Available online at www.sciencedirect.com
ScienceDirect
Journal of Complexity 23 (2007) 516–527
Journal of COMPLEXITY
www.elsevier.com/locate/jco

Pseudorandom numbers and entropy conditions

István Berkes[a,1], Walter Philipp[b,✠], Robert F. Tichy[c,*,2]

[a] *Department of Statistics, Technical University Graz, Steyrergasse 17/IV, A-8010 Graz, Austria*
[b] *Department of Statistics, University of Illinois, 725 S. Wright Street, Champaign, IL 61820, USA*
[c] *Institute of Mathematics A, Technical University Graz, Steyrergasse 30, A-8010 Graz, Austria*

Received 12 July 2006; accepted 20 December 2006
Available online 26 January 2007

Abstract

We investigate measures of pseudorandomness of finite sequences (x_n) of real numbers. Mauduit and Sárközy introduced the "well-distribution measure", depending on the behavior of the sequence (x_n) along arithmetic subsequences (x_{ak+b}). We extend this definition by replacing the class of arithmetic progressions by an arbitrary class $\mathscr{A}$ of sequences of positive integers and show that the so obtained measure is closely related to the metric entropy of the class $\mathscr{A}$. Using standard probabilistic techniques, this fact enables us to give precise bounds for the pseudorandomness measure of classical constructions. In particular, we will be interested in "truly" random sequences and sequences of the form $\{n_k\omega\}$, where $\{\cdot\}$ denotes fractional part, (n_k) is a given sequence of integers and $\omega \in [0, 1)$.

Keywords: Pseudorandomness; Discrepancy; Well-distribution measure; Metric entropy

1. Introduction

Computer generated pseudorandom numbers are used in many algorithms of applied mathematics (Monte Carlo methods, simulation, etc.) and the performance of such algorithms depends in an essential way on the properties of the random numbers used. A simple but important concept in the study of pseudorandomness is the discrepancy, characterizing how close the distribution of a finite

✠Deceased.

* Corresponding author.
E-mail addresses: berkes@tugraz.at (I. Berkes), tichy@tugraz.at (R.F. Tichy).

[1] The research of István Berkes was supported by OTKA grants T 43037, K 61052 and FWF grant S9603-N13.
[2] The research of Robert F. Tichy was supported by FWF grant S9603-N13.

0885-064X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2006.12.002

sequence is to the uniform distribution. The discrepancy D_N of a finite sequence $(x_1, \dots, x_N)$ in the unit interval $[0, 1)$ is defined by

$$D_N = D_N(x_1, \dots, x_N) := \sup_{0 \leqslant t \leqslant 1} \left| \frac{1}{N} \operatorname{card}(k \leqslant N : x_k \leqslant t) - t \right|. \tag{1}$$

An infinite sequence (x_n) in $[0, 1)$ is called uniformly distributed in the sense of Weyl if $D_N(x_1, \dots, x_N) \to 0$ as $N \to \infty$. Uniform distribution and discrepancy are particularly useful tools in connection with Monte Carlo and quasi-Monte Carlo integration, since by a well-known inequality of Koksma and its multi-dimensional generalizations (see e.g. [13, p. 143 and 151]), the error term in such procedures depends on the discrepancy of the pseudorandom sequence used. However, uniform distribution catches only one aspect of randomness and so called low discrepancy sequences may have rather poor performance with respect to other algorithms, such as simulation. Recall that if (ξ_n) is a sequence of i.i.d. random variables uniformly distributed in $[0, 1)$, then by the Chung–Smirnov LIL (see e.g. [22, p. 504]) we have

$$\limsup_{N \to \infty} \frac{ND_N(\xi_1, \dots, \xi_N)}{\sqrt{N \log \log N}} = \frac{1}{\sqrt{2}} \quad \text{a.s.} \tag{2}$$

In other words, the discrepancy of "truly" independent sequences has the precise order of magnitude $O(N^{-1/2}(\log \log N)^{1/2})$ with probability 1. On the other hand, if $\eta_n = \{n\omega\}$ where ω is a random variable uniformly distributed in $[0, 1)$, then by a result of Kesten [11] we have

$$ND_N(\eta_1, \dots, \eta_N) \sim \frac{2}{\pi^2} \log N \log \log N \quad \text{in probability.} \tag{3}$$

Here $\{t\}$ denotes the fractional part of t. Thus the sequence (η_n) gives a better remainder term in Monte Carlo integration than the "truly" i.i.d. sequence (ξ_n), but obviously its fluctuations are quite different from those of i.i.d. sequences and this makes (η_n) unsuitable for simulation purposes. A sequence resembling i.i.d. sequences not only has to have small discrepancy, but it must share several other properties with random sequences as well. Such properties can be used as "tests" for pseudorandomness, see Knuth [12] for a detailed discussion. For example, an i.i.d. sequence $(e_1, \dots, e_n) \in \{-1, 1\}^n$ has the *normality* property meaning that not too long strings of ± 1 occur in it with the "proper" frequency, it must be *well-distributed* relative to arithmetic progressions in the sense that the sums $\sum_{j=1}^{r} e_{a+bj}$ with $a \in \mathbb{Z}$, $b \in \mathbb{N}$ and subject to $1 \leqslant a + b \leqslant a + br \leqslant n$ are uniformly small compared with n (in fact, roughly $O(n^{1/2})$), it must have small multiple correlations, etc. In a series of papers (see e.g. [14–16]), Mauduit and Sárközy give a detailed study of these properties, in particular, they investigate the well-distribution and correlation measure of several concrete constructions of pseudorandom sequences. In the context of sequences in $[0, 1)$, they define the well-distribution measure by

$$W_N(x_1, \dots, x_N) := \sup_{(p_k) \in \mathcal{L}} \left| \sum_{p_k \leqslant N} \big(1(x_{p_k} \leqslant 1/2) - 1/2\big) \right|, \tag{4}$$

where $1(B)$ denotes the indicator function of the set B and $\mathcal{L}$ is the class of arithmetic progressions $p_k = a + bk$, $k = 1, 2, \dots$ with integers $a \geqslant 0$, $b \geqslant 1$. Both W_N and ND_N are suprema of sums of centered indicator functions $1(x_j \leqslant t) - t$, but they have a completely different behavior. For example, the order of magnitude of ND_N for an infinite sequence (x_k) in $[0, 1)$ can be as small as $O(\log N)$, an order of magnitude which is in fact the smallest possible by a classical result of

Schmidt (see e.g. [6,13]). In contrast, by a result of Roth [21], for any sequence $(x_1, \dots, x_N)$ we have

$$W_N(x_1, \dots, x_N) \geqslant cN^{1/4},$$

where c is an absolute constant. The discrepancy $D_N(x_1, \dots, x_N)$ can be fairly sharply estimated in terms of the exponential sums $S_N(h) = \sum_{k=1}^{N} e^{2\pi i h x_k}$ by using the Erdős–Turán and Koksma inequalities (see e.g. [6,13]), reducing the study of D_N to an analytic problem for which powerful tools exist. On the other hand, the computation of W_N leads to difficult combinatorial problems which are still unsolved in many important cases.

The purpose of the present paper is to give a detailed analysis of the well distribution measure W_N in (4); we will be specifically interested in the order of magnitude of $W_N(x_1, \dots, x_N)$ for i.i.d. sequences (x_n) and sequences of the type $x_k = \{n_k\omega\}$, where (n_k) is an increasing sequence of positive integers. The sequence $\{n_k\omega\}$ provides a particularly simple example for a uniformly distributed sequence in the sense of Weyl and it has been investigated extensively in the literature. Apart from technical simplifications, using the class $\mathcal{L}$ of arithmetic progressions in (4) has no particular significance; for example, for "not too large" classes $\mathcal{A}$ of sequences of positive integers and for i.i.d. sequences (x_n) we will be able to give sharp bounds for the more general quantity

$$W_N^{(\mathcal{A})}(x_1, \dots, x_N) := \sup_{(p_k)\in\mathcal{A}} \sup_{0\leqslant t\leqslant 1} \left| \sum_{p_k \leqslant N} \big(1(x_{p_k} \leqslant t) - t\big) \right|. \tag{5}$$

We will see that the order of magnitude of $W_N^{(\mathcal{A})}$ is intimately connected with the geometric properties of the class $\mathcal{A}$, namely, its metric entropy function $\kappa(\mathcal{A}; \delta, N)$ and related quantities. Metric entropy plays an important role in uniformity problems in the law of large numbers, CLT and LIL for random variables indexed by sets (see e.g. Dudley [7,8], Dudley and Philipp [9], Pollard [20]), but no such connection has been studied when uniformity is meant over subsequences of integers as in (5). In analogy with the existing probabilistic results on uniformity in the CLT, LIL and other limit theorems, it can be expected that metric entropy type quantities provide not only upper, but also lower estimates for $W_N^{(\mathcal{A})}$, thereby reducing the study of $W_N^{(\mathcal{A})}$ to the computation of metric entropy numbers.

Before formulating our results, it will be useful to review existing results on the ordinary discrepancy and well-distribution measure of the sequence $\{n_k\omega\}$. By a classical result of Weyl [23], for any increasing sequence (n_k) of integers, $\{n_k\omega\}$ is uniformly distributed for every $\omega \in [0, 1)$, except for a set of Lebesgue measure 0. Kesten's result cited above shows that

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

in measure. Another case where the order of magnitude of the discrepancy of $\{n_k\omega\}$ is known is when (n_k) grows very rapidly. Philipp [18] proved that if (n_k) satisfies the Hadamard gap condition

$$n_{k+1}/n_k \geqslant q > 1, \quad k = 1, 2, \dots \tag{6}$$

then we have for almost all $\omega \in [0, 1)$

$$\frac{1}{4} \leqslant \limsup_{N\to\infty} \frac{ND_N(\{n_k\omega\})}{\sqrt{N \log\log N}} \leqslant C(q), \tag{7}$$

where $C(q) \ll 1/(q-1)$. Recalling that the precise order of magnitude of the discrepancy of i.i.d. uniform sequences is $O(N^{-1/2}(\log\log N)^{1/2})$ with probability 1, the result of Philipp shows that, in the sense of discrepancy, the sequence $\{n_k\omega\}$ behaves exactly like an i.i.d. sequence. For subexponentially growing (n_k) the behavior of $D_N(\{n_k\omega\})$ is much more complicated and depends sensitively on the number-theoretic properties of the sequence (n_k); see Berkes et al. [3] for a detailed analysis of the arithmetic effect. In [3] it is also shown that in a certain statistical sense, for "most" subexponential sequences (n_k) the discrepancy $D_N(\{n_k\omega\})$ still satisfies (7). Passing to general sequences (n_k), Baker [1] proved, improving earlier results of Cassels [5] and Erdős and Koksma [10], that for any increasing sequence (n_k) of positive integers we have

$$ND_N(\{n_k\omega\}) = O(N^{1/2}(\log N)^{3/2+\varepsilon}) \quad \text{a.e.} \tag{8}$$

for any $\varepsilon > 0$. On the other hand, one can construct examples such that

$$ND_N(\{n_k\omega\}) \geqslant cN^{1/2}(\log N)^{1/2} \quad \text{a.e. for infinitely many } N$$

(see e.g. Berkes and Philipp [2]). This means that there exist sequences $\{n_k\omega\}$ whose discrepancy $D_N(\{n_k\omega\})$ exceeds the discrepancy of i.i.d. sequences, but the excess factor can be at most a power of $\log N$.

The previous results give a fairly satisfactory picture of the metric discrepancy of sequences $\{n_k\omega\}$ in a number of important cases. In contrast, relatively little is known on the well-distribution measure W_N of $\{n_k\omega\}$. Mauduit and Sárközy [15,16] showed that in the case $n_k = k$ we have

$$W_N(\{k\omega\}) \ll N^{1/2}(\log N)^{1+\varepsilon}$$

for almost every $\omega \in [0, 1)$, and that the exponent of the log can be replaced by $1/2$ if the partial quotients of the continued fraction expansion of ω remain bounded. They also proved that

$$W_N(\{k\omega\}) \gg N^{1/2}$$

for every irrational ω. Thus for almost all ω the order of magnitude of $W_N(\{k\omega\})$ is roughly $O(N^{1/2})$, which, as Theorem 1 in combination with the estimate (21) below will show, is very close to the order of magnitude of the well-distribution measure of "true" i.i.d. sequences. As noted, however, $ND_N(\{k\omega\})$ is much smaller than $O(N^{1/2})$, indicating a very complicated probabilistic behavior of the sequence $\{k\omega\}$.

Except for the sequence $\{k\omega\}$, no precise estimates for the well-distribution measure of $\{n_k\omega\}$ seem to be known. For the sequence $\{k^r\omega\}$ $(r = 2, 3, \ldots)$, Mauduit and Sárközy [15,16] proved that for almost every ω

$$W_N(\{k^r\omega\}) \ll N^{1-\alpha_r}$$

with some (explicitly computed) constant $\alpha_r > 0$. In particular,

$$W_N(\{k^2\omega\}) \ll N^{3/5}(\log N)^{2/5+\varepsilon} \quad \text{a.e.}$$

Philipp and Tichy [19] proved that for any increasing sequence (n_k) of integers we have

$$W_N(\{n_k\omega\}) \ll N^{2/3}(\log N)^{1+\varepsilon} \quad \text{a.e.} \tag{9}$$

It is possible that, in analogy with Baker's result (8), the factor $N^{2/3}$ here can be replaced by $N^{1/2}$, but this remains open.

2. Results

We are now ready to formulate our main results. Let (η_n) be any sequence of random variables with values in $[0, 1)$, and let $\mathcal{A}$ be a class of subsequences of $\mathbb{N}$. Our purpose is to estimate the quantity

$$W_N^{(\mathcal{A})}(\eta_1, \dots, \eta_N) := \sup_{(p_k)\in\mathcal{A}} \sup_{0\leqslant t\leqslant 1} \left| \sum_{p_k\leqslant N} \big(1(\eta_{p_k}\leqslant t) - t\big) \right|. \tag{10}$$

Our main interest will be the case when η_k are independent random variables or $\eta_k = \eta_k(\omega) = \{n_k\omega\}$, a sequence of random variables defined on the interval $[0, 1)$ endowed with Lebesgue measure. When the sequence (η_k) is understood, we simply write $W_N(\mathcal{A})$ instead of $W_N^{(\mathcal{A})}(\eta_1, \dots, \eta_N)$. Clearly, for any $\mathcal{A}$ and (η_k) we have

$$W_N^{(\mathcal{A})}(\eta_1, \dots, \eta_N) \leqslant N$$

and for "large" $\mathcal{A}$ this estimate cannot be substantially improved even if η_k are i.i.d. random variables. For example, if η_k are independent r.v.'s taking the values 0 and 2/3 with probability $1/2 - 1/2$ and $\mathcal{A}$ is the class of all increasing sequences in $\mathbb{N}$, then

$$W_N^{(\mathcal{A})}(\eta_1, \dots, \eta_N) \geqslant N/4.$$

Indeed, if for each ω we let $p_1(\omega) < p_2(\omega) < \cdots$ denote those indices such that $\eta_{p_k}(\omega) = 0$ then either (p_k) or its complement in the segment $[1, 2, \dots, N]$ has cardinality at least $N/2$. Consequently, we have for all ω

$$\sup_{(p_k)\in\mathcal{A}} \left| \sum_{p_k\leqslant N} \big(1(\eta_{p_k}\leqslant 1/2) - 1/2\big) \right| \geqslant N/4.$$

In the case when $\mathcal{A}$ consists of a single sequence and (η_n) is an i.i.d. uniform sequence of r.v.'s, we have

$$W_N^{(\mathcal{A})}(\eta_1, \dots, \eta_N) = o(N) \quad \text{a.s.} \tag{11}$$

by the Glivenko–Cantelli theorem of probability theory. (Actually, in this case the right hand side of (11) can be improved to $O\left((N \log\log N)^{1/2}\right)$ by the Chung–Smirnov law of the iterated logarithm.) If relation (11) holds for a larger class $\mathcal{A}$, this means a certain uniformity in the Glivenko–Cantelli theorem with respect to a class of subsequences of integers. Uniformity in the Glivenko–Cantelli theorem with respect to subsets of the Euclidean space $\mathbb{R}^d$ has been investigated extensively in the literature. Let (η_n) be a sequence of i.i.d. random variables, uniformly distributed over the unit cube $\mathbb{K}^d$ of $\mathbb{R}^d$, and let $\mathcal{C}$ be a class of Borel sets $\subseteq \mathbb{K}^d$. Put

$$Z_N(C) = \sum_{k\leqslant N} (1(\eta_k \in C) - \mu(C)), \quad C \in \mathcal{C},$$

where μ is the Lebesgue measure. As it turns out, the validity of the uniform strong law and LIL, i.e.

$$\lim_{N\to\infty} \sup_{C\in\mathcal{C}} \frac{1}{N} |Z_N(C)| = 0 \quad \text{a.s.} \tag{12}$$

and

$$\limsup_{N\to\infty} \frac{\sup_{C\in\mathcal{C}} |Z_N(C)|}{\sqrt{N \log\log N}} < \infty \quad \text{a.s.} \tag{13}$$

are closely connected with the geometry of the class $\mathcal{C}$, namely how closely the elements of $\mathcal{C}$ can be approximated by "special" sets. Specifically, let $\mathcal{N}_I(\delta, \mathcal{C})$ denote the smallest number r of measurable sets $A_1, \dots, A_r$ in $\mathbb{K}^d$ such that for every $C \in \mathcal{C}$ there exist A_i, A_j, $1 \leqslant i < j \leqslant r$ such that $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < \delta$ ("metric entropy with inclusion"). Then the validity of the uniform LIL and CLT is closely related to the finiteness of the entropy integral

$$\int_0^1 (\log \mathcal{N}_I(x^2, \mathcal{C}))^{1/2}\, dx.$$

(See e.g. Dudley [7,8], Dudley and Philipp [9].) Another important geometric property relevant for the uniform strong law (12), discovered by Vapnik and Červonenkis, is how finite sets $\{x_1, \dots, x_N\}$ in $\mathbb{R}^d$ can be "shattered" by $\mathcal{C}$, i.e. how many different sets of the form $\{x_1, \dots, x_N\} \cap C$, $C \in \mathcal{C}$ exist. In fact, a necessary and sufficient condition for (12) can be given in terms of this quantity; see e.g. Pollard [20, p. 22].

The purpose of this paper is to develop similar entropy concepts in the space of subsequences of $\mathbb{N}$ and apply them to prove uniform Glivenko–Cantelli laws of the type (11), together with rates of convergence, in particular, uniform laws of the iterated logarithm. Let $\mathcal{A}$ be a class of subsequences of $\mathbb{N}$ such that $\mathbb{N} \in \mathcal{A}$. For each $N \geqslant 1$ let $\mathcal{A}_N$ denote the collection of the restrictions of these subsequences to the segment $[1, 2, \dots, N]$ of the first N positive integers, i.e.

$$\mathcal{A}_N := \{A \cap [1, 2, \dots, N] : A \in \mathcal{A}\}.$$

Clearly

$$\mathcal{A}_N = \bigcup_{r\geqslant 1} \mathcal{A}_N(r),$$

where $\mathcal{A}_N(r)$ denotes the class of sets $A \in \mathcal{A}_N$ for which $N2^{-r} < \operatorname{card} A \leqslant N2^{-(r-1)}$. We call

$$\psi(\mathcal{A}; N, r) := \operatorname{card} \mathcal{A}_N(r) \tag{14}$$

the entropy function of the class $\mathcal{A}$.

Next, let $(\eta_k,\ k \geqslant 1)$ be a sequence of random variables with each η_k having uniform distribution over $[0, 1)$, i.e.

$$P(\eta_k \leqslant t) = t, \quad 0 \leqslant t \leqslant 1,\ k \geqslant 1. \tag{15}$$

In Theorems 2–5 we permit η_k to have asymptotically uniform distribution over $[0, 1)$.

Theorem 1. *Let (η_k) be a sequence of independent random variables with uniform distribution* (15) *over* [0,1). *Let $\mathcal{A}$ be a class of subsequences of $\mathbb{N}$ with entropy function ψ satisfying*

$$\psi(\mathcal{A}; N, r) \leqslant \exp\left(B \cdot 2^{r/2} \log\log N\right), \quad r \geqslant 0, \quad N \geqslant 10 \tag{16}$$

for some constant $B > 0$. Then with probability 1

$$\frac{1}{4} \leqslant \limsup_{N\to\infty} (N \log\log N)^{-1/2} W_N(\mathcal{A}) \leqslant C$$

for some constant C, depending only on the constant B in (16).

Next, let (n_k) be a sequence of real numbers satisfying the Hadamard gap condition

$$n_{k+1}/n_k \geqslant q > 1, \quad k = 1, 2, \ldots . \tag{17}$$

Then the sequence

$$\eta_k(\omega) := \{n_k \omega\} \tag{18}$$

defined on the unit interval $[0, 1)$ endowed with Lebesgue measure, is a sequence of random variables having asymptotically uniform distribution over $[0, 1)$.

Theorem 2. *Let (n_k) be a sequence of real numbers satisfying the Hadamard gap condition* (17) *and let $\eta_k = \eta_k(\omega) = \{n_k\omega\}$. Let $\mathcal{A}$ be a class of subsequences of $\mathbb{N}$ with entropy function satisfying*

$$\psi(\mathcal{A}; N, r) \leqslant B \cdot 2^{r^\beta} \tag{19}$$

for some constants $B > 0$ and $\beta > 0$. Then with probability 1

$$\frac{1}{4} \leqslant \limsup_{N\to\infty} (N \log\log N)^{-1/2} W_N(\mathcal{A}) \leqslant C$$

for some constant $C > 0$, depending only on B, β and q.

The second entropy concept is based on the Hamming distance of sequences of integers. For $N \geqslant 1$ we define the distance of two sequences A and B of positive integers by

$$d(A, B; N) = \frac{1}{N} \sum_{n\leqslant N} |1(n \in A) - 1(n \in B)|.$$

Given a class $\mathcal{A}$ of increasing sequences of positive integers we define the entropy function κ by

$$\kappa(\mathcal{A}; \delta, N) := \sup\left\{m : \text{there exist } A_1, \ldots, A_m \in \mathcal{A} \text{ such that } d(A_i, A_j; N) > \delta \text{ for all } i \neq j\right\}. \tag{20}$$

Clearly κ is a non-increasing function of $\delta \geqslant 0$.

Theorem 3. *Let (η_k) be a sequence of independent random variables with uniform distribution* (15) *over* $[0, 1)$. *Let $\mathcal{A}$ be a class of increasing sequences of positive integers with entropy function $\kappa(\mathcal{A}; \delta, N)$ growing not faster than a polynomial in $1/\delta$ (depending only on $\mathcal{A}$). Then with probability* 1

$$W_N(\mathcal{A}) \ll \sqrt{N}\left(\log \kappa\left(\mathcal{A}; N^{-\alpha}, N\right) + (\log\log N)^{1/2}\right) \quad \text{for any } \alpha > 1/2.$$

The same result holds if $\eta_k = \{n_k\omega\}$, where (n_k) is a sequence of real numbers satisfying the Hadamard gap condition (17).

As an example consider a Vapnik–Červonenkis (VC) class $\mathcal{A}$ in the set $\mathbb{N}$ of positive integers. For any finite set $F \subset \mathbb{N}$, let $\Delta^{\mathcal{A}}(F)$ be the number of different subsets $F \cap A$, $A \in \mathcal{A}$. For $n = 1, 2, \ldots$ let

$$m^{\mathcal{A}}(n) := \max\left(\Delta^{\mathcal{A}}(F) : \text{ card } F = n\right).$$

Clearly $m^{\mathcal{A}}(n) \leqslant 2^n$. Let

$$v = V(\mathcal{A}) := \begin{cases} \inf\{n : m^{\mathcal{A}}(n) < 2^n\} \\ \quad +\infty \text{ if } m^{\mathcal{A}}(n) = 2^n \text{ for all } n. \end{cases}$$

If $V(\mathcal{A}) < +\infty$ then $\mathcal{A}$ is called *a* VC class in $\mathbb{N}$. We recall a result of Dudley [7, Lemma 7.13] or Dudley [8, p. 105], measuring the size of VC classes. Let Γ be the set of all laws on $\mathbb{N}$ of the form

$$n^{-1} \sum_{j=1}^{n} \delta_{x(j)}$$

for unit point masses $\delta_{x(j)}$ at $x(j) \in \mathbb{N}$, $j = 1, 2, \ldots, n$; $n = 1, 2, \ldots$ where the $x(j)$ need not be distinct. For $\delta > 0$ and $\gamma \in \Gamma$ let

$$\kappa^*(\mathcal{A}, \gamma; \delta) := \sup\left\{m : \text{ there exist } A_1, \ldots, A_m \in \mathcal{A} \text{ such that } \gamma(A_i \Delta A_j) > \delta \text{ for } i \neq j\right\}$$

and

$$\kappa^*(\mathcal{A}; \delta) := \sup\{\kappa(\mathcal{A}, \gamma; \delta) : \gamma \in \Gamma\}.$$

Lemma 1 (*Dudley [7,8]*). *If $\mathcal{A}$ is a VC class in $\mathbb{N}$ with $V(\mathcal{A}) = v$, then there is a constant K depending only on v such that*

$$\kappa^*(\mathcal{A}; \delta) \leqslant K \delta^{-v} |\log \delta|^v \quad \textit{for all } \delta > 0.$$

Hence if $\mathcal{A}$ is a VC class in $\mathbb{N}$, the entropy function κ defined in (20) does not grow faster than a polynomial in $1/\delta$.

Corollary 1. *Let (η_k) be a sequence of independent random variables with uniform distribution* (15) *over* $[0, 1)$ *or $\eta_k = \{n_k \omega\}$ with a Hadamard lacunary (n_k). Then if $\mathcal{A}$ is a VC class in $\mathbb{N}$, with probability* 1 *we have*

$$W_N(\mathcal{A}) \ll \sqrt{N} \log N.$$

In the following two results we consider the case when $\eta_k = \{n_k \omega\}$ with an arbitrary increasing sequence (n_k) of positive integers. If $\mathcal{L}$ denotes the collection of all integer valued arithmetic progressions $p_k = a + bk$, $k = 1, 2, \ldots, a \geqslant 0, b \geqslant 1$, then it is easy to see that the entropy function ψ satisfies

$$\psi(\mathcal{L}; N, r) \leqslant 2^{2r}, \quad r = 1, 2, \ldots . \tag{21}$$

Theorem 4. *Let (n_k) be an increasing sequence of positive integers and let $\eta_k = \eta_k(\omega) = \{n_k \omega\}$. Let $\mathcal{A}$ be a class of subsequences of $\mathbb{N}$ with entropy function ψ satisfying* (19) *for some positive constants β and B. Then with probability* 1 *for any $\varepsilon > 0$*

$$\begin{aligned} W_N(\mathcal{A}) &\ll N^{\frac{\beta}{1+\beta}} (\log N)^{\frac{3}{1+\beta}+\varepsilon} \quad \textit{if } \beta > 1, \\ &\ll N^{\frac{1}{2}} (\log N)^{2+\varepsilon} \quad \textit{if } \beta = 1, \\ &\ll N^{\frac{1}{2}} (\log N)^{\frac{3}{2}+\varepsilon} \quad \textit{if } \beta < 1. \end{aligned}$$

Remark. The case $\beta = 2$ and (21) yield Theorem 1 in [19].

Theorem 5. *Let (n_k) be an increasing sequence of positive integers and let $\eta_k = \eta_k(\omega) = \{n_k\omega\}$. Let $\mathcal{A}$ be a class of increasing sequences of positive integers with entropy function $\kappa(\mathcal{A}; \delta, N) \leqslant C\delta^{-v}$ for some $v \geqslant 0$, where C depends only on $\mathcal{A}$. Then with probability* 1

$$W_N(\mathcal{A}) \ll N^{\frac{v+1}{v+2}} (\log N)^{\frac{3}{v+2}+\varepsilon}, \quad \varepsilon > 0.$$

3. Proofs

In what follows, we will prove Theorem 4 and outline the idea of the proof of Theorem 3 in the lacunary case, which is typical for the proof of the remaining results. Complete proofs of all results and a number of further results will be given in our forthcoming paper [4].

Assume the conditions of Theorem 4. Fix $N \geqslant 1$, $r \geqslant 1$ and let (p_k) be a fixed sequence in $[1, N]$ such that $(p_k) \in \mathcal{A}_N(r)$. By the Erdős–Turán inequality (see e.g. [6, p. 15], or [13, p. 114]) we have for any $1 \leqslant Q \leqslant N$

$$\sup_{0\leqslant t\leqslant 1} \left| \sum_{p_k \leqslant Q} \big(1(\eta_{p_k}(\omega) \leqslant t) - t\big) \right| \leqslant \frac{6R}{H} + 6 \sum_{1\leqslant h\leqslant H} \frac{1}{h} \left| \sum_{p_k\leqslant Q} \mathrm{e}(hn_{p_k}\omega) \right|.$$

Here $R = \#\{k : p_k \leqslant Q\}$, $\mathrm{e}(x) = \exp(2\pi i x)$ and $H \geqslant 1$ is arbitrary. Clearly $R \leqslant N$ and thus

$$\max_{Q\leqslant N} \sup_{0\leqslant t\leqslant 1} \left| \sum_{p_k\leqslant Q} \big(1(\eta_{p_k}(\omega)\leqslant t) - t\big) \right|^2 \leqslant \frac{72N^2}{H^2} + 72 \left(\sum_{1\leqslant h\leqslant H} \frac{1}{h} \max_{Q\leqslant N} \left| \sum_{p_k\leqslant Q} \mathrm{e}(hn_{p_k}\omega) \right| \right)^2.$$

By Hunt's inequality (see e.g. [17]) we have

$$E\left(\max_{Q\leqslant N} \left| \sum_{p_k\leqslant Q} \mathrm{e}(hn_{p_k}\omega) \right|^2 \right) \leqslant C \sum_{p_k\leqslant N} 1 \leqslant CN2^{-(r-1)}$$

and thus choosing $H = N$ and using Minkowski's inequality we get

$$E\left(\max_{Q\leqslant N} \sup_{0\leqslant t\leqslant 1} \left| \sum_{p_k\leqslant Q} \big(1(\eta_{p_k}\leqslant t) - t\big) \right|^2 \right) \ll N2^{-r}\log^2 N + 1 \ll N2^{-r}\log^2 N. \tag{22}$$

(To justify the last step, we note that without loss of generality we can assume that $N2^{-(r-1)} \geqslant 1$, since otherwise $\mathcal{A}_N(r)$ is empty.) Since the number of sequences $(p_k) \in \mathcal{A}_N(r)$ is at most $B \cdot 2^{r\beta}$ by the assumptions of Theorem 4, we have for any $\alpha > 0$, $\tau > 0$ (to be chosen suitably later),

$$P\left(\max_{(p_k)\in\mathcal{A}_N(r)} \max_{Q\leqslant N} \sup_{0\leqslant t\leqslant 1} \left| \sum_{p_k\leqslant Q} \big(1(\eta_{p_k}\leqslant t) - t\big) \right| \geqslant 2N^{\alpha}(\log N)^{\tau} \right) \ll N^{1-2\alpha}(\log N)^{2-2\tau} \cdot 2^{r(\beta-1)}. \tag{23}$$

Without loss of generality we can assume that $N2^{-(r-1)} \geqslant N^{\alpha}(\log N)^{\tau}$, i.e.

$$2^r \leqslant 2N^{1-\alpha}(\log N)^{-\tau}, \tag{24}$$

since otherwise the absolute value of the sum in (23) would be less than $N^{\alpha}(\log N)^{\tau}$. Summing the probability bounds in (23) over all r subject to (24) and choosing α and τ according to the following table:

β	α	τ
> 1	$\beta/(1+\beta)$	$(3+\varepsilon)/(1+\beta)$
$= 1$	$\frac{1}{2}$	$2+\varepsilon$
< 1	$\frac{1}{2}$	$\frac{3}{2}+\varepsilon$

we obtain letting $N = 2^{m+1}$, $m = 1, 2, \ldots$

$$P\left(\max_{2^m < Q \leqslant 2^{m+1}} \max_{(p_k)\in\mathcal{A}} \sup_{0\leqslant t\leqslant 1} \left|\sum_{p_k \leqslant Q} \big(1(\eta_{p_k} \leqslant t) - t\big)\right| \geqslant C^* 2^{m\alpha} m^{\tau}\right)$$
$$\ll (\log 2^m)^{-(1+\varepsilon')} \ll m^{-(1+\varepsilon')}$$

for some $C^* > 0$, $\varepsilon' > 0$. We apply the convergence part of the Borel–Cantelli lemma and obtain the conclusion of Theorem 4.

For the proof of Theorem 3 in the lacunary case define, for $0 \leqslant s < t \leqslant 1$,

$$x_n(s,t) := 1(s \leqslant \eta_n < t) - (t-s).$$

We state the following maximal inequality.

Proposition 1. *Let $N \geqslant 1$ be an integer and let $R \geqslant 1$. Suppose that $\ell := t - s \geqslant N^{-3/2}$. Then for some constant $A \geqslant 1$ depending only on q and for any $\beta > 0$ we have as $N \to \infty$*

$$P\left(\max_{Q\leqslant N} \left|\sum_{k=1}^{Q} x_k(s,t)\right| \geqslant AR\ell^{1/32}(N\log\log N)^{1/2}\right)$$
$$\ll \exp(-16R\ell^{-1/32}\log\log N) + R^{-8\beta}N^{-2\beta},$$

where the constant implied by $\ll$ only depends on q and β.

An exponential bound of this kind is a crucial ingredient of all discrepancy estimates of LIL type. The proof depends on a martingale approximation argument and can be modelled after the proof of [18, Proposition 4.2.1]. The details are, however, long and technical and will be given in [4].

To deduce Theorem 3 from Proposition 1, fix $1/2 < \alpha < 1$ and $0 \leqslant s < t \leqslant 1$. By the hypotheses of the theorem, we can choose $\beta > 0$ such that

$$\kappa(\mathcal{A}; \delta, N) \ll \delta^{-\beta/2}, \tag{25}$$

where the constant implied by $\ll$ depends only on $\mathcal{A}$. For simplicity we set

$$\kappa(\delta) := \kappa(\mathcal{A}; \delta^{\alpha}, [1/\delta]). \tag{26}$$

By Proposition 1 we have for any sequence $(p_k) \subset \mathbb{N}$ and $R \geqslant 1$, $0 < \varepsilon \leqslant 1/32$ and $t - s \geqslant 2^{-3r/2}$ as $r \to \infty$

$$P\left(\max_{Q \leqslant 2^r} \left|\sum_{p_k \leqslant Q} x_{p_k}(s,t)\right| \geqslant AR2^{\frac{1}{2}r}(t-s)^{\varepsilon}(\log \kappa(2^{-r}) + (\log r)^{\frac{1}{2}})\right)$$
$$\ll \begin{cases} \exp(-16R(t-s)^{-\varepsilon}\log\kappa(2^{-r})\log^{\frac{1}{2}} r) + R^{-8\beta}2^{-2r\beta} & \text{if } \log\kappa(2^{-r}) > \log^{\frac{1}{2}} r, \\ \exp(-16R(t-s)^{-\varepsilon}\log r) + R^{-8\beta}2^{-2r\beta} & \text{if } \log\kappa(2^{-r}) \leqslant \log^{\frac{1}{2}} r \end{cases} \quad (27)$$

for some constant $A \geqslant 1$. (In the case of the first line of (27) we apply Proposition 1 with R replaced by $R \log \kappa(2^{-r})(\log r)^{-1/2}$.) Let

$$\delta := AR(t-s)^{\varepsilon}2^{-r/2} \quad (28)$$

and $\mathcal{B} = \{(p_k^{(1)}), \ldots, (p_k^{(M)})\}$ a maximal set of sequences in $\mathcal{A}$ with pairwise distance $> \delta$ with respect to the Hamming distance $d(\cdot,\cdot,2^r)$. Then

$$M = \kappa(\mathcal{A}, \delta, 2^r) \leqslant \kappa(\mathcal{A}; 2^{-\alpha r}, 2^r) = \kappa(2^{-r})$$

since

$$\delta \geqslant (t-s)^{\varepsilon}2^{-r/2} \geqslant 2^{-r(3\varepsilon/2+1/2)} \geqslant 2^{-\alpha r}$$

provided we choose $\varepsilon > 0$ so small that $3\varepsilon/2 + 1/2 < \alpha$. Clearly, for any $(q_k) \in \mathcal{A}$ there is a $(p_k) \in \mathcal{B}$ with $d((p_k),(q_k),2^r) \leqslant \delta$, which implies that for any $Q \leqslant 2^r$ the sums $\sum_{p_k \leqslant Q} x_{p_k}(s,t)$ and $\sum_{q_k \leqslant Q} x_{q_k}(s,t)$ differ at most by $\delta 2^r = AR(t-s)^{\varepsilon}2^{r/2}$. Hence using (27) we get

$$P\left(\max_{(q_k)\in\mathcal{A}} \max_{Q \leqslant 2^r} \left|\sum_{q_k \leqslant Q} x_{q_k}(s,t)\right| \geqslant 2AR2^{\frac{1}{2}r}(t-s)^{\varepsilon}(\log\kappa(2^{-r}) + \log^{\frac{1}{2}} r)\right)$$
$$\ll \exp\left(-8R(t-s)^{-\varepsilon}(\log\kappa(2^{-r}) + \log^{\frac{1}{2}} r)\log^{\frac{1}{2}} r + \log\kappa(2^{-r})\right) + R^{-8\beta}\kappa(2^{-r})2^{-2r\beta}$$
$$\ll \exp(-4R(t-s)^{-\varepsilon}\log r) + R^{-8\beta}2^{-3r\beta/2}$$

by distinguishing the cases $\log\kappa(2^{-r}) > \log^{\frac{1}{2}} r$ and $\log\kappa(2^{-r}) \leqslant \log^{\frac{1}{2}} r$ and by using (25) in the estimate of the very last term.

The proof of Theorem 3 can now be completed by a chaining argument similar to that in [18].

Note added in proof

With great sadness, we inform the reader that Walter Philipp passed away on July 19, 2006, at the age of 69, near Graz, Austria.–I. Berkes and R.F. Tichy.

References

[1] R.C. Baker, Metric number theory and the large sieve, J. London Math. Soc. 24 (2) (1981) 34–40.
[2] I. Berkes, W. Philipp, The size of trigonometric and Walsh series and uniform distribution mod 1, J. London Math. Soc. 50 (2) (1994) 454–464.
[3] I. Berkes, W. Philipp, R.F. Tichy, Empirical processes in probabilistic number theory: the LIL for the discrepancy of $(n_k\omega)$ mod 1, Illinois J. Math. 50 (2006) 107–145.
[4] I. Berkes, W. Philipp, R.F. Tichy, Entropy conditions for subsequences of random variables with applications to empirical processes, Monatshefte Math., to appear.

[5] J.W.S. Cassels, Some metrical theorems of Diophantine approximation III, Proc. Cambridge Philos. Soc. 46 (1950) 219–225.
[6] M. Drmota, R.F. Tichy, Sequences, discrepancies and applications, Lecture Notes in Mathematics, vol. 1651, Springer, Berlin, 1997.
[7] R.M. Dudley, Central limit theorems for empirical measures, Ann. Probab. 6 (1978) 899–929, Ann. Probab. 7 (1979) 909–911 (Correction).
[8] R.M. Dudley, A course on empirical processes, Lecture Notes in Mathematics, vol. 1097, Springer, Berlin, 1984, pp. 1–142.
[9] R.M. Dudley, W. Philipp, Invariance principles for sums of Banach space valued random elements and empirical processes Z. Wahrscheinlichkeitstheorie verw. Geb. 62 (1983) 509–552.
[10] P. Erdös, J.F. Koksma, On the uniform distribution modulo 1 of sequences $(f(n, \vartheta))$, Proc. Kon. Nederl. Akad. Wetensch. 52 (1949) 851–854.
[11] H. Kesten, The discrepancy of random sequences $\{kx\}$, Acta Arith. 10 (1964/65) 183–213.
[12] D.E. Knuth, The Art of Computer Programming, vol. 2, second ed., Addison-Wesley, Reading, MA, 1981.
[13] L. Kuipers, H. Niederreiter, Uniform Distribution of Sequences, Wiley, New York, 1974.
[14] C. Mauduit, A. Sárközy, On finite pseudorandom binary sequences I. Measure of pseudorandomness, the Legendre symbol, Acta Arith. 82 (1997) 365–377.
[15] C. Mauduit, A. Sárközy, On finite pseudorandom binary sequences V. On $(n\alpha)$ and $(n^2\alpha)$ sequences, Monatshefte Math. 129 (2000) 197–216.
[16] C. Mauduit, A. Sárközy, On finite pseudorandom binary sequences VI. On $(n^k\alpha)$ sequences, Monatshefte Math. 130 (2000) 281–298.
[17] C.J. Mozzochi, On the pointwise convergence of Fourier series, Lecture Notes in Mathematics, vol. 199, Springer, Berlin, 1971.
[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, R.F. Tichy, Metric theorems for distribution measures of pseudorandom sequences, Monatshefte Math. 135 (2002) 321–326.
[20] D. Pollard, Convergence of Stochastic Processes, Springer, Berlin, 1984.
[21] K.F. Roth, Remark concerning integer sequences, Acta Arith. 9 (1964) 257–260.
[22] G.R. Shorack, J. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.
[23] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, Math. Ann. 77 (1916) 313–352.