30$ we get

$$\begin{aligned} \mathbb{P}(G^*(n, j, h, m, \ell)) &\ll \exp(-CD^2 2^{(\delta/30)h} 2^{(1/2-\delta/30)(n-\ell)} \log \log 2^n) \\ &\quad + 2^{(-1+\delta/15)(n-\ell)} 2^{(\delta/15)h} 2^{(-1/2-\delta/8)\ell}, \end{aligned}$$

and consequently

$$\begin{aligned} \mathbb{P}(G_n^*) &\ll \sum_{n/2 \leq \ell \leq n} \sum_{h=0}^{\infty} \exp(h + n - \ell - CD^2 2^{(\delta/30)h} 2^{(1/2-\delta/30)(n-\ell)} \log \log 2^n) \\ &\quad + \sum_{n/2 \leq \ell \leq n} \sum_{h \leq \ell/2} 2^h 2^{n-\ell} 2^{(-1+\delta/15)(n-\ell)} 2^{(\delta/15)h} 2^{(-1/2-\delta/8)\ell}. \end{aligned} \quad (27)$$

Here $h + n - \ell - CD^2 2^{(\delta/30)h} 2^{(1/2-\delta/30)(n-\ell)} \log \log 2^n \leq -h - (n - \ell) - 2 \log \log 2^n$ provided D is large enough, therefore the first sum in (27) is $\ll \exp(-2 \log \log 2^n) \ll n^{-2}$. Since the second sum in (27) is $\ll 2^{-(\delta/80)n} \ll n^{-2}$, we have $\mathbb{P}(G_n^*) \ll n^{-2}$.

It follows that with the exception of a set with probability $\ll n^{-2}$, the right hand side of (25) is

$$\begin{aligned} &\ll \sum_{h \geq A} 2^{-(\delta/30)h} \sqrt{2^n \log \log 2^n} + 2^{n/2} + \sum_{n/2 \leq \ell \leq n} \sum_{h \geq A} 2^{-(\delta/30)(h+n-\ell)} \sqrt{2^n \log \log 2^n} \\ &\ll \Delta^{\delta/30} \sqrt{2^n \log \log 2^n} + 2^{n/2}. \end{aligned}$$

□

7 Proof of Theorem 2

In Lemma 4 we proved that $C(f, g)$ is bilinear, symmetric and positive semidefinite in $f, g \in \mathcal{F}$. In particular, $\Gamma(s, t)$ is symmetric and positive semidefinite. From the triangle inequality we get that

$$\begin{aligned} |\mathbb{E}(f_{[0,s]}(U) - f_{[0,s']}(U))f_{[0,t]}(U + S_k)| &\leq 2|s - s'|, \\ |\mathbb{E}f_{[0,s]}(U)(f_{[0,t]}(U + S_k) - f_{[0,t']}(U + S_k))| &\leq 2|t - t'| \end{aligned}$$

for any $s, s', t, t' \in [0, 1]$ and any $k \geq 1$. Hence $\mathbb{E}f_{[0,s]}(U)f_{[0,t]}(U + S_k)$ is continuous on the unit square for any fixed $k \geq 1$. As observed in the proof of Lemma 4 (i), each term in the series defining $\Gamma(s, t)$ has absolute value at most $\psi(k) \ll k^{-(1+\delta)}$, and so the continuity of $\Gamma(s, t)$ follows.

Lemma 10. *For any fixed $0 \leq t_1 < t_2 < \dots < t_r \leq 1$ the random vector $\mathbf{Y}_k = (f_{[0,t_1]}(S_k), f_{[0,t_2]}(S_k), \dots, f_{[0,t_r]}(S_k))$ satisfies*

$$N^{-1/2} \sum_{k=1}^N \mathbf{Y}_k \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma),$$

where $\mathcal{N}(\mathbf{0}, \Sigma)$ denotes the multidimensional normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma = (\Gamma(t_i, t_j))_{1 \leq i, j \leq r}$.

Proof. Let $c = 1 - \delta/8$ and $c' = (c + \delta/2)/(1 + \delta)$. Note that $c' < c$. Let us decompose the set of positive integers into consecutive blocks $H_1, J_1, H_2, J_2, \dots$ such that $|H_i| = \lceil i^c \rceil$ and $|J_i| = \lceil i^{c'} \rceil$ for all $i \geq 1$. Put $f_j = f_{[0, t_j]}$, $j = 1, 2, \dots, r$. Following the notation of Section 4 let $\mathbf{T}_i = (T_i^{(f_1)}, T_i^{(f_2)}, \dots, T_i^{(f_r)})$ and $\mathbf{T}_i^* = (T_i^{(f_1)*}, T_i^{(f_2)*}, \dots, T_i^{(f_r)*})$, and let \mathbf{D}_i and \mathbf{D}_i^* be defined analogously. For any positive integer N let $R = R(N)$ be such that $N \in H_R \cup J_R$. Note that $R = \Theta(N^{1/(c+1)})$. Since $|\mathbf{T}_i| \ll |H_i| \ll R^c \ll N^{1/2-\delta/32}$ for all $1 \leq i \leq R$ and the same holds for $|\mathbf{D}_i|$, the contribution of any individual block is negligible, and we get

$$\sum_{k=1}^N \mathbf{Y}_k = \sum_{i=2}^R (\mathbf{T}_i + \mathbf{D}_i) + O(N^{1/2-\delta/32}). \quad (28)$$

Applying Lemma 6 to f_1, f_2, \dots, f_r we get $\mathbb{E} \left| \sum_{i=2}^R (\mathbf{T}_i + \mathbf{D}_i - \mathbf{T}_i^* - \mathbf{D}_i^*) \right|^2 \ll R^{1-\delta/2}$, and hence from the Chebyshev inequality and the Borel–Cantelli lemma $\left| \sum_{i=2}^R (\mathbf{T}_i + \mathbf{D}_i - \mathbf{T}_i^* - \mathbf{D}_i^*) \right| \ll R^{1-\delta/8}$ a.s. Thus replacing \mathbf{T}_i by \mathbf{T}_i^* and \mathbf{D}_i by \mathbf{D}_i^* , the right hand side of (28) changes by $\ll R^{1-\delta/8} \ll N^{1/2-\delta/32}$, and so

$$\sum_{k=1}^N \mathbf{Y}_k = \sum_{i=2}^R (\mathbf{T}_i^* + \mathbf{D}_i^*) + O(N^{1/2-\delta/32}) \quad \text{a.s.} \quad (29)$$

Thus it will be enough to prove the multidimensional CLT for the right hand side of (29).

Recall that $\mathbf{T}_2^*, \mathbf{T}_3^*, \dots$ are independent, mean $\mathbf{0}$ random vectors. From (14) we deduce

$$\text{Cov}(T_i^{(f_j)*}, T_i^{(f_k)*}) = \frac{1}{4} \left(\mathbb{E} T_i^{(f_j+f_k)*2} - \mathbb{E} T_i^{(f_j-f_k)*2} \right) \sim C(f_j, f_k) |H_i|.$$

Here $C(f_j, f_k) = \Gamma(t_j, t_k)$, thus letting Σ_i denote the covariance matrix of \mathbf{T}_i^* , we have $\Sigma_i \sim |H_i| \Sigma$. Since $\sum_{i=2}^R |H_i| \sim N$, we have $N^{-1} \sum_{i=2}^R \Sigma_i \rightarrow \Sigma$. Note that $|\mathbf{T}_i^*| \ll |H_i| \ll R^c \ll N^{1/2-\delta/32}$ for all $2 \leq i \leq R$. Thus $\mathbf{T}_2^*, \mathbf{T}_3^*, \dots$ satisfies the Lindeberg condition, and by the classical multidimensional CLT we have

$$N^{-1/2} \sum_{i=2}^R \mathbf{T}_i^* \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma). \quad (30)$$

Similarly, $\mathbf{D}_2^*, \mathbf{D}_3^*, \dots$ are independent, mean $\mathbf{0}$ random vectors. From (14) we deduce

$$\sum_{i=2}^R \mathbb{E} |\mathbf{D}_i^*|^2 \ll \sum_{i=2}^R |J_i| \ll R^{1+c'} = o(N),$$

therefore $N^{-1/2} \sum_{i=2}^R \mathbf{D}_i^* \xrightarrow{\mathbb{P}} \mathbf{0}$. The last relation together with (30) imply $N^{-1/2} \sum_{i=2}^R (\mathbf{T}_i^* + \mathbf{D}_i^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$, as claimed. \square

Lemma 10 implies the convergence of the finite dimensional distributions of the sequence $\sqrt{N}(F_N(t) - t)$ in (6) to those of $K(t)$, and to prove Theorem 2 it remains to prove the tightness of the sequence in $D[0, 1]$. To this end, fix $0 < \varepsilon < 1$ and let

m be the positive integer such that $1/(m+1) \leq \varepsilon < 1/m$. Applying Lemma 8 with $a = j/m$, $j = 0, 1, \dots, m-2$ and $\Delta = 2/m$, it follows that

$$\mathbb{P} \left(\max_{0 \leq j \leq m-2} \sup_{0 < y \leq 2/m} \left| \sum_{k=1}^N f_{[j/m, j/m+y)}(S_k) \right| \gg m^{-\delta/30} \sqrt{N} + N^{1/2-\delta/30} \right) \ll m^{-3} + mN^{-\delta/30}.$$

Note that for any $s, t \in [0, 1]$ with $s < t \leq s + \varepsilon$ we have $\sqrt{N}|(F_N(t) - t) - (F_N(s) - s)| = N^{-1/2} \sum_{k=1}^N f_{[s, t)}(S_k)$, and that $[s, t)$ is a subset of $[j/m, (j+2)/m)$ for some $0 \leq j \leq m-2$. This implies that with the exception of a set with probability $\ll \varepsilon^3 + \varepsilon^{-1} N^{-\delta/30}$ the fluctuation of the process $\sqrt{N}(F_N(t) - t)$ over any subinterval of $[0, 1]$ with length $\leq \varepsilon$ is $\ll \varepsilon^{\delta/30} + N^{-\delta/30}$. By Theorem 15.2 of Billingsley [8, Chapter 3], the sequence $\sqrt{N}(F_N(t) - t)$ is tight in $D[0, 1]$. This completes the proof of Theorem 2.

8 Proof of Theorem 3

Lemma 11. Fix $0 \leq t_1 < t_2 < \dots < t_r \leq 1$, and consider the random vector $\mathbf{Y}_k = (f_{[0, t_1)}(S_k), f_{[0, t_2)}(S_k), \dots, f_{[0, t_r)}(S_k))$. With probability 1, the set of limit points of the sequence $(2N \log \log N)^{-1/2} \sum_{k=1}^N \mathbf{Y}_k$ is the ellipsoid

$$\{\Sigma \mathbf{x} \in \mathbb{R}^r : \langle \mathbf{x}, \Sigma \mathbf{x} \rangle \leq 1\} \quad (31)$$

in the range $\{\Sigma \mathbf{x} : \mathbf{x} \in \mathbb{R}^r\}$ of the matrix $\Sigma = (\Gamma(t_i, t_j))_{1 \leq i, j \leq r}$.

Proof. Lemma 11 can be proved in the same way as Lemma 10. The only difference is that after deducing (29), instead of the multidimensional CLT we apply the multidimensional version of Kolmogorov's LIL due to Berning [7]. This way we get that with probability 1, the set of limit points of the sequence $(2N \log \log N)^{-1/2} \sum_{i=2}^R \mathbf{T}_i^*$ is the ellipsoid (31), and that $(2N \log \log N)^{-1/2} \sum_{i=2}^R \mathbf{D}_i^* \rightarrow \mathbf{0}$ a.s. \square

Using Lemma 11 and Lemma 9, the proof of Theorem 3 can be easily completed. Let

$$\alpha_N(t) = \frac{N(F_N(t) - t)}{\sqrt{2N \log \log N}}, \quad 0 \leq t \leq 1,$$

and let L denote the (random) class of limit functions of subsequences of α_N in the $D[0, 1]$ topology. Similarly to the proof of the functional CLT, which requires tightness and the convergence of the finite dimensional distributions, to prove the functional LIL in Theorem 3 we will show that with probability 1 the sequence α_N is relatively compact in the $D[0, 1]$ topology, and that for any fixed finite subset $T \subset [0, 1]$ we have $\{f|_T : f \in L\} = \{f|_T : f \in B(\Gamma)\}$ a.s.

To prove relative compactness, fix $0 < \varepsilon < 1$ and let m be the positive integer such that $1/(m+1) \leq \varepsilon < 1/m$. Applying Lemma 9 with $a = j/m$, $j = 0, 1, \dots, m-2$ and $\Delta = 2/m$, it follows that

$$\mathbb{P} \left(\max_{0 \leq j \leq m-2} \max_{2^n \leq N < 2^{n+1}} \sup_{0 < y \leq 2/m} \left| \sum_{k=1}^N f_{[j/m, j/m+y)}(S_k) \right| \gg m^{-\delta/30} \sqrt{2^n \log \log 2^n} + \sqrt{2^n} \right) \ll m/n^2.$$

By the Borel–Cantelli lemma this implies that with probability 1, for any $0 < \varepsilon < 1$ there exists a (random) integer N_0 such that for any $N \geq N_0$ the fluctuation of the process $\alpha_N(t)$ over any subinterval of $[0, 1]$ of length $\leq \varepsilon$ is $\ll \varepsilon^{\delta/30}$. Applying Lemma 9 with $a = 0$ and $\Delta = 1$, from the Borel–Cantelli lemma we similarly deduce that with probability 1, $\sup_{t \in [0, 1]} |\alpha_N(t)| \ll 1$. By Theorem 14.3 of Billingsley [8, Chapter 3], with probability 1 the sequence $\alpha_N(t)$ is relatively compact in $D[0, 1]$. Further, it is easy to see from the fluctuation estimate that with probability 1, any subsequence of $\alpha_N(t)$ convergent in $D[0, 1]$ has a continuous limit function. Hence with probability 1, L is a compact subset of $C[0, 1]$ and any subsequence of α_N convergent in $D[0, 1]$ is also uniformly convergent.

Next, fix a finite set $T \subset [0, 1]$. Let $H(\Gamma)$ and $H(\Gamma|_{T \times T})$ denote the reproducing kernel Hilbert spaces determined by the covariance functions Γ and $\Gamma|_{T \times T}$, respectively. Then $H(\Gamma|_{T \times T}) = \{f|_T : f \in H(\Gamma)\}$; moreover, $\|g\|_{H(\Gamma|_{T \times T})} = \min\{\|f\|_{H(\Gamma)} : f|_T = g\}$, see Aronszajn [1, Section I.5]. In particular, the closed unit ball of $H(\Gamma|_{T \times T})$ is $\{f|_T : f \in B(\Gamma)\}$. Lemma 11 thus shows that with probability 1, the set of limit points of $\alpha_N|_T$ is identical with $\{f|_T : f \in B(\Gamma)\}$. The relative compactness of α_N implies that in fact $\{f|_T : f \in L\} = \{f|_T : f \in B(\Gamma)\}$ a.s., as claimed.

Now fix finite sets $T_1 \subset T_2 \subset \dots \subset [0, 1]$ such that $\bigcup_{n=1}^{\infty} T_n$ is dense in $[0, 1]$. Observe that if K and K' are compact subsets of $C[0, 1]$ such that $\{f|_{T_n} : f \in K\} = \{f|_{T_n} : f \in K'\}$ for all $n \geq 1$, then $K = K'$. By Lemma 3 of Oodaira [18], the continuity of $\Gamma(s, t)$ implies that $B(\Gamma)$ is compact in $C[0, 1]$. The arguments above show that with probability 1, L is a compact subset of $C[0, 1]$ and $\{f|_{T_n} : f \in L\} = \{f|_{T_n} : f \in B(\Gamma)\}$ for all $n \geq 1$. Therefore $L = B(\Gamma)$ a.s. This finishes the proof of Theorem 3.

9 Proof of Theorem 1

The proof of Theorem 1 is similar to, but is actually much simpler than that of Theorems 2 and 3. The claims on $C(f, g)$ were proved in Lemma 4. If in the proof of Lemmas 10 and 11 we replace the vector-valued random variables \mathbf{Y}_k , \mathbf{T}_i , \mathbf{T}_i^* , \mathbf{D}_i and \mathbf{D}_i^* by the real-valued random variables $f(S_k)$, $T_i^{(f)}$, $T_i^{(f)*}$, $D_i^{(f)}$ and $D_i^{(f)*}$, respectively, where $f \in \mathcal{F}$ is fixed, we get that the sum $\sum_{k=1}^N f(S_k)$ satisfies the CLT and the LIL.

References

- [1] Aronszajn, N.: *Theory of reproducing kernels*. Trans. Amer. Math. Soc. 68 (1950), 337–404.
- [2] Bazarova, A., Berkes, I. and Raseta, M.: *On the discrepancy of random walks on the circle*. Unif. Distrib. Theory 14 (2019), 73–86.
- [3] Berkes, I. and Borda, B.: *Berry–Esseen bounds and Diophantine approximation*. Anal. Math. 44 (2018), 149–161.
- [4] Berkes, I. and Borda, B.: *On the discrepancy of random subsequences of $\{n\alpha\}$* . Acta Arith. 191 (2019), 383–416.

- [5] Berkes, I. and Raseta, M.: *On the discrepancy and empirical distribution function of $\{n_k\alpha\}$* . Unif. Distrib. Theory 10 (2015), 1–17.
- [6] Berkes, I. and Weber, M.: *On the convergence of $\sum c_k f(n_k x)$* . Mem. Amer. Math. Soc. 201 (2009), no. 943, viii+72 pp.
- [7] Berning, J.: *On the multivariate law of the iterated logarithm*. Ann. Probab. 7 (1979), 980–988.
- [8] Billingsley, P.: *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [9] Bobkov, S.: *Central limit theorem and Diophantine approximations*. J. Theoret. Probab. 31 (2018), 2390–2411.
- [10] Bobkov, S.: *Khinchin’s theorem and Edgeworth approximations for weighted sums*. Ann. Statist. 47 (2019), 1616–1633.
- [11] Bobkov, S. and Ledoux, M.: *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*. Mem. Amer. Math. Soc. 261 (2019), no. 1259, v+126 pp.
- [12] Diaconis, P.: *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics Lecture Notes–Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [13] Donsker, M.: *Justification and extension of Doob’s heuristic approach to the Kolmogorov–Smirnov theorems*. Ann. Math. Statistics 23 (1952), 277–281.
- [14] Finkelstein, H.: *The law of the iterated logarithm for empirical distributions*. Ann. Math. Statist. 42 (1971), 607–615.
- [15] Hensley, D. and Su, F.: *Random walks with badly approximable numbers*. Unusual applications of number theory, 95–101, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 64, Amer. Math. Soc., Providence, RI, 2004.
- [16] Kuipers, L. and Niederreiter, H.: *Uniform Distribution of Sequences*. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
- [17] Lévy, P.: *L’addition des variables aléatoires définies sur un circonférence*. Bull. Soc. Math. France 67 (1939), 1–41.
- [18] Oodaira, H.: *On Strassen’s version of the law of the iterated logarithm for Gaussian processes*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 21 (1972), 289–299.
- [19] Philipp, W.: *A functional law of the iterated logarithm for empirical distribution functions of weakly dependent random variables*. Ann. Probability 5 (1977), 319–350.
- [20] Schatte, P.: *On the asymptotic uniform distribution of the n -fold convolution mod 1 of a lattice distribution*. Math. Nachr. 128 (1986) 233–241.
- [21] Schatte, P.: *On a law of the iterated logarithm for sums mod 1 with application to Benford’s law*. Probab. Theory Related Fields 77 (1988), 167–178.
- [22] Schatte, P.: *On a uniform law of the iterated logarithm for sums mod 1 and Benford’s law*. Lithuanian Math. J. 31 (1991), 133–142.
- [23] Strassen, V.: *An invariance principle for the law of the iterated logarithm*. Z. Wahrsch. verw Gebiete 3 (1964), 211–226.

- [24] Su, F.: *Convergence of random walks on the circle generated by an irrational rotation*. Trans. Amer. Math. Soc. 350 (1998), 3717–3741.