On the distribution of the van der Corput sequence in arbitrary base

Bence Borda
Department of Mathematics, Rutgers University
110 Frelinghuysen Road, Piscataway, NJ-08854, USA
Email: bordabence85@gmail.com

Keywords: van der Corput sequence, $L^p$ discrepancy, central limit theorem, large deviations
Mathematics Subject Classification (2010): 11K31, 11K38, 60F05, 60F10

Abstract

A central limit theorem with explicit error bound, and a large deviation result are proved for a sequence of weakly dependent random variables of a special form. As a corollary, under certain conditions on the function $f : [0, 1] \to \mathbb{R}$ a central limit theorem and a large deviation result are obtained for the sum $\sum_{n=0}^{N-1} f(x_n)$, where $x_n$ is the base $b$ van der Corput sequence for an arbitrary integer $b \geq 2$. Similar results are also proved for the $L^p$ discrepancy of the same sequence for $1 \leq p < \infty$. The main methods used in the proofs are the Berry–Esseen theorem and Fourier analysis.

1 Introduction

For an integer $b \geq 2$ the base $b$ van der Corput sequence $x_n$ is defined the following way. If the base $b$ representation of the integer $n \geq 0$ is $n = \sum_{i=1}^{m} a_i b^{i-1}$ for some digits $a_i \in \{0, 1, \ldots, b-1\}$, then

$$x_n = \sum_{i=1}^{m} \frac{a_i}{b^i}.$$

The main importance of this sequence is that it is of low discrepancy. Indeed, the discrepancy function of the base $b$ van der Corput sequence

$$\Delta_N(x) = |\{0 \leq n < N : x_n < x\}| - Nx,$$

defined for nonnegative integers $N$, and $x \in [0, 1]$, satisfies

$$0 \leq \Delta_N(x) \leq \frac{b}{4} \log_b N + b.$$
The precise value of
\[ \limsup_{N \to \infty} \sup_{x \in [0,1]} \frac{\Delta_N(x)}{\log N} \]
in terms of the base \( b \) was found by Faure (\[4\] Theorem 1, Theorem 2 and Sections 5.5.1–5.5.3).

In this article we study the random aspects of the base \( b \) van der Corput sequence. Let
\[ \Phi(\lambda) = \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \]
denote the distribution function of the standard normal distribution. Our main result is that the sum
\[ S(N) = \sum_{n=0}^{N-1} \left( \frac{1}{2} - x_n \right) \]
satisfies the following central limit theorem.

**Theorem 1.** Let \( x_n \) be the base \( b \) van der Corput sequence, where \( b \geq 2 \) is an arbitrary integer. Then for any integer \( M > b^2 \) and any real number \( \lambda \) we have
\[ \frac{1}{M} \left| \left\{ 0 \leq N < M : \frac{S(N) - c(b) \log_b N}{\sqrt{d(b) \log_b N}} < \lambda \right\} \right| = \Phi(\lambda) + O \left( \frac{\sqrt{\log \log_b M}}{\sqrt{\log_b M}} \right), \]
where \( c(b) = \frac{b^2 - 1}{12b} \) and \( d(b) = \frac{b^4 + 120b^3 - 480b^2 + 600b - 241}{720b^2} \). The implied constant in the error term is absolute.

The following large deviation result complements Theorem 1.

**Theorem 2.** Let \( x_n \) be the base \( b \) van der Corput sequence, where \( b \geq 2 \) is an arbitrary integer. For any integer \( M > b \) and any real number \( \lambda \geq 3 \) we have
\[ \frac{1}{M} \left| \left\{ 0 \leq N < M : \left| S(N) - \frac{b^2 - 1}{12b} \log_b M \right| \geq 25\lambda b \sqrt{\log_b M + 1} \right\} \right| \leq \frac{4\sqrt{\lambda}}{e^{\sqrt{\lambda} - 2}} + \frac{1}{b \sqrt{\log_b M - 2}}. \]
Since
\[ \int_0^1 \Delta_N(x) \, dx = S(N), \] (1)
we have that \( S(N) = O(b \log_b M) \), therefore Theorem 2 is meaningful only when applied with \( \lambda = O\left(\sqrt{\log_b M}\right) \). Note that for all such values of \( \lambda \) the error term \( \frac{1}{b \sqrt{\log_b M} \log_b N} \) is of smaller order of magnitude than \( \frac{4}{e^{\sqrt{\log_b M}} - 1} \). The question of whether the upper bound in Theorem 2 can be improved to \( O\left(e^{-d \lambda}\right) \) or to \( O\left(e^{-d \lambda^2}\right) \) for some constant \( d > 0 \) is left open.

Observation (1) gives the idea that the sum \( S(N) \) is related to the \( L^p \) norm \( \|\Delta_N\|_p = \left( \int_0^1 |\Delta_N(x)|^p \, dx \right)^{\frac{1}{p}} \) of the discrepancy function. As simple corollaries to Theorem 1 and Theorem 2 we thus obtain that \( \|\Delta_N\|_p \) satisfies the same central limit theorem and large deviation result as \( S(N) \).

**Theorem 3.** Let \( x_n \) be the base \( b \) van der Corput sequence, where \( b \geq 2 \) is an arbitrary integer. Let \( 1 \leq p < \infty \) be an arbitrary real. Then for any integer \( M > b^2 \) and any real number \( \lambda \) we have

\[
\frac{1}{M} \left\{ 0 \leq N < M : \left\| \Delta_N \right\|_p - c(b) \log_b N \sqrt{d(b) \log_b N} < \lambda \right\} = \Phi(\lambda) + O\left(\frac{\sqrt{\log \log_b M}}{\sqrt{\log_b M}}\right),
\]

where \( c(b) = \frac{b^2 - 1}{12b} \) and \( d(b) = \frac{b^4 + 120b^3 - 480b^2 + 600b - 241}{720b^2} \). The implied constant in the error term depends only on \( p \).

**Theorem 4.** Let \( x_n \) be the base \( b \) van der Corput sequence, where \( b \geq 2 \) is an arbitrary integer. Let \( 1 \leq p < \infty \) be an arbitrary real. There exists a positive constant \( A \) depending only on \( p \) such that for any integer \( M > b^2 \) and any real number \( \lambda \geq 1 \) we have

\[
\frac{1}{M} \left\{ 0 \leq N < M : \left| \left\| \Delta_N \right\|_p - \frac{b^2 - 1}{12b} \log_b N \right| \geq A\lambda b \sqrt{\log_b N} \right\} \leq e^{-\sqrt{\lambda}}.
\]

Similar central limit theorems concerning the distribution of the van der Corput sequence have already appeared in the literature. In \[3\] Theorem 3 is proved in the special case when \( b = 2 \) with an error term \( o(1) \) of unspecified order of magnitude.
In Section 1.3 of [1] Theorem 1 is proved, again in the special case \( b = 2 \), with an error term \( O \left( \frac{\log \log M}{\sqrt{\log M}} \right) \). Our proof of Theorem 1 is the generalization of the proof in Section 1.3 of [1]. In a doctoral dissertation ([7] Theorem 4.1.1.) a central limit theorem for the supremum norm \( \| \Delta_N \|_\infty \) of the discrepancy function in the case of an arbitrary base \( b \geq 2 \), similar to Theorem 3 is proved. The main difference is that \( c(b) \) is to be replaced by \( c_\infty(b) = \frac{2b-1}{12} \) and \( d(b) \) is to be replaced by

\[
d_\infty(b) = \frac{4b^7 - 10b^6 + 10b^5 + 14b^4 - 77b^3 + 127b^2 - 68b + 8}{720b^2(b-1)^2(b+1)}.
\]

Moreover, the theorem is stated only in the special case when \( M \) is a power of the base \( b \), and the error term is of an unspecified order of magnitude \( o(1) \). In [3] and [7] central limit theorems for various generalizations of the van der Corput sequence are also studied. Large deviation results have not yet been obtained.

Finally, we give a method to generalize Theorem 1 and Theorem 2 for sums of the form \( \sum_{n=0}^{N-1} f(x_n) \), where the function \( f : [0, 1] \to \mathbb{R} \) is sufficiently nice, and \( x_n \) is the base \( b \) van der Corput sequence. Since the discrepancy satisfies

\[
\sup_{x \in [0,1]} |\Delta_N(x)| = O \left( b \log_b N \right),
\]

the Koksma inequality ([5] Chapter 2 Theorem 5.1) implies that if \( f : [0, 1] \to \mathbb{R} \) is of bounded variation, then

\[
\sum_{n=0}^{N-1} f(x_n) = N \int_0^1 f(x) \, dx + O \left( \log N \right),
\]

as \( N \to \infty \), with an implied constant depending only on \( b \) and the total variation of \( f \). Under more restrictive assumptions on the function \( f \) the error term actually satisfies a central limit theorem and a large deviation result. The following proposition reduces the problem of studying the distribution of \( \sum_{n=0}^{N-1} f(x_n) \) to that of \( S(N) \).

**Proposition 5.** Let \( f : [0, 1] \to \mathbb{R} \) be twice differentiable with \( f'' \in L^1([0, 1]) \), and let \( x_n \) denote the base \( b \) van der Corput sequence, where \( b \geq 2 \) is an arbitrary integer. For any integer \( N > 0 \) we have

\[
\left| \sum_{n=0}^{N-1} f(x_n) - N \int_0^1 f(x) \, dx + (f(1) - f(0)) S(N) \right| \leq \frac{b}{3} \| f'' \|_1.
\]

The natural interpretation of the quantity \( f(1) - f(0) \) is that the periodic extension of \( f \) on \( \mathbb{R} \) with period 1 has jumps of this size.
In Section 2 we derive the normalizing factors \( c(b) \) and \( d(b) \) of Theorem 1. Section 3 is devoted to the proofs of Theorem 1 and Theorem 2, while the proofs of Theorem 3, Theorem 4 and Proposition 5 are given in Section 4.

2 The expected value and the variance of \( S(N) \)

We start by deriving a formula for the sum \( S(N) \) in terms of the base \( b \) digits of \( N \) as follows.

**Proposition 6.** Let \( b \geq 2 \) be an integer and let \( N = \sum_{i=1}^{m} a_i b^{i-1} \) be the base \( b \) representation of an integer \( N \geq 0 \), where \( a_i \in \{0, 1, \ldots, b-1\} \). Then

\[
S(N) = \sum_{i=1}^{m} \frac{(b+1)a_i - a_i^2}{2b} - \sum_{1 \leq i < j \leq m} \frac{a_ia_j}{b^{j-i+1}}.
\]

**Proof.** By splitting the sum \( S(N) \) we get

\[
S(N) = \sum_{n=0}^{a_mb^{m-1}-1} \left( \frac{1}{2} - x_n \right) + \sum_{n=a_mb^{m-1}}^{N-1} \left( \frac{1}{2} - x_n \right).
\]  

Since

\[
\{x_n : 0 \leq n < a_mb^{m-1}\} = \left\{ \frac{k}{b^{m-1}} + \frac{a}{b^m} : 0 \leq k < b^{m-1}, \ 0 \leq a < a_m \right\},
\]

we obtain that the first sum in (2) is

\[
\sum_{n=0}^{a_mb^{m-1}-1} \left( \frac{1}{2} - x_n \right) = \sum_{k=0}^{b^{m-1}-1} \sum_{a=0}^{a_m-1} \left( \frac{1}{2} - \frac{k}{b^{m-1}} - \frac{a}{b^{m}} \right) = \frac{(b+1)a_m - a_m^2}{2b}.
\]

To compute the second sum in (2) note that for any \( a_mb^{m-1} \leq n < N \) the first base \( b \) digit of \( n \) is \( a_m \), and hence \( x_n = x_{n-a_mb^{m-1}} + a_m \). Therefore by reindexing the sum we obtain

\[
\sum_{n=a_mb^{m-1}}^{N-1} \left( \frac{1}{2} - x_n \right) = \sum_{n=0}^{N-a_mb^{m-1}-1} \left( \frac{1}{2} - x_n - \frac{a_m}{b^m} \right)
\]

\[
= S(N - a_mb^{m-1}) - a_m \frac{b^{m} - N}{b}.
\]
Using the base $b$ representation of $N$ we thus find the recursion

$$
S \left( \sum_{i=1}^{m} a_i b^{i-1} \right) = \frac{(b+1)a_m - a_m^2}{2b} - \frac{m-1}{b^m - i+1} + S \left( \sum_{i=1}^{m-1} a_i b^{i-1} \right).
$$

(3)

Applying the recursion (3) $m$ times finishes the proof.

If $N$ is a random variable uniformly distributed in $\{0, 1, \ldots, b^m - 1\}$ for some integers $b \geq 2$ and $m \geq 1$, then the base $b$ digits $a_1, \ldots, a_m$ of $N$ are independent random variables, each uniformly distributed in $\{0, 1, \ldots, b - 1\}$. Therefore Proposition 6 can be used to find the expected value and the variance of the sum $S(N)$. Here and from now on the expected value and the variance of a real valued random variable $X$ are denoted by $E(X)$ and $\text{Var}(X)$, respectively.

**Proposition 7.** Let $N$ be a random variable which is uniformly distributed in $\{0, 1, \ldots, b^m - 1\}$ for some integers $b \geq 2$ and $m \geq 1$. Then

$$
\left| E \left( S(N) \right) - \frac{b^2 - 1}{12b^m} \right| \leq \frac{1}{4},
$$

$$
\text{Var} \left( S(N) \right) = \frac{b^4 + 120b^3 - 480b^2 + 600b - 241}{720b^2} m + O(b).
$$

The implied constant in the error term is absolute.

**Proof.** Using the independence of the base $b$ digits $a_1, \ldots, a_m$ of $N$, from Proposition 6 we get that the expected value of $S(N)$ is

$$
E \left( S(N) \right) = \sum_{i=1}^{m} \frac{(b+1)E(a_i) - E(a_i^2)}{2b} - \sum_{1 \leq i < j \leq m} \frac{E(a_i)E(a_j)}{b^{j-i+1}} = \frac{b^2 - 1}{12b^m} + \frac{1}{4} - \frac{1}{4b^m}.
$$

To find the variance of $S(N)$, first let us use the independence of $a_1, \ldots, a_m$ again to obtain

$$
\text{Var} \left( \sum_{i=1}^{m} \frac{(b+1)a_i - a_i^2}{2b} \right) = \sum_{i=1}^{m} \text{Var} \left( \frac{(b+1)a_i - a_i^2}{2b} \right) = \frac{b^4 + 55b^2 - 56}{720b^2} m.
$$

(4)
\[
\text{Var} \left( \sum_{1 \leq i < j \leq m} \frac{a_i a_j}{b^{j-i+1}} \right) = \sum_{1 \leq i_1 < j_1 \leq m} \sum_{1 \leq i_2 < j_2 \leq m} \left( E(a_{i_1}a_{j_1}a_{i_2}a_{j_2}) - \frac{(b-1)^4}{16} \right) \frac{1}{b^{j_1-i_1+1}b^{j_2-i_2+1}}.
\]

(5)

We will group the terms according to the size of \(\{i_1, j_1\} \cap \{i_2, j_2\}\). If \(\{i_1, j_1\} \cap \{i_2, j_2\}\) is the empty set, then \(a_{i_1}, a_{j_1}, a_{i_2}, a_{j_2}\) are independent, and therefore the contribution is zero.

If \(\{i_1, j_1\} \cap \{i_2, j_2\}\) has size 1, then

\[
E(a_{i_1}a_{j_1}a_{i_2}a_{j_2}) - \frac{(b-1)^4}{16} = \frac{(b-1)^3(b+1)}{48}.
\]

Let \(s > 0, t > 0\) and \(1 \leq A \leq m - s - t\) be integers. The sum of over all \(1 \leq i_1 < j_1 \leq m\) and \(1 \leq i_2 < j_2 \leq m\) such that \(\{i_1, j_1\} \cup \{i_2, j_2\} = \{A, A+s, A+s+t\}\) is \(\frac{2}{b^{2s+t+2}} + \frac{2}{b^{s+t+2}} + \frac{2}{b^{s+2t+2}}\), hence we have that the contribution of this case in (5) is

\[
\sum_{s,t>0} \frac{(b-1)^3(b+1)}{48} (m - s - t) \left( \frac{2}{b^{2s+t+2}} + \frac{2}{b^{s+t+2}} + \frac{2}{b^{s+2t+2}} \right) = \frac{b^2 + 2b - 3}{24b^2} m + O(1).
\]

If \(\{i_1, j_1\} \cap \{i_2, j_2\}\) has size 2, then \(i_1 = i_2\) and \(j_1 = j_2\), and hence

\[
E(a_{i_1}a_{j_1}a_{i_2}a_{j_2}) - \frac{(b-1)^4}{16} = \frac{(7b^2 - 12b + 5)(b^2 - 1)}{144}.
\]

Therefore the contribution of this case in (5) is

\[
\sum_{1 \leq i < j \leq m} \frac{1}{b^{2j-2i+2}} = \frac{7b^2 - 12b + 5}{144b^2} m + O(1).
\]

Altogether we find that

\[
\text{Var} \left( \sum_{1 \leq i < j \leq m} \frac{a_i a_j}{b^{j-i+1}} \right) = \frac{13b^2 - 13}{144b^2} m + O(1).
\]

(6)

Finally, it is easy to see that two times the covariance of the sums in question is
\[
2 \sum_{1 \leq i_1 \leq m \atop 1 \leq i_2 < j_2 \leq m} E \left( \left( \frac{(b+1)a_{i_1}-a_{i_1}^2}{2b} - \frac{b^2+3b-4}{12b} \right) \left( \frac{(b-1)^2}{4} - a_{i_2}a_{j_2} \right) \frac{1}{b_{j_2}} \right)
\]

\[
= \frac{b^3 - 5b^2 + 5b - 1}{6b^2} (m-1) + O(1), \quad (7)
\]

by noticing that the terms for which \(i_1 \notin \{i_2, j_2\}\) are all zero. Adding (1), (3) and (7), we obtain the desired formula for \(\text{Var}(S(N))\).

\[
\square
\]

### 3 Proofs of Theorem 1 and Theorem 2

Let \(N\) be a random variable again, uniformly distributed in \(\{0, 1, \ldots, b^m-1\}\). Proposition 6 expresses \(S(N)\) in terms of independent random variables \(a_1, \ldots, a_m\).

In this Section we prove a general central limit theorem and a large deviation result for random variables expressed in terms of independent variables in a similar way. These general results fit into the subject of weakly dependent random variables. The proof of Theorem 9 below is the generalization of the proof in Section 1.3 of [1].

For positive integers \(a\) and \(m\) let \([m]\) denote the set \(\{1, 2, \ldots, m\}\), and let

\[
\left( [m] \atop \leq a \right) = \{A \subseteq [m] : |A| \leq a\}.
\]

For a finite set \(A\) of integers let \(\text{diam} A = \max A - \min A\), and for random variables \(X_1, \ldots, X_m\) let \(X_A = (X_i : i \in A)\) for any \(A \subseteq [m]\).

We are going to use the fact that for any real numbers \(\lambda\) and \(x\) we have

\[
\Phi(\lambda + x) = \Phi(\lambda) + O(|x|), \quad (8)
\]

\[
\Phi(\lambda(1 + x)) = \Phi(\lambda) + O(|x|). \quad (9)
\]

Note that \(\Phi(\lambda + x) - \Phi(\lambda)\) is the integral of \(\frac{1}{\sqrt{2\pi}} e^{-t^2/2}\) over an interval of length \(|x|\), therefore (8) in fact holds with implied constant \(\frac{1}{\sqrt{2\pi}}\). Since \(0 \leq \Phi \leq 1\), (8) holds for any \(|x| > \frac{1}{2}\) with implied constant 2. If \(|x| \leq \frac{1}{2}\), then for \(\lambda \geq 0\) \(\Phi(\lambda(1 + x)) - \Phi(\lambda)\) is an integral over an interval of length \(|\lambda x|\), moreover this interval is contained in \([\lambda/2, 3\lambda/2]\), therefore the integrand is at most \(\frac{1}{\sqrt{2\pi}} e^{-\lambda^2/2}\).

Hence
\[ |\Phi(\lambda(1 + x)) - \Phi(\lambda)| \leq |\lambda x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \]

and clearly the same is true for \( \lambda < 0 \). Note that \( \frac{|\lambda|}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{8}} \) is bounded on \( \mathbb{R} \), in fact the maximum is attained at \( \lambda = \pm 2 \) with maximum value less than 2. Thus altogether (9) holds with implied constant 2.

**Proposition 8.** Let \( 2 \leq a \leq m \) be integers, and let \( X_1, X_2, \ldots, X_m \) be independent real valued random variables. For every \( A \in \binom{[m]}{\leq a} \) let \( f_A : \mathbb{R}^{|A|} \to \mathbb{R} \) be Borel measurable. Suppose that for every \( A \in \binom{[m]}{\leq a} \) we have

1. \( E f_A(X_A) = 0 \),
2. \( |f_A(X_A)| \leq e^{-c \cdot \text{diam} A} \)

for some constant \( c > 0 \). Let \( q = \left( \frac{2}{1 - e^{-c}} \right)^{a+\frac{1}{2}} \) and \( g(x) = \sum_{k=0}^{\infty} \frac{x^{2ak}}{(2ak)!} \).

1. For any integer \( k \geq 1 \) we have

\[ E \left( \sum_{A \in \binom{[m]}{\leq a}} f_A(X_A) \right)^{2k} \leq q^{2k}(2ak)! \cdot m^k. \]

2. For any real number \( \lambda \geq 1 \) we have

\[ \Pr \left( \left| \sum_{A \in \binom{[m]}{\leq a}} f_A(X_A) \right| \geq \lambda q \sqrt{m} \right) \leq \frac{\sqrt{\lambda}}{g(\sqrt{\lambda} - 1)}. \]

**Proof.** (1) Let \( L \) denote the left hand side of the claim. By expanding we get

\[ L = E \left( \sum_{A \in \binom{[m]}{\leq a}} f_A(X_A) \right)^{2k} = \sum_{A_1, \ldots, A_{2k} \in \binom{[m]}{\leq a}} E \prod_{i=1}^{2k} f_{A_i}(X_{A_i}). \]

For each ordered \( 2k \)-tuple \((A_1, \ldots, A_{2k}) \in \binom{[m]}{\leq a}^{2k}\) consider the hypergraph \( \mathcal{H} \) on \([m]\) with edges \( A_1, \ldots, A_{2k} \). In this proof by a hypergraph we mean an unordered collection of subsets of \([m]\), called edges, with possible repetitions. Let \( p \) denote the number of connected components of \( \mathcal{H} \), where \( 1 \leq p \leq 2k \). Note that if
\( p > k \), then there exists an isolated edge in \( \mathcal{H} \), which using the independence of \( X_1, \ldots, X_m \) and condition (i) implies that

\[
E \prod_{i=1}^{2k} f_{A_i}(X_{A_i}) = 0.
\]

Suppose now that \( 1 \leq p \leq k \). Let \( C_1, \ldots, C_p \) be the connected components of \( \mathcal{H} \), and let \( d_j = \text{diam} \bigcup C_j \). The main observation is that the connectedness implies

\[
d\bigcup C_j \leq \sum_{A \in C_j} \text{diam} A,
\]

\[
\sum_{j=1}^{p} d_j \leq \sum_{i=1}^{2k} \text{diam} A_i,
\]

\[
\left| \prod_{i=1}^{2k} f_{A_i}(X_{A_i}) \right| \leq \exp \left( -c \cdot \sum_{i=1}^{2k} \text{diam} A_i \right) \leq \exp \left( -c \cdot \sum_{j=1}^{p} d_j \right). \tag{11}
\]

Let \( M_j = \min \bigcup C_j \). Then \( \bigcup C_j \subseteq [M_j, M_j + d_j] \). We are going to group the terms of (10) according to the values \( p, M_1, \ldots, M_p, d_1, \ldots, d_p \) associated with the corresponding hypergraph \( \mathcal{H} \). For given \( p, M_1, \ldots, M_p, d_1, \ldots, d_p \) all the sets \( A_1, \ldots, A_{2k} \) have to be a subset of the set

\[
\bigcup_{j=1}^{p} [M_j, M_j + d_j]
\]

of size at most \( \sum_{j=1}^{p} d_j + p \). The number of ordered \( 2k \)-tuples \((A_1, \ldots, A_{2k}) \in \binom{[m]}{\leq a}^{2k}\) for which the corresponding hypergraph \( \mathcal{H} \) has associated values \( p, M_1, \ldots, M_p, d_1, \ldots, d_p \) is therefore at most

\[
\left( \sum_{j=1}^{p} d_j + p \right)^{2ak}.
\]

This together with (11) implies that in (10) we have

\[
L \leq \sum_{p=1}^{k} \sum_{M_1, \ldots, M_p} \sum_{d_1, \ldots, d_p=0}^{\infty} \left( \sum_{j=1}^{p} d_j + p \right)^{2ak} \exp \left( -c \sum_{j=1}^{p} d_j \right).
\]

Let \( d = \sum_{j=1}^{p} d_j \). It is known that the number of representations of a given nonnegative integer \( d \) in this form is \( \binom{d+p-1}{p-1} \), therefore we get
\[ L \leq \sum_{p=1}^{k} \sum_{d=0}^{\infty} \left( \frac{d + p - 1}{p - 1} \right) (d + p)^{2ak} e^{-cd} m^p \leq \sum_{p=1}^{k} \sum_{d=0}^{\infty} \frac{\prod_{j=1}^{2ak+p-1} (d + j)}{(p - 1)!} e^{-cd} m^p. \]

The series over \( d \) is in fact the well-known Taylor series
\[ \sum_{d=0}^{\infty} (d + \ell) \cdots (d + 2)(d + 1)x^d = \frac{\ell!}{(1 - x)^{\ell+1}} \]
with \( \ell = 2ak + p - 1 \) and \( x = e^{-c} \), thus we have
\[ L \leq \sum_{p=1}^{k} \frac{(2ak + p - 1)!}{(p - 1)!} \cdot \frac{m^p}{(1 - e^{-c})^{2ak+p}} = \sum_{p=1}^{k} \frac{(2ak + p - 1)}{2ak} (2ak)! \cdot \frac{m^p}{(1 - e^{-c})^{2ak+p}}. \]
Here for every \( 1 \leq p \leq k \) we have
\[ \frac{m^p}{(1 - e^{-c})^{2ak+p}} \leq \frac{m^k}{(1 - e^{-c})^{(2a+1)k}}. \]

We can also use the combinatorial identity and trivial estimate
\[ \binom{n}{n} + \binom{n+1}{n} + \cdots + \binom{n+k-1}{n} = \binom{n+k}{n+1} \leq 2^{n+k} \]
with \( n = 2ak \) to finally obtain
\[ L \leq 2^{(2a+1)k}(2ak)! \cdot \frac{m^k}{(1 - e^{-c})^{(2a+1)k}} = q^{2k}(2ak)!m^k. \]

(2) Let \( P \) denote the probability in the claim. Note that \( g(x) \) is monotone increasing on \([0, \infty)\). Therefore for any real number \( 0 < \alpha < 1 \) we have
\[ P = \Pr \left( \left\| \sum_{A \in \binom{[m]}{\alpha m}} f_A (X_A) \right\| \geq \lambda \sqrt{m} \right) \]
\[ = \Pr \left( g \left( \frac{\alpha}{q^\frac{\alpha}{\sqrt{m}}} \sum_{A \in \binom{[m]}{\alpha m}} f_A (X_A) \right)^{\frac{1}{\alpha}} \right) \geq g \left( \alpha \lambda^{\frac{1}{\alpha}} \right). \]

Applying Markov’s inequality and Lebesgue’s monotone convergence theorem we obtain that
\[ P \leq \frac{1}{g(\alpha \lambda^2)} \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{q^{2k} m^{k}(2ak)!} E \left( \sum_{A \in \binom{[m]}{\leq a}} f_A(X_A) \right)^{2k}. \]

Proposition 8 (1) yields the upper bound

\[ P \leq \frac{1}{g(\alpha \lambda^2)} \sum_{k=0}^{\infty} \alpha^{2k} = \frac{1}{1 - \alpha^{2a}} \cdot \frac{1}{g(\alpha \lambda^2)}. \]

Choosing \( \alpha = 1 - \frac{1}{\lambda} - a \) and noticing \( 1 - \alpha^{2a} \geq 1 - \alpha = \lambda^{-\frac{1}{2}} \) finishes the proof.

\[ \Box \]

**Theorem 9.** Let \( 2 \leq a \leq m \) be integers, and let \( X_1, X_2, \ldots, X_m \) be independent real valued random variables. For every \( A \in \binom{[m]}{\leq a} \) let \( f_A : \mathbb{R}^{|A|} \rightarrow \mathbb{R} \) be Borel measurable. Suppose that for every \( A \in \binom{[m]}{\leq a} \) we have

(i) \( E f_A(X_A) = 0 \),

(ii) \( |f_A(X_A)| \leq e^{-c \cdot \text{diam} A} \),

(iii) \( \sigma_m^2 = E \left( \sum_{A \in \binom{[m]}{\leq a}} f_A(X_A) \right)^2 > 0 \)

for some constant \( c > 0 \). Then for any real number \( \lambda \) we have

\[ \text{Pr}\left( \frac{1}{\sigma_m} \sum_{A \in \binom{[m]}{\leq a}} f_A(X_A) < \lambda \right) = \Phi(\lambda) + O\left( \frac{1}{\sqrt{\log m}} \cdot \frac{m^{\frac{3}{4}}}{\sigma_m^2} \right). \]

The implied constant in the error term depends only on \( a \) and \( c \).

Note that Proposition 8 (1) with \( k = 1 \) implies that \( \sigma_m^2 = O(m) \). The smallest attainable error term in Theorem 9 is therefore \( O\left( \frac{1}{\sqrt{\log m}} \cdot \frac{m^{\frac{3}{4}}}{\sigma_m^2} \right) \), which holds whenever \( \sigma_m^2 > d \cdot m \) for some constant \( d > 0 \).

**Proof.** Throughout this proof the implied constants in the \( O \) notation will depend only on \( a \) and \( c \). We may assume \( \sigma_m^2 \geq m^\frac{3}{4} \), otherwise the error term is larger than \( 1 \). We start by partitioning the set \( [m] \) into \( m_0 \) intervals of integers \( I_1, I_2, \ldots, I_{m_0} \), in such a way that \( \max I_i = \min I_{i+1} - 1 \) and \( |I_i| = \Theta\left( \frac{m}{m_0} \right) \) for any \( i \). Assume \( |I_i| > \frac{6}{c} \log m \) for all \( i \). Let
\[ Y_i = \sum_{A \in \binom{I_i}{\leq a}} f_A(X_A), \]

\[ Z_j = \sum_{A \in \binom{I_j}{\leq a}} f_A(X_A). \]

Then the random variable we are interested in can be written as

\[ \sum_{A \in \binom{I_i}{\leq a}} f_A(X_A) = \sum_{i=1}^{m_0} Y_i + \sum_{j=1}^{m_0-1} Z_j + W, \quad (12) \]

where the random variable \( W \) is defined by (12). Then \( Y_1, \ldots, Y_{m_0} \) are independent, and the assumption \( |I_i| > \frac{3}{2} \log m \) implies that \( Z_1, \ldots, Z_{m_0-1} \) are also independent.

Since the number of sets \( A \in \binom{I_i}{\leq a} \) such that \( \text{diam } A = d \) is at most \( m \cdot (d+1)^a \), condition (ii) implies that

\[ |W| \leq \sum_{\substack{A \in \binom{I_i}{\leq a} \\text{diam } A > \frac{3}{2} \log m}} e^{-c \text{diam } A} \leq \sum_{d>\frac{3}{2} \log m} m(d+1)^a e^{-cd} \]

\[ = O \left( m \log^a m \cdot e^{-c \frac{3}{2} \log m} \right) = O \left( \frac{1}{m} \right). \quad (13) \]

Similarly,

\[ |Y_i| \leq \sum_{A \in \binom{I_i}{\leq a}} e^{-c \text{diam } A} \leq \sum_{d=0}^{\infty} |I_i|(d+1)^a e^{-cd} = O \left( |I_i| \right) = O \left( \frac{m}{m_0} \right). \quad (14) \]

The number of sets \( A \in \binom{I_j}{\leq a} \) with \( A \subseteq [\max I_j - d, \max I_j + d] \) is at most \((2d+1)^a\), therefore condition (ii) implies

\[ |Z_j| \leq \sum_{d=0}^{\infty} (2d+1)^a e^{-cd} = O(1). \quad (15) \]

Finally, note that the number of sets \( A \in \binom{I_j}{\leq a} \) such that \( \text{diam } A = d_1 \) which intersect \([\max I_j - d_2, \max I_j + d_2]\) is at most \((2d_1+2d_2+1)^a\), thus from conditions (i) and (ii) we obtain that for any \( i \) and \( j \) we have
\[ |E (Y_i Z_j)| \leq \sum_{d_1, d_2 \geq 0} (2d_1 + 2d_2 + 1)^a (2d_2 + 1)^a e^{-cd_1} e^{-cd_2} = O(1). \quad (16) \]

By taking the variance of (12) we get

\[
\sigma_m^2 = \sum_{i=1}^{m_0} \text{Var} (Y_i) + \sum_{j=1}^{m_0-1} \text{Var} (Z_j) + 2 \sum_{i=1}^{m_0} \sum_{j=1}^{m_0-1} E (Y_i Z_j) + 2 \sum_{i=1}^{m_0} E (Y_i W) + 2 \sum_{j=1}^{m_0-1} E (Z_j W) + \text{Var} (W).
\]

By noticing that \( E (Y_i Z_j) = 0 \) unless \( i = j \) or \( i = j + 1 \), the bounds (13)–(16) imply

\[
\sigma_m^2 = \sum_{i=1}^{m_0} \text{Var} (Y_i) + O (m_0). \quad (17)
\]

We now want to apply the Berry–Esseen theorem to the sum \( \sum_{i=1}^{m_0} Y_i \) of independent random variables. Applying Proposition 8 (1) with \( k = 2 \) we obtain

\[
E Y_i^4 = O (|I_i|^3) = O \left( \frac{m^2}{m_0^2} \right),
\]

therefore the Hölder inequality implies

\[
\sum_{i=1}^{m_0} E |Y_i|^3 \leq \sum_{i=1}^{m_0} (E Y_i^4)^\frac{3}{4} = O \left( \frac{m^2}{\sqrt{m_0}} \right).
\]

As long as \( m_0 = o (\sigma_m^2) \), we can see from (17) that

\[
\left( \sum_{i=1}^{m_0} \text{Var} (Y_i) \right)^\frac{3}{2} = \sigma_m^3 (1 + o(1)).
\]

Therefore the Berry–Esseen theorem ([2] Section 9.1 Theorem 3) implies that

\[
\Pr \left( \frac{1}{\sqrt{\sum_{i=1}^{m_0} \text{Var} (Y_i)}} \sum_{i=1}^{m_0} Y_i < \lambda \right) = \Phi (\lambda) + O \left( \frac{\sum_{i=1}^{m_0} E |Y_i|^3}{(\sum_{i=1}^{m_0} \text{Var} (Y_i))^\frac{3}{2}} \right)
\]

\[
= \Phi (\lambda) + O \left( \frac{m^\frac{3}{2}}{\sigma_m^3 \sqrt{m_0}} \right). \quad (18)
\]

14
From (17) we obtain
\[ \frac{1}{\sqrt{\sum_{i=1}^{m_0} \text{Var}(Y_i)}} = \frac{1}{\sigma_m} \left( 1 + O \left( \frac{m_0}{\sigma_m^2} \right) \right). \]

Therefore we can use (9) with \( x = O \left( \frac{m_0}{\sigma_m} \right) \) to replace the normalizing factor in the probability in (18) by \( \frac{1}{\sigma_m} \) to get
\[ \text{Pr} \left( \frac{1}{\sigma_m} \sum_{i=1}^{m_0} Y_i < \lambda \right) = \Phi(\lambda) + O \left( \frac{m_0^2}{\sigma_m^3 \sqrt{m_0}} + \frac{m_0}{\sigma_m^2} \right). \] (19)

Recall that a simple version of the Chernoff bound states that if \( \zeta_1, \ldots, \zeta_n \) are independent random variables such that \( \text{E}(\zeta_j) = 0 \) and \( |\zeta_j| \leq 1 \) for every \( 1 \leq j \leq n \), then for any \( t > 0 \) we have
\[ \text{Pr} \left( \left| \sum_{j=1}^{n} \zeta_j \right| > t \sqrt{n} \right) \leq 2e^{-\frac{t^2}{2}}. \]

According to (15) there exists a constant \( K > 0 \) such that \( |Z_j| \leq K \) for all \( j \). Condition (i) ensures that \( \text{E}(Z_j) = 0 \) for all \( j \). Therefore we can apply the Chernoff bound to \( \zeta_j = Z_j/K \) with \( n = m_0 - 1 \) and \( t = \sqrt{\log m} \) to obtain
\[ \text{Pr} \left( \frac{1}{\sigma_m} \sum_{j=1}^{m_0-1} Z_j > K \sqrt{\log m - \frac{m_0-1}{\sigma_m}} \right) \leq \frac{2}{\sqrt{m}}. \] (20)

From (12), (13) and (20) we get
\[ \text{Pr} \left( \frac{1}{\sigma_m} \sum_{A \in \binom{[m]}{\leq s}} f_A(X_A) < \lambda \right) = \text{Pr} \left( \frac{1}{\sigma_m} \sum_{i=1}^{m_0} Y_i < \lambda + O \left( \sqrt{\log m - \frac{m_0}{\sigma_m} + \frac{1}{\sigma_m m}} \right) \right) + O \left( \frac{1}{\sqrt{m}} \right). \]

Combining (19) and (8) with \( x = O \left( \sqrt{\log m - \frac{m_0}{\sigma_m} + \frac{1}{\sigma_m m}} \right) \) we finally obtain
\[ \Pr \left( \frac{1}{\sigma_m} \sum_{A \in \binom{[m]}{\leq 2}} f_A(X_A) < \lambda \right) = \Phi(\lambda) + O\left( \frac{m^2}{\sigma^2_m \sqrt{m}} + \frac{m_0}{\sigma^2_m} + \sqrt{\log m} \sqrt{m_0 \sigma_m} + \frac{1}{\sigma_m m} + \frac{1}{\sqrt{m}} \right). \]

The optimal choice for \( m_0 \) is when the first and the third error terms are equal, which holds when

\[ m_0 = \Theta\left( \frac{m^2}{\log m \cdot \sigma^2_m} \right). \]

Using \( \sigma^2_m \geq m^2 \) it is easy to check that for this choice of \( m_0 \) both our assumptions \(|I_i| > 6 \log m c\) and \( m_0 = o(\sigma^2_m) \) hold.

\[ \square \]

**Proof of Theorem** Suppose that \( M = b^m \) for some integer \( m \geq 2 \). Let \( N \) be a random variable uniformly distributed in \([0,1,\ldots,b^m - 1]\). Then the base \( b \) digits \( a_1,\ldots,a_m \) of \( N \) are independent random variables. Let \( K > 0 \) be a constant for which

\[ \left| \frac{(b+1)a_i - a^2_i}{2b} - \mathbb{E} \left( \frac{(b+1)a_i - a^2_i}{2b} \right) \right| \leq Kb, \]

\[ \left| \frac{a_ia_j}{b} - \mathbb{E} \left( \frac{a_ia_j}{b} \right) \right| \leq Kb \]

for any \( 1 \leq i < j \leq m \). Using Proposition we can write \( S(N) \) in the form

\[ S(N) - \mathbb{E} (S(N)) = Kb \sum_{A \in \binom{[m]}{\leq 2}} f_A(a_A), \]

where \( f_0 = 0, f_{\{i\}} (x) = \frac{(b+1)x-x^2}{2Kb^2} - \mathbb{E} \left( \frac{(b+1)a_i - a^2_i}{2Kb^2} \right) \) and for \( 1 \leq i < j \leq m \)

\[ f_{\{i,j\}} (x, y) = -\left( \frac{xy}{Kb^2} - \mathbb{E} \left( \frac{a_ia_j}{Kb^2} \right) \right) \cdot \frac{1}{b^{j-i}}. \]

Then the conditions of Theorem are satisfied with \( a = 2 \) and \( c = \log 2 \). According to Proposition we have \( \sigma^2_m = \frac{1}{K^2b^2} \operatorname{Var} (S(N)) = \Theta(m) \), hence we obtain

\[ \Pr \left( \frac{S(N) - \mathbb{E} (S(N))}{\sqrt{\operatorname{Var} (S(N))}} < \lambda \right) = \Phi(\lambda) + O\left( \sqrt{\log m} \right). \]
Since $d(b) = \Theta(b^2)$, from Proposition 7 we can see that

$$\frac{1}{\sqrt{\text{Var}(S(N))}} = \frac{1}{\sqrt{d(b)m}} \left(1 + O\left(\frac{1}{bm}\right)\right),$$

$$\frac{E(S(N))}{\sqrt{d(b)m}} = \frac{c(b)m}{\sqrt{d(b)m}} + O\left(\frac{1}{b\sqrt{m}}\right).$$

Hence if we replace $\text{Var}(S(N))$ by $d(b)m$, and then $E(S(N))$ by $c(b)m$ in the probability, then using (9) with $x = O\left(\frac{1}{bm}\right)$ and (8) with $x = O\left(\frac{1}{b\sqrt{m}}\right)$ the error we make is $O\left(\frac{1}{bm} + \frac{1}{b\sqrt{m}}\right)$. Thus

$$\text{Pr}\left(\frac{S(N) - c(b)m}{\sqrt{d(b)m}} < \lambda\right) = \Phi(\lambda) + O\left(\frac{\sqrt{\log m}}{\sqrt{m}}\right). \quad (21)$$

We now show that (21) holds for any $M > b^2$. Let $M = \sum_{i=1}^{m} c_i b^{i-1}$ be the base $b$ representation of $M$, where $c_i \in \{0, 1, \ldots, b - 1\}$ and $c_m > 0$. Let

$${M^* = \sum_{m \log m - 1 \leq i \leq m} c_i b^{i-1}.}$$

Let $N$ be a random variable uniformly distributed in $\{0, 1, \ldots, M^* - 1\}$, and consider its base $b$ representation $N = \sum_{i=1}^{m} a_i b^{i-1}$. Note that we allow $a_m$ to be zero. Then the random variables $(a_i : 1 \leq i < m - \log m - 1)$ are independent, and each is uniformly distributed in $\{0, 1, \ldots, b - 1\}$. Let us introduce new random variables $a_j^*$ for every $m - \log m - 1 \leq j \leq m$, such that

$$(a_i, a_j^* : 1 \leq i < m - \log m - 1 \leq j \leq m)$$

are identically distributed independent random variables. Let

$$N^* = \sum_{1 \leq i < m - \log m - 1} a_i b^{i-1} + \sum_{m - \log m - 1 \leq j \leq m} a_j^* b^{j-1}.$$  

Then $S(N^*)$ satisfies (21). Note that there are $O(\log m)$ base $b$ digits at which $N$ and $N^*$ differ. According to the formula in Proposition 6 if a single base $b$ digit of $N$ is changed, $S(N)$ can change by at most $O(b)$. Hence $S(N^*) = S(N) + O(b \log m)$. Using (8) with $x = O\left(\frac{\log m}{\sqrt{m}}\right)$, the error of replacing $S(N^*)$ in (21) by $S(N)$ is $O\left(\frac{\sqrt{\log m}}{\sqrt{m}}\right)$, therefore

$$\frac{1}{M^*} \left\{0 \leq N < M^* : \frac{S(N) - c(b)m}{\sqrt{d(b)m}} < \lambda\right\} = \Phi(\lambda) + O\left(\frac{\sqrt{\log m}}{\sqrt{m}}\right).$$
Here the error of replacing $M^*$ by $M$ is

$$O \left( \frac{M - M^*}{M} \right) = O \left( \frac{b^m - \log m - 1}{b^m - 1} \right) = O \left( \frac{\sqrt{\log m}}{\sqrt{m}} \right).$$

Finally, note that $\frac{M}{m} \leq N \leq M$ with probability $1 - O \left( \frac{1}{m} \right)$, and for all such $N$ we have $\log_b N = m + O (\log m)$. Using (8) with $x = O \left( \frac{\log m}{\sqrt{m}} \right)$, the error of replacing $c(b)m$ by $c(b) \log_b N$ is $O \left( \frac{\log m}{\sqrt{m}} \right)$. Using (9) with $x = O \left( \frac{\log m}{m} \right)$, the error of replacing $\sqrt{d(b)m}$ by $\sqrt{d(b) \log_b N}$ is $O \left( \frac{\log m}{m} \right)$. Hence we get

$$\frac{1}{M} \left\{ 0 \leq N < M : \frac{S(N) - c(b) \log_b N}{\sqrt{d(b) \log_b N}} < \lambda \right\} = \Phi(\lambda) + O \left( \frac{\sqrt{\log m}}{\sqrt{m}} \right).$$

The error term can be expressed in terms of $M$ by noting $m \geq \log b M$.

**Proof of Theorem**

First, assume $M = b^m$ for some integer $m \geq 2$. Let $N$ be a random variable uniformly distributed in $\{0, 1, \ldots, b^m - 1\}$, and let $N = \sum_{i=1}^{m} a_i b^{i-1}$ be the base $b$ representation of $N$, where $a_1, \ldots, a_m$ are independent random variables, each uniformly distributed in $\{0, 1, \ldots, b - 1\}$. Note that for any $1 \leq i < j \leq m$ we have

$$\left| \frac{(b + 1)a_i - a_j^2}{2b} - \frac{(b + 1)E(a_i) - E(a_j^2)}{2b} \right| \leq \frac{3}{4} b,$$

$$\left| \frac{a_i a_j}{b} - \frac{E(a_i)E(a_j)}{b} \right| \leq \frac{3}{4} b.$$

Using Proposition 6 we can write $S(N)$ in the form

$$S(N) - E (S(N)) = \frac{3}{4} b \sum_{A \in \{\{0\} \cup \{i\} \cup \{i,j\} \cup \{\{i,j\}\}} f_A (a_A),$$

where $f_\emptyset = 0$, $f_{\{i\}} (x) = \frac{4}{3b} \frac{(b+1)x-x^2}{2b} - \frac{4}{3b} E \left( \frac{(b+1)a_i - a_i^2}{2b} \right)$ and for $1 \leq i < j \leq m$

$$f_{\{i,j\}} (x, y) = -\frac{4}{3b} \left( \frac{xy}{b} - E \left( \frac{a_i a_j}{b} \right) \right) \cdot \frac{1}{b^{j-i}}.$$

Then the conditions of Proposition 8 (2) are satisfied with $a = 2$, $c = \log 2$, $q = 32$ and

$$g(x) = \sum_{k=0}^{\infty} \frac{x^{4k}}{(4k)!} = e^x + e^{-x} \frac{\cos x}{4} + \frac{\cos x}{2} \geq \frac{e^x - 2}{4}.$$
Therefore Proposition 8 (2) yields

\[
\Pr \left( |S(N) - \mathbb{E}(S(N))| \geq 24\lambda b \sqrt{m} \right) = \Pr \left( \left| \sum_{A \in \binom{[m]}{\leq 2}} f_A (a_A) \right| \geq 32\lambda \sqrt{m} \right) \leq \frac{4\sqrt{\lambda}}{e^{\sqrt{\lambda} - 1} - 2}.
\] (22)

Now we prove (22) holds for any integer \( M > b \). Let \( M = \sum_{i=1}^{m} c_i b^{i-1} \) be the base \( b \) representation of \( M \), where \( c_i \in \{0, 1, \ldots, b-1\} \) and \( c_m > 0 \). Let

\[
N^* = \sum_{m-\sqrt{m} + 1 \leq i \leq m} c_i b^{i-1}.
\]

Let \( N \) be a random variable uniformly distributed in \( \{0, 1, \ldots, M^* - 1\} \), and consider its base \( b \) representation \( N = \sum_{i=1}^{m} a_i b^{i-1} \). Then \( (a_i : 1 \leq i < m - \sqrt{m} + 1) \) are independent random variables, each uniformly distributed in \( \{0, 1, \ldots, b-1\} \). Let us introduce new random variables \( a^*_j \) for \( m - \sqrt{m} + 1 \leq j \leq m \) such that

\[
(a_i, a^*_j : 1 \leq i < m - \sqrt{m} + 1 \leq j \leq m)
\]

are identically distributed independent random variables. Let

\[
N^* = \sum_{1 \leq i < m - \sqrt{m} + 1} a_i b^{i-1} + \sum_{m - \sqrt{m} + 1 \leq j \leq m} a^*_j b^{j-1}.
\]

Then \( S(N^*) \) satisfies (22). Using Proposition 6 and Proposition 7 we get the following estimates:

\[
\left| \mathbb{E}(S(N^*)) - \frac{b^2 - 1}{12b} m \right| \leq \frac{1}{4} \leq \frac{\lambda b \sqrt{m}}{24 \sqrt{2}},
\]

\[
\left| \frac{b^2 - 1}{12b} m - \frac{b^2 - 1}{12b} \log_b M \right| \leq \frac{b^2 - 1}{12b} \leq \frac{\lambda b \sqrt{m}}{36 \sqrt{2}},
\]

\[
|S(N) - S(N^*)| \leq \frac{(b+1)^2}{8b} \sqrt{m} + 2\sqrt{m} \leq \frac{41}{96} \lambda b \sqrt{m}.
\]

Since

\[
24 + \frac{1}{24 \sqrt{2}} + \frac{1}{36 \sqrt{2}} + \frac{41}{96} < 25,
\]

these estimates imply
$$\frac{1}{M^*} \left\{ 0 \leq N < M^* : \left| S(N) - \frac{b^2 - 1}{12b} \log_b M \right| \geq 25\lambda b \sqrt{\log_b M + 1} \right\} \leq \frac{4\sqrt{\lambda}}{e^{\sqrt{\lambda}} - 2}. $$

Finally, note that the error of replacing $M^*$ by $M$ is at most

$$\frac{M - M^*}{M} \leq \frac{b^{m-\sqrt{m}+1}}{b^{m-1}} \leq \frac{1}{b^{\sqrt{\log_b M} - 2}}. $$

\[\square\]

4 Proofs of Theorem 3 and Theorem 4

In this Section the proofs of Theorem 3, Theorem 4 and Proposition 5 are given.

We start by estimating an exponential sum in terms of the base $b$ van der Corput sequence. Proposition 10 below is a special case of Lemma 3 in [6]. Nevertheless, for the sake of completeness a proof is included.

**Proposition 10.** Let $b \geq 2$ be an integer and let $x_n$ denote the base $b$ van der Corput sequence. If $\ell$ is an integer such that $b^s \nmid \ell$ for some positive integer $s$, then for any positive integer $N$ we have

$$\left|\sum_{n=0}^{N-1} e^{2\pi i \ell x_n} \right| < b^s. $$

**Proof.** Let $N = \sum_{j=1}^{m} a_j b^{j-1}$ be the base $b$ representation of $N$ with base $b$ digits $a_j \in \{0, 1, \ldots, b-1\}$ with $a_m \neq 0$. By splitting the sum we get

$$\left|\sum_{n=0}^{N-1} e^{2\pi i \ell x_n} \right| \leq \left|\sum_{n=0}^{a_m b^{m-1}-1} e^{2\pi i \ell x_n} \right| + \left|\sum_{n=a_m b^{m-1}}^{N-1} e^{2\pi i \ell x_n} \right|. \tag{23} $$

Note that for any $a_m b^{m-1} \leq n < N$ the base $b$ representation of $n$ starts with the digit $a_m$. From the definition of the base $b$ van der Corput sequence we know that for any such $n$ we have $x_n = x_{n-a_m b^{m-1}} + \frac{a_m}{b^m}$, therefore we can reindex the second sum to obtain

$$\left|\sum_{n=a_m b^{m-1}}^{N-1} e^{2\pi i \ell x_n} \right| = \left|\sum_{n=0}^{N-a_m b^{m-1}-1} e^{2\pi i \ell \frac{a_m}{b^m} x_{n+a_m b^{m-1}}} \right|. $$

Using the base $b$ representation of $N$, repeated application of the triangle inequality in (23) yields
\[
\left| \sum_{n=0}^{N-1} e^{2\pi i t x_n} \right| \leq \sum_{j=1}^{m} \left| \sum_{n=0}^{a_j b^{j-1} - 1} e^{2\pi i t x_n} \right|.
\]  

(24)

For any \(1 \leq j \leq m\) we have

\[
\{x_n : 0 \leq n < a_j b^j\} = \left\{ \frac{k}{b^{j-1}} + \frac{a}{b^j} : 0 \leq k < b^{j-1}, \ 0 \leq a < a_j \right\},
\]

therefore

\[
\left| \sum_{n=0}^{a_j b^j - 1} e^{2\pi i t x_n} \right| = \left| \sum_{k=0}^{b^{j-1} - 1} e^{2\pi i \frac{t}{b^{j-1}} k} \right| \cdot \left| \sum_{a=0}^{a_j - 1} e^{2\pi i \frac{t a}{b^j}} \right|.
\]

The assumption \(b^s \nmid \ell\) implies that the first factor is zero whenever \(s \leq j - 1\). Thus we get from (24) that

\[
\left| \sum_{n=0}^{N-1} e^{2\pi i t x_n} \right| \leq \sum_{j=1}^{s} \left| \sum_{n=0}^{a_j b^j - 1} e^{2\pi i t x_n} \right| \leq \sum_{j=1}^{s} a_j b^j - 1 < b^s.
\]

\(\Box\)

**Proof of Theorem 3.** It is enough to prove the theorem in the special case when \(p\) is a positive even integer. Indeed, if \(p \geq 1\) is arbitrary, we can choose a positive even integer \(p' > p\). Observation (1) then implies

\[
S(N) \leq \|\Delta_N\|_p \leq \|\Delta_N\|_{p'}.
\]

Theorem 1 and Theorem 3 with \(p'\) thus imply Theorem 3 with \(p\).

From now on we assume \(p\) is a positive even integer. Every implied constant in the \(O\) notation will depend only on \(p\). From the alternative form of the discrepancy function

\[
\Delta_N(x) = \sum_{n=0}^{N-1} \left( \chi([x_n,1]) - x \right),
\]

where \(\chi\) denotes the characteristic function of a set, one obtains via routine integration that for any integer \(\ell \neq 0\) we have

\[
\int_{0}^{1} \Delta_N(x)e^{-2\pi i t x} \, dx = \frac{1}{2\pi i \ell} \sum_{n=0}^{N-1} e^{-2\pi i t x_n}.
\]

21
Therefore Parseval’s formula and observation (1) yield

\[ \int_0^1 (\Delta_N(x) - S(N))^2 \, dx = \sum_{\ell \neq 0} \frac{1}{4\pi^2 \ell^2} \left| \sum_{n=0}^{N-1} e^{-2\pi i \ell x_n} \right|^2. \]

Let \( N = \sum_{j=1}^{m} a_j b^{j-1} \) be the base \( b \) representation of \( N \), where \( a_j \in \{0, 1, \ldots, b-1\} \) and \( a_m > 0 \). Note \( N < b^m \). Let \( b^s \parallel \ell \) denote the fact that \( b^s \mid \ell \) but \( b^{s+1} \nmid \ell \).

By splitting the sum according to the highest power of \( b \) dividing \( \ell \), and applying Proposition 10 and a trivial estimate we obtain

\[ \int_0^1 (\Delta_N(x) - S(N))^2 \, dx \]

\[ = \sum_{s=0}^{m-2} \sum_{\ell \neq 0, b^s \parallel \ell} \frac{1}{4\pi^2 \ell^2} \left| \sum_{n=0}^{N-1} e^{-2\pi i \ell x_n} \right|^2 + \sum_{\ell \neq 0, b^{m-1} \parallel \ell} \frac{1}{4\pi^2 \ell^2} \left| \sum_{n=0}^{N-1} e^{-2\pi i \ell x_n} \right|^2 \]

\[ \leq \sum_{s=0}^{m-2} \sum_{\ell \neq 0, b^s \parallel \ell} \frac{1}{4\pi^2 \ell^2} 4^{2s+2} + \sum_{\ell \neq 0, b^{m-1} \parallel \ell} \frac{1}{4\pi^2 \ell^2} b^{2m} \]

\[ \leq \sum_{s=0}^{m-1} \sum_{\ell \neq 0, b^s \parallel \ell} \frac{b^2}{4\pi^2 \ell^2} = \frac{b^2}{12} m \leq \frac{b^2}{12} (\log_b N + 1). \quad (25) \]

For a positive even integer \( p \) consider the binomial formula

\[ \Delta_N(x)^p = S(N)^p + pS(N)^{p-1} (\Delta_N(x) - S(N)) \]

\[ + \sum_{k=2}^{p} \binom{p}{k} S(N)^{p-k} (\Delta_N(x) - S(N))^k. \]

By integrating on \([0, 1]\) we get

\[ \int_0^1 \Delta_N(x)^p \, dx = S(N)^p + \sum_{k=2}^{p} \binom{p}{k} S(N)^{p-k} \int_0^1 (\Delta_N(x) - S(N))^k \, dx. \]

Using the facts that \( \Delta_N(x) = O(b (\log_b N + 1)) \) and \( S(N) = O(b (\log_b N + 1)) \), we get from (25) that for any \( 2 \leq k \leq p \)
\[
\int_0^1 (\Delta_N(x) - S(N))^k \, dx \leq \sup_{x \in [0, 1]} |\Delta_N(x) - S(N)|^{k-2} \int_0^1 (\Delta_N(x) - S(N))^2 \, dx
\]

\[
= O \left( b^k (\log_b N + 1)^{k-1} \right).
\]

Thus we have

\[
\int_0^1 \Delta_N(x)^p \, dx = S(N)^p + O \left( b^p (\log_b N + 1)^{p-1} \right).
\]

(26)

Now we prove the theorem. Let \( M > b^2 \), and let \( N \) be a random variable uniformly distributed in \( \{0, 1, \ldots, M - 1\} \). We know from Theorem 1 that the event

\[
\frac{S(N) - c(b) \log_b N}{\sqrt{d(b) \log_b N}} > -\frac{c(b)}{4 \sqrt{d(b)}} \sqrt{\log_b M}
\]

has probability

\[
1 - \Phi \left( -\frac{c(b)}{4 \sqrt{d(b)}} \sqrt{\log_b M} \right) = 1 - O \left( \frac{\sqrt{\log \log_b M}}{\sqrt{\log_b M}} \right).
\]

The event \( M^{3/4} \leq N < M \) also has probability

\[
1 - O \left( \frac{1}{\sqrt{M}} \right) = 1 - O \left( \frac{\sqrt{\log \log_b M}}{\sqrt{\log_b M}} \right).
\]

Therefore it is enough to consider the intersection of these two events, on which

\[
S(N) > c(b) \left( \log_b N - \frac{1}{4} \sqrt{\log_b N \log_b M} \right)
\]

\[
\geq c(b) \left( \frac{3}{4} \log_b M - \frac{1}{4} \sqrt{\log_b M \log_b M} \right) = \frac{1}{2} c(b) \log_b M
\]

holds. For every such \( N \) we get from (26) that

\[
\int_0^1 \Delta_N(x)^p \, dx = S(N)^p \left( 1 + O \left( \frac{1}{\log_b M} \right) \right),
\]

\[
\|\Delta_N\|_p = S(N) \left( 1 + O \left( \frac{1}{\log_b M} \right) \right) = S(N) + O(b).
\]
Theorem 3 is thus reduced to Theorem 1.

\[ \Box \]

**Proof of Theorem 4.** Similarly to the proof of Theorem 3 we may assume that \( p \) is a positive even integer. Since \( \| \Delta_N \|_p = O \left( b \left( \log_b N + 1 \right) \right) \), by choosing \( A \) large enough we may assume that \( \lambda \leq \sqrt{\log_b M} \). Recall (26) from the proof of Theorem 3:

\[
\int_0^1 \Delta_N(x)^p \, dx = S(N)^p + O \left( b^p \left( \log_b N + 1 \right)^{p-1} \right)
\]

for any \( N > 0 \). Let \( N \) be a random variable which is uniformly distributed in \( \{0, 1, \ldots, M - 1\} \). We know from Theorem 2 that \( S(N) > \frac{1}{2} c(b) \log_b M \) with probability

\[
1 - O \left( e^{-c \sqrt{\log_b M}} + \frac{1}{b \sqrt{\log_b M} - 2} \right)
\]

for some constant \( c > 0 \). We also have \( \frac{M}{b \sqrt{\log_b M}} \leq N < M \) with probability at least \( 1 - O \left( \frac{1}{b \sqrt{\log_b M}} \right) \). For all such \( N \) we have \( \| \Delta_N \|_p = S(N) + O(b) \) and

\[
\log_b N = \log_b M + O \left( \sqrt{\log_b M} \right),
\]

\[
\sqrt{\log_b N} = \sqrt{\log_b M} + O(1).
\]

These estimates together with Theorem 2 yield

\[
\frac{1}{M} \left| \left\{ 0 \leq N < M : \| \Delta_N \|_p - \frac{b^2 - 1}{12b} \log_b N \geq A \lambda b \sqrt{\log_b N} \right\} \right|
= O \left( \frac{\sqrt{\lambda}}{e^{\sqrt{\lambda} - 1} - 2} + e^{-c \sqrt{\log_b M}} + \frac{1}{b \sqrt{\log_b M} - 2} \right)
\]

for any \( \lambda \geq 3 \) with some constant \( A > 0 \) depending only on \( p \). By replacing \( A \) by a larger constant we can simplify the upper bound to \( e^{-\sqrt{\lambda}} \) and relax the condition \( \lambda \geq 3 \) to \( \lambda \geq 1 \).

\[ \Box \]

**Proof of Proposition 5.** Let us write \( f(x) \) in the form

\[
f(x) = \int_0^1 f(t) \, dt + (f(1) - f(0)) \left( x - \frac{1}{2} \right) + g(x), \quad (27)
\]
where \( g : [0, 1] \to \mathbb{R} \) is defined via (27). Then we have \( \int_0^1 g(x) \, dx = 0 \) and \( g(0) = g(1) \). Note that (27) is the expansion of \( f(x) \) with respect to the Bernoulli polynomials with an explicit remainder term. For any integer \( N > 0 \) we have

\[
\sum_{n=0}^{N-1} f(x_n) = N \int_0^1 f(t) \, dt - (f(1) - f(0)) S(N) + \sum_{n=0}^{N-1} g(x_n).
\]

We now have to show that the last sum is negligible. Since \( g \) is twice differentiable on \([0, 1]\) and \( g(0) = g(1) \), we have that the periodic extension of \( g \) to \( \mathbb{R} \) with period 1 is Lipschitz, therefore its Fourier series converges to \( g \):

\[
g(x) = \sum_{\ell \in \mathbb{Z}} \hat{g}(\ell) e^{2\pi i \ell x}
\]

for any \( x \in [0, 1] \), where

\[
\hat{g}(\ell) = \int_0^1 g(x) e^{-2\pi i \ell x} \, dx.
\]

We have \( \hat{g}(0) = 0 \), because \( \int_0^1 g(x) \, dx = 0 \). Since \( g(0) = g(1) \), integration by parts yields that for any integer \( \ell \neq 0 \)

\[
\hat{g}(\ell) = \frac{g'(1) - g'(0)}{4\pi^2 \ell^2} - \int_0^1 g''(x) \frac{e^{-2\pi i \ell x}}{4\pi^2 \ell^2} \, dx = \frac{1}{4\pi^2 \ell^2} \int_0^1 g''(x) (1 - e^{-2\pi i \ell x}) \, dx,
\]

\[
|\hat{g}(\ell)| \leq \frac{1}{2\pi^2 \ell^2} \int_0^1 |g''(x)| \, dx = \frac{\|f''\|_1}{2\pi^2 \ell^2}.
\]

Therefore

\[
\left| \sum_{n=0}^{N-1} g(x_n) \right| = \left| \sum_{n=0}^{N-1} \sum_{\ell \neq 0} \hat{g}(\ell) e^{2\pi i \ell x_n} \right| \leq \sum_{\ell \neq 0} \frac{\|f''\|_1}{2\pi^2 \ell^2} \left| \sum_{n=0}^{N-1} e^{2\pi i \ell x_n} \right|.
\]

We can split up the sum according to the highest power of \( b \) dividing \( \ell \). Proposition 10 hence gives

\[
\left| \sum_{n=0}^{N-1} g(x_n) \right| \leq \sum_{s=0}^{\infty} \sum_{t \neq 0} \frac{\|f''\|_1}{2\pi^2 t^2} \left| \sum_{n=0}^{N-1} e^{2\pi i \ell x_n} \right| \leq \sum_{s=0}^{\infty} \sum_{t \neq 0} \frac{\|f''\|_1}{2\pi^2 b^s t^2} b^{s+1} = \frac{b^2}{6(b-1)} \|f''\|_1 \leq \frac{b}{3} \|f''\|_1.
\]

\( \square \)
References

[1] J. Beck: Probabilistic Diophantine Approximation. Randomness in Lattice Point Counting. Springer Monographs in Mathematics. Springer, Cham (2014) xvi+487 pp. ISBN: 978-3-319-10740-0.

[2] Y. S. Chow, H. Teicher: Probability Theory. Independence, Interchangeability, Martingales. Third edition. Springer Texts in Statistics. Springer-Verlag, New York (1997) xxii+488 pp. ISBN: 0-387-98228-0.

[3] M. Drmota, G. Larcher, F. Pillichshammer: Precise distribution properties of the van der Corput sequence and related sequences. Manuscripta Math. 118, no. 1 (2005), 11–41

[4] H. Faure: Discrépances de suites associées à un système de numération (en dimension un). Bull. Soc. Math. France 109, no. 2 (1981), 143–182

[5] L. Kuipers, H. Niederreiter: Uniform Distribution of Sequences. Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney (1974) xiv+390 pp. ISBN: 0-471-51045-9.

[6] P. Proinov, V. Grozdanov: On the diaphony of the van der Corput–Halton sequence. J. Number Theory 30, no. 1 (1988), 94–104

[7] A. Wohlfarter: Distribution Properties of Generalized van der Corput Sequences (Doctoral dissertation). TU Wien (2009)