Forschungsschwerpunkt
Algorithmen und mathematische Modellierung

Metric discrepancy results for sequences $\{n_k x\}$ and diophantine equations

István Berkes, Walter Philipp, and Robert Tichy

Institut für Analysis und Computational Number Theory (Math A)
Institut für Statistik

Report 2005-6, November 2005

Metric discrepancy results for sequences $\{n_k x\}$ and diophantine equations

István Berkes, Walter Philipp and Robert F. Tichy

Dedicated to Professor Wolfgang M. Schmidt on the occasion of his 70^{th} birthday

Abstract

We establish a law of the iterated logarithm for the discrepancy of sequences $(n_k x)$ mod 1 where (n_k) is a sequence of integers satisfying a sub-Hadamard growth condition and such that one and four-term Diophantine equations in the variables n_k do not have too many solutions. The conditions are discussed, the probabilistic details of the proof are given elsewhere. As a corollary to our results, the asymptotic behavior of sums $\sum f(n_k x)$ is obtained.

1 Introduction

Let (n_k) be an increasing sequence of positive integers. For $0 \le x \le 1$ set

$$\eta_k = \eta_k(x) := n_k x \qquad (\text{mod } 1)\,. \tag{1}$$

The discrepancy of the first N elements of the sequence (η_k) is defined as

$$D_N = D_N(x) := \sup_{0 \le s \le 1} \left| \frac{1}{N} \operatorname{card}\,(k \le N : \eta_k(x) \le s) - s \right| . \tag{2}$$

In his fundamental paper on uniform distribution mod 1 H. Weyl [23] proved, among many other things, that $D_N(x) \to 0$ for almost all $x \in (0,1)$, i.e. that $(n_k x)$ is uniformly distributed mod 1 for all $x \in (0,1)$ except for a set of Lebesgue measure zero. This was later improved independently by Cassels [5] and by Erdős and Koksma [7] who proved that for almost all $x \in (0,1)$

$$ND_N(x) = O(N^{\frac{1}{2}} (\log N)^{\frac{5}{2}+\varepsilon}), \qquad \varepsilon > 0.$$

The best result so far has been achieved by R.C. Baker [1] who reduced the exponent $\frac{5}{2}$ of the logarithm to $\frac{3}{2}$. The exact exponent of the logarithm is still an open problem, except for the fact that it cannot be less than $\frac{1}{2}$, as was shown by Berkes and Philipp [4].

Determining the exact order of magnitude of $D_N(x)$ for a concrete sequence (n_k) is generally a hard problem and a satisfactory solution exists only in a few special cases. In the case $n_k = k$, Kesten [15] proved that

$$ND_N(x) \sim \frac{2}{\pi^2} \log N \log\log N$$

in measure. (For the remainder term, see Schoissengeier [20].) Another important case when a sharp bound for the magnitude of $D_N(x)$ is known is the case when (n_k) is a lacunary sequence. Philipp [17], [18] proved that if (n_k) satifies the Hadamard gap condition

$$\frac{n_{k+1}}{n_k} \geq 1+\rho, \quad \rho > 0,\ k=1,2,\dots \tag{3}$$

then we have for almost all x

$$\frac{1}{4} \leq \limsup_{N\to\infty} \frac{ND_N(x)}{\sqrt{N\log\log N}} \leq C(\varrho), \tag{4}$$

where $C(\varrho) \ll \frac{1}{\varrho}$. This result has an obvious probabilistic flavor. If (x_n) is a sequence of independent random variables uniformly distributed (in the probabilistic sense) over $(0,1)$, then by the classical Chung-Smirnov law of the iterated logarithm for empirical distribution functions (see e.g. Shorack and Wellner [21], p. 504), the discrepancy D_N^* of $(x_n, n \leq N)$ satisfies

$$\limsup_{N\to\infty} \frac{ND_N^*}{\sqrt{N\log\log N}} = \frac{1}{\sqrt{2}} \tag{5}$$

with probability one. Thus, roughly speaking, under the Hadamard gap condition (3) the sequence $n_k x$ (mod 1) behaves like a sequence of independent random variables. This heuristics, which plays an important role in harmonic analysis and is the key for understanding a number of interesting phenomena (see e.g. Kac [13]), should be used, however, with great care. Berkes and Philipp [4] have constructed sequences (n_k) satisfying (3), for which the lower bound $\frac{1}{4}$ in (4) can be improved to $c\log\log\frac{1}{\varrho}$ with an absolute constant c. Hence there cannot be an upper bound (4), independent of ϱ, that works for all sequences (n_k) satisfying a Hadamard gap condition (3). A deeper analysis of the problem shows that under the Hadamard gap condition (3), the behavior of D_N is determined by a delicate interplay between the speed of growth and the number-theoretic properties of (n_k). If (n_k) grows extremely rapidly, then $\{n_k x\}$ is indeed a nearly i.i.d. (independent, identically distributed) sequence of random variables as one can easily see from the mixing relation

$$\lim_{n\to\infty} |x \in (\alpha,\beta) : \{nx\} \leq t| = (\beta-\alpha)t \qquad (0 \leq \alpha < \beta \leq 1),$$

where $|\cdot|$ stands for the Lebesgue measure. See Philipp [18], Lemma 4.2.1, where a remainder term is also given. Specifically, if

$$\sum_{k=1}^{\infty} n_k/n_{k+1} < \infty,$$

then the limsup in (4) is $1/\sqrt{2}$, in accordance with (5). (This follows easily using the approximation method in Berkes [2].) If, however, n_{k+1}/n_k is bounded, then the arithmetic structure of (n_k) comes into play. The number-theoretic effect becomes particularly clear if in (2) we compute the right hand side for a single s only (without the sup), i.e., if we study the behavior of the sum

$$\sum_{k\leq N} f(n_k x), \tag{6}$$

where $f = I_{(0,s)} - s$, extended with period 1. (Here, and in the sequel, $I_A(\cdot)$ denotes the indicator function of the set A.) In the case $n_k = 2^k$ the sum in (6) is asymptotically normally distributed, as shown by Kac [12]. The corresponding LIL (law of the iterated logarithm)

$$\limsup_{N\to\infty} \frac{1}{\sqrt{N\log\log N}} \sum_{k\le N} f(n_k x) = \gamma \quad \text{a.e.} \tag{7}$$

is also valid, where $\gamma = \gamma(s)$ is an explicitly computable constant (see e.g. Berkes and Philipp [3]). Thus in this case the behavior of (6) is the same as that of sums of independent random variables. On the other hand, Erdős and Fortet showed (see [13], p. 646) that for $n_k = 2^k - 1$ both the central limit theorem and the LIL (7) break down; in fact, the limsup in (7) is not any more a constant almost everywhere. This interesting phenomenon was cleared up by Gaposhkin [11], who showed that the sum in (6) satisfies the central limit theorem for all "nice" functions f if and only if for any fixed nonzero integers a, b, c the number of solutions of the Diophantine equation

$$an_\nu + bn_\mu = c$$

is bounded by a constant $C = C(a, b)$. For concrete sequences (n_k), this criterion is usually not easy to verify, but one can give simple sufficient criteria for its validity. For example, the above Diophantine condition is satisfied if for any rational r and any positive integer sequences k_m, l_m tending to infinity, the limit relation

$$\lim_{m\to\infty} n_{k_m}/n_{l_m} = r$$

can hold only if the fraction on the left side *equals* r for $m \ge m_0$. In particular, the criterion is satisfied in each of the following cases:

(a) $\lim_{k\to\infty} n_{k+1}/n_k = \infty$
(b) $n_k | n_{k+1}$ for any k
(c) $\lim_{k\to\infty} n_{k+1}/n_k = \alpha$, where α^r is irrational for $r = 1, 2, \ldots$

In the case when Gaposhkin's Diophantine condition is satisfied, the corresponding LIL (7) is also valid (see Berkes and Philipp [3]).

The previous results give a fairly complete picture on the discrepancy $D_N(x)$ and the underlying probabilistic structure of $n_k x$ mod 1 in the case when (n_k) satisfies the Hadamard gap condition (3). The purpose of this paper is to study the same problem for sub-Hadamard sequences, i.e. when $n_{k+1}/n_k \to 1$. This problem is considerably harder than the Hadamard case and very few results are known here. Philipp [19] proved, verifying a conjecture of R. C. Baker, that the LIL (4) holds for all Hardy-Littlewood-Pólya sequences (n_k). These are defined as follows. Let $(q_1, q_2, \ldots, q_\tau)$ be a finite set of coprime positive integers and let (n_k) be the multiplicative semigroup generated by $(q_1, q_2, \ldots, q_\tau)$ and arranged in increasing order. Thus

$$(n_k)_{k=1}^\infty = (q_1^{\alpha_1} q_2^{\alpha_2} \ldots q_\tau^{\alpha_\tau},\ \alpha_i \ge 0,\ 1 \le i \le \tau).$$

Then relation (4) holds with a constant $C(r)$ on the right side depending only on the number r of primes involved in the prime factorization of $q_1 \ldots, q_\tau$. Note that Hardy-Littlewood-Pólya sequences grow fairly rapidly: they are subexponential, but satisfy the gap condition

$$n_{k+1} - n_k \ge \frac{n_k}{(\log n_k)^\gamma} \qquad (k = 1, 2, \ldots)$$

for some $\gamma > 0$. (See Tijdeman [22].) The proof of Lemma 3 below will also show (cf. relation (15)) that $n_k \geq \exp(c_1 k^\alpha)$ for some $0 < \alpha < 1$.

In the opposite direction one can show (see Berkes and Philipp [4]) that for any $\varepsilon_k \to 0$, there exists a sequence (n_k) of integers satisfying

$$n_{k+1}/n_k \geq 1 + \varepsilon_k, \quad k = 1, 2, \dots$$

such that

$$\limsup_{N\to\infty} \frac{ND_N(x)}{\sqrt{N \log\log N}} = \infty$$

for almost every x. Hence no subexponential speed of growth can guarantee, by itself, the law of the iterated logarithm (4) for $D_N(x)$ and thus in the sub-Hadamard domain the LIL is an individual affair: its validity depends on the specific properties of the sequence (n_k). While the Hardy-Littlewood-Pólya sequences are the only known examples for LIL behavior (4) in the subexponential domain, in this paper we will see that they represent the rule rather than the exception. Indeed, we will show that the "majority" of sequences (n_k) (in a suitable statistical sense) with a minimal subexponential speed of growth satisfies the LIL, and this is the case for all 'sufficiently irregular' (n_k). These results will follow from Theorem 1 below, the main result of our paper, which gives explicit Diophantine conditions on (n_k) guaranteeing the validity of (4). To formulate the theorem, we introduce the following conditions. We will say that a sequence (n_k) satisfies

Condition A, if for any fixed nonzero integers a, b, c the number of solutions of the Diophantine equation

$$an_\nu + bn_\mu = c$$

is bounded by a constant $C = C(a, b)$.

Condition A*, if (n_k) satisfies Condition **A** with a constant $C(a, b)$ independent of a, b.

Condition B, if there exist constants $0 < \alpha < \frac{1}{2}$ and $C > 0$ such that for each positive integer b and for each $R \geq 1$ the number of solutions (h, n_ν) of the Diophantine equation

$$hn_\nu = b$$

with $h \in \mathbb{N}$, $1 \leq h \leq R$ does not exceed CR^α.

Condition C, if there exist constants $0 < \beta < \frac{3}{2}$ and $C > 0$ such that for each $N \geq 1$ and for all fixed integers h_i with $0 < |h_i| \leq N^4$, $\quad i = 1, 2, 3, 4$ the number of non-degenerate solutions of the Diophantine equation

$$h_1 n_{\nu_1} + h_2 n_{\nu_2} + h_3 n_{\nu_3} + h_4 n_{\nu_4} = 0 \tag{8}$$

subject to

$$1 \leq \nu_i \leq N, \qquad i = 1, 2, 3, 4 \tag{9}$$

does not exceed CN^β.

Condition G, if there exists a constant $0<\eta<1$ such that for all $k \ge k_0 = k_0(\eta)$ we have

$$n_{k+k^{1-\eta}}/n_k \ge k. \tag{10}$$

Here, and in the sequel, n_j is meant as $n_{[j]}$ if j is not an integer. Condition **A** is Gaposhkin's necessary and sufficient condition for the nearly independent behavior of $\{n_k x\}$ in the Hadamard case; Condition **A*** is the uniform version of **A**. Conditions **B** and **C** are analogous Diophantine conditions which play a basic role in the subexponential domain. Condition **G** is the growth condition we will assume throughout this paper; it implies $n_k \ge \exp(k^\eta)$ and thus it restricts our investigations to a zone under the exponential speed. It is easy to see that condition **G** is implied by the gap condition

$$n_{k+1}/n_k \ge 1 + ck^{-\alpha} \qquad (k=1,2,\ldots) \tag{11}$$

for any $0<\eta<1-\alpha$ but, unlike (11), relation (10) does not require that all individual gaps $n_{k+1}-n_k$ are large.

With the above notations, we can formulate now our main result.

Theorem 1 *Assume (n_k) is an increasing sequence of positive integers satisfying conditions* **B**, **C** *and* **G**. *Then there is a constant D, depending only on the constants α, β, η and C appearing in these conditions such that for almost all $x \in (0,1)$*

$$\frac{1}{4} \le \limsup_{N\to\infty} \frac{ND_N(x)}{\sqrt{N \log\log N}} \le D.$$

In Section 2 we give some comments on the conditions **B**, **C** and **G**. We will show that they are satisfied by a wide class of sequences including the Hardy-Littlewood-Pólya sequences as well as 'almost all' sequences (in some natural sense). This requires application of a recent version the subspace theorem due to Evertse, Schlickewei and Schmidt [8]. The details of the proof including the whole probabilistic machinery, such as martingale inequalities and chaining arguments will be given in a subsequent paper, which will also extend and simplify the methods used in Philipp [19].

The Diophantine condition in Theorem 1 involves equations of the type

$$a_1 n_{\nu_1} + \ldots + a_p n_{\nu_p} = b \tag{12}$$

with arbitrary nonzero coefficients $a_1, \ldots, a_p$. Equations of this type with coefficients ± 1 play an important role in the study of lacunary exponential sums $\sum e^{2\pi i h n_k x}$; the idea goes back to Sidon (see e.g. Kahane [14], Gaposhkin [10]). Bounds on $\sum e^{2\pi i h n_k x}$ lead, in turn, to bounds on the discrepancy $D_N(x)$, in view of the Erdős-Turán inequality. Hence bounding the number of solutions of the Diophantine equation (12) with coefficients ± 1 will also lead to bounds for the discrepancy $D_N(x)$ of $n_k x$, although these bounds are necessarily cruder than the LIL (4), due to the "usual defect" of the Erdős-Turán inequality. However, for comparison with classical Sidon theory, we formulate here one such result.

Theorem 2 *Let (n_k) be an increasing sequence of positive integers and let $p \geq 2$. Assume that there exists a constant $C > 0$, depending on p and the sequence (n_k), such that the number of solutions of the Diophantine equation*

$$\pm n_{\nu_1} \pm \ldots \pm n_{\nu_p} = b$$

is at most C for any $b \neq 0$. Then

$$ND_N(x) = O(N^{1/2}(\log N)^{1+\frac{1}{p}})$$

for almost every x.

As a byproduct of the proof of Theorem 1, we can get precise asymptotic results for lacunary sums (6) not only for centered indicator functions f, but for any function f of bounded variation (BV). In particular, it is easy to get the central limit theorem and the law of the iterated logarithm (with a precise constant) for such sums. These results can be proved much simpler than Theorem 1 itself. For example, the LIL for the sums (6) expresses the LIL for the discrepancy of $(n_k x)$, where we consider a single interval $(0, s)$, instead of taking the sup for all intervals in (2). This special case can be handled relatively easily; the main difficulty in the proof of Theorem 1 is to get the uniformity over all subintervals of $(0, 1)$. Since sums $\sum f(n_k x)$ received considerable attention in harmonic analysis, it is worth formulating such a corollary of Theorem 1 and comparing it with the earlier theory.

Theorem 3 *Let (n_k) be an increasing sequence of positive integers satisfying conditions* **C** *and* **G**. *Assume that*

$$f(x+1) = f(x), \quad \int_0^1 f(x)dx = 0, \quad f \in BV(0,1)$$

and

$$\int_0^1 \left(\sum_{k \leq N} f(n_k x) \right)^2 dx \sim \sigma^2 N$$

for some $\sigma > 0$. Then

$$\frac{1}{\sqrt{N}} \sum_{k \leq N} f(n_k x) \to \mathcal{N}(0, \sigma^2) \qquad \textit{in distribution}$$

and

$$\limsup_{N \to \infty} \frac{1}{(N \log \log N)^{1/2}} \sum_{k \leq N} f(n_k x) = \sigma \sqrt{2} \quad \text{a.e.}$$

As we noted before, the asymptotic behavior of $\sum_{k \leq N} f(n_k x)$ is completely known in the case when (n_k) satisfies the Hadamard gap condition (3). In particular, a necessary and sufficient condition for the CLT (Central Limit Theorem) is the Diophantine condition **A**. On the other hand, practically nothing is known in the sub-Hadamard case, with the exception of a recent CLT of Fukuyama and Petit [9] concerning the Hardy-Littlewood-Pólya sequence. Theorem 3 gives a fairly complete description of the asymptotics of $\sum_{k \leq N} f(n_k x)$ under the sub-Hadamard growth condition **G**. Note that instead of the two-term Diophantine condition **A** we assumed here the four-term condition **C**, but it is easy to see that Theorem 3 remains valid if condition **C** is replaced by $\mathbf{A}^*$. Hence a condition very close to **A** is essentially the right Diophantine assumption for the CLT also in the subexponential domain. A more detailed investigation of the situation will be given elsewhere.

In conclusion we make some comments on the permutation-invariance of our results. Changing the order of terms of a sequence (x_n) leads generally to a drastic change of its discrepancy (see e.g. Kuipers and Niederreiter [16]). Thus it is unclear what happens in Theorem 1 if we permute the terms of the sequence (n_k). On the other hand, the usual heuristics behind our theorems says that the sequence $n_k x$ mod 1 behaves like a sequence of independent, identically distributed random variables. As the i.i.d. character of a sequence is permutation-invariant, it is natural to expect that our results remain valid after any permutation of (n_k). In the case when (n_k) satisfies the Hadamard gap condition (3), this is indeed the case: the proof in Philipp [17] uses the multiplicative orthogonality of lacunary trigonometric series and thus it is permutation-invariant. However, the martingale method in the proof of Theorem 1 uses the increasing character of (n_k) in an essential way, and thus the permutation-invariance of Theorem 1 remains open. Using different methods, permutation-invariance can be proved under a stronger form of the Diophantine condition **C**; we shall prove this in a subsequent paper.

2 Comments on conditions B, C and G

We first show that the Hardy-Littlewood-Pólya sequences satisfy conditions **B**, **C** and **G**.

Lemma 1 *Let (n_k) be a Hardy-Littlewood-Pólya sequence. Then there is a constant $C > 0$ such that for each positive integer b and for each $R \geq 1$ the number of solutions (h, n_ν) of the Diophantine equation*

$$hn_\nu = b \tag{13}$$

with $h \in \mathbb{N}$, $1 \leq h \leq R$ does not exceed $C(\log R)^r$. Here r is the number of primes involved in the prime factorization of $q_1, \ldots, q_\tau$.

Proof: Let $p_1, \ldots, p_r$ be the primes appearing in the prime factorization of $q_1, \ldots, q_\tau$ and write $b = p_1^{\alpha_1} \ldots p_r^{\alpha_r} M$, where M is not divisible by $p_1, \ldots, p_r$. Then n_ν in (13) has the form $n_\nu = p_1^{\beta_1} \ldots p_r^{\beta_r}$ with integers $\beta_i \geq 0$ and thus (13) implies that $\beta_i \leq \alpha_i$, $i = 1, \ldots, r$ and

$$h = p_1^{\alpha_1 - \beta_1} \ldots p_r^{\alpha_r - \beta_r} M = p_1^{\delta_1} \ldots p_r^{\delta_r} M$$

with integers $\delta_i \geq 0$. Now $h \leq R$ implies $p_1^{\delta_1} \ldots p_r^{\delta_r} \leq R/M \leq R$ and consequently $\delta_i \leq \log R / \log 2$, $i = 1, \ldots, r$. Thus the number of r-tuples $(\delta_1, \ldots, \delta_r)$ and consequently the number of h's that can possibly yield a candidate h for a solution (h, n_ν) of (13) is at most $(1 + \log R / \log 2)^r$. As h in (13) determines ν uniquely, Lemma 1 is proved.

■

Lemma 2 *A Hardy-Littlewood-Pólya sequence satisfies condition* **C** *with* $\beta = 1$ *and*

$$C = \exp(18^9(\tau + 1)),$$

where τ *denotes the number of generating elements of the sequence.*

Proof: The number of choices for ν_4 in (8) is N and thus the lemma follows from Theorem 1.1 of Evertse-Schlickewei-Schmidt [8] upon fixing ν_4 and dividing (8) by $h_4 n_{\nu_4}$.

■

Lemma 3 *A Hardy-Littlewood-Pólya sequence* (n_k) *satisfies condition* **G**.

Proof: Let $q_1, \ldots, q_\tau$ be the generating elements of (n_k). Clearly, an element $n_j = q_1^{\delta_1} \ldots q_\tau^{\delta_\tau}$ of the sequence (n_k) satisfies $n_j \le R$ iff

$$\delta_1 \log q_1 + \ldots + \delta_\tau \log q_\tau \le \log R$$

and thus the number $A(R)$ of elements of (n_k) in $[0, R]$ equals the number of lattice points $(\delta_1, \ldots, \delta_\tau)$ in the τ-dimensional 'tetrahedron'

$$x_1 \log q_1 + \ldots + x_\tau \log q_\tau \le \log R, \qquad x_1 \ge 0, \ldots, x_\tau \ge 0.$$

The volume of the tetrahedron is $c_1(\log R)^\tau$ where

$$c_1 = c_1(\tau) = \frac{1}{\tau! \log q_1 \ldots \log q_\tau}$$

and thus by a well known argument in analysis we have, as $R \to \infty$,

$$A(R) = c_1(\log R)^\tau + O((\log R)^{\tau-1}). \tag{14}$$

From (14) and the trivial relation $A(n_k) = k$ we get

$$\log n_k \sim \left(\frac{k}{c_1}\right)^{1/\tau}. \tag{15}$$

Formulas (14) and (15) and $\log kn_k \sim \log n_k$ imply that the number of n_j's in the interval $[n_k, kn_k]$ is

$$c_1[(\log kn_k)^\tau - (\log n_k)^\tau] + O((\log kn_k)^{\tau-1}) \sim c_1\tau(\log k)(\log n_k)^{\tau-1} \sim c_2 k^{(\tau-1)/\tau} \log k$$

as $k \to \infty$. Thus for $k \ge k_0$ we have

$$n_{k+2c_2k^{(\tau-1)/\tau}\log k} \ge kn_k$$

and consequently (10) holds with any $\eta < 1/\tau$.

■

We now show that, in some sense, almost all sequences (n_k) growing like a polynomial with a fixed large degree will satisfy conditions **B** and **C**. We shall construct these sequences by induction. Let $n_1 = 1$ and suppose that $n_1 < n_2 < \ldots < n_{k-1}$ have already been constructed and satisfy

$$(j-1)^{50} < n_j \le j^{50} \qquad j = 1, 2, \ldots, k-1. \tag{16}$$

Then the cardinality of the set of integers of the form

$$a_1 n_{\mu_1} + a_2 n_{\mu_2} + a_3 n_{\mu_3}$$

with $1 \le \mu_1, \mu_2, \mu_3 \le k-1$, $|a_1|, |a_2|, |a_3| \le k^{11}$ is at most $(2k^{11}+1)^3 (k-1)^3 = O(k^{36})$. Hence the cardinality of the set of integers included in the set of rational numbers

$$\frac{1}{a}(a_1 n_{\mu_1} + a_2 n_{\mu_2} + a_3 n_{\mu_3}), \quad a \in \mathbb{Z} - \{0\},\ |a| \le k^{11} \tag{17}$$

subject to $1 \le \mu_1, \mu_2, \mu_3 \le k-1,\ |a_1|, |a_2|, |a_3| \le k^{11}$ is $O(k^{47})$. Thus, the interval $((k-1)^{50}, k^{50}]$ contains at most that many integers of the form (17). This number is at most $O(1/k^2)$ times the total number of integers in the interval. Calling these numbers "bad", we choose now n_k from the "good" integers (which constitute an overwhelming majority for k large), and note that (16) is satisfied also for $j = k$. This construction yields an infinite increasing sequence (n_k) with the property that for $k \ge k_0$ the Diophantine equation

$$a_1 n_{\mu_1} + a_2 n_{\mu_2} + a_3 n_{\mu_3} + a_4 n_{\mu_4} = 0 \tag{18}$$

with $1 \le \mu_i \le k,\ i = 1, 2, 3, 4$ and $\max(|a_i|, i = 1, 2, 3, 4) \le k^{11}$ has no solution if one of the indices, say μ_4, equals k and the corresponding factor $a_4 \ne 0$, while the other three indices μ_i are strictly less than k. Call this property **NS** (for "no solutions").

We now show that the constructed sequence (n_k) satisfies condition **C**. Let $N \ge N_0$ be given and consider (8) subject to $0 < |h_i| \le N^4,\ i = 1, 2, 3, 4$, fixed and $1 \le \nu_1 \le \nu_2 \le \nu_3 \le \nu_4 \le N$. We can assume without loss of generality that $\nu_4 > 3N^{\frac{4}{11}}$, since otherwise (8) can have only $(3N^{\frac{4}{11}})^4 = O(N^\beta)$ solutions where $\beta = 16/11 < 3/2$. We now distinguish cases regarding the relative size of the indices ν_i. If $\nu_4 > \nu_3$, we set $k = \nu_4$. Then using property **NS** if follows that (8) has no solutions subject to $0 < |h_i| \le N^4$, since then $|h_i| \le N^4 < k^{11}$ by $k = \nu_4 > 3N^{\frac{4}{11}}$. (Note that the validity of **NS** has been established only for $k \ge k_0$, but by $k = \nu_4 > 3N^{\frac{4}{11}}$ this is satisfied if $N \ge N_0$. For the finitely many remaining values $1 \le N < N_0$ condition **C** is trivially satisfied.) If $\nu_4 = \nu_3 > \nu_2$ then (8) reduces to

$$h_1 n_{\nu_1} + h_2 n_{\nu_2} + h^* n_{\nu_4} = 0 \tag{19}$$

with $h^* = h_3 + h_4 \ne 0$ since otherwise the proper subsum $h_3 n_{\nu_3} + h_4 n_{\nu_4}$ would vanish and thus the solution $(n_{\nu_1}, n_{\nu_2}, n_{\nu_3}, n_{\nu_4})$ of (8) would be degenerate. In (18) we set $a_1 = h_1, a_2 = h_2, a_3 = 0,\ a_4 = h^*$ and $\mu_4 = k$, and conclude that by property **NS**, (19) has no solutions since $|h^*| \le 2N^4 < k^{11}$. If $\nu_4 = \nu_3 = \nu_2 > \nu_1$, then (8) reduces to

$$h_1 n_{\nu_1} + h^{**} n_{\nu_4} = 0 \tag{20}$$

with $h^{**} = h_2 + h_3 + h_4 \ne 0$ since otherwise we would have $h_1 = 0$, contrary to the assumption. In (18) we set $a_1 = h_1, a_2 = a_3 = 0, a_4 = h^{**}$ and $\mu_4 = k$, and conclude that (20) cannot have a solution since $|h^{**}| \leq 3N^4 < k^{11}$. Finally, if $\nu_4 = \nu_3 = \nu_2 = \nu_1$ then there are only N possibilities for the 4-tuple $(\nu_1, \nu_2, \nu_3, \nu_4)$ and thus for fixed h_i the number of solutions of (8) is at most N, regardless of the restrictions on h_i.

To verify that (n_k) also satisfies condition **B**, let $R \geq 1$, $b \geq 1$ be given and consider the equation $hn_\nu = b$ with $h \in \mathbb{N}$, $1 \leq h \leq R$. If this equation has a solution (h, n_ν) at all, let (l, n_μ) denote its solution with the largest μ. This means we have to study the Diophantine equation

$$hn_\nu - ln_\mu = 0 \tag{21}$$

subject to $\nu \leq \mu$ and $1 \leq h \leq R$. Set $k = \mu$. If $k \leq R^{\frac{1}{4}}$ then $\nu \leq R^{\frac{1}{4}}$, and since ν uniquely determines h in (21), in this case the number of solutions (h, n_ν) of (21) does not exceed $R^{\frac{1}{4}}$, regardless of the restriction on h. If $k > R^{\frac{1}{4}}$, then by property **NS**, equation (21) has no solutions other than $h = l$, $\nu = \mu$, since from $\nu < \mu$, $1 \leq h \leq R$ it follows $1 \leq l \leq h \leq R \leq k^4 \leq k^{11}$. Thus the impossibility of a solution follows by setting in (18) $a_1 = a_2 = 0$, $a_3 = h$ and $a_4 = -l$.

If instead of (16) we require $n_j \in I_j$, where $I_1, I_2, \ldots$ are disjoint intervals on the positive line, each lying to the right of the preceding one, then the same construction will work as long as the length $|I_j|$ of the interval I_j satisfies $|I_j| \geq j^{49}$. Specifically, if $n_1 < n_2 < \ldots < n_{k-1}$ are given, the number of "bad" choices for n_k in the interval I_k is $O(1/k^2)$ times the total number of integers in the interval, and thus if we choose n_k at random, uniformly among all integers in the interval I_k, then the Borel-Cantelli lemma shows that with probability one, all choices for $k \geq k_0$ will be "good". Hence the above construction yields the following

Corollary Let $I_1, I_2, \ldots$ be disjoint intervals on the positive line, each lying to the right of the preceding one, such that $|I_k| \geq k^{49}$, $k = 1, 2, \ldots$ and let (n_k) be a random sequence such that n_k is uniformly distributed over the integers of the interval I_k. Then (n_k) satisfies conditions **B** and **C** with probability one.

With a proper choice of the intervals I_k we can "regulate" the speed of growth of (n_k); in fact, we can guarantee an arbitrarily prescribed speed of growth provided this exceeds that of k^γ, γ large. Specifically, if $\phi(k)$, $k \geq 0$ is a sequence of integers with $\phi(0) = 0$ and $\phi(k) - \phi(k-1) > 2k^{49}$, $k = 1, 2, \ldots$, then choosing $I_k = [\phi(k) - k^{49}, \phi(k) + k^{49}]$ will imply $n_k \sim \phi(k)$. In particular, we can guarantee the validity of Condition **G** as well.

References

[1] R. C. Baker, Metric number theory and the large sieve, J. London Math. Soc. (2), **24** (1981), 34-40.

[2] I. Berkes, On almost i.i.d. subsequences of the trigonometric system, Proc. Funct. Anal. Sem., University of Texas at Austin, Lecture Notes Math. **1332** (1987), 54-63.

[3] I. Berkes and W. Philipp, An a.s. invariance principle for lacunary series $f(n_k x)$, Acta Math. Acad. Sci. Hung. **34** (1979), 141-155.

[4] I. Berkes and W. Philipp, The size of trigonometric and Walsh series and uniform distribution mod 1, J. London Math. Soc. (2), **50** (1994), 454-464.

[5] J. W. S. Cassels, Some metrical theorems in Diophantine approximation III, Proc. Cambridge Philos. Soc. **46** (1950), 219-225.

[6] M. Drmota and R. F. Tichy, Sequences, Discrepancies and Applications, Lecture Notes Math. **1651**, Springer, Berlin, Heidelberg, New York, 1997.

[7] P. Erdős and J. F. Koksma, On the uniform distribution modulo 1 of sequences $(f(n, \vartheta))$, Proc. Kon. Nederl. Akad. Wetensch. **52** (1949), 851-854.

[8] J.-H. Evertse, R. H.-P. Schlickewei and W. M. Schmidt, Linear equations in variables which lie in a multiplicative group, Ann. of Math. (2) **155** (2002), no. 3, 807–836.

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

[10] V. F. Gaposhkin, Lacunary series and independent functions (in Russian), Uspehi Mat. Nauk **21/6** (1966), 3-82.

[11] V. F. Gaposhkin, On the central limit theorem for some weakly dependent sequences (in Russian), Teor. Verojatn. Prim. **15** (1970), 666-684.

[12] M. Kac, On the distribution of values of sums of type $\sum f(2^k t)$, Ann. Math. **47** (1947), 33-49.

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

[14] J. Kahane, Some Random Series of Functions, Second Edition, Cambridge University Press, 1985.

[15] H. Kesten, The discrepancy of random sequences $\{kx\}$, Acta Arith. **10** (1964/65), 183-213.

[16] L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences, Wiley, New York, 1974.

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

[18] W. Philipp, A functional law of the iterated logarithm for empirical distribution functions of weakly dependent random variables, Ann. Probab. **5** (1977), 319-350.

[19] 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** (1994), 707-727.

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

[21] R. Shorack and J. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.

[22] R. Tijdeman, On integers with many small prime factors, Compositio Math. **26** (1973), 319-330.

[23] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, Math. Ann. **77** (1916), 313-352.

I. Berkes, Institut für Statistik, Technische Universität Graz, Steyrergasse 17,
A-8010 Graz, Austria

W. Philipp, Department of Statistics, University of Illinois at Urbana-Champaign,
725 S. Wright Street, Champaign, IL 61820, USA

R.F. Tichy, Institut für Mathematik A, Technische Universität Graz, Steyrergasse 30,
A-8010 Graz, Austria