The central limit theorem for subsequences in probabilistic number theory

Christoph Aistleitner*, Christian Elsholtz†

Abstract

Let $(n_k)_{k\geq 1}$ be an increasing sequence of positive integers, and let $f(x)$ be a real function satisfying

$$f(x+1) = f(x), \qquad \int_0^1 f(x)\ dx = 0, \qquad \mathrm{Var}_{[0,1]} f < \infty. \tag{1}$$

If $\lim_{k\to\infty} \frac{n_{k+1}}{n_k} = \infty$ the distribution of

$$\frac{\sum_{k=1}^N f(n_k x)}{\sqrt{N}} \tag{2}$$

converges to a Gaussian distribution. In the case

$$1 < \liminf_{k\to\infty} \frac{n_{k+1}}{n_k}, \qquad \limsup_{k\to\infty} \frac{n_{k+1}}{n_k} < \infty$$

there is a complex interplay between the analytic properties of the function f, the number-theoretic properties of $(n_k)_{k\geq 1}$, and the limit distribution of (2).

In this paper we prove that any sequence $(n_k)_{k\geq 1}$ satisfying $\limsup_{k\to\infty} \frac{n_{k+1}}{n_k} = 1$ contains a nontrivial subsequence $(m_k)_{k\geq 1}$ such that for any function satisfying (1) the distribution of

$$\frac{\sum_{k=1}^N f(m_k x)}{\sqrt{N}}$$

converges to a Gaussian distribution. This result is best possible: for any $\varepsilon > 0$ there exists a sequence $(n_k)_{k\geq 1}$ satisfying $\limsup_{k\to\infty} \frac{n_{k+1}}{n_k} \leq 1+\varepsilon$ such that for every nontrivial subsequence $(m_k)_{k\geq 1}$ of $(n_k)_{k\geq 1}$ the distribution of (2) does not converge to a Gaussian distribution for some f.

Our result can be viewed as a Ramsey type result: a sufficiently dense increasing integer sequence contains a subsequence having a certain requested number-theoretic property.

*Graz University of Technology, Institute of Mathematics A, Steyrergasse 30, 8010 Graz, Austria. e-mail: aistleitner@math.tugraz.at. Research supported by the Austrian Research Foundation (FWF), Project S9603-N23. This paper was written during a stay of the author at the Hausdorff Institute of Mathematics in Bonn, Germany, in the framework of the "HIM Junior Trimester Program on Stochastics". The author wants to thank the HIM for its hospitality.

†Graz University of Technology, Institute of Mathematics A, Steyrergasse 30, 8010 Graz, Austria. e-mail: elsholtz@math.tugraz.at.

Mathematics Subject Classification: Primary 60F05, 42A55. Secondary: 11D04, 05C55, 11K06

Keywords: central limit theorem, lacunary sequences, linear Diophantine equations, Ramsey type theorem

1 Introduction and statement of results

1.1 Revision of results in Ramsey theory

Ramsey theory has often been summarized with the words of T. Motzkin "Complete disorder is impossible". The principle underlying Ramsey theory has often been observed in mathematics, and has for example been discussed by Burkill and Mirsky [9] and Nešetřil [23]. Let $\mathcal{S}$ be a family of objects and let P be a property that an element of $S \in \mathcal{S}$ may possess. Which additional property Q, perhaps in some quantitative or qualitative form, does guarantee that all $S \in \mathcal{S}$ with property Q do also have property P? Let us briefly recall two well known examples of this principle:

1. Let $\mathcal{S}$ be the family of 2-coloured complete finite graphs. Let P_n be the property that a graph $S \in \mathcal{S}$ contains a monochromatic complete subgraph K_n on n vertices. The main problem of Ramsey theory is, which size t on the original complete graph $K_t \in \mathcal{S}$ does guarantee that K_t has property P_n? It is known (for example), that a complete 2-coloured graph on $t = \binom{2n}{n}$ vertices contains a monochromatic subgraph K_n. A large proportion of P. Erdős' papers, and those of the Hungarian combinatorics school, are devoted to this circle of problems of extremal combinatorics. Methodically, probabilistic methods play an important role in this area.

2. Similarly, let $\mathcal{S}$ denote the family of subsets of the positive integers. Let S have property P_k, if S contains k integers in arithmetic progression. Szemerédi's theorem states that a set of $S \in \mathcal{S}$ of positive upper density has property P_k. The main open problem is to determine the correct density condition which guarantees that $S \in \mathcal{S}$ has property P_k. Many eminent mathematicians, such as Roth, Szemerédi, Fürstenberg, Bourgain, Gowers, Tao, Green, Sanders have worked on this problem; for the most recent work see Sanders [28], with references to the relevant literature. The methods include combinatorics, harmonic analysis and ergodic theory. In particular the Szemerédi regularity lemma (a version of which is part of Szemerédi's proof of his theorem [31]) has had enormous impact in graph theory and theoretical computer science.

Informally speaking, one says that by this Ramsey principle, no complete disorder is possible, as for example for a sufficiently large arbitrarily coloured complete graph a complete *monochromatic* induced subgraph (and hence a highly regular substructure) exists. Similarly, an arbitrary subset S of the positive integers contains a highly regular substructure, namely an arithmetic progression of length k, if only the set is dense enough.
Burkill and Mirsky [9] and Nešetřil [23] discussed this principle in a wider context, giving many further examples in different areas, including for example finite and infinite matrices, functions etc.
Todays' vast amount of literature on Ramsey type results includes quantitative and qualitative aspects. The former perhaps mainly coming from the harmonic analysis approach, due to Roth and Gowers, the latter perhaps primarily coming from the ergodic approach (Fürstenberg).
In spite of the numerous work in the literature, we are not aware that this principle has been studied in the context of limit distributions, which itself has an enormous body of literature in classical probabilistic number theory. Informally speaking, in this paper we show that for an arbitrary sequence, that only satisfies some "subexponential" gap condition, namely

$\limsup_k \frac{n_{k+1}}{n_k} \leq 1$, there exists some at most exponentially growing subsequence (m_k) such that simultaneously for a large class of real periodic functions f the distribution of

$$\frac{\sum_{k=1}^N f(m_k x)}{\sqrt{N}}$$

converges to a Gaussian distribution. Moreover we show that our result is best possible (a precise statement follows below). Our proof is based on the observation that a sufficiently dense integer sequence offers enough choice to find a subsequence having a certain requested number-theoretic property. On the other hand, if the original sequence is too thin, it is not possible to find an appropriate subsequence. Our theorem can be seen as a counterpart of the so-called subsequence principle, a general informal principle in probability theory which asserts that a sequence of random variables always contains a (possibly extremely thin) subsequence which behaves like a sequence of independent random variables (cf. Chatterji [10]). Utilizing the Ramsey principle we show that if the original sequence contains sufficiently many elements, then it is possible to restrict the growth of the subsequence to at most exponential growth.
It can be hoped for, that this investigation encourages other researchers to find more examples of the Ramsey principle in areas that have not traditionally been studied from the view point of Ramsey theory.

1.2 Revision of relevant results on limit distributions in probabilistic number theory

A sequence $(x_k)_{k\geq 1}$ of real numbers from the unit interval is called *uniformly distributed modulo one* (u.d. mod 1) if for all subintervals $[a,b)$ of the unit interval

$$\frac{1}{N}\sum_{k=1}^N \mathbb{1}_{[a,b)}(x_k) \to (b-a) \qquad \text{as} \quad N \to \infty.$$

The "quality" of the distribution of a sequence can be measured by the so-called discrepancy function D_N. The discrepancy $D_N(x_1,\dots,x_N)$ of the first N elements of $(x_k)_{k\geq 1}$ is defined as

$$D_N(x_1,\dots,x_N) = \sup_{0\leq a<b\leq 1} \left| \frac{1}{N}\sum_{k=1}^N \mathbb{1}_{[a,b)}(x_k) - (b-a) \right|.$$

It is easy to see that a sequence is u.d. mod 1 if and only if its discrepancy tends to zero as $N\to\infty$ (we refer to [11] and [22] for an introduction to uniform distribution theory and discrepancy theory).

In his seminal paper of 1916, Hermann Weyl showed that a sequence $(x_k)_{k\geq 1}$ is u.d. mod 1 if and only if

$$\frac{1}{N}\sum_{k=1}^N \cos 2\pi h x_k \to 0, \qquad \frac{1}{N}\sum_{k=1}^N \sin 2\pi h x_k \to 0, \qquad \text{for all } h\in\mathbb{Z},\ h\neq 0.$$

This criterion can be used for an easy proof of the fact that the sequence $(\langle kx\rangle)_{k\geq 1}$, where $\langle\cdot\rangle$ denotes the fractional part, is u.d. mod 1 for irrational x. In many cases it is very difficult to decide whether a certain sequence is u.d. or not; famous examples are the sequence $(\langle (3/2)^k \rangle)_{k\geq 1}$ and $(\langle 2^k \sqrt{2}\rangle)_{k\geq 1}$. By a general principle of Weyl, sequences of the form $(\langle n_k x\rangle)_{k\geq 1}$, where $(n_k)_{k\geq 1}$ is a fixed sequence of distinct integers, are u.d. mod 1 for almost all values of x (in the sense of Lebesgue measure). For fast growing $(n_k)_{k\geq 1}$ much more is true: in this case the sequence of functions $(\langle n_k x\rangle)_{k\geq 1}$, where $x \in [0,1]$ (these functions may be seen as random variables over the probability space $([0,1], \mathcal{B}([0,1]), \lambda_{[0,1]})$), exhibits many asymptotic properties which are typical for sequences of independent and identically distributed (i.i.d.) random variables.

In this context Weyl's theorem that $(\langle n_k x\rangle)_{k\geq 1}$ is u.d. mod 1 for almost all x (which implies that the discrepancy of $(\langle n_k x\rangle)_{k\geq 1}$ tends to zero for almost all x) can be either seen as a variant of the Glivenko-Cantelli theorem for the random variables $(\langle n_k x\rangle)_{k\geq 1}$ or as a strong law of large numbers for $(\cos 2\pi n_k x)_{k\geq 1}$ and $(\sin 2\pi n_k x)_{k\geq 1}$. As we mentioned before, for fast growing $(n_k)_{k\geq 1}$ much more is true. For example, classical results of Salem and Zygmund [26],[27] and Erdős and Gál [13] show that for any increasing sequence of positive integers $(n_k)_{k\geq 1}$ satisfying the growth condition

$$\frac{n_{k+1}}{n_k} > q > 1, \qquad k \geq 1, \tag{3}$$

the system $(\cos 2\pi n_k x)_{k\geq 1}$ satisfies the central limit theorem (CLT)

$$\lambda \left\{ x \in (0,1) : \frac{\sum_{k=1}^N \cos 2\pi n_k x}{\sqrt{N/2}} \leq t \right\} \to \Phi(t), \quad t \in \mathbb{R},$$

where λ denotes the Lebesgue measure and Φ the standard normal distribution function. A sequence satisfying (3) is called a "lacunary" sequence.

If the function $\cos 2\pi\cdot$ is replaced by a more general 1-periodic function f, the situation becomes much more complicated, and the asymptotic behaviour of the distribution of

$$\frac{\sum_{k=1}^N f(n_k x)}{\sqrt{N}} \tag{4}$$

is controlled by a complex interplay between analytic properties of f and number-theoretic properties of $(n_k)_{k\geq 1}$. The sequence (4) does not necessarily possess a limit distribution, and if such a limit distribution exists it may be non-Gaussian. For example, Erdős and Fortet (see [20, p.646]) showed that in the case

$$f(x) = \cos 2\pi x + \cos 4\pi x, \qquad n_k = 2^k + 1,$$

the distribution of (4) converges to a non-Gaussian limit distribution. Typically it is assumed that f satisfies

$$f(x+1) = f(x), \qquad \int_0^1 f(x)\, dx = 0, \qquad \mathrm{Var}_{[0,1]} f(x) < \infty. \tag{5}$$

For functions satisfying these conditions, the asymptotic distribution of (4) is a Gaussian distribution if, for example (cf. [18], [32]),

- $$\frac{n_{k+1}}{n_k} \to \infty \qquad \text{as} \qquad k \to \infty \tag{6}$$

- $\dfrac{n_{k+1}}{n_k}$ is an integer for $k \geq 1$
- $\dfrac{n_{k+1}}{n_k} \to \theta$ for some θ satisfying $\theta^r \notin \mathbb{Q},\ r = 1, 2, \ldots$

Gaposhkin [19] observed that the asymptotic behaviour of (4) has an intimate connection with the number of solutions (k, l) of Diophantine equations of the form

$$an_k \pm bn_l = c, \qquad a, b, c \in \mathbb{Z}, \tag{7}$$

and Aistleitner and Berkes [3] found a necessary and sufficient condition, formulated in terms of the number of solutions of Diophantine equations of the type (7), which guarantees that the distribution of (4) converges to a Gaussian distribution.

There are only few precise results in the case of sub-lacunary growing sequences $(n_k)_{k\geq1}$. Generally speaking, in the lacunary case the behaviour of $(f(n_kx))_{k\geq1}$ is somewhat similar to the behaviour of sequences of i.i.d. random variables, whereas this is not necessarily true for sub-lacunary $(n_k)_{k\geq1}$. In the sub-lacunary case one needs either very strong number-theoretic conditions (cf. [1], [5], [6], [16], [17], [25]), or obtains only results for "almost all" sequences $(n_k)_{k\geq1}$ in an appropriate statistical sense (cf. [4], [8], [14], [15]). An exception is the sequence $n_k = k,\ k \geq 1$, where very precise results are known due to the fact that the behaviour of $(\langle kx \rangle)_{k\geq1}$ is intimately connected with the properties of the continued fraction expansion of x (cf. [12], [21], [29], [30]).

In this paper we will show, roughly speaking, the following principle: if $(n_k)_{k\geq1}$ is a sub-lacunary sequence, then it always contains a subsequence $(m_k)_{k\geq1}$ such that for any f satisfying (5) the distribution of

$$\frac{\sum_{k=1}^N f(m_kx)}{\sqrt{N}} \tag{8}$$

converges to a Gaussian distribution. On the other hand, if $(n_k)_{k\geq1}$ is already lacunary, then it may happen that (8) does not converge to a Gaussian distribution for every possible nontrivial subsequence $(m_k)_{k\geq1}$ of $(n_k)_{k\geq1}$. In this informal statement we call a subsequence "nontrivial" if it is not superlacunary (that means $\limsup_{k\to\infty} n_{k+1}/n_k = \infty$ is not allowed). This restriction is necessary, since trivially an arbitrary sequence $(n_k)_{k\geq1}$ contains a superlacunary growing subsequence, for which by (6) the CLT is always true.

We mention that a similar problem has been considered in a more general context by Bobkov and Götze [7]. Let $X_1, X_2, \ldots$ be a sequence of uncorrelated random variables. Then under certain weak assumptions, such as e.g.

$$\max_{1\leq k\leq N} |X_k| = o(\sqrt{N}), \qquad \text{and} \qquad \frac{X_1^2 + \cdots + X_N^2}{N} \to 1 \quad \text{in probability,}$$

the sequence $X_1, X_2, \ldots$ contains a subsequence $X_{i_1}, X_{i_2}, \ldots$ for which the distribution of

$$\frac{X_{i_1} + \cdots + X_{i_N}}{\sqrt{N}}$$

converges to the standard normal distribution. In this result the sequence $(i_k)_{k\geq1}$ can be chosen to grow slowly, in the sense that for any prescribed $(j_k)_{k\geq1}$ satisfying $j_k/k \to \infty$ it is possible to have $i_k \le j_k$ for sufficiently large k. We note that this result of Bobkov and Götze does not apply in our situation, since our random variables $(f(n_k x))_{k\ge1}$ are (generally) *not* uncorrelated.

1.3 Statement of results

We will prove the following two theorems:

Theorem 1 *Let $(n_k)_{k\ge1}$ be a strictly increasing sequence of positive integers. If*

$$\limsup_{k\to\infty} \frac{n_{k+1}}{n_k} = 1, \tag{9}$$

then there exists a subsequence $(m_k)_{k\ge1}$ of $(n_k)_{k\ge1}$, satisfying

$$q_1 \le \frac{m_{k+1}}{m_k} \le q_2, \quad k \ge 1, \qquad \textit{for some} \qquad 1 < q_1 < q_2 < \infty,$$

such that for all functions f satisfying (5) and for all $t \in \mathbb{R}$

$$\lambda \left\{ x \in (0,1) : \frac{\sum_{k=1}^N f(m_k x)}{\sqrt{\|f\|_2^2 N}} \le t \right\} \to \Phi(t)$$

holds.

In the formulation of this theorem,

$$\|f\|_2 = \left(\int_0^1 f(x)^2 \, dx \right)^{1/2}.$$

Theorem 2 shows that condition (9) in Theorem 1 is optimal:

Theorem 2 *Let $\varepsilon > 0$. Then there exists a strictly increasing sequence $(n_k)_{k\ge1}$ of positive integers, satisfying*

$$\frac{n_{k+1}}{n_k} \le 1 + \varepsilon, \qquad k \ge 1,$$

such that for every subsequence $(m_k)_{k\ge1}$ of $(n_k)_{k\ge1}$, which satisfies

$$\limsup_{k\to\infty} \frac{m_{k+1}}{m_k} < \infty,$$

there exists a trigonometric polynomial f such that the distribution of

$$\frac{\sum_{k=1}^N f(m_k x)}{\sqrt{N}}$$

does not converge to a Gaussian distribution.

The similar problem concerning the law of the iterated logarithm (LIL) seems to be much more complicated. In this context we formulate the following
Open problem: Under which assumptions does the sequence $(n_k)_{k\geq 1}$ contain a nontrivial subsequence $(m_k)_{k\geq 1}$ for which $(f(m_k x))_{k\geq 1}$ satisfies the (exact) law of the iterated logarithm, for all functions f satisfying (5)?
This problem is related to the problem of finding the best possible condition, formulated in terms of Diophantine equations of the form (7), which guarantees the (exact) law of the iterated logarithm for $f(n_k x)$ for lacunary $(n_k)_{k\geq 1}$ (cf. [2]). As in the case of the CLT, if $(n_k)_{k\geq 1}$ is already lacunary, there does in general not necessarily exist a subsequence having the required properties concerning the LIL. On the other hand, it is unclear if $\limsup n_{k+1}/n_k = 1$ is sufficient for the existence of a nontrivial subsequence $(m_k)_{k\geq 1}$ for which the exact LIL is satisfied (a "not exact" version of the LIL is true for an arbitrary lacunary sequence, see [24]).

2 Proof of Theorem 1

Let $(n_k)_{k\geq 1}$ be given, and assume that

$$\limsup_{k\to\infty} \frac{n_{k+1}}{n_k} = 1. \tag{10}$$

For $r \geq 1$ we set

$$I_r = \left[2^{r_0+4r}, 2^{r_0+4r+1}\right), \tag{11}$$

where r_0 is fixed and sufficiently large, such that we can find a positive nondecreasing integer-valued function $g(r)$ such that

$$g(r) \geq 3, \qquad g(r) \to \infty \quad \text{as} \quad r \to \infty$$

and

$$\#\{k \in \mathbb{N} : \ n_k \in I_r\} \geq g(r), \qquad r \geq 1,$$

and a positive nondecreasing function $h(r)$ such that $h(r) \to \infty$ as $r \to \infty$ and

$$h(r) \geq 1, \qquad \frac{r}{h(r)} \geq \frac{r-1}{h(r-1)}, \qquad \lceil h(r) \rceil^4 < g(r), \quad r \geq 1. \tag{12}$$

By (10) it is possible to find such r_0, g, h. In particular (12) holds if h is growing very slowly.

Now we construct the sequence $(m_k)_{k\geq 1}$ inductively. For m_1 we choose one of the values of $(n_k)_{k\geq 1}$ in the interval I_1, for m_2 we choose one of the values of $(n_k)_{k\geq 1}$ in I_2, and generally, in the r-th step, we choose for m_r one of the values of $(n_k)_{k\geq 1}$ in I_r. Additionally, $(m_k)_{k\geq 1}$ shall have the following property:

- for every $N \geq 1$ and all integers a, b satisfying $1 \leq a < b \leq h(N),\ a \neq b,\ \log_2(b/a) \geq 2$, and all integers c the number of solutions $(k, l),\ k, l \leq N,\ k \neq l$, of the Diophantine equation

  $$a m_k - b m_l = c$$

  where

  $$a \leq h(k),\ b \leq h(l), \tag{13}$$

$$\text{and} \quad k-l = \left\| \frac{\log_2(b/a)}{4} \right\|, \tag{14}$$

is at most $N/h(N)$ (in this statement $\|\cdot\|$ denotes the nearest integer, i.e. for $y = \lfloor y \rfloor + \langle y \rangle$ we set $\|y\| = \lfloor y \rfloor$ or $\|y\| = \lceil y \rceil$, depending on whether $\langle y \rangle < 1/2$ or $\langle y \rangle \geq 1/2$).

To show that it is always possible to find inductively an appropriate m_r, we assume that $m_1, \dots, m_{r-1}$ are already constructed, and $(m_k)_{1 \leq k \leq r-1}$ satisfies the above conditions. For fixed $a < b$, there are at most $\lceil h(r) \rceil^2$ values of c for which the number of solutions $(k,l),\ k,l \leq r-1,\ k \neq l$, of

$$am_k - bm_l = c, \qquad \text{subject to conditions} \quad (13),(14),$$

is greater than or equal to $(r-1)/\lceil h(r) \rceil^2$. Since there are at least $g(r) > \lceil h(r) \rceil^4$ elements of $(n_k)_{k \geq 1}$ in I_r, and since there are at most $h(r)^2$ possible choices for a, b such that $1 \leq a < b \leq h(r),\ \log_2(b/a) \geq 2$, there exists at least one number m_r equal to one of the elements of $(n_k)_{k \geq 1}$ in the interval I_r, for which for all $1 \leq a < b \leq h(r),\ \log_2(b/a) \geq 2$, and all $c \in \mathbb{Z}$,

$$\begin{aligned}
& \#\{(k,l),\ k,l \leq r-1,\ k \neq l,\ \text{satisfying } (13),(14) \text{ and } am_k - bn_l = c\} \\
& +\mathbb{1}(am_r - bm_{r - \|\log_2(b/a)/4\|} = c) \\
\leq\ & (r-1)/\lceil h(r) \rceil^2 + 1 \\
\leq\ & (r-1)/h(r)^2 + 1 \\
\leq\ & r/h(r)
\end{aligned}$$

(where the last inequality follows from $h(r) \geq 1$).

Lemma 1 *The sequence $(m_k)_{k \geq 1}$ constructed as above satisfies*

$$8 \leq \frac{m_{k+1}}{m_k} \leq 32, \qquad k \geq 1, \tag{15}$$

and for all fixed positive integers a, b and all integers c

$$\#\{(k,l):\ 1 \leq k,l \leq N,\ am_k - bm_l = c\} = o(N) \qquad \text{as} \qquad N \to \infty, \tag{16}$$

uniformly in c (with the exception of the trivial solutions $k = l$ in the case $a = b,\ c = 0$).

Theorem 1 follows from Lemma 1 and the following theorem, which can be found in [3]:

Theorem 3 *Let $(m_k)_{k \geq 1}$ be a lacunary sequence of positive integers, and let f be a function satisfying (5). Assume that for all fixed positive integers a, b and for all integers c*

$$\#\{1 \leq k,l \leq N:\ an_k - bn_l = c\} = o(N) \quad \text{as} \quad N \to \infty,$$

uniformly in c (with the exception of the trivial solutions $k = l$ in the case $a = b,\ c = 0$). Then

$$\lambda \left\{ x \in (0,1):\ \frac{\sum_{k=1}^N f(m_k x)}{\sqrt{\|f\|_2^2 N}} \leq t \right\} \to \Phi(t)$$

for all $t \in \mathbb{R}$.

It remains to prove Lemma 1. Equation (15) is true by construction. In fact, since $m_k \in I_k$, and $m_{k+1} \in I_{k+1}$, clearly

$$\frac{m_{k+1}}{m_k} \in \left[2^3, 2^5\right], \qquad k \geq 1$$

(the intervals I_k were defined in (11)).

Thus $(m_k)_{k\geq 1}$ is a lacunary sequence (and $(am_k)_{k\geq 1}$ for $a \geq 1$ is also a lacunary sequence), which implies that (16) is true in the case $a = b$, since for every lacunary sequence $(\mu_k)_{k\geq 1}$ the number of solutions $(k,l),\ k \neq l$ of

$$\mu_k - \mu_l = c$$

is bounded by a constant, uniformly in $c \in \mathbb{Z}$ (cf. [33, p. 203]).

Now assume $a < b$ (which is also sufficient for the case $a > b$, since $am_k - bm_l = c$ is equivalent to $bm_l - am_k = -c$). If $\log_2(b/a) < 2$, then by (15) the sequences $(am_k)_{k\geq 1}$ and $(bm_k)_{k\geq 1}$ have no elements in common, and the sequence containing all numbers $am_k,\ k \geq 1$ and $bm_k,\ k \geq 1$ is a lacunary sequence. Therefore (16) holds in this case.

If $\log_2(b/a) \geq 2$ and $k - l < \|\log_2(b/a)/4\|$, then $4(k-l)+2 \leq \log_2(b/a)$ and

$$\frac{am_k}{bm_l} \leq 2^{4(k-l)+1-\log_2(b/a)} \leq \frac{1}{2}.$$

On the other hand, if $\log_2(b/a) \geq 2$ and $k - l > \|\log_2(b/a)/4\|$, then $4(k-l)-2 > \log_2(b/a)$ and

$$\frac{am_k}{bm_l} \geq 2^{4(k-l)-1-\log_2(b/a)} > 2.$$

This implies that in the case $\log_2(b/a) \geq 2$ the existence of (k_1,l_1) and (k_2,l_2), satisfying $k_i - l_i \neq \|\log_2(b/a)/4\|,\ i = 1,2$, and

$$am_{k_1} - bm_{l_1} = am_{k_2} - bm_{l_2}$$

implies $k_1 = k_2,\ l_1 = l_2$. Therefore for given positive integers $a < b,\ \log_2(b/a) \geq 2$, and any $c \in \mathbb{Z}$ there is at most one solution (k,l) of the Diophantine equation

$$am_k - bm_l = c \tag{17}$$

for which $k - l \neq \|\log_2(b/a)/4\|$.

The number of solutions $(k,l),\ k,l \leq N$, of (17), for which $k - l = \|\log_2(b/a)/4\|$ is $o(N)$, uniformly in c (for any fixed $a < b$), by the construction of $(m_k)_{k\geq 1}$ (more precisely, it is bounded by $N/h(N)$, where h is the function in (12)). This proves Lemma 1.

3 Proof of Theorem 2

Let $\varepsilon > 0$. We construct a sequence $(n_k)_{k\geq 1}$ satisfying

$$1 + \frac{\varepsilon}{2} \leq \frac{n_{k+1}}{n_k} \leq 1 + \varepsilon, \qquad k \geq 1,$$

such that for every subsequence $(m_k)_{k\geq 1}$ of $(n_k)_{k\geq 1}$ with

$$\limsup_{k\to\infty} \frac{m_{k+1}}{m_k} < \infty$$

there exists a trigonometric polynomial f for which the distribution of

$$\frac{\sum_{k=1}^N f(m_k x)}{\sqrt{N}} \tag{18}$$

does not converge to a Gaussian distribution.

First we choose coprime integers p, q such that

$$1 + \frac{5\varepsilon}{8} \leq \frac{p}{q} \leq 1 + \frac{7\varepsilon}{8}, \tag{19}$$

and define a sequence $(\nu_k)_{k\geq 1}$ by

$$\nu_k = \left\lceil \frac{p^k}{q^k} \right\rceil - \left\lceil \frac{2p}{p-q} \right\rceil, \qquad k \geq 1.$$

Write k_0 for the smallest index for which

$$\nu_k \geq 1, \quad \text{and} \quad 1 + \frac{\varepsilon}{2} \leq \frac{\nu_{k+1}}{\nu_k} \leq 1 + \varepsilon \qquad \text{for all} \quad k \geq k_0,$$

(it is possible to find such a k_0 because of (19)).

Define

$$n_k = \nu_{k+k_0}, \qquad k \geq 1.$$

Then $(n_k)_{k\geq 1}$ satisfies

$$1 + \frac{\varepsilon}{2} \leq \frac{n_{k+1}}{n_k} \leq 1 + \varepsilon \quad \text{for all} \quad k \geq 1. \tag{20}$$

Let $(m_k)_{k\geq 1}$ be an arbitrary subsequence of $(n_k)_{k\geq 1}$ satisfying

$$\limsup_{k\to\infty} \frac{m_{k+1}}{m_k} < \infty. \tag{21}$$

We want to show that there exists a trigonometric polynomial f such that the distribution of (18) does not converge to a Gaussian distribution.

As a consequence of (21) there exists a number $q_2 > 1$ such that

$$\frac{m_{k+1}}{m_k} \leq q_2, \qquad k \geq 1.$$

For consecutive elements of $(m_k)_{k\geq 1}$ we define a function

$$d(m_k, m_{k+1}) = w - v,$$

where v and w are the (uniquely defined) indices for which $m_k = n_v$, $m_{k+1} = n_w$ (i.e. d measures the difference of the indices of the numbers m_k and m_{k+1} within the original sequence $(n_k)_{k\geq 1}$). Let

$$s = \min \left\{ h \geq 1 : \limsup_{N\to\infty} \left(\frac{1}{N} \sum_{k=1}^{N} \mathbb{1}(d(m_k, m_{k+1}) = h) \right) > 0 \right\},$$

and define

$$f(x) = \cos 2\pi p^s x + \cos 2\pi q^s x.$$

Since by the definition of s and the orthogonality of the trigonometric system

$$\int_0^1 \left(\frac{1}{\sqrt{N}} \sum_{\substack{1\leq k\leq N, \\ d(m_k,m_{k+1})<s}} f(m_k x) \right)^2 dx \leq \frac{4}{N} \sum_{\substack{1\leq k\leq N, \\ d(m_k,m_{k+1})<s}} 1 \to 0 \quad \text{as} \quad N \to \infty,$$

we have

$$\frac{1}{\sqrt{N}} \sum_{\substack{1\leq k\leq N, \\ d(m_k,m_{k+1})<s}} f(m_k x) \to 0 \quad \text{in distribution,}$$

and therefore we can assume w.l.o.g. that $d(m_k, m_{k+1}) \geq s$ for all elements of $(m_k)_{k\geq 1}$.

If $d(m_k, m_{k+1}) = s$, then by the definition of $(\nu_k)_{k\geq 1}$ the numbers m_k and m_{k+1} are of the form

$$\left\lceil \frac{p^r}{q^r} \right\rceil - \left\lceil \frac{2p}{p-q} \right\rceil, \qquad \left\lceil \frac{p^{r+s}}{q^{r+s}} \right\rceil + \left\lceil \frac{2p}{p-q} \right\rceil$$

for some $r \geq 1$, and therefore

$$q^s m_{k+1} - p^s m_k = q^s \left(\left\lceil \frac{p^{r+s}}{q^{r+s}} \right\rceil - \left\lceil \frac{2p}{p-q} \right\rceil \right) - p^s \left(\left\lceil \frac{p^r}{q^r} \right\rceil - \left\lceil \frac{2p}{p-q} \right\rceil \right).$$

Thus, if $d(m_k, m_{k+1}) = s$,

$$\begin{aligned} q^s m_{k+1} - p^s m_k &\geq q^s \frac{p^{r+s}}{q^{r+s}} - q^s \frac{2p}{p-q} - q^s - p^s \frac{p^r}{q^r} - p^s + p^s \frac{2p}{p-q} \\ &\geq (p^s - q^s) \frac{2p}{p-q} - 2p^s > 0, \end{aligned}$$

and

$$\begin{aligned} q^s m_{k+1} - p^s m_k &\leq q^s \frac{p^{r+s}}{q^{r+s}} + q^s - q^s \frac{2p}{p-q} - p^s \frac{p^r}{q^r} + p^s \frac{2p}{p-q} + p^s \\ &\leq 2p^s + \frac{2p^{s+1}}{p-q}. \end{aligned}$$

In other words, if $d(m_k, m_{k+1}) = s$, then

$$q^s m_{k+1} - p^s m_k \in \left[1, 2p^s + \frac{2p^{s+1}}{p-q} \right]. \tag{22}$$

It is easy to see that the sequence consisting of all elements of the form $p^s m_k,\ k \geq 1$, and all elements of the form

$$q^s m_k \quad k \geq 1, \qquad \text{except those } k \text{ for which} \quad d(m_k, m_{k+1}) = s,$$

sorted in increasing order, is a lacunary sequence. Since for every lacunary sequence $(\mu_k)_{k \geq 1}$ the number of solutions (k,l) of

$$\mu_k - \mu_l = c$$

is bounded by a constant, uniformly in $c \in \mathbb{Z}$ (cf. [33, p. 203]), this proves that the number of solutions $(k,l),\ k \neq l$, of

$$p^s m_k - q^s m_l = c,$$

subject to $d(m_l, m_{l+1}) \neq s$, is bounded by a constant, uniformly in c. If $d(m_l, m_{l+1}) = s$, then by (22)

$$p^s m_k - q^s m_l = c$$

implies that

$$p^s m_k - p^s m_{l+1} \in \left[c - 2p^s + \frac{2p^{s+1}}{p-q}, c + 2p^s + \frac{2p^{s+1}}{p-q}\right].$$

This equation has only finitely many solutions (k,l) for which $k \neq l+1$ (since $(m_k)_{k \geq 1}$ is lacunary). Therefore the number of solutions (k,l) of

$$p^s m_k - q^s m_l = c, \quad \text{where either} \quad d(m_l, m_{l+1}) \neq s \quad \text{or} \quad k \neq l+1, \tag{23}$$

is bounded by a constant, uniformly in c, and in the case $d(m_l, m_{l+1}) = s$ and $k = l+1$ we always have

$$p^s m_k - q^s m_l \in \left[1, 2p^s + \frac{2p^{s+1}}{p-q}\right].$$

Divide $\mathbb{N}$ into consecutive blocks $\Delta_1, \Delta_1', \Delta_2, \Delta_2', \ldots$, where

$$|\Delta_i| = \lceil \log \log i \rceil, \qquad |\Delta_i'| = \left\lceil \log_{(1+\varepsilon/2)} 4p^s \right\rceil$$

($|\cdot|$ denotes the number of elements of a set, and $\log x$ should be understood as $\max(1, \log x)$). Then by (20) for arbitrary $i_1 > i_2 \geq 1$ and $k_1 \in \Delta_{i_1},\ k_2 \in \Delta_{i_2}$ we have

$$\frac{m_{k_1}}{m_{k_2}} \geq \left(1 + \frac{\varepsilon}{2}\right)^{k_1 - k_2} \geq \left(1 + \frac{\varepsilon}{2}\right)^{\left\lceil \log_{(1+\varepsilon/2)} 4p^s \right\rceil} \geq 4p^s. \tag{24}$$

Since for every $r \in \mathbb{Z}$

$$\frac{1}{\sum_{h=1}^N |\Delta_h|} \# \left\{ (k,l),\ k,l \in \bigcup_{h=1}^N \Delta_h : p^s m_k - q^s m_l = r \right\} \in [0,1], \quad N \geq 1,$$

by the Bolzano-Weierstrass theorem it is possible to choose a subsequence $(h_j)_{j \geq 1}$ of $\mathbb{N}$ such that for all $r,\ 1 \leq r \leq 2p^s + \frac{2p^{s+1}}{p-q}$,

$$\frac{1}{\sum_{h=1}^{N_j} |\Delta_h|} \# \left\{ (k,l),\ k,l \in \bigcup_{h=1}^{N_j} \Delta_h : p^s m_k - q^s m_l = r \right\} \to \gamma_r \quad \text{as} \quad j \to \infty \tag{25}$$

for some appropriate constants $\gamma_r,\ 1 \leq r \leq 2p^s + \frac{2p^{s+1}}{p-q}$. For these constants γ_r necessarily

$$\sum_{r=1}^{2p^s+\frac{2p^{s+1}}{p-q}} \gamma_r = 1,$$

since by (22) and (23)

$$\sum_{r=1}^{2p^s+\frac{2p^{s+1}}{p-q}} \frac{1}{\sum_{h=1}^N |\Delta_h|} \ \# \left\{ (k,l),\ k,l \in \bigcup_{h=1}^N \Delta_h : p^s m_k - q^s m_l = r \right\} \to 1 \quad \text{as} \quad N \to \infty.$$

Let $t \in \mathbb{R}$, and define

$$\psi(x) = \frac{1 + 2\sum_{r=1}^{2p^s+\frac{2p^{s+1}}{p-q}} \gamma_r \cos 2\pi r x}{2}.$$

The functions $\psi(x)$ and $e^{-t^2\psi(x)/2}$ are Lipschitz-continuous, and therefore, as some simple calculations show, for every positive integer w

$$\int_0^1 e^{-t^2\psi(x)/2} \cos 2\pi w x \ dx \ll w^{-1}$$

(here and in the sequel "$\ll$" is the Vinogradov symbol). Using standard trigonometric identities we can write the function

$$\prod_{h=1}^{N_j} \left(1 + \frac{it \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right) \tag{26}$$

in the form

$$\begin{aligned} &\sum\nolimits^* \left(\frac{it}{2\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right)^{\chi_1+\dots+\chi_{N_j}} \times \\ &\times \left(\prod_{h=1}^{N_j} (4|\Delta_h|)^{\chi_h - 1} \right) \cos 2\pi \left(\pm \chi_1 \mu_1 m_{k_1} \pm \dots \pm \chi_{N_j} \mu_{N_j} m_{N_j} \right) x, \end{aligned} \tag{27}$$

where $\sum^*$ contains all the sums

$$\sum_{k_1 \in \Delta_1} \cdots \sum_{k_{N_j} \in \Delta_{N_j}} \ \sum_{(\chi_1,\dots,\chi_j) \in \{0,1\}^{N_j}} \ \sum_{(\mu_1,\dots,\mu_{N_j}) \in \{p^s,q^s\}^{N_j}} \ \sum_{\pm},$$

where

$$\sum_{\pm}$$

is meant as a sum over all the 2^{N_j} many possible choices of signs "+" and "−" inside the cos-function.

The factor $2^{-(\chi_1+\dots+\chi_{N_j})}$ in (27) comes from $\cos x = (\cos x + \cos(-x))/2$ and the iterative use of the formula $\cos x \cos y = 2^{-1}(\cos(x+y) + \cos(x-y))$, while the norming factors in the product come from the surplus contribution of the sums $\sum_{k_i} \sum_{\mu_i} \sum_{\pm}$ in the case $\chi_i = 0$. For the vector $(\chi_1, \dots, \chi_{N_j}) \neq (0, \dots, 0)$, write $\hat{h}$ for the largest index of a nonzero element of this vector (i.e. $\hat{h}(\chi_1, \dots, \chi_{N_j}) = \max\{h : \ 1 \leq h \leq N_j, \ \chi_h = 1\}$). By (24) the order of magnitude of the sum

$$\pm \chi_1 \mu_1 m_{k_1} \pm \dots \pm \chi_{N_j} \mu_{N_j} m_{N_j}$$

is determined by

$$\mu_{k_{\hat{h}}} m_{k_{\hat{h}}}.$$

More precisely, it follows from (24) that, independent of the choice of signs + and −, the frequency of

$$\cos 2\pi \left(\pm \chi_1 \mu_1 m_{k_1} \pm \dots \pm \chi_{N_j} \mu_{N_j} m_{N_j} \right) x$$

lies in the interval

$$\left[\mu_{k_{\hat{h}}} m_{k_{\hat{h}}}/2, \ 2\mu_{k_{\hat{h}}} m_{k_{\hat{h}}} \right],$$

which means that in this case

$$\int_0^1 e^{-t^2\psi(x)/2} \cos 2\pi \left(\pm \chi_1 m_{k_1} \pm \dots \pm \chi_{N_j} m_{N_j} \right) x \ dx \ll m_{k_{\hat{h}}}^{-1}.$$

We assume w.l.o.g. $|t| \leq |\Delta_{N_j}|$, which is true for sufficiently large j, and which implies

$$\left| \frac{t}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right| \ll N_j^{-1/2}. \tag{28}$$

For a fixed vector $(\chi_1, \dots, \chi_{N_j}) \neq (0, \dots, 0)$, the sum (27) is a sum of

$$\left(\prod_{h=1}^{N_j} |\Delta_h| \right) 2^{2N_j}$$

cos-functions, all of which have coefficient

$$\left(\frac{it}{2\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right)^{\chi_1+\dots+\chi_{N_j}} \left(\prod_{h=1}^{N_j} (4|\Delta_h|)^{\chi_h - 1} \right)$$

and a frequency in

$$\left[q^s \left(\min_{k \in \Delta_{\hat{h}}} m_k \right) /2, 2p^s \left(\max_{k \in \Delta_{\hat{h}}} m_k \right) \right]$$

(where $\hat{h}(\chi_1, \dots, \chi_{N_j})$ is defined as above).
Thus by (28)

$$\left| \int_0^1 e^{-t^2\psi(x)/2} \sum_{k_1 \in \Delta_1} \cdots \sum_{k_{N_j} \in \Delta_{N_j}} \sum_{(\mu_1, \dots, \mu_{N_j}) \in \{p^s, q^s\}^{N_j}} \sum_{\pm} \times \right.$$

$$
\times \left(\frac{it}{2\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right)^{\chi_1+\cdots+\chi_{N_j}} \left(\prod_{h=1}^{N_j} (4|\Delta_h|)^{\chi_h - 1} \right) \times
$$

$$
\times \cos\left(2\pi\left(\pm\chi_1\mu_1 m_{k_1} \pm \cdots \pm \chi_{N_j}\mu_{N_j} m_{N_j}\right) x\right)\, dx \Bigg| \tag{29}
$$

$$
\begin{aligned}
&\ll \left(\prod_{h=1}^{N_j} |\Delta_h| \right) 2^{2N_j} \left(\prod_{h=1}^{N_j} (4|\Delta_h|)^{\chi_h - 1} \right) \left(\min_{k \in \Delta_{\hat{h}}} m_k \right)^{-1} \left(2^{-1} N_j^{-1/2} \right)^{\chi_1+\cdots+\chi_{N_j}} \\
&\ll \left(\prod_{h=1}^{N_j} (4|\Delta_h|)^{\chi_h} \right) \left(\min_{k \in \Delta_{\hat{h}}} m_k \right)^{-1} \left(2^{-1} N_j^{-1/2} \right)^{\chi_1+\cdots+\chi_{N_j}} \\
&\ll \left(2|\Delta_{\hat{h}}| \right)^{\hat{h}} N_j^{-1/2} \left(\min_{k \in \Delta_{\hat{h}}} m_k \right)^{-1}.
\end{aligned}
$$

For the vector $(\chi_1, \ldots, \chi_{N_j}) = (0, \ldots, 0)$, which corresponds to multiplying N_j times the factor 1 in the product (26), the integral (29) gives

$$
\int_0^1 e^{-t^2\psi(x)/2}\, dx.
$$

Since there are at most 2^v vectors $(\chi_1, \ldots, \chi_{N_j})$ for which the index $\hat{h}$ of the largest nonzero element is v, this implies

$$
\begin{aligned}
&\left| \int_0^1 e^{-t^2\psi(x)/2} \prod_{h=1}^{N_j} \left(1 + \frac{it \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right) dx - \int_0^1 e^{-t^2\psi(x)/2}\, dx \right| \\
\ll\ & N_j^{-1/2} \sum_{v=1}^{N_j} 2^v \left(2|\Delta_v| \right)^v \left(\min_{k \in \Delta_v} m_k \right)^{-1} \\
\ll\ & N_j^{-1/2}.
\end{aligned} \tag{30}
$$

In the sequel, the symbol $\mathbb{E}$ denotes the expected value with respect to x and $\lambda_{[0,1]}$.

Writing $e(x) = e^x$ and using the well-known estimate

$$
e(ix) = (1 + ix) e^{-x^2/2 + w(x)}, \quad \text{where} \quad |w(x)| \leq |x^3|,
$$

we have

$$
\begin{aligned}
& \mathbb{E} \left(e \left(\frac{it \sum_{h=1}^{N_j} \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right) \right) \\
=\ & \mathbb{E} \left(\prod_{h=1}^{N_j} e \left(\frac{it \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right) \right) \\
=\ & \mathbb{E} \left(\prod_{h=1}^{N_j} \left(1 + \frac{it \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right) e \left(- \frac{t^2 \sum_{h=1}^{N_j} \left(\sum_{k \in \Delta_h} f(m_k x) \right)^2}{2 \sum_{h=1}^{N_j} |\Delta_h|} \right) e^{W_j} \right),
\end{aligned}
$$

where, using $|f(x)| \leq 2$ and $|t| \leq |\Delta_{N_j}|$, we have

$$\begin{aligned}
|W_j| &:= \sum_{h=1}^{N_j} w\left(\frac{t\sum_{k\in\Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j}|\Delta_h|}}\right)\\
&\leq \sum_{h=1}^{N_j} |t|^3 \frac{8|\Delta_h|^3}{\left(\sum_{h=1}^{N_j}|\Delta_h|\right)^{3/2}}\\
&\ll N_j(\log\log N_j)^6 N_j^{-3/2}\\
&\ll N_j^{-1/4}. \qquad (31)
\end{aligned}$$

Using $1+x \leq e^x$, we have

$$\begin{aligned}
\left|\prod_{h=1}^{N_j}\left(1+\frac{it\sum_{k\in\Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j}|\Delta_h|}}\right)\right| &\leq \prod_{h=1}^{N_j}\left|\left(1+\frac{it\sum_{k\in\Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j}|\Delta_h|}}\right)\right|\\
&\leq e\left(\frac{t^2\sum_{h=1}^{N_j}\left(\sum_{k\in\Delta_h} f(m_k x)\right)^2}{2\sum_{h=1}^{N_j}|\Delta_h|}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left(e\left(\frac{it\sum_{h=1}^{N_j}\sum_{k\in\Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j}|\Delta_h|}}\right)\right) - \mathbb{E}e^{-t^2\psi(x)/2}\right|\\
\ll &\left|\mathbb{E}\left(\prod_{h=1}^{N_j}\left(1+\frac{it\sum_{k\in\Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j}|\Delta_h|}}\right)\times\right.\right.\\
&\quad\left.\left.\times e\left(-\frac{t^2\sum_{h=1}^{N_j}\left(\sum_{k\in\Delta_h} f(m_k x)\right)^2}{2\sum_{h=1}^{N_j}|\Delta_h|}\right)\left(e^{W_j}-1\right)\right)\right|\\
&+\left|\mathbb{E}\left(\prod_{h=1}^{N_j}\left(1+\frac{it\sum_{k\in\Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j}|\Delta_h|}}\right)\times\right.\right.\\
&\quad\left.\left.\times\left(e\left(-\frac{t^2\sum_{h=1}^{N_j}\left(\sum_{k\in\Delta_h} f(m_k x)\right)^2}{2\sum_{h=1}^{N_j}|\Delta_h|}\right)-e^{-t^2\psi(x)/2}\right)\right)\right| + N_j^{-1/2}\\
\ll &\ \mathbb{E}|e^{W_j}-1| + N_j^{-1/2}\\
&+\mathbb{E}\left|1-e\left(\frac{t^2\sum_{h=1}^{N_j}\left(\sum_{k\in\Delta_h} f(m_k x)\right)^2}{2\sum_{h=1}^{N_j}|\Delta_h|} - t^2\psi(x)/2\right)\right|. \qquad (32)
\end{aligned}$$

Using (31) we obtain

$$\mathbb{E}|e^{W_j}-1| \ll N_j^{-1/4}.$$

The function

$$\psi_j(x) := \sum_{h=1}^{N_j} \left(\sum_{k \in \Delta_h} f(m_k x) \right)^2$$

is a sum of

$$\sum_{h=1}^{N_j} (2|\Delta_h|)^2$$

cos-functions. Since

$$\begin{aligned}
& \sum_{h=1}^{N_j} \left(\sum_{k \in \Delta_h} f(m_k x) \right)^2 \\
= & \sum_{h=1}^{N_j} \sum_{k_1,k_2 \in \Delta_h} \left(\cos 2\pi p^s m_{k_1} x + \cos 2\pi q^s m_{k_1} x\right) \left(\cos 2\pi p^s m_{k_2} x + \cos 2\pi q^s m_{k_2} x\right) \\
= & \sum_{h=1}^{N_j} \sum_{k_1,k_2 \in \Delta_h} \Big(\cos 2\pi (p^s m_{k_1} + p^s m_{k_2}) x + \cos 2\pi (p^s m_{k_1} - p^s m_{k_2}) x \\
& +2 \cos 2\pi (p^s m_{k_1} + q^s m_{k_2}) x + 2 \cos 2\pi (p^s m_{k_1} - q^s m_{k_2}) x \\
& + \cos 2\pi (q^s m_{k_1} + q^s m_{k_2}) x + \cos 2\pi (q^s m_{k_1} - q^s m_{k_2}) x \Big),
\end{aligned}$$

and since the equations

$$\begin{aligned}
p^s m_{k_1} + p^s m_{k_2} &= c \\
p^s m_{k_1} + q^s m_{k_2} &= c \\
q^s m_{k_1} + q^s m_{k_2} &= c
\end{aligned}$$

trivially have only finitely many solutions (k_1, k_2), uniformly in $c \in \mathbb{Z}$, since the equations

$$\begin{aligned}
p^s m_{k_1} - p^s m_{k_2} &= c \\
q^s m_{k_1} - q^s m_{k_2} &= c
\end{aligned}$$

only have finitely many solutions (k_1, k_2), uniformly in $c \in \mathbb{Z}$ (except the trivial solutions $k_1 = k_2$ in case $c = 0$), and since

$$p^s m_{k_1} - q^s m_{k_2} = c$$

has only finitely many solutions, uniformly in $c \in \mathbb{Z} \backslash \left\{ 1, \dots, 2p^s + \frac{2p^{s+1}}{p-q} \right\}$, this implies that for

$$\psi_j(x) = \sum_{r=0}^{\infty} a_r \cos 2\pi r x = \underbrace{\sum_{r=0}^{2p^s + \frac{2p^{s+1}}{p-q}} a_r \cos 2\pi r x}_{=: \psi_j^{(1)}(x)} + \underbrace{\sum_{r = 2p^s + \frac{2p^{s+1}}{p-q} + 1}^{\infty} a_r \cos 2\pi r x}_{=: \psi_j^{(2)}(x)}$$

we have for $r > 2p^s + \frac{2p^{s+1}}{p-q}$

$$|a_r| \ll 1, \quad \text{uniformly in } r, \qquad \text{and therefore} \quad \|\psi_j^{(2)}\| \ll \left(\sum_{h=1}^{N_j} |\Delta_h|^2 \right)^{1/2}. \tag{33}$$

By (25),

$$\frac{\psi_j^{(1)}(x)}{\sum_{h=1}^{N_j} |\Delta_h|} \to \psi(x), \quad \text{as} \quad j \to \infty, \qquad \text{uniformly in } x \in [0,1].$$

Now for the last term in (32) we have

$$\begin{aligned}
& \mathbb{E} \left| 1 - e \left(\frac{t^2 \sum_{h=1}^{N_j} \left(\sum_{k \in \Delta_h} f(m_k x) \right)^2}{2 \sum_{h=1}^{N_j} |\Delta_h|} - \frac{t^2 \psi(x)}{2} \right) \right| \\
= \; & \mathbb{E} \left| 1 - e \left(\frac{t^2 \left(\psi_j^{(1)}(x) + \psi_j^{(2)}(x) \right)}{2 \sum_{h=1}^{N_j} |\Delta_h|} - \frac{t^2 \psi(x)}{2} \right) \right| \\
= \; & \mathbb{E} \left| 1 - e \left(\frac{t^2}{2} \left(\left(\frac{\psi_j^{(1)}(x)}{\sum_{h=1}^{N_j} |\Delta_h|} - \psi(x) \right) + \frac{\psi_j^{(2)}(x)}{2 \sum_{h=1}^{N_j} |\Delta_h|} \right) \right) \right|.
\end{aligned} \tag{34}$$

Since

$$\frac{\psi_j^{(2)}(x)}{\sum_{h=1}^{N_j} |\Delta_h|} \le 4 |\Delta_{N_j}|, \quad \text{uniformly in } x,$$

and since by (33)

$$\mathbb{P} \left\{ x \in (0,1) : \ \psi_j^{(2)}(x) > \left(\sum_{h=1}^{N_j} |\Delta_h| \right)^{3/4} \right\} \ll N_j^{-1/2},$$

we obtain that (34) is

$$\ll \mathbb{E} \left| 1 - e \left(\frac{t^2}{2} \left(\underbrace{\left| \frac{\psi_j^{(1)}(x)}{\sum_{h=1}^{N_j} |\Delta_h|} - \psi(x) \right| + \frac{\left(\sum_{h=1}^{N_j} |\Delta_h| \right)^{3/4}}{\sum_{h=1}^{N_j} |\Delta_h|}}_{\to 0 \text{ uniformly in } x} \right) \right) \right| + \underbrace{e^{t^2 |\Delta_{N_j}|} N_j^{-1/2}}_{\to 0}.$$

Combining all our estimates, we have shown that for every fixed t

$$\mathbb{E} \left(e \left(\frac{it \sum_{h=1}^{N_j} \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}} \right) \right) \to e^{-t^2 \psi(x)/2}$$

as $j \to \infty$.

It is easy to see that

$$\frac{\sum_{h=1}^{N_j} \sum_{k \in \Delta'_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} (|\Delta_h| + |\Delta'_h|)}} \to 0 \qquad \text{in distribution.}$$

Thus

$$\frac{\sum_{h=1}^{N_j} \sum_{k \in \Delta_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} |\Delta_h|}}$$

and consequently also

$$\frac{\sum_{h=1}^{N_j} \sum_{k \in \Delta_h \cup \Delta'_h} f(m_k x)}{\sqrt{\sum_{h=1}^{N_j} (|\Delta_h| + |\Delta'_h|)}}$$

converge in distribution to a non-Gaussian distribution (as $j \to \infty$). Therefore it is not possible that the distribution of

$$\frac{\sum_{k=1}^{N} f(m_k x)}{\sqrt{N}}$$

converges to a Gaussian distribution, which proves Theorem 2.

References

[1] C. Aistleitner. Diophantine equations and the LIL for the discrepancy of sublacunary sequences. *Illinois J. Math.*, 53(3):785–815, 2009.

[2] C. Aistleitner. On the law of the iterated logarithm for the discrepancy of Lacunary sequences. *Trans. Amer. Math. Soc.*, 362(11):5967–5982, 2010.

[3] C. Aistleitner and I. Berkes. On the central limit theorem for $f(n_k x)$. *Probab. Theory Related Fields*, 146(1-2):267–289, 2010.

[4] I. Berkes. A central limit theorem for trigonometric series with small gaps. *Z. Wahrsch. Verw. Gebiete*, 47(2):157–161, 1979.

[5] I. Berkes, W. Philipp, and R. F. Tichy. Empirical processes in probabilistic number theory: the LIL for the discrepancy of $(n_k \omega)$ mod 1. *Illinois J. Math.*, 50(1-4):107–145 (electronic), 2006.

[6] I. Berkes, W. Philipp, and R. F. Tichy. Metric discrepancy results for sequences $\{n_k x\}$ and Diophantine equations. In *Diophantine approximation*, volume 16 of *Dev. Math.*, pages 95–105. SpringerWienNewYork, Vienna, 2008.

[7] S. G. Bobkov and F. Götze. On the central limit theorem along subsequences of noncorrelated observations. *Teor. Veroyatnost. i Primenen.*, 48(4):745–765, 2003.

[8] S. G. Bobkov and F. Götze. Concentration inequalities and limit theorems for randomized sums. *Probab. Theory Related Fields*, 137(1-2):49–81, 2007.

[9] H. Burkill and L. Mirsky. Monotonicity. *J. Math. Anal. Appl.*, 41:391–410, 1973.

[10] S. D. Chatterji. A principle of subsequences in probability theory: the central limit theorem. *Advances in Math.*, 13:31–54; correction, ibid. 14 (1974), 266–269, 1974.

[11] M. Drmota and R. F. Tichy. *Sequences, discrepancies and applications*, volume 1651 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1997.

[12] Y. Dupain and V. T. Sós. On the one-sided boundedness of discrepancy-function of the sequence $\{n\alpha\}$. *Acta Arith.*, 37:363–374, 1980.

[13] P. Erdös and I. S. Gál. On the law of the iterated logarithm. I, II. *Nederl. Akad. Wetensch. Proc. Ser. A.* **58** = *Indag. Math.*, 17:65–76, 77–84, 1955.

[14] K. Fukuyama. A central limit theorem for trigonometric series with bounded gaps. *Probab. Theory Related Fields*, to appear.

[15] K. Fukuyama. A central limit theorem and a metric discrepancy result for sequences with bounded gaps. In *Dependence in Probability, Analysis and Number Theory*, pages 233–246. Kendrick Press, 2010.

[16] K. Fukuyama and K. Nakata. A metric discrepancy result for the Hardy-Littlewood-Pólya sequences. *Monatsh. Math.*, 160(1):41–49, 2010.

[17] K. Fukuyama and B. Petit. Le théorème limite central pour les suites de R. C. Baker. *Ergodic Theory Dynam. Systems*, 21(2):479–492, 2001.

[18] V. F. Gapoškin. Lacunary series and independent functions. *Uspehi Mat. Nauk*, 21(6 (132)):3–82, 1966.

[19] V. F. Gapoškin. The central limit theorem for certain weakly dependent sequences. *Teor. Verojatnost. i Primenen.*, 15:666–684, 1970.

[20] M. Kac. Probability methods in some problems of analysis and number theory. *Bull. Amer. Math. Soc.*, 55:641–665, 1949.

[21] H. Kesten. The discrepancy of random sequences $\{kx\}$. *Acta Arith.*, 10:183–213, 1964/1965.

[22] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences.* Wiley-Interscience [John Wiley & Sons], New York, 1974. Pure and Applied Mathematics.

[23] J. Nešetřil. Ramsey theory. In *Handbook of combinatorics, Vol. 1, 2*, pages 1331–1403. Elsevier, Amsterdam, 1995.

[24] W. Philipp. Limit theorems for lacunary series and uniform distribution mod 1. *Acta Arith.*, 26(3):241–251, 1974/75.

[25] W. Philipp. Empirical distribution functions and strong approximation theorems for dependent random variables. A problem of Baker in probabilistic number theory. *Trans. Amer. Math. Soc.*, 345(2):705–727, 1994.

[26] R. Salem and A. Zygmund. On lacunary trigonometric series. *Proc. Nat. Acad. Sci. U. S. A.*, 33:333–338, 1947.

[27] R. Salem and A. Zygmund. La loi du logarithme itéré pour les séries trigonométriques lacunaires. *Bull. Sci. Math. (2)*, 74:209–224, 1950.

[28] T. Sanders. On Roth's theorem on progressions. *Ann. of Math., to appear.*

[29] J. Schoissengeier. On the discrepancy of $(n\alpha)$. *Acta Arith.*, 44(3):241–279, 1984.

[30] J. Schoissengeier. A metrical result on the discrepancy of $(n\alpha)$. *Glasgow Math. J.*, 40(3):393–425, 1998.

[31] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[32] S. Takahashi. A gap sequence with gaps bigger than the Hadamards. *Tôhoku Math. J. (2)*, 13:105–111, 1961.

[33] A. Zygmund. *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. Reprint of the 1979 edition.