Berry–Esseen bounds and Diophantine approximation

István Berkes* and Bence Borda†

Dedicated to the memory of Jean-Pierre Kahane

Abstract

Let $S_N, N = 1, 2, \dots$ be a random walk on the integers, let α be an irrational number and let $Z_N = \{S_N \alpha\}$, where $\{\cdot\}$ denotes fractional part. Then $Z_N, N = 1, 2, \dots$ is a random walk on the circle, and from classical results of probability theory it follows that the distribution of Z_N converges weakly to the uniform distribution. We determine the precise speed of convergence, which, in addition to the distribution of the elementary step X of the random walk S_N, depends sensitively on the rational approximation properties of α.

1 Introduction

Let $X_1, X_2, \dots$ be i.i.d. integer valued random variables and $S_N = \sum_{n=1}^N X_n$. Assume X_1 is nondegenerate, that is, there does not exist a constant c such that $\mathbb{P}(X_1 = c) = 1$. Let α be an irrational number and put $Z_N = \{S_N \alpha\}$, where $\{\cdot\}$ denotes fractional part. Then $Z_N, N = 1, 2, \dots$ is a random walk on the circle and from classical results of probability theory (see e.g. [8]) it follows that the distribution of Z_N converges weakly to $U(0,1)$, the uniform distribution on $(0,1)$. The speed of convergence in $Z_N \overset{d}{\to} U(0,1)$, i.e. the order of magnitude of the quantity

$$\Delta_N := \sup_{0 \le x \le 1} |\mathbb{P}(\{S_N \alpha\} < x) - x|$$

was first investigated by Schatte [15]. It is easy to see that Δ_N depends sensitively on the Diophantine approximation properties of α. Indeed, if α is very close to a rational number p/q, then as long as $|S_N|$ is small, $S_N \alpha$ is close to an integer multiple of $1/q$ and thus the distribution of $\{S_N \alpha\}$ is markedly different from $U(0,1)$. By a standard definition (see e.g. [7, p. 121]), the *type* γ of an irrational number α is the supremum of all c such that

$$\liminf_{q \to \infty} q^c \|q\alpha\| = 0,$$

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

†A. Rényi Institute of Mathematics, 1053 Budapest, Reáltanoda u. 13-15, Hungary. e-mail: bordabence85@gmail.com.

where $\|t\|$ denotes the distance of a real number t from the nearest integer. Schatte [15] proved that if $\mathbb{E}|X_1|^3 < \infty$ and α is of finite type $\gamma > 1$, then

$$\Delta_N = O(N^{-1/(2\gamma)+\varepsilon}), \quad \Delta_N = \Omega(N^{-1/(2\gamma)-\varepsilon}) \tag{1.1}$$

holds for any $\varepsilon > 0$. Note that for two sequences $a_N \in \mathbb{R}$ and $b_N > 0$ the notation $a_N = \Omega(b_N)$ means that $\limsup_{N\to\infty} |a_N|/b_N > 0$.

The purpose of the present paper is to give a sharp estimate of Δ_N for a large class of i.i.d. integer valued sequences (X_n) and irrational numbers α. Our results will cover all (X_n) with $\mathbb{E}X_1^2 < \infty$ and also a large class of heavy-tailed random variables X_1 with $\mathbb{P}(|X_1| > x)$ having order of magnitude $x^{-\beta}$ with some $0 < \beta < 2$. Concerning α, we will assume that

$$0 < \liminf_{q\to\infty} q^\gamma \|q\alpha\| < \infty \tag{1.2}$$

for some $\gamma \geq 1$. If (1.2) holds, we will say that α has *strong type* γ. Note the difference between ordinary and strong type: relation (1.2) means that for a sufficiently large constant C the approximation

$$\left|\alpha - \frac{p}{q}\right| < \frac{C}{q^{\gamma+1}}$$

holds for infinitely many fractions p/q, while for a sufficiently small C it holds only for finitely many p/q. In contrast, if the (ordinary) type of α is γ, we only know that the approximation

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{\gamma+1+\varepsilon}}$$

holds for infinitely many fractions p/q if $\varepsilon < 0$ and finitely many fractions p/q if $\varepsilon > 0$. For example, almost all irrational α (in the Lebesgue sense) have type 1, while α has strong type 1 if and only if the continued fraction of α has bounded partial quotients. Such numbers are called *badly approximable.*

For the class of irrational α of a given type, estimates of Δ_N that are sharp up to a factor of N^ε, as in (1.1), are thus best possible. The first estimate of Δ_N sharp up to logarithmic factors is also due to Schatte [15]: if $\mathbb{E}|X_1|^3 < \infty$ and α is badly approximable, then

$$\Delta_N = O(N^{-1/2}\log N), \quad \Delta_N = \Omega(N^{-1/2}\log^{-1/2} N).$$

Using elaborate arithmetic and combinatorial tools, Su [16] proved that if $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ and α is a quadratically irrational number, then

$$C_1 N^{-1/2} \leq \Delta_N \leq C_2 N^{-1/2} \tag{1.3}$$

with some positive constants $C_1, C_2 > 0$, yielding the exact order of magnitude of Δ_N. According to the theorem of Lagrange, α is quadratically irrational if and only if the partial quotients in the continued fraction of α are eventually periodic. In particular, quadratically irrational numbers are badly approximable. The method used by Su relies heavily on this periodicity, and thus is not applicable to all badly approximable numbers.

We now formulate our results. The main message of our first theorem is that (1.3) holds under much more general circumstances. In particular, it is enough to assume the boundedness instead of the periodicity of the partial quotients in the continued fraction of α.

Theorem 1.1. *Let $X_1, X_2, \dots$ be i.i.d. integer valued, nondegenerate random variables with $\mathbb{E}X_1^2 < \infty$, and let $S_N = \sum_{n=1}^N X_n$. If α is badly approximable, then*

$$C_1 N^{-1/2} \le \Delta_N \le C_2 N^{-1/2} \tag{1.4}$$

for every $N \in \mathbb{N}$ with some constants $C_1, C_2 > 0$ depending only on α and the distribution of X_1.

As we will see, the upper bound in (1.4) remains valid assuming only that X_1 is a nondegenerate random variable.

Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}X_1 = 0, \mathbb{E}X_1^2 = 1$ and $\mathbb{E}|X_1|^{2+\delta} < \infty$ for some $0 < \delta \le 1$ and let $S_N = \sum_{n=1}^N X_n$. By the classical Berry–Esseen estimate (see e.g. [12], p. 151) we have

$$\widetilde{\Delta}_N := \sup_{x \in \mathbb{R}} |\mathbb{P}(S_N/\sqrt{N} < x) - \Phi(x)| = O(N^{-\delta/2})$$

where $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x e^{-t^2/2} dt$ is the standard normal distribution function. The remainder term here cannot be improved in general. Thus we see that while in the case of ordinary i.i.d. sums we need finite third moments for the convergence speed $O(N^{-1/2})$ in the CLT, in the case of mod 1 sums the nondegeneracy of X_1 suffices to this purpose.

We now turn to the case of an irrational α of strong type $\gamma > 1$, when we need some additional technical assumptions on X_1. For an integer valued random variable Y let $\operatorname{supp} Y = \{k \in \mathbb{Z} : \mathbb{P}(Y = k) > 0\}$ denote the support of (the distribution of) Y.

Theorem 1.2. *Let $X_1, X_2, \dots$ be i.i.d. integer valued, nondegenerate random variables with $\mathbb{E}X_1^2 < \infty$, and let $S_N = \sum_{n=1}^N X_n$. Suppose that $\operatorname{supp} X_1$ is a (finite or infinite) arithmetic progression, and that there exists a constant $K > 0$ such that for any large enough $N \in \mathbb{N}$ the sequence $\mathbb{P}(S_N = k)$, $k \in \operatorname{supp} S_N$ is nonincreasing for $k > \mathbb{E}S_N + K\sqrt{N}$ and nondecreasing for $k < \mathbb{E}S_N - K\sqrt{N}$. If α is of strong type $\gamma > 1$, then*

$$\Delta_N = O(N^{-1/(2\gamma)}), \qquad \Delta_N = \Omega(N^{-1/(2\gamma)})$$

with implied constants depending only on α and the distribution of X_1.

Again, the upper bound for Δ_N is valid assuming only that X_1 is nondegenerate. The monotonicity assumption on the sequence $\mathbb{P}(S_N = k)$, $k \in \operatorname{supp} S_N$ is particularly simple to check if S_N has a *unimodal* distribution, that is, $\mathbb{P}(S_N = k)$, $k \in \operatorname{supp} S_N$ is nondecreasing for some $k < k^*$ and nonincreasing for $k > k^*$. For example, if $\operatorname{supp} X_1$ has cardinality 2, then S_N has a binomial, hence unimodal distribution. Verifying a conjecture of Brockett and Kemperman [2], Odlyzko and Richmond [10] proved that if the support of X_1 is the set $\{0, 1, \dots, d\}$ for some $d \ge 1$, then the distribution of S_N is unimodal for $N \ge N_0$.

In the previous two theorems we assumed that X_1 has a finite variance. Let us now consider a random variable X_1 with a "heavy-tailed" distribution, that is, with $\mathbb{E}X_1^2 = \infty$. For the sake of simplicity we will assume that the tail distribution of $|X_1|$ is a power function, namely

$$\mathbb{P}(|X_1| \ge x) \sim c x^{-\beta} \quad \text{as } x \to \infty \tag{1.5}$$

with some constants $c > 0$ and $0 < \beta < 2$. By classical results of probability theory (see e.g. [4], Chapter XVII.5), relation (1.5) and the additional assumption

$$\lim_{x\to\infty} \mathbb{P}(X_1 \geq x)/\mathbb{P}(|X_1| \geq x) \quad \text{exists} \tag{1.6}$$

imply that that for a suitable centering factor a_N we have

$$(S_N - a_N)/N^{1/\beta} \xrightarrow{d} G_\beta \tag{1.7}$$

where G_β is a stable law with index β. Moreover, (1.5) and (1.6) together are also necessary for (1.7). We also note that for $0 < \beta < 1$ we can choose $a_N = 0$ and for $1 < \beta < 2$ (in which case $\mathbb{E}X_1$ exists), we can choose $a_N = \mathbb{E}S_N = N\mathbb{E}X_1$. The case $\beta = 1$ is exceptional: for symmetric X_1 we can choose $a_N = 0$, but e.g. if $X_1 > 0$ and $\mathbb{P}(X_1 = k) = 6/(\pi^2 k^2)$ $(k = 1, 2, \ldots)$, then (1.7) holds with $a_N = \frac{6}{\pi^2} N \log N$.

We can now formulate the analogue of Theorem 1.1 for heavy-tailed distributions.

Theorem 1.3. *Let $X_1, X_2, \ldots$ be i.i.d. integer valued random variables and let $S_N = \sum_{n=1}^N X_n$. Suppose that (1.5) and (1.6) hold. If α is badly approximable, then*

$$C_1 N^{-1/\beta} \leq \Delta_N \leq C_2 N^{-1/\beta} \tag{1.8}$$

for every $N \in \mathbb{N}$ with some constants $C_1, C_2 > 0$ depending only on α and the distribution of X_1.

As we will see, for the upper bound in (1.8) we need only (1.5), but not (1.6). The proof of the lower bound will use essentially the limit relation (1.7) (and thus both of (1.5) and (1.6)), but the centering factor a_N in (1.7) does not appear in (1.8). We note also that by choosing β sufficiently close to 0, Δ_N will converge to 0 at an arbitrarily fast polynomial speed.

Finally, we give an analogue of Theorem 1.2 for heavy-tailed distributions.

Theorem 1.4. *Let $X_1, X_2, \ldots$ be i.i.d. integer valued random variables, let $S_N = \sum_{n=1}^N X_n$ and assume that (1.7) holds with some centering factor a_N. Suppose, moreover, that* $\operatorname{supp} X_1$ *is an arithmetic progression, and that there exists a constant $K > 0$ such that for any large enough $N \in \mathbb{N}$ the sequence $\mathbb{P}(S_N = k)$, $k \in \operatorname{supp} S_N$ is nonincreasing for $k > a_N + KN^{1/\beta}$ and nondecreasing for $k < a_N - KN^{1/\beta}$. If α is of strong type $\gamma > 1$, then*

$$\Delta_N = O(N^{-1/(\beta\gamma)}), \qquad \Delta_N = \Omega(N^{-1/(\beta\gamma)}) \tag{1.9}$$

with implied constants depending only on α and the distribution of X_1.

As in the case of Theorem 1.3, the upper bound in (1.9) is valid under assuming only (1.5), while the proof of the lower bound will make an essential use of (1.7), i.e. both (1.5) and (1.6).

It is worth comparing Theorems 1.3, 1.4 with the corresponding classical results for the speed of convergence of centered and normed sums of i.i.d. random variables to a stable law. Assume (1.7), let F denote the distribution function of X_1 and

$$\Delta_N^* = \sup_{x\in\mathbb{R}} |\mathbb{P}((S_N - a_N)/N^{1/\beta} < x) - G_\beta(x)|. \tag{1.10}$$

Satybaldina [13], [14] proved that under the additional assumption

$$\int_{\mathbb{R}} |x|^{\lfloor\beta\rfloor}|F(x) - G_\beta(x)|dx < \infty \tag{1.11}$$

where $\lfloor\beta\rfloor$ denotes the greatest integer smaller or equal to β, we have

$$\Delta_N^* = \begin{cases} O(N^{-(2/\beta-1)}) & \text{if} \quad 1 \le \beta < 2 \\ O(N^{-(1/\beta-1)}) & \text{if} \quad 0 < \beta < 1. \end{cases} \tag{1.12}$$

Hall [5] proved that without the assumption (1.11) these estimates are generally not valid and under some monotonicity assumptions for the distribution of X_1 he gave necessary and sufficient conditions for weaker polynomial estimates of Δ_N^*. For remainder term estimates for independent, not identically distributed random variables X_k we refer to Paulauskas [11] and the references therein. Just as in the case of mod 1 sums, choosing β sufficiently close to 0, Δ_N^* will converge to 0 at an arbitrarily fast polynomial speed.

If in the definition of Δ_N we replace the distribution of $\{S_N\alpha\}$ with the corresponding empirical measure, i.e. $N^{-1}\sum_{n=1}^N \delta_{\{S_n\alpha\}}$, where δ_x denotes the probability measure concentrated at x, then Δ_N becomes the star discrepancy D_N^* of the first N terms of the sequence $\{S_n\alpha\}$, i.e.

$$D_N^* := \sup_{0\le x\le 1} \left| \frac{1}{N}\sum_{n=1}^N \left(I_{[0,x)}(\{S_n\alpha\}) - x\right)\right|$$

where $I_{[0,x)}$ is the indicator function of the interval $[0,x)$. The discrepancy D_N of the first N terms of the sequence $\{S_n\alpha\}$ is defined by taking the supremum over all subintervals $[x,y) \subset [0,1]$, i.e.

$$D_N := \sup_{0\le x<y\le 1} \left| \frac{1}{N}\sum_{n=1}^N \left(I_{[x,y)}(\{S_n\alpha\}) - (y-x)\right)\right|.$$

These two quantities also provide a natural measure of the distance of the distribution of the sequence $\{S_n\alpha\}$ from the uniform distribution, and are widely used in analysis and number theory. Note that D_N^* and D_N are random variables. In [1] we gave estimates of D_N for the same class of random walks S_N and irrational α as in the present paper. Estimating D_N, however, is considerably harder than estimating Δ_N since instead of using Fainleib's inequality employed below, we need the Erdős–Turán inequality leading to the estimation of exponential sums and rather hard combinatorics. As a consequence, the results in [1] are slightly less precise than those in the present paper and are also of a different character.

2 Upper estimates

In this section we prove the upper estimates in Theorems 1.1–1.4 in a somewhat stronger form. The proof will be based on the Fainleib inequality (see e.g. [3], [9]) which states that for any $H \in \mathbb{N}$ we have

$$\Delta_N \le \frac{4}{H} + \frac{4}{\pi}\sum_{h=1}^H \frac{|\varphi(2\pi h\alpha)|^N}{h} \tag{2.1}$$

where φ denotes the characteristic function of X_1. Note that the Fainleib inequality is basically an Erdős–Turán-type inequality for Δ_N instead of the discrepancy D_N. It is thus natural to prove upper estimates for Δ_N under certain conditions for φ.

Proposition 2.1. *Let $X_1, X_2, \dots$ be i.i.d. random variables and let $S_N = \sum_{n=1}^N X_n$. Suppose that there exist real constants $0 < \beta \leq 2$, $c > 0$, and an integer $d > 0$ such that $|\varphi(2\pi x)| \leq 1 - c\|dx\|^\beta$ for any $x \in \mathbb{R}$. If an irrational α satisfies $\|q\alpha\| \geq Cq^{-\gamma}$ for every $q \in \mathbb{N}$ with some constants $C > 0$ and $\gamma \geq 1$, then $\Delta_N = O\left(N^{-1/(\beta\gamma)}\right)$ with an implied constant depending only on α and the distribution of X_1.*

Note that if X_1 is integer valued and nondegenerate, then its characteristic function φ satisfies the conditions of Proposition 2.1 with $\beta = 2$, a suitable $c > 0$ and with $d > 0$ denoting the greatest common divisor of supp $(X_1 - X_2)$. Furthermore, if there exist constants $K, x_0 > 0$ such that

$$\mathbb{E}(X_1^2 I_{\{|X_1| \leq x\}}) \geq K x^{2-\beta} \quad \text{for } x \geq x_0, \tag{2.2}$$

then the conditions of Proposition 2.1 are satisfied with the same β, $d > 0$ denoting the greatest common divisor of supp $(X_1 - X_2)$ and some $c > 0$. For a proof of these simple facts see e.g. [1, Proposition 3.2]. This shows that the upper bounds in Theorems 1.1, 1.2 remain valid valid under the sole assumption that X_1 is nondegenerate. Note also that (2.2) follows from (1.5) by integration by parts and thus the upper estimates in Theorems 1.3, 1.4 are valid assuming only (1.5).

Proof. Let us apply the Fainleib inequality (2.1) with $H = [N^{1/(\beta\gamma)}]$. Using the estimate

$$|\varphi(2\pi h\alpha)|^N \leq \left(1 - c\|hd\alpha\|^\beta\right)^N \leq e^{-c\|hd\alpha\|^\beta N},$$

it will thus be enough to prove

$$\sum_{h=1}^{[N^{1/(\beta\gamma)}]} \frac{e^{-c\|hd\alpha\|^\beta N}}{h} = O\left(N^{-1/(\beta\gamma)}\right). \tag{2.3}$$

We wish to use summation by parts in (2.3). To this end, let $s_h = \sum_{j=1}^h e^{-c\|jd\alpha\|^\beta N}$ for any $1 \leq h \leq [N^{1/(\beta\gamma)}]$. Let $K = (hd)^\gamma / C$ (where C is the constant in the Proposition) and let $a_j \in (-1/2, 1/2]$ be the unique number equivalent to $jd\alpha$ mod 1. On the one hand, since $\|jd\alpha\| \geq C(hd)^{-\gamma}$, we have $a_j \notin (-1/K, 1/K)$ for every $1 \leq j \leq h$. On the other hand, for any $1 \leq j, j' \leq h$, $j \neq j'$ we have

$$|a_j - a_{j'}| \geq \|(j - j')d\alpha\| \geq C(hd)^{-\gamma} = 1/K,$$

and thus each interval of the form $[k/K, (k+1)/K)$ or $(-(k+1)/K, -k/K]$, $k = 1, 2, \dots$ contains a_j for at most one index j. Therefore

$$s_h \leq 2 \sum_{k=1}^\infty e^{-c(k/K)^\beta N} = 2 \sum_{k=1}^\infty e^{-aNk^\beta / h^{\beta\gamma}}$$

with a constant $a = cC^\beta / d^{\beta\gamma}$. Note that here the $k = 1$ term dominates. Indeed, using the fact that $N/h^{\beta\gamma} \geq 1$ we can further estimate s_h as

$$s_h \leq 2e^{-aN/h^{\beta\gamma}} \sum_{k=1}^\infty e^{-aN(k^\beta - 1)/h^{\beta\gamma}} \leq 2e^{-aN/h^{\beta\gamma}} \sum_{k=1}^\infty e^{-a(k^\beta - 1)}.$$

The value of this convergent series depends only on a and β, hence $s_h = O(e^{-aN/h^{\beta\gamma}})$. Applying summation by parts to the left hand side of (2.3) we thus obtain

$$\sum_{h=1}^{[N^{1/(\beta\gamma)}]} \frac{e^{-c\|hd\alpha\|^\beta N}}{h} = \sum_{h=1}^{[N^{1/(\beta\gamma)}]-1} \frac{s_h}{h(h+1)} + \frac{s_{[N^{1/(\beta\gamma)}]}}{[N^{1/(\beta\gamma)}]}$$
$$= O\left(\sum_{h=1}^{\infty} \frac{e^{-aN/h^{\beta\gamma}}}{h^2} + N^{-1/(\beta\gamma)} \right).$$

By checking that the terms in this series are increasing on $1 \le h \le (a\beta\gamma N/2)^{1/(\beta\gamma)}$ we finally get

$$\sum_{h=1}^{[N^{1/(\beta\gamma)}]} \frac{e^{-c\|hd\alpha\|^\beta N}}{h} = O\left(N^{1/(\beta\gamma)} \frac{e^{-2/(\beta\gamma)}}{N^{2/(\beta\gamma)}} + \sum_{h > (a\beta\gamma N/2)^{1/(\beta\gamma)}} \frac{1}{h^2} + N^{-1/(\beta\gamma)} \right)$$
$$= O(N^{-1/(\beta\gamma)}).$$

□

3 Lower estimates

In this section we prove the lower estimates in Theorems 1.1–1.4. First, note that the lower estimates in Theorems 1.1 and 1.3 follow easily from the local limit theorem [6, Theorem 4.2.1] for i.i.d. sums. Indeed, in Theorem 1.1 $(S_N - \mathbb{E}S_N)/\sqrt{N}$ converges weakly to a normal law and under the conditions of Theorem 1.3 we have (1.7) with a suitable centering factor a_N. By [6, Theorem 4.2.1], for a suitable integer k we have $\mathbb{P}(S_N = k) \ge C_1/\sqrt{N}$ and $\mathbb{P}(S_N = k) \ge C_1/N^{1/\beta}$, respectively, with some constant $C_1 > 0$ depending only on the distribution of X_1. Hence the distribution of $\{S_N \alpha\}$ has an atom with weight at least $C_1/\sqrt{N}$ resp. $C_1/N^{1/\beta}$, so by the continuity of the uniform distribution we have $\Delta_N \ge C_1/(2\sqrt{N})$ and $\Delta_N \ge C_1/(2N^{1/\beta})$, respectively, for $N \ge N_0$. Note that in particular the lower estimates in Theorems 1.1 and 1.3 hold for any irrational α regardless of its Diophantine character, with a constant $C_1 > 0$ independent of α.

The lower estimates in Theorems 1.2 and 1.4 are deduced in the following common form.

Proposition 3.1. *Let $X_1, X_2, \ldots$ be i.i.d. integer valued random variables and let $S_N = \sum_{n=1}^N X_n$. Let $0 < \beta \le 2$, and suppose that there exists a sequence $E_N \in \mathbb{R}$ for which $\mathbb{P}(|S_N - E_N| \ge tN^{1/\beta}) \to 0$ uniformly in N as $t \to \infty$. Suppose, moreover, that $\operatorname{supp} X_1$ is a (finite or infinite) arithmetic progression, and that there exists a constant $K > 0$ such that for any large enough $N \in \mathbb{N}$ the sequence $\mathbb{P}(S_N = k)$, $k \in \operatorname{supp} S_N$ is nonincreasing on $k > E_N + KN^{1/\beta}$ and nondecreasing on $k < E_N - KN^{1/\beta}$. If α is of strong type $\gamma > 1$, then $\Delta_N = \Omega(N^{-1/(\beta\gamma)})$ with an implied constant depending only on α and the distribution of X_1.*

If $\mathbb{E}X_1^2 < \infty$, then $\mathbb{P}(|S_N - \mathbb{E}S_N| \ge t\sqrt{N}) \to 0$ uniformly in N as $t \to \infty$ because of the Chebyshev inequality. The lower estimate in Theorem 1.2 thus follows from Proposition 3.1 with $\beta = 2$. Under the conditions of Theorem 1.4 there exists a sequence $E_N \in \mathbb{R}$ such that $(S_N - E_N)/N^{1/\beta}$ converges to a stable distribution of index β which implies that

$$\mathbb{P}(|S_N - E_N| \geq tN^{1/\beta}) \to 0 \text{ uniformly in } N \text{ as } t \to \infty. \tag{3.1}$$

Thus the lower estimate in Theorem 1.4 also follows.

Proof. We may assume that X_1 is nondegenerate, otherwise the claim is trivial. Let $d > 0$ be the difference of $\operatorname{supp} X_1$, that is, $\operatorname{supp} X_1 = \{d_0 + kd : k \in I\}$ with some integer d_0 and interval I of integers of the form $I = [0, i]$, $I = (-\infty, 0]$, $I = [0, \infty)$ or $I = \mathbb{Z}$. Note that $\operatorname{supp} S_N$ is also an arithmetic progression with difference $d > 0$.

Let $\varepsilon > 0$ be an arbitrary number, to be chosen later. We claim that there exist constants $N_0 > 0$ and $a > 0$ depending only on ε and the distribution of X_1 such that for any $N \geq N_0$ and any $k \in \operatorname{supp} S_N$, $|k - E_N| \geq aN^{1/\beta}$ we have $\mathbb{P}(S_N = k) < \varepsilon/|k - E_N|$. Indeed, using the monotonicity assumption, for any $k \in \operatorname{supp} S_N$, $k - E_N > 2KN^{1/\beta}$ we have

$$\begin{aligned} \left\lfloor \frac{k - E_N}{2d} \right\rfloor \mathbb{P}(S_N = k) &\leq \sum_{\ell=1}^{\lfloor (k-E_N)/(2d) \rfloor} \mathbb{P}(S_N = k - d\ell) \\ &\leq \mathbb{P}\left(S_N - E_N \geq \frac{k - E_N}{2} \right) \to 0 \end{aligned}$$

when $(k - E_N)/N^{1/\beta} \to \infty$. A similar estimate holds for $k - E_N < -2KN^{1/\beta}$. The existence of $N_0 > 0$ and $a > 0$ as in the claim clearly follow.

The definition (1.2) of strong type implies the existence of a constant $C > 0$ depending only on α such that $\|q\alpha\| < Cq^{-\gamma}$ for infinitely many $q \in \mathbb{N}$. For every such q let $N = \lfloor q^{\beta\gamma}/b \rfloor$, where $b > 0$ is a large constant to be chosen later, depending on α, the distribution of X_1, $\varepsilon > 0$ and $a > 0$ from the previous claim. We may assume $N \geq N_0$.

Let $f(x) = \mathbb{P}(\{S_N\alpha\} < x) - x$. By considering all possible values $k \in \operatorname{supp} S_N$

$$f(x) = \sum_{k \in \operatorname{supp} S_N} \mathbb{P}(S_N = k) \left(I_{[0,x)}(\{k\alpha\}) - x \right). \tag{3.2}$$

Let p denote the integer closest to $q\alpha$. For any $k \in \operatorname{supp} S_N$, $|k - E_N| < q^\gamma/(3C)$ we have

$$\left| k\alpha - E_N \left(\alpha - \frac{p}{q} \right) - \frac{kp}{q} \right| = |k - E_N| \cdot \left| \alpha - \frac{p}{q} \right| < \frac{q^\gamma}{3C} \cdot \frac{Cq^{-\gamma}}{q} = \frac{1}{3q}.$$

This means that the distance of $k\alpha$ from the set $E_N(\alpha - p/q) + (1/q)\mathbb{Z}$ is less than $1/(3q)$, in other words, $k\alpha$ does not fall into the middle third interval between any two consecutive points of the arithmetic progression $E_N(\alpha - p/q) + (1/q)\mathbb{Z}$. Consider such a middle third interval in $[0, 1]$. More precisely, let $J = [u, v) \subseteq [0, 1]$ be an interval of length $1/(3q)$ such that $u \in E_N(\alpha - p/q) + 1/(3q) + (1/q)\mathbb{Z}$. Then for any $k \in \operatorname{supp} S_N$, $|k - E_N| < q^\gamma/(3C)$ we have $\{k\alpha\} \notin J$. Therefore, using (3.2) we can write $f(v) - f(u)$ in the form

$$f(v) - f(u) = \sum_{k \in \operatorname{supp} S_N} \mathbb{P}(S_N = k) \left(I_J(\{k\alpha\}) - \frac{1}{3q} \right)$$
$$= \mathbb{P}\left(|S_N - E_N| < \frac{q^\gamma}{3C} \right) \frac{-1}{3q} + \sum_{\substack{k \in \operatorname{supp} S_N \\ |k - E_N| \geq q^\gamma/(3C)}} \mathbb{P}(S_N = k) \left(I_J(\{k\alpha\}) - \frac{1}{3q} \right). \quad (3.3)$$

By choosing $b > 0$ large enough we can ensure that the probability in the first term in (3.3) is at least $1/2$ (see (3.1)), and hence the term itself is at most $-1/(6q)$. To prove the proposition it will therefore be enough to show that the second term in (3.3) is less than or equal to $1/(12q)$. Indeed, this would imply

$$\sup_{0 \leq x \leq 1} |f(x)| \geq \frac{|f(v) - f(u)|}{2} \geq \frac{1}{24q} = \Omega(N^{-1/(\beta\gamma)}).$$

We will only estimate the terms $k \in \operatorname{supp} S_N$, $k - E_N \geq q^\gamma/(3C)$ in the second term of (3.3). The proof for $k - E_N \leq -q^\gamma/(3C)$ is analogous. Let k_0 be the largest integer in $\operatorname{supp} S_N$ such that $k_0 - E_N < q^\gamma/(3C)$. (Note we may have $k_0 < 0$.) Since $\operatorname{supp} S_N$ is an arithmetic progression with difference d, we wish to estimate

$$M := \sum_{\substack{k > k_0 \\ k \equiv k_0 \pmod d}} \mathbb{P}(S_N = k) \left(I_J(\{k\alpha\}) - \frac{1}{3q} \right).$$

We will use summation by parts to estimate M. To this end, for any $k > k_0$, $k \equiv k_0 \pmod d$ let

$$A_k = \sum_{\substack{k_0 < \ell \leq k \\ \ell \equiv k_0 \pmod d}} \left(I_J(\{\ell\alpha\}) - \frac{1}{3q} \right).$$

By the definition of discrepancy, $|A_k|$ is at most $(k - k_0)/d$ times the discrepancy of the first $(k - k_0)/d$ terms of the sequence $\{nd\alpha + k_0\alpha\}$, $n = 1, 2, \ldots$. The translation by $k_0\alpha$ modulo 1 does not affect the discrepancy, and $d\alpha$ is also of strong type $\gamma > 1$. From classical estimates of the discrepancy of Kronecker sequences (see e.g. [7, Lemma 3.2 p. 122, Exercise 3.12 p. 131]) we thus have $|A_k| \leq B(k - k_0)^{1-1/\gamma}$ for some constant $B > 0$ depending only on α and the distribution of X_1 (in fact, the value of d).

By choosing $b > 0$ large enough, we can ensure $q^\gamma/(3C) > aN^{1/\beta}$. Then for every $k > k_0$ we have $\mathbb{P}(S_N = k) < \varepsilon/(k - E_N)$. In particular, $\mathbb{P}(S_N = k)|A_k| \to 0$ as $k \to \infty$, therefore we can apply summation by parts to the infinite series defining M to obtain

$$M = \sum_{\substack{k > k_0 \\ k \equiv k_0 \pmod d}} A_k \left(\mathbb{P}(S_N = k) - \mathbb{P}(S_N = k + d) \right).$$

For any integer $\ell \geq 0$ consider the terms for which $2^\ell \leq k - k_0 < 2^{\ell+1}$. Observe that after applying the triangle inequality, we obtain a telescoping sum because of the monotonicity assumption on $\mathbb{P}(S_N = k)$. Using $|A_k| \leq B(k-k_0)^{1-1/\gamma}$ and $\mathbb{P}(S_N = k) < \varepsilon/(k - E_N)$ we thus obtain

$$\begin{aligned}\left| \sum_{\substack{2^\ell \leq k-k_0 < 2^{\ell+1} \\ k \equiv k_0 \pmod d}} A_k \left(\mathbb{P}(S_N = k) - \mathbb{P}(S_N = k+d)\right) \right| \\ \leq B 2^{(\ell+1)(1-1/\gamma)} \sum_{\substack{2^\ell \leq k-k_0 < 2^{\ell+1} \\ k \equiv k_0 \pmod d}} \left(\mathbb{P}(S_N = k) - \mathbb{P}(S_N = k+d)\right) \\ \leq 2B 2^{\ell(1-1/\gamma)} \frac{\varepsilon}{2^\ell + k_0 - E_N}.\end{aligned}$$

Here $k_0 - E_N \geq q^\gamma/(3C) - d$, and we may assume $q^\gamma/(3C) - d \geq q^\gamma/(6C)$. Hence by summing over $\ell \geq 0$ we get

$$|M| \leq 2\varepsilon B \sum_{\ell=0}^{\infty} \frac{2^{\ell(1-1/\gamma)}}{2^\ell + q^\gamma/(6C)}.$$

Estimating the terms $2^\ell \leq q^\gamma/(6C)$ and $2^\ell > q^\gamma/(6C)$ separately, we finally obtain

$$\begin{aligned}|M| &\leq 2\varepsilon B \left(\sum_{2^\ell \leq q^\gamma/(6C)} \frac{2^{\ell(1-1/\gamma)}}{q^\gamma/(6C)} + \sum_{2^\ell > q^\gamma/(6C)} 2^{-\ell/\gamma} \right) \\ &\leq 2\varepsilon B \left(\frac{(6C)^{1/\gamma}}{1 - 2^{1/\gamma - 1}} + \frac{(6C)^{1/\gamma}}{1 - 2^{-1/\gamma}} \right) \frac{1}{q}.\end{aligned}$$

By choosing $\varepsilon > 0$ small enough in terms of B, C and γ (in particular, depending only on α and the distribution of X_1), we can ensure $|M| < 1/(24q)$. Similarly, in the second term of (3.3) the sum over $k - E_N < -q^\gamma/(3C)$ will be less than $1/(24q)$. Hence $|f(v) - f(u)| \geq 1/(12q)$, and we are done.

□

Acknowledgement. The authors are indebted to the referee for valuable comments.

References

[1] I. Berkes and B. Borda, On the discrepancy of random subsequences of $\{n\alpha\}$. Submitted for publication.

[2] P. L. Brockett and J. H. B. Kemperman, On the unimodality of high convolutions. Ann. Probability 10 (1982), 270–277.

[3] A. S. Fainleib, A generalization of Esseen's inequality and its application in probabilistic number theory. Izv. Akad. Nauk SSSR Ser. Mat. 32 (1968), 859–879.

[4] W. Feller, An introduction to probability theory and its applications, Vol II. Wiley, 1971.

[5] P. Hall, Two-sided bounds on the rate of convergence to a stable law. Z. Wahrscheinlichkeitstheorie verw. Gebiete 57 (1981), 349–364.

[6] I. A. Ibragimov and Yu. V. Linnik, Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen, 1971.

[7] L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences. Pure and Applied Mathematics. Wiley, 1974.

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

[9] H. Niederreiter and W. Philipp, Berry–Esseen bounds and a theorem of Erdős and Turán on uniform distribution mod 1, Duke Math. J. 40(1973), 633–649.

[10] A. M. Odlyzko and L. B. Richmond, On the unimodality of high convolutions of discrete distributions, Ann. Probability 13 (1985), 299–306.

[11] V. Paulauskas, Estimates of the remainder term in limit theorems in the case of stable limit law. Lithuanian Math. J. 14 (1974), 127–146.

[12] V. V. Petrov, Limit theorems of probability theory. Sequences of independent random variables. Clarendon Press, 1995.

[13] K. I. Satybaldina, Absolute estimates of the rate of convergence to stable laws. Theory Probab. Appl. 17 (1972), 726–728.

[14] K. I. Satybaldina, On the estimation of the rate of convergence in a limit theorem with a stable stable limit law. Theory Probab. Appl. 18 (1973), 202–204.

[15] P. Schatte, On the asymptotic uniform distribution of the n-fold convolution mod 1 of a lattice distribution, Math. Nachr. 128 (1986) 233–241.

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