On the discrepancy of random walks on the circle

Alina Bazarova[1], István Berkes[2] and Marko Raseta[3]

Abstract

Let $X_1, X_2, \ldots$ be i.i.d. absolutely continuous random variables, let $S_k = \sum_{j=1}^k X_j \pmod 1$ and let D_N^* denote the star discrepancy of the sequence $(S_k)_{1 \le k \le N}$. We determine the limit distribution of $\sqrt{N} D_N^*$ and the weak limit of the sequence $\sqrt{N}(F_N(t) - t)$ in the Skorohod space $D[0,1]$, where $F_N(t)$ denotes the empirical distribution function of the sequence $(S_k)_{1 \le k \le N}$.

1 Introduction

Let $X_1, X_2, \ldots$ be i.i.d. absolutely continuous random variables and let $S_k = \sum_{j=1}^k X_j \pmod 1$. By a classical result of Lévy [6], the distribution of S_k

2010 Mathematics Subject Classification. 11K38, 60G50, 60F17

Keywords: i.i.d. sums mod 1; discrepancy

[1)] University of Birmingham, Centre for Computational Biology, Institute of Cancer and Genomic Sciences. Email: a.bazarova@bham.ac.uk.

[2)] A. Rényi Institute of Mathematics, Reáltanoda u. 13–15, 1053 Budapest, Hungary. Email: berkes.istvan@renyi.mta.hu. Research supported by NKFIH Grant K 125569.

[3)] University of Keele, Research Institute for Primary Care and Health Sciences and Research Institute for Applied Clinical Sciences. Email: m.raseta@keele.ac.uk.

converges weakly to the uniform distribution on $(0, 1)$. Schatte [8] proved that the speed of convergence is exponential, and letting

$$D_N^* := \sup_{0 \le a < 1} \left| \frac{1}{N} \sum_{k=1}^{N} I_{(0,a)}(S_k) - a \right| \quad (N = 1, 2, \ldots)$$

denote the star discrepancy of the sequence $(S_k)_{1 \le k \le N}$, he also proved in [9] the law of the iterated logarithm

$$\limsup_{N \to \infty} \sqrt{\frac{N}{\log \log N}} D_N^* = \gamma \quad \text{a.s.} \tag{1}$$

where

$$\gamma = \sup_{x \in [0,1)} \sigma^2(x)$$

with

$$\sigma^2(x) = x - x^2 + 2 \sum_{j=1}^{\infty} \left(\mathbb{E} I_{(0,x)}(U) I_{(0,x)}(U + X_j) - x^2 \right). \tag{2}$$

Here U is a uniform (0, 1) random variable independent of the sequence $(X_n)_{n \ge 1}$ and for $0 \le a < b \le 1$, $I_{(a,b)}$ denotes the indicator function of (a, b), extended with period 1. Letting

$$F_N(t) = F_N(t, \omega) = \frac{1}{N} \sum_{k=1}^{N} I_{(0,t)}(S_k) \qquad (0 \le t \le 1) \tag{3}$$

denote the empirical distribution function of the first N terms of the sequence $(S_n)_{n \ge 1}$, Berkes and Raseta [2] proved a Strassen type functional LIL for $F_N(t)$, yielding precise asymptotics for several functionals of the empirical process. The purpose of the present paper is to prove the following result, determining the limit distribution of $\sqrt{N} D_N^*$.

Theorem 1. *Let $X_1, X_2, \ldots$ be i.i.d. random variables and assume X_1 is bounded with bounded density. Let*

$$\Gamma(s, t) = s(1 - t) + \sum_{k=1}^{\infty} \mathbb{E} f_s(U) f_t(U + S_k) + \sum_{k=1}^{\infty} \mathbb{E} f_t(U) f_s(U + S_k), \tag{4}$$

where U is a $U(0,1)$ variable independent of $(X_n)_{n\in\mathbb{N}}$ and $f_s = I_{(0,s)} - s$. Then the series in (4) are absolutely convergent and

$$\sqrt{N} D_N^* \xrightarrow{d} \sup_{0\le t\le 1} |K(t)|, \tag{5}$$

where $K(s)$ a mean zero Gaussian process with covariance function $\Gamma(s,t)$.

Actually, Theorem 1 will be deduced from a more general functional result describing the weak limit behavior of the empirical distribution function $F_N(t)$.

Theorem 2. *Under the conditions of Theorem 1 we have*

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

Relation (6) expresses weak convergence in the Skorohod space $D[0,1]$, see Billingsley [4] for basic definitions and facts for weak convergence of probability measures on metric spaces.

By a classical result of Donsker [5], if $X_1, X_2, \dots$ are i.i.d. random variables with distribution function F and F_N denotes the empirical distribution function of the sample $(X_1, \dots, X_N)$, then

$$\sqrt{N}(F_N(t) - F(t)) \xrightarrow{D[0,1]} B(F(t))$$

where B is Brownian bridge. Note the substantial difference caused by considering mod 1 sums in the present case.

If X_1 has a lattice distribution, the situation changes essentially. For example, in [1] it shown that if α is irrational and X_1 takes the values α and 2α with probability $1/2 - 1/2$, then up to logarithmic factors, the order of magnitude of D_N^* is $O(N^{-1/2})$ or $O(N^{-1/\gamma})$ according as $\gamma < 2$ or $\gamma > 2$, where γ is the Diophantine rank of α, i.e. the supremum of numbers c such that $|\alpha - p/q| < q^{-c-1}$ holds for infinitely many fractions p/q. The asymptotic distribution of D_N^* in this case remains open.

2 Proofs

The proof of our theorems uses, similarly to that of the functional LIL in [2], a traditional blocking argument combined with a coupling lemma of Schatte, see Lemma 1 below. The substantial new difficulty is to prove the tightness of the sequence $\sqrt{N}(F_N(t) - t)$, since the standard maximal inequalities (e.g. Billingsley's inequalities in [4], Section 2.12) are not applicable here. We circumvent this difficulty by proving a Chernoff type exponential bound (Lemma 6) for the considered partial sums which, combined with the chaining method of Philipp [7], yields the desired fluctuation inequality (Lemma 7).

Lemma 1. *Let $\ell \geq 1$ and let $I_1, I_2, \ldots$ be disjoint blocks of integers with $\geq \ell$ integers between consecutive blocks. Then there exists a sequence $\delta_1, \delta_2, \ldots$ of random variables such that*

$$|\delta_n| \leq Ce^{-\lambda \ell}$$

with some positive constants C, λ and the random vectors

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

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

For the proof, see [2]. The uniformity statement is implicit in the proof; see also Lemma 4.3 of [3].

In what follows, $C, \lambda, \gamma, \gamma' \ldots$ will denote positive constants, possibly different at different places, depending (at most) on the distribution of X_1. The relation $\ll$ will mean the same as the big O notation, with a constant depending on the distribution of X_1.

Let $\mathcal{F}$ denote the class of functions f of the form $f = I_{(a,b)} - (b - a)$ $(0 \leq a < b \leq 1)$, extended with period 1. For $f \in \mathcal{F}$ we put

$$A^{(f)} := \|f\|^2 + 2\sum_{k=1}^{\infty} \mathbb{E} f(U) f(U + S_k) \tag{7}$$

where U is a uniform $(0, 1)$ random variable, independent of $(X_j)_{j\geq 1}$ and $\|f\|$ denotes the $L^2(0, 1)$ norm of f. Put further

$$\widetilde{m}_k = \sum_{j=1}^{k} \lfloor j^{1/2} \rfloor, \quad \widehat{m}_k = \sum_{j=1}^{k} \lfloor j^{1/4} \rfloor$$

and let $m_k = \widetilde{m}_k + \widehat{m}_k$. Using Lemma 1 we can construct sequences $(\Delta_k)_{k\in\mathbb{N}}$, $(\Pi_k)_{k\in\mathbb{N}}$ of random variables such that $\Delta_0 = 0$, $\Pi_0 = 0$,

$$|\Delta_k| \le Ce^{-\lambda k^{1/4}}, \qquad |\Pi_k| \le Ce^{-\lambda\sqrt{k}} \tag{8}$$

and

$$T_k^{(f)} := \sum_{j=m_{k-1}+1}^{m_{k-1}+\lfloor\sqrt{k}\rfloor} f(S_j - \Delta_{k-1}) \qquad k = 1, 2, \dots$$
$$T_k^{*(f)} := \sum_{j=m_{k-1}+\lfloor\sqrt{k}\rfloor+1}^{m_k} f(S_j - \Pi_{k-1}) \qquad k = 1, 2, \dots$$

are sequences of independent random variables. Since $\int_0^1 f(x)dx = 0$ for $f \in \mathcal{F}$, the uniformity statement in Lemma 1 implies that $\mathbb{E}T_k^{(f)} = \mathbb{E}T_k^{*(f)} = 0$ for $k \ge 2$. The following asymptotic estimates for the variances of $T_k^{(f)}$ and $T_k^{*(f)}$ are from [2].

Lemma 2. *For $f \in \mathcal{F}$ we have*

$$\sum_{k=1}^{n} \mathbb{V}\mathrm{ar}(T_k^{(f)}) \sim A^{(f)}\widetilde{m}_n \qquad \sum_{k=1}^{n} \mathbb{V}\mathrm{ar}(T_k^{*(f)}) \sim A^{(f)}\widehat{m}_n,$$

where $A^{(f)}$ is defined by (7).

Since

$$\mathrm{Cov}(T_k^{(f)}, T_k^{(g)}) = \frac{1}{4}\left(\mathbb{V}\mathrm{ar}\,(T_k^{(f+g)}) - \mathbb{V}\mathrm{ar}\,(T_k^{(f-g)})\right),$$

Lemma 2 implies

$$\sum_{k=1}^{n} \mathrm{Cov}(T_k^{(f)}, T_k^{(g)}) \sim \frac{1}{4}\left(A^{(f+g)} - A^{(f-g)}\right)\widetilde{m}_n \tag{9}$$

and

$$\sum_{k=1}^{n} \mathrm{Cov}(T_k^{*(f)}, T_k^{*(g)}) \sim \frac{1}{4}\left(A^{(f+g)} - A^{(f-g)}\right)\widehat{m}_n. \tag{10}$$

From (7) it follows that

$$A^{(f+g)}-A^{(f-g)} = 4\langle f,g\rangle+4\sum_{k=1}^{\infty}\mathbb{E}f(U)g(U+S_k)+4\sum_{k=1}^{\infty}\mathbb{E}g(U)f(U+S_k). \quad (11)$$

Lemma 3. *Let $f \in \mathcal{F}$, $h > 0$ and let ξ be a random variable with $|\xi| < h$. Then for any $n \geq 1$ we have*

$$\mathbb{E}|f(S_n+\xi) - f(S_n)|^2 \leq Ch.$$

Proof. Since X_1 is bounded with bounded density, Theorem 1 of [8] implies that the sums $S_n = \sum_{k=1}^n X_k \pmod 1$ have a uniformly bounded density and thus

$$\mathbb{P}(S_n \in J) \leq C|J| \quad \text{for any interval } J. \quad (12)$$

Now if $f = I_{(a,b)}-(b-a)$, then $|f(S_n+\xi)-f(S_n)| = |I_{(a,b)}(S_n+\xi)-I_{(a,b)}(S_n)|$ is different from 0 only if one of $S_n+\xi$ and S_n lies in (a,b) and the other does not, which, in view of $|\xi| < h$, implies that S_n lies closer to the boundary of (a,b) than h, i.e. $S_n \in (a, a+h)$ or $S_n \in (b-h, b)$. Since $|f(S_n+\xi) - f(S_n)| \leq 2$, Lemma 3 follows from (12).

Lemma 4. *For $f \in \mathcal{F}$ and any $M \geq 0$, $N \geq 1$ we have*

$$\mathbb{E}\left(\sum_{k=M+1}^{M+N} f(S_k)\right)^2 \leq C\|f\|N. \quad (13)$$

Proof. We first show

$$|\mathbb{E}f(S_k)f(S_\ell)| \leq Ce^{-\lambda(\ell-k)}\|f\| \qquad (k < \ell). \quad (14)$$

Indeed, by the proof of Lemma 1 in [2], there exists a r.v. Δ with $|\Delta| \leq Ce^{-\lambda(\ell-k)}$ such that $S_\ell - \Delta$ is a uniform r.v. independent of S_k. Hence

$$\mathbb{E}f(S_\ell - \Delta) = \int_0^1 f(t)dt = 0$$

and thus

$$\mathbb{E}f(S_k)f(S_\ell - \Delta) = \mathbb{E}f(S_k)\mathbb{E}f(S_\ell-\Delta) = 0. \quad (15)$$

On the other hand,

$$\begin{aligned}
&|\mathbb{E}f(S_k)f(S_\ell) - \mathbb{E}f(S_k)f(S_\ell - \Delta)| \\
&\le \mathbb{E}\big(|f(S_k)|\,|f(S_\ell) - f(S_\ell - \Delta)|\big) \le \\
&\big(\mathbb{E}f^2(S_k)\big)^{1/2}\big(\mathbb{E}|f(S_\ell) - f(S_\ell - \Delta)|^2\big)^{1/2}.
\end{aligned} \tag{16}$$

Using (12) we get

$$\mathbb{E}f^2(S_k) \le C\int_0^1 f^2(t)dt = C\|f\|^2. \tag{17}$$

Also, $|\Delta| \le Ce^{-\lambda(\ell-k)}$ and Lemma 3 imply

$$\mathbb{E}|f(S_\ell) - f(S_\ell - \Delta)|^2 \le Ce^{-\lambda(\ell-k)} \tag{18}$$

which, together with (16)–(18), gives

$$|\mathbb{E}f(S_k)f(S_\ell) - \mathbb{E}f(S_k)f(S_\ell - \Delta)| \le Ce^{-\lambda(\ell-k)}.$$

Thus using (15) we get (14). Now by (14)

$$\left|\sum_{M+1\le k<\ell\le M+N} \mathbb{E}f(S_k)f(S_\ell)\right| \le CN\|f\|\sum_{\ell\ge 1} e^{-\lambda\ell} \le CN\|f\|$$

which, together with (17), completes the proof of Lemma 4. □

Let $0 < t_1 < \ldots < t_r \le 1$ and put

$$\mathbf{Y}_k = (f_{(0,t_1)}(S_k), f_{(0,t_2)}(S_k), \ldots, f_{(0,t_r)}(S_k))$$

where $f_{(a,b)} = I_{(a,b)} - (b-a)$, with the indicator $I_{(a,b)}$ extended with period 1, as before.

Lemma 5. *We have*

$$N^{-1/2}\sum_{k=1}^N \mathbf{Y}_k \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}), \tag{19}$$

where

$$\mathbf{\Sigma} = (\Gamma(t_i, t_j))_{1\le i,j\le r}.$$

Proof. Let

$$\mathbf{T}_k = \left(T_k^{(f_{(0,t_1)})}, \ldots, T_k^{(f_{(0,t_r)})}\right), \qquad \mathbf{T}_k^* = \left(T_k^{*(f_{(0,t_1)})}, \ldots, T_k^{*(f_{(0,t_r)})}\right).$$

and let $\mathbf{\Sigma_k}$ denote the covariance matrix of the vector $\mathbf{T}_k$. From (9), (10) and (11) it follows that

$$m_n^{-1}\left(\mathbf{\Sigma_1} + \ldots + \mathbf{\Sigma_n}\right) \longrightarrow \mathbf{\Sigma}.$$

Clearly

$$|\mathbf{T}_k| \le C_r k^{1/2} = o(m_k^{1/2})$$

where C_r is a constant depending on r, showing that the sequence $(\mathbf{T}_k)_{k\ge1}$ of independent, mean $\mathbf{0}$ random vectors satisfies the Lindeberg condition and thus

$$m_n^{-1/2}\sum_{k=1}^{n}\mathbf{T}_k \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}). \tag{20}$$

A similar statement holds for the sequence $(\mathbf{T}_k^*)_{k\ge1}$, implying that

$$\left|\sum_{k=1}^{n}\mathbf{T}_k^*\right| = O_P(\widehat{m}_n) = o_P(m_n^{1/2}). \tag{21}$$

and consequently

$$m_n^{-1/2}\sum_{k=1}^{n}(\mathbf{T}_k + \mathbf{T}_k^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}). \tag{22}$$

Now using (8) and Lemma 3 we get

$$\|T_k^{(f)} - \sum_{j=m_{k-1}+1}^{m_{k-1}+\lfloor\sqrt{k}\rfloor} f(S_j)\| \ll \sqrt{k}e^{-\lambda k^{1/4}}$$

and

$$\|T_k^{*(f)} - \sum_{j=m_{k-1}+\lfloor\sqrt{k}\rfloor+1}^{m_k} f(S_j)\| \ll k^{1/4}e^{-\lambda k^{1/2}}$$

and thus

$$\|\sum_{k=1}^{m_n}\mathbf{Y}_k - \sum_{k=1}^{n}(\mathbf{T}_k + \mathbf{T}_\mathbf{k}^*)\| = O(1).$$

Together with (22) this shows that (19) holds for the indices $N = m_n$. To get (19) for all N, observe that $m_k \sim ck^{3/2}$ and thus for $m_k \le N < m_{k+1}$ we have

$$\left| \sum_{j=1}^{N} \mathbf{Y}_j - \sum_{j=1}^{m_k} \mathbf{Y}_j \right| = O(m_{k+1} - m_k) = O(k^{1/2}) = O(m_k^{1/3}) = O(N^{1/3}).$$

This completes the proof of Lemma 5.

Lemma 6. *For $f \in \mathcal{F}$, any $N \ge 1$, $t \ge 1$ and $\|f\| \ge \frac{1}{5} N^{-5/18}$ we have*

$$\mathbb{P}\left\{ \Big| \sum_{k=1}^{N} f(S_k) \Big| \ge t \|f\|^{1/4} \sqrt{N} \right\} \ll \exp\left(-Ct\|f\|^{-7/20}\right) + t^{-2} \exp(-CN^{1/3}). \tag{23}$$

Remark. The constants $1/5, 5/18, 1/4, 7/20, 1/3$ in (23) are not sharp and the inequality could be easily improved. However, the present form of Lemma 6 will suffice for the chaining argument in Lemma 7.

Proof. Put

$$\psi(n) = \sup_{0 \le x \le 1} |\mathbb{P}(S_n \le x) - x|.$$

By Theorem 1 of [8] we have

$$\psi(n) \le Ce^{-\gamma n} \qquad (n \ge 1)$$

for some constant $\gamma > 0$. Divide the interval $[1, N]$ into subintervals $I_1, \dots, I_L$, with $L \sim N^{2/3}$, where each interval I_ν contains $\sim N^{1/3}$ terms. We set

$$\sum_{k=1}^{N} f(S_k) = \eta_1 + \cdots + \eta_L$$

where

$$\eta_\nu = \sum_{k \in I_\nu} f(S_k).$$

We deal with the sums $\sum \eta_{2j}$ and $\sum \eta_{2j+1}$ separately. Since there is a separation $\sim N^{1/3}$ between the even block sums η_{2j}, we can apply Lemma 1 to get

$$\eta_{2j} = \eta_{2j}^* + \eta_{2j}^{**}$$

where

$$\eta_{2j}^{*} = \sum_{k\in I_{2j}} f(S_k - \Delta_j)$$
$$\eta_{2j}^{**} = \sum_{k\in I_{2j}} (f(S_k) - f(S_k - \Delta_j)).$$

Here the Δ_j are r.v.'s with $|\Delta_j| \leq \psi(N^{1/3}) \leq C\exp(-\gamma N^{1/3})$ and the r.v.'s η_{2j}^{*} $j = 1, 2, \ldots$ are independent. Conditionally on Δ_j, the distribution of S_k in a term of η_{2j}^{**} is the same as the (unconditional) distribution of an $S_{k_1} + c$ with $k_1 < k$ and a constant c and thus by Lemma 3, the L_2 norm of each summand in η_{2j}^{**} is $\leq C\psi^{1/2}(N^{1/3}) \leq C\exp(-\gamma N^{1/3})$ and thus for $\|f\| \geq N^{-1}$ we have

$$\begin{aligned}\|\eta_{2j}^{**}\| &\leq CN\exp(-\gamma N^{1/3}) \leq C\|f\|N^2\exp(-\gamma N^{1/3})\\ &\leq C\|f\|\exp(-\gamma' N^{1/3}).\end{aligned} \tag{24}$$

Thus

$$\Big\|\sum \eta_{2j}^{**}\Big\| \leq C\|f\|\exp(-\gamma'' N^{1/3})$$

and therefore by the Markov inequality

$$\begin{aligned}&\mathbb{P}\left(\Big|\sum \eta_{2j}^{**}\Big| \geq t\|f\|^{1/4}\sqrt{N}\right)\\ &\leq Ct^{-2}\|f\|^{-1/2}N^{-1}\|f\|^2\exp(-2\gamma'' N^{1/3}) \leq Ct^{-2}\exp(-2\gamma'' N^{1/3}).\end{aligned}$$

Let now $|\lambda| \leq dN^{-1/3}$ with a sufficiently small constant $d > 0$. Then $|\lambda\eta_{2j}^{*}| \leq 1/2$ for all N and thus using $e^x \leq 1 + x + x^2$ for $|x| \leq 1/2$ we get, using $E\eta_{2j}^{*} = 0$ for $j \geq 2$,

$$\begin{aligned}\mathbb{E}\left(\exp\lambda\left(\sum_j \eta_{2j}^{*}\right)\right) &= \prod_j \mathbb{E}\left(e^{\lambda\eta_{2j}^{*}}\right) \leq \prod_j \mathbb{E}(1 + \lambda\eta_{2j}^{*} + \lambda^2\eta_{2j}^{*2})\\ &= \prod_j (1 + \lambda^2\mathbb{E}\eta_{2j}^{*2}) \leq \exp\left(\lambda^2\sum_j \mathbb{E}\eta_{2j}^{*2}\right).\end{aligned} \tag{25}$$

Here, and in the rest of the proof of the lemma, the sums and products are extended for $j \geq 2$. By Lemma 4

$$\|\eta_{2j}\| \leq C\|f\|^{1/2}N^{1/6},$$

which, together with (24) and the Minkowski inequality, implies

$$\|\eta_{2j}^*\| \le C\|f\|^{1/2}N^{1/6}.$$

Thus the last expression in (25) cannot exceed

$$\exp\left(\lambda^2 C\|f\| \sum_j N^{1/3}\right) \le \exp(\lambda^2 C\|f\|N).$$

We choose now

$$\lambda = \frac{d}{2}N^{-1/2}\|f\|^{-3/5}$$

with the number d introduced before and note that by $\|f\| \ge \frac{1}{5}N^{-5/18}$ we have $|\lambda| \le dN^{-1/3}$. Thus using the Markov inequality, we get

$$\mathbb{P}\left\{\left|\sum_j \eta_{2j}^*\right| \ge t\|f\|^{1/4}\sqrt{N}\right\} \le 2\exp\left\{-\lambda t\|f\|^{1/4}\sqrt{N} + \lambda^2 C\|f\|N\right\}$$
$$= 2\exp\left(-\|f\|^{-7/20}t + C\|f\|^{-1/5}\right) \le 2\exp\left(-C\|f\|^{-7/20}t\right).$$

Recall that the sum here is extended for $j \ge 2$. However, the term corresponding to $j = 1$ is $O(N^{1/3})$ and since $\|f\|^{1/4}\sqrt{N} \ge N^{0.4}$ for $N \ge N_0$ by the assumptions of the lemma, the last chain of estimates remains valid by including the term $j = 1$ in the sum in the first probability and changing t to $2t$. A similar argument applies for the odd blocks η_{2j+1}^* (note that $\mathbb{E}\eta_1^*$ can be different from 0, but this causes no problem), completing the proof of Lemma 6.

Lemma 7. *For any* $N \ge 1$, $0 < \delta < 1$ *we have*

$$\mathbb{P}\left(\sup_{0\le a\le\delta}\left|\sum_{k=1}^N (I_{(0,a)}(S_k) - a)\right| \gg \delta^{1/8}\sqrt{N} + N^{4/9}\right) \ll \delta^4 + N^{-2}.$$

Proof. For any $h \ge 1$, $1 \le j \le 2^h$ let $\varphi_h^{(j)}$ denote the indicator function of the interval $[(j-1)2^{-h}, j2^{-h})$ and put

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

Clearly $\|\varphi_h^{(j)}\| = 2^{-h/2}$. We observe that if $0 \le a \le 1$ has the dyadic expansion

$$a = \sum_{j=1}^{\infty} \varepsilon_j 2^{-j} \qquad \varepsilon_j = 0, 1$$

and $H \ge 1$ is an arbitrary integer, then $g_a = I_{(0,a)}$ satisfies

$$\sum_{h=1}^{H} \varrho_h(x) \le g_a(x) \le \sum_{h=1}^{H} \varrho_h(x) + \sigma_H(x) \tag{26}$$

where ϱ_h is the indicator function of $\Big[\sum_{j=1}^{h-1} \varepsilon_j 2^{-j}, \sum_{j=1}^{h} \varepsilon_j 2^{-j}\Big)$ and σ_H is the indicator function of $\Big[\sum_{j=1}^{H} \varepsilon_j 2^{-j}, \sum_{j=1}^{H} \varepsilon_j 2^{-j} + 2^{-H}\Big)$. For $\varepsilon_h = 0$ clearly $\varrho_h \equiv 0$ and thus (26) 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 some of the $\varphi_H^{(j)}$. It follows that

$$g_a(x) - a \le \sum_{1 \le h \le H}^{*} (\varrho_h(x) - \varepsilon_h 2^{-h}) + (\sigma_H(x) - 2^{-H}) + 2^{-H}$$

and

$$g_a(x) - a \ge \sum_{1 \le h \le H}^{*} (\varrho_h(x) - \varepsilon_h 2^{-h}) - 2^{-H}$$

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

$$\sum_{k \le N} (g_a(S_k) - a) \le \sum_{k \le N} \sum_{1 \le h \le H}^{*} (\varrho_h(S_k) - \varepsilon_h 2^{-h}) + \sum_{k \le N} (\sigma_H(S_k) - 2^{-H}) + N 2^{-H}$$

and

$$\sum_{k \le N} (g_a(S_k) - a) \ge \sum_{k \le N} \sum_{1 \le h \le H}^{*} (\varrho_h(S_k) - \varepsilon_h 2^{-h}) - N 2^{-H}.$$

Hence it follows that for any $N \ge 1$, $H \ge 1$ there exist suitable integers $1 \le j_h \le 2^h$, $1 \le h \le H$ such that

$$\left| \sum_{k \le N} (g_a(S_k) - a) \right| \le 2 \sum_{h \le H} \left| \sum_{k \le N} \varphi_h^{(j_h)}(S_k) - 2^{-h} \right| + N 2^{-H}$$

$$= 2 \sum_{h \le H} F(N, j_h, h) + N 2^{-H}. \tag{27}$$

Introduce the events

$$G(N,j,h) = \left\{ F(N,j,h) \ge 2^{-h/8}\sqrt{N} \right\},$$
$$G_N = \bigcup_{A \le h \le B \log_2 N} \bigcup_{j \le 2^h} G(N,j,h)$$

with $A = \log_2 \frac{1}{a}$, $B = \frac{5}{9}$. For $h \le B \log_2 N$ we have $\|\varphi_j^{(h)} - 2^{-h}\| \ge 2^{-h/2} - 2^{-h} \ge 2^{-h/2}(1 - 1/\sqrt{2}) \ge \frac{1}{5} N^{-5/18}$ and thus applying (23) with $t = 1$, we get

$$\mathbb{P}(G(N,h,j)) \ll \exp(-C 2^{7h/40}) + N^{-3}$$

and consequently

$$\mathbb{P}(G_N) \ll \sum_{h=A}^{\infty} 2^h \exp(-C 2^{7h/40}) + N^{-3} \sum_{h \le B \log_2 N} 2^h. \tag{28}$$

Clearly, the second term on the right hand side of (28) is $\ll N^{-2}$. On the other hand, the terms of the first sum in (28) decrease superexponentially and thus the sum can be bounded by a constant times its first term, i.e. the sum is

$$\ll 2^A \exp(-C 2^{7A/40}) \ll 2^{-4A} \ll a^4.$$

Hence

$$\mathbb{P}(G_N) \ll a^4 + N^{-2}.$$

Note that when breaking the interval $(0, a)$ into dyadic intervals of length 2^{-h} we automatically have $h \ge \log \frac{1}{a} = A$ and thus choosing

$$H = [B \log_2 N],$$

it follows that with the exception of a set with probability $\ll a^4 + N^{-2}$, for any $0 < a \le \delta$ the expression in the second line of (27) is

$$\ll \sum_{A \le h \le H} 2^{-h/8}\sqrt{N} + N 2^{-H} \ll 2^{-A/8}\sqrt{N} + N^{4/9} \ll a^{1/8}\sqrt{N} + N^{4/9}$$
$$\ll \delta^{1/8}\sqrt{N} + N^{4/9}.$$

This proves Lemma 7.

Lemma 5 implies the convergence of the finite dimensional distributions of the sequence $\sqrt{N}(F_N(t)-t)$ in (5) to those of K and to prove Theorem 2 it remains to prove the tightness of the sequence in $D[0,1]$. To this end, fix $\varepsilon > 0$ and choose h so that $2^{-h} \le \varepsilon < 2^{-(h-1)}$. Note that for $j = 0, 1, \dots, 2^h - 1$ we have

$$\mathbb{P}\left(\sup_{0\le a\le 2^{-h}}\left|\sum_{k=1}^{N}(f_{j2^{-h}+a}(S_k) - f_{j2^{-h}}(S_k))\right| \gg 2^{-h/8}\sqrt{N} + N^{4/9}\right) \ll 2^{-4h} + N^{-2}. \tag{29}$$

For $j = 0$ relation (29) is identical with Lemma 7 and for $j = 1, 2, \dots$ the proof is the same. It follows that

$$\mathbb{P}\left(\max_{0\le j\le 2^h-1}\sup_{0\le a\le 2^{-h}}\left|\sum_{k=1}^{N}(f_{j2^{-h}+a}(S_k) - f_{j2^{-h}}(S_k))\right| \gg 2^{-h/8}\sqrt{N} + N^{4/9}\right) \ll 2^{-3h} + 2^h N^{-2}. \tag{30}$$

Then (30) implies that with the exception of a set with probability

$$\ll 2^{-3h} + 2^h N^{-2} \ll \varepsilon^3 + N^{-2}\varepsilon^{-1}$$

the fluctuation of the process $\sqrt{N}(F_N(t) - t)$ over any subinterval of $(0,1)$ with length $\le \varepsilon$ is

$$\ll \varepsilon^{1/8} + N^{-1/18}.$$

By Theorem 15.5 of Billingsley [4, Chapter 3], the sequence $\sqrt{N}(F_N(t) - t)$ is tight in $D[0,1]$. This completes the proof of Theorem 2; Theorem 1 follows immediately from Theorem 2.

References

[1] Berkes, I. and Borda, B.: On the discrepancy of random subsequences of $\{n\alpha\}$. Acta Arithmetica, to appear.

[2] Berkes, I. and Raseta, M.: On the discrepancy and empirical distribution function of $\{n_k\alpha\}$. Unif. Distr. Theory 10 (2015), 1–17.

[3] 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.

[4] Billingsley, P.: Convergence of probability measures. Wiley, 1968.

[5] Donsker, M.: Justification and extension of Doob's heuristic approach to the Kolmogorov- Smirnov theorems, Ann. Math. Statist. 23 (1952), 277–281.

[6] Lévy, P.: L'addition des variables aléatoires définies sur une circonference. Bull. Soc. Math. France 67 (1939), 1–40.

[7] Philipp, W.: Limit theorems for lacunary series and uniform distribution mod 1. Acta Arith. 26 (1975), 241–251.

[8] Schatte, P.: On the asymptotic uniform distribution of sums reduced mod 1. Math. Nachr. 115 (1984), 275–281.

[9] 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.