On irregularities of distribution of binary sequences relative to arithmetic progressions, I. (General results)

Cécile Dartyge, Katalin Gyarmati, András Sárközy

On irregularities of distribution of binary sequences relative to arithmetic progressions, I. (General results)

Cécile Dartyge Katalin Gyarmati
András Sárközy

January 20, 2016

Abstract

In 1964 K. F. Roth initiated the study of irregularities of distribution of binary sequences relative to arithmetic progressions and since that numerous papers have been written on this subject. In the applications one needs binary sequences which are well distributed relative to arithmetic progressions, in particular, in cryptography one needs binary sequences whose short subsequences are also well-distributed relative to arithmetic progressions. Thus we introduce weighted measures of pseudorandomness of binary sequences to study this property.

2010 Mathematics Subject Classification: primary 11K38; secondary 11B25, 11K45.

Key words and phrases: arithmetic progressions, well-distribution, binary sequence.

Research partially supported by the Hungarian National Foundation for Scientific Research, grants K100291 and NK104183 and the grant ANR-10-BLAN 0103 MUNUM

We study the typical and minimal values of this measure for binary sequences of a given length.

1 Introduction

K. F. Roth [13] was the first who studied the irregularities of distribution of sequences relative to arithmetic progressions in 1964. Among others, it follows from his results that

Theorem 1 (Roth [13]) *If $N, Q \in \mathbb{N}$ with $Q \le N^{1/2}$ and $E_N = (e_1, e_2, \dots, e_N) \in \{-1, 1\}^N$, then there are integers a, t, q such that*

$$1 \le a \le a + (t-1)q \le N,\ q \le Q \tag{1}$$

and

$$\Big|\sum_{j=0}^{t-1} e_{a+jq}\Big| > c_1 Q^{1/2}$$

with some absolute constant c_1.

Taking here $Q = [N^{1/2}]$ we get

Corollary 1 (Roth) *If $N \in \mathbb{N}$ and $E_N = (e_1, e_2, \dots, e_N) \in \{-1, 1\}^N$, then there are integers a, t, q such that* (1) *holds,*

$$q \le N^{1/2}$$

and

$$\Big|\sum_{j=0}^{t-1} e_{a+jq}\Big| > c_2 N^{1/4}.$$

Since that numerous papers have been written on related problems; see the most recent papers [4], [12], [15], [16], [17] and the reference lists at the end of these papers. In particular, improving on a result of Beck [3], Matoušek and Spencer [10] proved

Theorem 2 (Matoušek and Spencer [10]) *If $N \in \mathbb{N}$, then there exists a sequence $E_n = (e_1, e_2, \ldots, e_N) \in \{-1, 1\}^N$ such that for every a, t, q satisfying* (1) *we have*

$$\Big| \sum_{j=0}^{t-1} e_{a+jq} \Big| < c_3 N^{1/4}$$

with some absolute constant c_3.

This shows that Theorem 1 is sharp apart from the constant factor c_1.

Binary sequences with strong pseudorandom properties play a crucial role in cryptography ; e.g., they are used as key in the frequently used encrypting system called Vernam cipher. Thus in [11] Mauduit and Sárközy initiated a new constructive and quantitative approach to study pseudorandom binary sequences

$$E_N = (e_1, e_2, \ldots, e_N) \in \{-1, 1\}^N. \tag{2}$$

In particular, they introduced the following measures of pseudorandomness of sequences of this type:

Definition 1 *The* **well-distribution measure** *of the sequence* (2) *is defined by*

$$W(E_N) = \max_{a,b,t} \Big| \sum_{j=0}^{t-1} e_{a+jb} \Big|$$

where $a, b, t \in \mathbb{N}$ and $1 \le a \le a + (t-1)b \le N$.

Definition 2 *For $k \in \mathbb{N}$, $k \le N$ the* **correlation measure of order** *k of the sequence* (2) *is defined as*

$$C_k(E_N) = \max_{M,D} \Big| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \cdots e_{n+d_k} \Big|$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ and $M \in \mathbb{N}$ such that $0 \le d_1 < d_2 < \cdots < d_k \le N - M$.

Then the sequence $E_N \in \{-1, 1\}^N$ is said to possess strong pseudorandom properties or, briefly, it is considered a "good" PR sequence

if both $W(E_N)$ and $C_k(E_N)$ (at least for "small" k) are small. This terminology is justified by the fact that for a "truly" random sequence $E_N \in \{-1,1\}^N$, i. e. , for choosing each $E_N \in \{-1,1\}^N$ with probability $\frac{1}{2^N}$ both $W(E_N)$ and (for fixed k) $C_k(E_N)$ are "small": they are expected to be around $N^{1/2}$ which is much smaller than the trivial upper bound N. This was proved by Cassaigne, Mauduit and Sárközy [5]:

Theorem 3 (Cassaigne, Mauduit and Sárközy [5]) *For all $\varepsilon > 0$ there are numbers $N_0 = N_0(\varepsilon)$ and $\delta = \delta(\varepsilon)$ such that for $N > N_0$ we have*

$$P(W(E_N) > \delta N^{1/2}) > 1 - \varepsilon$$

and

$$P(W(E_N) > 6(N \log N)^{1/2}) < \varepsilon.$$

Theorem 4 (Cassaigne, Mauduit and Sárközy [5]) *For all $k \in \mathbb{N}$, $k \geq 2$ and $\varepsilon > 0$ there are numbers $N_0 = N_0(\varepsilon, k)$ and $\delta = \delta(\varepsilon, k)$ such that for $N > N_0$ we have*

$$P(C_k(E_N) > \delta N^{1/2}) > 1 - \varepsilon$$

and

$$P(C_k(E_N) > 5(kN \log N)^{1/2}) < \varepsilon.$$

(Later these results have been sharpened by Alon, Kohayakawa, Mauduit, Moreira and Rödl [9], [2] and Aistleitner [1].)

In the last 15 years many papers have been written on the measures of pseudorandomness of binary sequences and many "good" PR binary sequences have been constructed; a survey of all these results has been presented by Gyarmati [8].

By using the notation introduced in Definition 1, Theorems 1 and 2 can be rewritten in the following form:

$$c_1 N^{1/4} < \min_{E_N \in \{-1,1\}^N} W(E_N) < c_3 N^{1/4}. \tag{3}$$

Comparing the upper bound here with Theorem 3, we may observe that the minimum of $W(E_N)$ (which is around $N^{1/4}$) is much smaller than its typical value (which is around $N^{1/2}$).

Note that the proof of the upper bound in (3) given by Matoušek and Spencer [10] is an existence proof, and no constructive proof is known. Indeed, the best known construction (presented in [7]) gives only

$$W(E_N) < c_4 N^{1/3} (\log N)^{2/3}. \tag{4}$$

In the sequel of this paper we will slightly improve on this construction.

In this paper our goal is to study the following problem:

Suppose we need a PR binary sequence of unknown length L. If we can estimate L reasonably well, say, we can find U such that $U < L < 2U$, then there is no problem: we construct a "good" PR sequence $E_N = (e_1, e_2, \ldots, e_N)$ with $2U < N < 4U$ (it is not too difficult to construct such a sequence), and then keeping only the first L elements of the sequence for any $U < L < 2U$: $E_L = (e_1, e_2, \ldots, e_L)$, we get a "good" PR sequence for any $U < L < 2U$. Namely it follows from the definitions of the measures W and C_k that if E_N is "good", $0 \le n < n + M \le N$ and $M \gg N$ (or just $M > N^{1-\varepsilon}$), then $(e_{n+1}, e_{n+2}, \ldots, e_{n+M})$ is also "good". If however, we cannot say more than, say, $U < L < U^{100}$ then this approach does not work; the problem is that if $M < N^{1/2}$ and $1 \le n < n + M \le N$ then the "good" PR properties of $(e_1, e_2, \ldots, e_N)$ are not enough to guarantee that $(e_{n+1}, e_{n+2}, \ldots, e_{n+M})$ is also "good"; indeed, even $e_{n+1} = e_{n+2} = \ldots = e_{n+M} = 1$ is possible.

This problem could be handled easily if we could construct sequences $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, 1\}^N$ such that for every $M > N^{\varepsilon}$, $1 \le n < n + M \le N$ the subsequence $(e_{n+1}, e_{n+2}, \ldots, e_{n+M})$ is "good", its PR measures W, C_k are less than $M^{1/2+\varepsilon}$, or just less than M^{1-c} would be a great step. But are there sequences E_N of this type? How far can we get in this direction ? Here we will study these ques-

tions focusing on the measure W; although the correlation measure also will get into the picture, we will focus on it in a subsequent paper.

First in Section 2 we will introduce a weighted version W_α of the measure W for studying these problems. In Section 3 we will estimate W_α for a "truly" random sequence $E_N \in \{-1,+\}^N$. In Section 4 we will formulate a conjecture on the minimal value of $W_\alpha(E_N)$ over all $E_N \in \{-1,1\}^N$, and in Sections 4 and 5 we will prove partial results towards the lower bound in this conjecture.

In the sequel of this paper we will present **constructive bounds** for $\min W_\alpha(E_N)$.

2 The weighted well-distribution measures

In the rest of this paper we will also use the following notations and definitions: if E_N is the binary sequence in (2), $n \in \{0,1,\dots,N-1\}$, $M \in \mathbb{N}$ and $0 \le n < n+M \le N$, then we write

$$E_N(n,M) = (e_{n+1}, e_{n+2}, \dots, e_{n+M}).$$

Definition 3 *If E_N is the binary sequence in (2) and $0 \le \alpha \le 1/2$, then the* **weighted α-well-distribution measure** *of E_N is defined as*

$$W_\alpha(E_N) = \max_{0\le n<n+M\le N} M^{-\alpha} W(E_N(n,M)).$$

Then clearly we have

$$W_0(E_N) = W(E_N). \tag{5}$$

For $0 \le \alpha \le 1/2$ we will write

$$m_\alpha(N) = \min_{E_N\in\{-1,1\}^N} W_\alpha(E_N).$$

Our main goals are to study $W_\alpha(E_N)$ for fixed α and a "truly" random $E_N \in \{-1,1\}^N$, and to estimate $m_\alpha(N)$ for fixed α. However, we will

also need a modified version of the measure introduced in Definition 3.

Consider again the binary sequence E_N in (2) and for $a, b, M \in \mathbb{N}$, $1 \le a \le a + (M-1)b \le N$, write

$$U(E_N, M, a, b) = \sum_{j=0}^{M-1} e_{a+jb}.$$

Definition 4 *If E_N is the binary sequence in (2) and $0 \le \alpha \le 1/2$, then the modified weighted α-well-distribution measure of E_N is defined as*

$$\widetilde{W}_\alpha(E_N) = \max_{0<M<N} \Big(M^{-\alpha} \max_{1\le a\le a+(M-1)b\le N} |U(E_N, M, a, b)| \Big).$$

For $0 \le \alpha \le 1/2$ we write

$$\widetilde{m}_\alpha(N) = \min_{E_N \in \{-1,1\}^N} \widetilde{W}_\alpha(E_N).$$

Clearly we have $W_0(E_N) = \widetilde{W}_0(E_N)$ for all $E_N \in \{-1,1\}^N$ and $m_0(N) = \widetilde{m}_0(N)$ for all $N \in \mathbb{N}$.

3 The weighted α-well-distribution measure for random binary sequences

We will show that Theorem 3 can be extended to $W_\alpha(E_N)$ with $0 \le \alpha \le 1/2$:

Theorem 5 *Assume that $0 \le \alpha \le 1/2$. Then for all $\varepsilon > 0$ there are numbers $N_0 = N_0(\varepsilon)$ and $\delta = \delta(\varepsilon)$ such that if $N > N_0$ then for a random sequence $E_N \in \{-1,1\}^N$ (i.e. choosing each $E_N \in \{-1,1\}^N$ with probability $1/2^N$) we have*

$$P(W_\alpha(E_N) > \delta N^{1/2-\alpha}) > 1 - \varepsilon \tag{6}$$

and

$$P(W_\alpha(E_N) > 6(N \log N)^{1/2} N^{-\alpha}) < \varepsilon. \tag{7}$$

Proof of Theorem 5. For $\alpha = 0$ the statement of the theorem holds by Theorem 3 and (5). Thus we may assume that

$$0 < \alpha \leq 1/2. \tag{8}$$

First we will prove (6). By the definitions of W and W_α we have

$$W_\alpha(E_N) \geq \max_{M \leq N} M^{-\alpha} W(E_N(0, M)) \geq N^{-\alpha} W(E_N) \geq N^{-\alpha} \Big| \sum_{j=1}^{N} e_j \Big|.$$

Thus it suffices to prove that

$$P\Big(N^{-\alpha} \Big| \sum_{j=1}^{N} e_j \Big| > \delta N^{1/2-\alpha}\Big) > 1 - \varepsilon$$

or, in equivalent form,

$$P\Big(\Big| \sum_{j=1}^{N} e_j \Big| > N^{1/2}\Big) > 1 - \varepsilon.$$

This is inequality (2.7) in [5] which was proved there (under the same conditions) and this completes the proof of (6).

Now we prove (7). This could be proved in an elementary manner like (2.2) in [5] but this would be rather lengthy; it is much simpler to use Chernoff's inequality [6] (see also [18]). We will apply the following special case of this inequality:

Lemma 1 *Let $X_1, X_2, \ldots, X_k$ be independent random variables with $P(X_i = 1) = P(X_i = -1) = 1/2$ for $i = 1, 2, \ldots, k$ and let $X = \sum_{i=1}^{k} X_i$. Then for all $A > 0$ we have*

$$P(|X| \geq A) \leq 2e^{-A^2/2k}.$$

(See the section "Better Chernoff bounds for some special cases" in [18].) By Definition 3 we have

$$
\begin{aligned}
&P(W_\alpha(E_N) > 6(N\log N)^{1/2}N^{-\alpha}) \\
&= P\Big(\max_{0\le n<n+M\le N} M^{-\alpha}W(E_N(n,M)) > 6(N\log N)^{1/2}N^{-\alpha}\Big) \\
&\le \sum_{0\le n<n+M\le N} P(M^{-\alpha}W(E_N(n,M)) > 6(N\log N)^{1/2}N^{-\alpha}) \\
&= \sum_{0\le n<n+M\le N} P(W(E_N(n,M)) > 6(N\log N)^{1/2}(M/N)^{\alpha}) \\
&= \sum_{0\le n<n+M\le N} P\Big(\max_{1\le a<a+(t-1)b\le M}\Big|\sum_{j=0}^{t-1} e_{n+a+jb}\Big| > 6(N\log N)^{1/2}(M/N)^{\alpha}\Big) \\
&\le \sum_{0\le n<n+M\le N}\ \sum_{1\le a<a+(t-1)b\le M} P\Big(\Big|\sum_{j=0}^{t-1} e_{n+a+jb}\Big| > 6(N\log N)^{1/2}(M/N)^{\alpha}\Big).
\end{aligned}
\tag{9}
$$

It remains to estimate the general term of this double sum. This can be done by using Lemma 1 with t, $e_{n+a+(i-1)b}$ (for $i = 1, 2, \ldots, t$) and $6(N\log N)^{1/2}(M/N)^{\alpha}$ in place of k, X_i and A respectively.

We obtain that

$$
P\Big(\Big|\sum_{j=0}^{t-1} e_{n+a+jb}\Big| > 6(N\log N)^{1/2}(M/N)^{\alpha}\Big) \le 2e^{-18N(\log N)(M/N)^{2\alpha}/t}.
\tag{10}
$$

It follows from our conditions on a, b and t that

$$
t-1 \le (t-1)b \le M-a \le M-1
$$

whence $t \le M$. Thus we get from (10) that

$$
\begin{aligned}
P\Big(\Big|\sum_{j=0}^{t-1} e_{n+a+jb}\Big| > 6(N\log N)^{1/2}(M/N)^{\alpha}\Big) &\le 2e^{-18(\log N)(M/N)^{2\alpha-1}} \\
&\le 2e^{-18\log N} \le \frac{1}{N^{17}}.
\end{aligned}
\tag{11}
$$

In (9) we have $0 \le n \le N$, $1 \le M \le N$ and $1 \le a, b, t \le M \le N$ so that each of the parameters n, M, a, b and t can be chosen in at most

N ways. Thus it follows from (9) and (11) that

$$P(W_\alpha(E_N) > 6(\log N)^{1/2}N^{-\alpha}) \le \sum_{0\le n<n+M\le N}\ \sum_{1\le a<a+(t-1)b\le M} \frac{1}{N^{17}}$$
$$\le \frac{1}{N^{12}} < \varepsilon$$

if N is large enough which proves (7) and this completes the proof of Theorem 5.

4 A conjecture on the minimum of $W_\alpha(E_N)$ and a related lower bound

By Corollary 1, there exists $c_2 > 0$ such that for any binary sequence satisfying (2) there are integers a, t, q such that (1) holds with $q \le N^{1/2}$ and

$$\Big|\sum_{j=0}^{t-1} e_{a+jq}\Big| > c_2 N^{1/4}.$$

Then we have for $\alpha \in [0, 1/2]$

$$\frac{1}{t^\alpha}\Big|\sum_{j=0}^{t-1} e_{a+jq}\Big| > c_2 N^{1/4} t^{-\alpha} \gg N^{1/4-\alpha}.$$

By this observation we deduce that:

$$m_\alpha(N) \gg N^{1/4-\alpha} \qquad \text{for } \alpha \in [0, 1/2]. \tag{12}$$

We conjecture that this estimate for the minimum of $W_\alpha(N)$ can be sharpened in the following way:

Conjecture 1 *For $0 \le \alpha \le 1/2$ we have*

$$c_5 N^{1/4-\alpha/2} < m_\alpha(N) < c_6 N^{1/4-\alpha/2}. \tag{13}$$

Note that by Corollary 1 and (5) this inequality holds for $\alpha = 0$. Unfortunately we have not been able to improve (12); the difficulty is that we have not been able to adapt Roth's method used in [13]. Thus

instead of estimating $W_\alpha(E_N)$ we will give a lower bound for $\widetilde{W}_\alpha(E_N)$ (which can be handled more easily) as a partial result. Some other partial results will be proved in the next section and in the sequel of this paper.

Adapting Roth's method we will prove

Theorem 6 *For $\alpha \in [0,1/2]$, $\alpha > 0$ and $N \in \mathbb{N}$, $N > N_0(\varepsilon)$, we have:*

$$\widetilde{m}_\alpha(N) \geq \Big(\frac{2}{\pi\sqrt{5}} - \varepsilon\Big) N^{1/4-\alpha/2}.$$

Proof of Theorem 6. The main tool is the following result of Sárközy ([14] Corollary 4), which was proved there by a generalization of Roth's argument.

Lemma 2 (Sárközy) *If $\varepsilon > 0$, $N > N_0(\varepsilon)$ is a positive integer and $s_1, s_2, \ldots, s_N$ a set of N complex numbers, then there exists integers n, q such that $1 \leq q \leq \sqrt{N}$ and*

$$|D(n,q,[\sqrt{N}/2])| \geq \Big(\frac{2}{\pi\sqrt{5}} - \varepsilon\Big)\Big(\frac{1}{N}\sum_{m=1}^{N} |s_m|^2\Big)^{1/2} N^{1/4}, \qquad (14)$$

where $D(n,q,k)$ is defined by

$$D(n,q,k) = s_n + s_{n+q} + \cdots + s_{n+(k-1)q},$$

with $s_i = 0$ if $i \notin \{1,\ldots,N\}$.

Let $\varepsilon > 0$. We apply this lemma with $N > N_0(\varepsilon)$ and any sequence $E_N = \{e_1, e_2, \ldots, e_N\} \in \{-1,+1\}^N$ (setting also $e_i = 0$ if $i \notin \{1,2,\ldots,N\}$) in place of $s_1, s_2, \ldots, s_N$. Then there exist integers n_0 and $q_0 \leq \sqrt{N}$ such that

$$\Big|D(n_0,q_0,[\sqrt{N}/2])\Big| = \left|\sum_{i=0}^{[\sqrt{N}/2]-1} e_{n_0+iq_0}\right| \geq \left(\frac{2}{\pi\sqrt{5}} - \varepsilon\right) N^{1/4}. \qquad (15)$$

Define the integers a, M by

$$\{n_0, n_0+q_0, n_0+2q_0, \ldots, n_0+([\sqrt{N}/2]-1)q_0\} \cap \{1,2,\ldots,N\} =$$
$$= \{a, a+q_0, a+2q_0, \ldots, a+(M-1)q_0\}. \qquad (16)$$

Then clearly we have

$$1 \le M \le \sqrt{N}/2, \tag{17}$$

$$1 \le a \le a + (M-1)q_0 \le N \tag{18}$$

and

$$e_{n_0+iq_0} = 0 \text{ if } i \in \mathbb{Z},\ n_0 + iq_0 \notin \{1, 2, \dots, N\}. \tag{19}$$

It follows from (16), (18) and (19) that

$$\begin{aligned}\Big|D(n_0, q_0, [\sqrt{N}/2])\Big| &= \left|\sum_{i=0}^{[\sqrt{N}/2]-1} e_{n_0+iq_0}\right| = \left|\sum_{j=0}^{M-1} e_{a+jq_0}\right| \\ &= |U(E_N, M, a, q_0)| \end{aligned} \tag{20}$$

(where $U(E_N, M, a, b)$ is the notation used in Definition 4). By (15), (17), (18), (20) and the definition of $\widetilde{W}_\alpha(E_N)$ we have

$$\begin{aligned}\widetilde{W}_\alpha(E_N) &\ge M^{-\alpha}\,|U(E_N, M, a, b)| \ge \left(\frac{\sqrt{N}}{2}\right)^{-\alpha}\left(\frac{2}{\pi\sqrt{5}} - \varepsilon\right) N^{1/4} \\ &\ge \left(\frac{2}{\pi\sqrt{5}} - \varepsilon\right) N^{1/4-\alpha/2}\end{aligned}$$

for every $E_N \in \{-1, +1\}^N$ which, by the definition of $\widetilde{m}_\alpha(N)$, completes the proof of the theorem.

5 Lower bound for $W_\alpha(E_N)$ for almost all E_N

In this section we will present a lower bound for $W_\alpha(E_N)$ for all α and $E_N \in \{-1, 1\}^N$, and from this we will deduce a lower bound for $W_\alpha(E_N)$ **for almost all** $E_N \in \{-1, 1\}^N$ which is smaller than the conjectured lower bound in Conjecture 1 by just a logarithm factor.

Theorem 7 *For $0 \le \alpha \le 1/2$ and $N \in \mathbb{N}$ we have*

$$W_\alpha(E_N) \ge \sqrt{\frac{2}{3}}\Big[\frac{N}{4C_2(E_N)}\Big]^{1/2-\alpha}. \tag{21}$$

Corollary 2 *For all $\varepsilon > 0$ there is a number N_0 such that for $N \in \mathbb{N}$, $N > N_0$ we have:*

$$P\Big(W_\alpha(E_N) \geq \sqrt{\frac{2}{3}}\Big[\frac{1}{20\sqrt{2}}\Big(\frac{N}{\log N}\Big)^{1/2}\Big]^{1/2-\alpha}\Big) \geq 1-\varepsilon. \tag{22}$$

Proof of Theorem 7. Define k by

$$k = \Big[\frac{1}{4}\frac{N}{C_2(E_N)}\Big],$$

and consider the sum

$$Z := \sum_{n=0}^{N-k} (e_{n+1} + e_{n+2} + \cdots + e_{n+k})^2.$$

It follows from the definition of $W_\alpha(E_N)$ that

$$\frac{|e_{n+1} + \cdots + e_{n+k}|}{k^\alpha} \leq W_\alpha(E_N),$$

for all $0 \leq n \leq N$. Thus we have

$$Z \leq (N-k+1)k^{2\alpha} W_\alpha(E_N)^2. \tag{23}$$

On the other hand, clearly we have:

$$\begin{aligned} Z &= \sum_{n=0}^{N-k} (e_{n+1} + \cdots + e_{n+k})^2 \\ &= \sum_{n=0}^{N-k} k + \sum_{n=0}^{N-k} \sum_{1\leq i\neq j\leq k} e_{n+i}e_{n+j} \\ &\geq (N-k+1)k - \sum_{1\leq i\neq j\leq k} \Big|\sum_{n=0}^{N-k} e_{n+i}e_{n+j}\Big| \\ &\geq (N-k+1)k - k^2 C_2(E_N). \end{aligned} \tag{24}$$

It follows from (23) and (24) that

$$(N-k+1)k^{2\alpha} W_\alpha(E_N)^2 \geq Z \geq (N-k+1)k - k^2 C_2(E_N)$$

whence

$$W_\alpha(E_N)^2 \geq k^{1-2\alpha} - k^{2-2\alpha}\frac{C_2(E_N)}{N-k+1}. \tag{25}$$

Here we have

$$\frac{C_2(E_N)}{N-k+1} = \frac{C_2(E_N)}{N+1-\left[\frac{N}{4C_2(E_N)}\right]} \le \frac{C_2(E_N)}{\frac{3N}{4}}$$
$$= \frac{4}{3}\frac{1}{\frac{N}{C_2(E_N)}} \le \frac{1}{3k}.$$

Thus we obtain from (25) that

$$W_\alpha(E_N)^2 \ge k^{1-2\alpha} - \frac{k^{1-2\alpha}}{3} = \frac{2}{3}k^{1-2\alpha}$$

whence

$$W_\alpha(E_N) \ge \sqrt{\frac{2}{3}}\Big[\frac{N}{4C_2(E_N)}\Big]^{1/2-\alpha},$$

which completes the proof of the theorem.

Proof of Corollary 2. By the second inequality in Theorem 4 for $N > N_0(\varepsilon)$ we have

$$P(C_2(E_N) \le 5\sqrt{2}(N\log N)^{1/2}) \ge 1-\varepsilon.$$

Thus it follows from (21) with probability greater than or equal to $1-\varepsilon$ that

$$W_\alpha(E_N) \ge \sqrt{\frac{2}{3}}\Big[\frac{N}{4\cdot 5\sqrt{2}(N\log N)^{1/2}}\Big]^{1/2-\alpha}$$
$$= \sqrt{\frac{2}{3}}\Big[\frac{1}{20\sqrt{2}}\Big(\frac{N}{\log N}\Big)^{1/2}\Big]^{1/2-\alpha}$$

which completes the proof of the corollary.

References

[1] C. Aistleitner, On the limit distribution of the well-distribution measure of random binary sequences, preprint.

[2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudorandomness for finite sequences: typical values, Proc. London Math. Soc. 95 (2007), 778-812.

[3] J. Beck, Roth's estimate of the discrepancy of integer sequences is nearly sharp, Combinatorica 1 (1981), 319-325.

[4] J. Beck, A. Sárközy and C. L. Stewart, On irregularities of distribution in shifts and dilatations of integer sequences, II, in: *Number Theory in Progress*, eds. K. Győry et al., Walter de Gruyter, Berlin-New York (1999), 633-638.

[5] J. Cassaigne, C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences VII: The measures of pseudorandomness, Acta Arith. 103 (2002), 97-118.

[6] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Stat. 23 (1952), 493-507.

[7] P. Erdős and A. Sárközy, Some solved and unsolved problems in combinatorial number theory, Math. Slovaca 28 (1978), 407-421.

[8] K. Gyarmati, Measures of pseudorandomness, in: *Finite fields and their Applications*, eds. P. Charpin et al., Radon Series on Computational and Applied Mathematics, de Gruyter, (2013), 43-64.

[9] Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudorandomness for finite sequences: minimum and typical values, in: *Proc. of WORDS '03*, vol. 27 of TUCS Gen. Publ. , Turku Cent. Comput. Sci., Turku, 2003; 159-169.

[10] J. Matoušek and J. Spencer, Discrepancy in arithmetic progression, J. Amer. Math. Soc. 9 (1996), 195-204.

[11] C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences, I. Measure of pseudorandomness, the Legendre symbol, Acta Arith. 82 (1997), 365-377.

[12] L. Mérai, The higher dimensional analogue of certain estimates of Roth and Sárközy, Period. Math. Hung., to appear

[13] K. F. Roth, Remark concerning integer sequences, Acta Arith. 9 (1964), 257-260.

[14] A. Sárközy, Some remarks concerning irregularities of distribution of sequences of integers in arithmetic progressions. IV, Acta Math. Acad. Sci. Hungar. 30 (1-2) (1977), 155-162

[15] A. Sárközy and C. L. Stewart, Irregularities of sequences relative to long arithmetic progressions, in: *Analytic Number Theory, Essays in Honour of Roth*, eds. W. W. R. Chen et al, Cambridge Univ. Press, Cambridge, (2009), 389-401.

[16] B. Valkó, On irregularities of sums of integers, Acta Arith. 92 (2000), 367-381.

[17] B. Valkó, Discrepancy of arithmetic progressions in higher dimensions, J. Number Theory 92 (2002), 117-130.

[18] Chernoff bound, Wikipedia, http://en.wikipedia.org/wiki/Chernoff-bound.

Cécile Dartyge

Institut Elie Cartan, Université de Lorraine, B. P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France

e-mail: cecile.dartyge@univ-lorraine.fr

Katalin Gyarmati

Eötvös Loránd University, Department of Algebra and Number Theory and MTA–ELTE Geometric and Algebraic Combinatorics Research Group, H-1117 Budapest, Pázmány Péter Sétány 1/C, Hungary

e-mail: gykati@cs.elte.hu

András Sárközy

Eötvös Loránd University, Department of Algebra and Number Theory, H-1117 Budapest, Pázmány Péter Sétány 1/C, Hungary

e-mail: sarkozy@cs.elte.hu