JOURNAL
de Théorie des Nombres
de BORDEAUX
anciennement Séminaire de Théorie des Nombres de Bordeaux

Christoph AISTLEITNER

On the limit distribution of the well-distribution measure of random binary sequences

Tome 25, nº 2 (2013), p. 245-259.

<http://jtnb.cedram.org/item?id=JTNB_2013__25_2_245_0>

Journal de Théorie des Nombres de Bordeaux **25** (2013), 245–259

On the limit distribution of the well-distribution measure of random binary sequences

par Christoph AISTLEITNER

Résumé. Nous prouvons l'existence d'une distribution limite de la mesure de bonne distribution normalisée $W(E_N)/\sqrt{N}$ (quand $N \to \infty$) pour des suites binaires aléatoires E_N. Par ce moyen, nous résolvons un problème posé par Alon, Kohayakawa, Mauduit, Moreira et Rödl.

Abstract. We prove the existence of a limit distribution of the normalized well-distribution measure $W(E_N)/\sqrt{N}$ (as $N \to \infty$) for random binary sequences E_N, by this means solving a problem posed by Alon, Kohayakawa, Mauduit, Moreira and Rödl.

1. Introduction and statement of results

Let $E_N = (e_n)_{1\leq n\leq N} \in \{-1,1\}^N$ be a finite binary sequence. For $M \in \mathbb{N}$, $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ set

$$U(E_N, M, a, b) = \sum \{e_{a+jb} :\ 1 \leq j \leq M,\ 1 \leq a + jb \leq N \text{ for all } j\}.$$

In other words, $U(E_N, M, a, b)$ is the discrepancy of E_N along an arithmetic progression in $\{1, \ldots, N\}$. The *well-distribution measure* $W(E_N)$ is then defined as

$$W(E_N) := \max \{|U(E_N, M, a, b)|,\ \text{where } 1 \leq a + b \text{ and } a + Mb \leq N\}.$$

The main result of the present paper is the following Theorem 1.1, which solves a problem posed by Alon, Kohayakawa, Mauduit, Moreira, and Rödl [2].

Theorem 1.1. *Let E_N denote random elements from $\{-1,1\}^N$, equipped with the uniform probability measure. There exists a limit distribution $F_W(t)$ of*

$$\left(\frac{W(E_N)}{\sqrt{N}}\right)_{N\geq 1}. \tag{1.1}$$

The function $F_W(t)$ is continuous and satisfies

$$\lim_{t\to\infty} \frac{t(1 - F_W(t))}{e^{-t^2/2}} = \frac{8}{\sqrt{2\pi}}. \tag{1.2}$$

It should be emphasized that the limit distribution of (1.1) is *not* the normal distribution. However, as a consequence of Theorem 1.1 and the Radon-Nikodŷm theorem, the limit distribution $F_W(t)$ has a density with respect to the Lebesgue measure. The tail estimate (1.2) in Theorem 1.1 should be compared to the corresponding asymptotic result for the tail probabilities $1-\Phi(t)$ of a standard normal random variable, for which

$$\lim_{t\to\infty} \frac{t(1-\Phi(t))}{e^{-t^2/2}} = \frac{1}{\sqrt{2\pi}}.$$

The measure W_N was introduced by Mauduit and Sárközy [11], together with two other measures of pseudorandomness. Again, let $E_N = (e_n)_{1\le n\le N} \in \{-1,1\}^N$ be a finite binary sequence. For $k\in\mathbb{N},\ M\in\mathbb{N},\ X\in\{-1,1\}^k$ and $D=(d_1,\dots,d_k)\in\mathbb{N}^k$ with $0\le d_1<\cdots<d_k<N$, we define

$$T(E_N,M,X) = \#\left\{n:\ n\le M,\ n+k\le N,\ (e_{n+1},\dots,e_{n+k})=X\right\},$$

$$V(E_N,M,D) = \sum\left\{e_{n+d_1}\dots e_{n+d_k}:\ 1\le n\le M,\ n+d_k\le N\right\}.$$

This means that $T(E_N,M,X)$ counts the number of occurrences of the pattern X in a certain part of E_N, and $V(E_N,M,D)$ quantifies the correlation among k segments of E_N, which are relatively positioned according to D.

The *normality measure* $\mathcal{N}(E_N)$ is defined as

$$\mathcal{N}(E_N) = \max_k \max_X \max_M \left| T(E_N,M,X) - \frac{M}{2^k}\right|,$$

where the maxima are taken over all $k\le \log_2 N,\ X\in\{-1,1\}^k,\ 0<M\le N+1-k$.
The *correlation measure* of order k, which is denoted by $C_k(E_N)$, is defined as

$$C_k(E_N) = \max\left\{|V(E_N,M,D)|:\ M,D \text{ satisfy } M+d_k\le N\right\}.$$

In [7] Cassaigne, Mauduit and Sárközy studied the “typical” values of $W(E_N)$ and $C_k(E_N)$ for random binary sequences E_N, and the minimal possible values of $W(E_N)$ and $C_k(E_N)$ for special sequences E_N. These investigations were extended by Alon, Kohayakawa, Mauduit, Moreira, and Rödl, who in [1] studied in detail the possible minimal and in [2] the “typical” values of $W(E_N)$, $\mathcal{N}(E_N)$ and $C_k(E_N)$ (see also [10] for an earlier survey paper). Among the results in [2] are the following two theorems. Here and throughout the rest of the present paper, E_N denotes *random* elements of $\{-1,1\}^N$, equipped with the uniform probability measure.

Theorem A. For any given $\varepsilon>0$, there exist numbers $N_0=N_0(\varepsilon)$ and $\delta=\delta(\varepsilon)>0$ such that for $N\ge N_0$

$$\delta\sqrt{N} < W(E_N) < \frac{\sqrt{N}}{\delta} \tag{1.3}$$

and

$$\delta\sqrt{N} < \mathcal{N}(E_N) < \frac{\sqrt{N}}{\delta}$$

with probability at least $1-\varepsilon$.

Theorem B. For any $\delta > 0$, there exist numbers $c(\delta) > 0$ and $N_0 = N_0(\delta)$ such that for any $N \geq N_0$

$$\mathbb{P}\left(W(E_N) < \delta\sqrt{N}\right) > c(\delta)$$

and

$$\mathbb{P}\left(\mathcal{N}(E_N) < \delta\sqrt{N}\right) > c(\delta).$$

In other words, Theorem A means that the pseudorandomness measures $W(E_N)$ and $\mathcal{N}(E_N)$ are of typical asymptotic order $\sqrt{N}$, while Theorem B means that the lower bounds in Theorem A are optimal. In [2] there are also theorems describing the typical asymptotic order of $C_k(E_N)$, which prove the existence of a limit distribution of $C_k(E_N)/\mathbb{E}\left(C_k(E_N)\right)$ in the case when $k = k(N)$ grows slowly in comparison with N (in this case the limit distribution is concentrated at a point). At the end of [2], Alon *et.al.* formulated the following open problem:

> *(Problem 33) Investigate the existence of the limiting distribution of*
>
> $$\left(W(E_N)/\sqrt{N}\right)_{N\geq 1}, \quad \left(\mathcal{N}(E_N)/\sqrt{N}\right)_{N\geq 1} \quad \textit{and} \quad \frac{C_k(E_N)}{\sqrt{N\log\binom{N}{k}}}.$$
>
> *Investigate these distributions.*

Subsequently they write: *"It is most likely that all three sequences in Problem 33 have limiting distributions".*

Theorem 1.1 proves the existence of a limit distribution of the normalized well-distribution measure of random binary sequences, by this means solving the first instance of Problem 33 above. The case of the normality measure $\mathcal{N}(E_k)$ seems to be much more difficult, and I could not obtain any satisfactory results. The case of the correlation measure $C_k(E_N)$ is considerably different from the cases of the well-distribution measure $W(E_N)$ and the normality measure $\mathcal{N}(E_N)$, since $C_k(E_N)$ depends on two parameters. It is reasonable to assume that the limiting distribution (provided that it exists) will depend on the choice of $k = k(N)$. As mentioned before, there already exist several results on the typical asymptotic order of $C_k(E_N)$, see [2, 3].

There exist several generalizations of the aforementioned pseudorandomness measures, for example to higher dimensions and to a continuous setting (see for example [4, 5, 9]); the problem concerning the typical asymptotic order and the existence of limit distributions is unsolved in many cases.

2. Auxiliary results

Lemma 2.1 (Hoeffding's inequality; see e.g. [12, Lemma 2.2.7])**.** *Let $(e_n)_{1\le n\le N}$ be independent random variables such that $e_n = 1$ and $e_n = -1$ with probability 1/2 each, for $n \ge 1$. Then*

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} e_n\right| > t\sqrt{N}\right) \le 2e^{-t^2/2}.$$

Lemma 2.2 (Donsker's theorem; see e.g. [6, Theorem 14.1])**.** *Let $(\xi_n)_{n\ge1}$ be a sequence of independent and identically distributed random variables with mean zero and variance σ^2. Define*

$$Y_N(s) = \frac{1}{\sigma\sqrt{N}} \sum_{n=1}^{\lfloor Ns\rfloor} \xi_n, \qquad 0 \le s \le 1.$$

Then

$$Y_N \Rightarrow Z,$$

where Z is the (standard) Wiener process and $\Rightarrow$ denotes weak convergence in the Skorokhod space $D([0,1])$.

A direct consequence of Donsker's theorem is the following Corollary 2.1:

Corollary 2.1. *Let $(e_n)_{n\ge1}$ be a sequence of independent random variables such that $e_n = 1$ and $e_n = -1$ with probability 1/2 each, for $n \ge 1$. Then for any $t \in \mathbb{R}$*

$$\mathbb{P}\left(\max_{1\le M_1\le M_2\le N}\left|\sum_{n=M_1}^{M_2} e_n\right| \le t\sqrt{N}\right) \to \mathbb{P}\left(\max_{0\le s_1\le s_2\le 1}|Z(s_2)-Z(s_1)| \le t\right)$$

as $N \to \infty$.

The quantity $\max_{0\le s_1\le s_2\le 1}|Z(s_2)-Z(s_1)|$ in Corollary 2.1 is called the *range* of the Wiener process. Its density $d(s)$ has been calculated by Feller [8] and is given by

$$d(s) = 8\sum_{k=1}^{\infty}(-1)^{k-1}k^2\phi(ks), \qquad s > 0, \tag{2.1}$$

where ϕ denotes the (standard) normal density function.

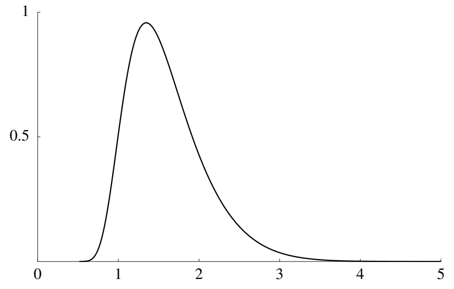

FIGURE 2.1. The density function $d(s)$ of the range of a standard Wiener process.

Lemma 2.3. *Let $(e_n)_{1\leq n\leq N}$ be independent random variables such that $e_n = 1$ and $e_n = -1$ with probability 1/2 each, for $n \geq 1$. Assume that N is of the form*

$$N = j2^m \qquad \text{for } j, m \in \mathbb{Z},\ 2^{10} < j \leq 2^{11} \text{ and } m \geq 1.$$

Then, if N is sufficiently large, for any $t > 2$

$$\mathbb{P}\left(\max_{1\leq M_1\leq M_2\leq N} \left| \sum_{n=M_1}^{M_2} e_n \right| > 1.38t\sqrt{N} \right) \leq 2^{24}e^{-t^2/2}.$$

Lemma 2.4. *Let $(e_n)_{1\leq n\leq N}$ be independent random variables such that $e_n = 1$ and $e_n = -1$ with probability 1/2 each, for $n \geq 1$. Then, if N is sufficiently large, for any $t > 2$*

$$\mathbb{P}\left(\max_{1\leq M_1\leq M_2\leq N} \left| \sum_{n=M_1}^{M_2} e_n \right| > 1.39t\sqrt{N} \right) \leq 2^{24}e^{-t^2/2}.$$

For an integer $B \geq 1$ we define modified well-distribution measures $W^{(\leq B)}$ and $W^{(>B)}$ by setting

$$W^{(\leq B)}(E_N) = \max\left\{ |U(E_N, M, a, b)| :\ b \leq B \text{ and } 1 \leq a+b,\ a+Mb \leq N \right\}$$

and

$$W^{(>B)}(E_N) = \max\left\{ |U(E_N, M, a, b)| :\ b > B \text{ and } 1 \leq a+b,\ a+Mb \leq N \right\}.$$

This means that for $W^{(\leq B)}$ we only consider arithmetic progressions having step size at most B, while for $W^{(>B)}$ we only consider arithmetic progressions of step size larger than B. Trivially an arithmetic progression with step size larger than B, which is contained in $\{1, \ldots, N\}$, cannot contain more than $\lceil N/(B+1) \rceil$ elements. The idea is that the limit distribution of

W is almost the same as the limit distribution of $W^{(\leq B)}$ for large B, while the contribution of $W^{(>B)}$ is almost negligible if B is large.

Lemma 2.5. *For any positive integer B there exists $N_0 = N_0(B)$ such that for all $N \geq N_0$ for any $t \in \mathbb{R}, t > 2$,*

$$\mathbb{P}\left(W^{(>B)}(E_N) > 1.4t\sqrt{N/(B+1)}\right) \leq 2^{28}(B+1)^2 e^{-t^2/2}. \tag{2.2}$$

Lemma 2.6. *For any integer $B \geq 1$ and any $t \in \mathbb{R}$ the limit*

$$F_W^{(\leq B)}(t) = \lim_{N\to\infty} \mathbb{P}\left(W^{(\leq B)}(E_N)N^{-1/2} \leq t\right)$$

exists.

We have to prove Lemmas 2.3, 2.4, 2.5 and 2.6. The proofs will be given in this order below. Lemmas 2.3 and 2.4 are a maximal form of Hoeffdings large deviations inequality (Lemma 2.1), and will be proved by using a classical dyadic decomposition method which is commonly used in probablity theory and probabilistic number theory. Using Lemma 2.4 we will prove Lemma 2.5, which essentially says that the probability that the discrepancy along any arithmetic progression with "large" step size B is of order $\sqrt{N}$ is very small. Finally using Donsker's invariance principle (Corollary 2.1) we will prove Lemma 2.6, which is the main ingredient in the proof of Theorem 1.1 in the next section.

Proof of Lemma 2.3: We use a modified version of a classical dyadic decomposition technique. By assumption N is of the form $j2^m$ for $j, m \in \mathbb{Z}, 2^{10} < j \leq 2^{11}$ and $m \geq 1$. We write $\mathcal{A}_{m+1}$ for the class of all sets of the form

$$\{j_1 2^m + 1, \dots, j_2 2^m\}, \qquad \text{where} \quad j_1, j_2 \in \{0, \dots, j\},\ j_1 < j_2.$$

Trivially, there exist at most 2^{22} sets of this form.

Furthermore, for every k, $0 \leq k \leq m$ we write $\mathcal{A}_k$ for the class of all sets of 2^k consecutive integers which start at position $j_1 2^k$ for some $j_1 \in \{0, \dots, j2^{m-k} - 1\}$. $\mathcal{A}_k$ contains exactly $j2^{m-k}$ sets of this form.

Then every set $\{k :\ 1 \leq M_1 \leq k \leq M_2 \leq N\}$ can be written as a disjoint union of at most one element of $\mathcal{A}_{m+1}$, and at most two elements of each of the classes $\mathcal{A}_k,\ 0 \leq k \leq m$.

For any set A_{m+1} from $\mathcal{A}_{m+1}$ we have by Hoeffdings inequality (Lemma 2.1)

$$\mathbb{P}\left(\left|\sum_{n \in A_{m+1}} e_n\right| > t\sqrt{N}\right) \leq 2e^{-t^2/2}.$$

Now assume that $k \in \{0, \dots, m\}$, and let A_k be any set from $\mathcal{A}_k$. By construction A_k contains $2^k \leq N2^{k-m}/2^{10}$ elements. By Hoeffding's inequality for any $t > 0$

$$\mathbb{P}\left(\left|\sum_{n\in A_k} e_n\right| > t\sqrt{2^k}\right) \le 2e^{-t^2/2},$$

which implies

$$\mathbb{P}\left(\left|\sum_{n\in A_k} e_n\right| > t\sqrt{(m-k+1)2^{k-m-10}}\sqrt{N}\right) \le 2e^{-(m-k+1)t^2/2}.$$

If we assume $t > 2$, then $e^{-t^2/2} \le 1/4$, and therefore

$$\mathbb{P}\left(\left|\sum_{n\in A_k} e_n\right| > 2^{-5}t\sqrt{(m-k+1)2^{k-m}}\sqrt{N}\right) \le 2e^{-t^2/2}\left(\frac{1}{4}\right)^{m-k}.$$

Now observe that

$$\sum_{k=0}^{m}\sqrt{(m-k+1)2^{k-m}} \le \sum_{k=0}^{\infty}\sqrt{(k+1)2^{-k}} \le 6,$$

and

$$2^{-5}\sum_{k=0}^{m}\sqrt{(m-k+1)2^{k-m}} \le 0.19. \tag{2.3}$$

Letting

$$\begin{aligned} A = &\left(\bigcup_{A_{m+1}\in\mathcal{A}_{m+1}}\left\{\left|\sum_{n\in A_{m+1}} e_n\right| > t\sqrt{N}\right\}\right)\cup \\ &\cup\left(\bigcup_{0\le k\le m}\ \bigcup_{A_k\in\mathcal{A}_k}\left\{\left|\sum_{n\in A_k} e_n\right| > 2^{-5}t\sqrt{(m-k+1)2^{k-m}}\sqrt{N}\right\}\right), \end{aligned}$$

this implies

$$\mathbb{P}(A) \le 2^{23}e^{-t^2/2} + \sum_{k=0}^{m} j2^{m-k}2e^{-t^2/2}\left(\frac{1}{4}\right)^{m-k} \le 2^{24}e^{-t^2/2}. \tag{2.4}$$

As mentioned before, every set $\{k : 1 \le M_1 \le k \le M_2 \le N\}$ can be written as a disjoint union of one set from $\mathcal{A}_{m+1}$ and at most two sets from each of the classes $\mathcal{A}_k,\ 0 \le k \le m$. By (2.3) we have on the complement of A

$$\begin{aligned} \max_{1\le M_1\le M_2\le N}\left|\sum_{k=M_1}^{M_2} e_n\right| &\le \left(1 + 2\left(2^{-5}\sum_{k=0}^{m}\sqrt{(m-k+1)2^{k-m}}\right)\right)\sqrt{N} \\ &\le 1.38\sqrt{N}, \end{aligned}$$

and thus by (2.4) for every $t \geq 2$

$$\mathbb{P}\left(\max_{1 \leq M_1 \leq M_2 \leq N} \left| \sum_{k=M_1}^{M_2} e_n \right| > 1.38 t \sqrt{N} \right) \leq \mathbb{P}(A) \leq 2^{24} e^{-t^2/2},$$

which proves the lemma. □

Proof of Lemma 2.4: Assume that N is *not* of the form described in Lemma 2.3. Write $\hat{N}$ for the smallest integer which is of this form, and which satisfies $\hat{N} \geq N$. Then, if N is sufficiently large, $\hat{N}/N \leq 2^{10} + 1/2^{10}$. Thus by Lemma 2.3 for $t > 2$

$$\begin{aligned}
&\mathbb{P}\left(\max_{1 \leq M_1 \leq M_2 \leq N} \left| \sum_{n=M_1}^{M_2} e_n \right| > 1.39 t \sqrt{N} \right) \\
&\leq \mathbb{P}\left(\max_{1 \leq M_1 \leq M_2 \leq \hat{N}} \left| \sum_{n=M_1}^{M_2} e_n \right| > 1.39 t \sqrt{N} \right) \\
&\leq \mathbb{P}\left(\max_{1 \leq M_1 \leq M_2 \leq \hat{N}} \left| \sum_{n=M_1}^{M_2} e_n \right| > 1.38 t \sqrt{\hat{N}} \right) \\
&\leq 2^{24} e^{-t^2/2}.
\end{aligned}$$

which proves Lemma 2.4. □

Proof of Lemma 2.5: Let $\mathcal{P} = \{a + b, \ldots, a + Mb\}$ be an arithmetic progression in $\{1, \ldots, N\}$. We say that $\mathcal{P}$ is of maximal length if $a < 0$ and $a + (M+1)b > N$. Denote the class of all arithmetic progressions, which are contained in the definition of $W^{(>B)}$ (that is, all arithmetic progressions in $\{1, \ldots, N\}$ with step size exceeding B) by $\hat{\mathcal{A}}$, and the class of all *maximal* arithmetic progressions among them by $\mathcal{A}$. Then for any $k \in \{B+1, \ldots, N\}$, the class $\mathcal{A}$ contains at most k different arithmetic progressions with step size k, and each of them has at most $\lceil N/k \rceil$ elements.

Let $\mathcal{P}, \hat{\mathcal{P}}$ denote arithmetic progressions from $\hat{\mathcal{A}}$. We write $\hat{\mathcal{P}} \subset \mathcal{P}$, if $\hat{\mathcal{P}} = \mathcal{P}$ or if $\hat{\mathcal{P}}$ can be obtained by removing a section from the beginning and/or from the end of $\mathcal{P}$. Then for any $\hat{\mathcal{P}} \in \hat{\mathcal{A}}$ there exists a least one

$\mathcal{P} \in \mathcal{A}$ for which $\hat{\mathcal{P}} \subset \mathcal{P}$. Thus

$$\begin{aligned} W^{(>B)}(E_N) &= \max_{\hat{\mathcal{P}} \in \hat{\mathcal{A}}} \left\{ \left| \sum_{n \in \hat{\mathcal{P}}} e_n \right| \right\} \\ &= \max_{\mathcal{P} \in \mathcal{A}} \max_{\hat{\mathcal{P}} \subset \mathcal{P}} \left\{ \left| \sum_{n \in \hat{\mathcal{P}}} e_n \right| \right\} \\ &= \max_{B < k \leq N} \max_{\substack{\mathcal{P} \in \mathcal{A}, \\ \mathcal{P} \text{ has step size } k}} \max_{\hat{\mathcal{P}} \subset \mathcal{P}} \left\{ \left| \sum_{n \in \hat{\mathcal{P}}} e_n \right| \right\}. \end{aligned}$$

To prove (2.2) it is obviously sufficient to consider those arithmetic progressions which contain at least $1.4\sqrt{N/B}$ elements. For these arithmetic progressions we can use Lemma 2.3 (provided N is sufficiently large), and obtain for any $t > 2$ and any $\mathcal{P}$ with step size k, using the estimate

$$\lceil N/k \rceil \leq \frac{1.4}{1.39} \frac{N}{k}$$

(which holds for sufficiently large N),

$$\mathbb{P}\left(\max_{\hat{\mathcal{P}} \subset \mathcal{P}} \left\{ \left| \sum_{n \in \hat{\mathcal{P}}} e_n \right| \right\} > 1.39t\sqrt{\lceil N/k \rceil} \right) \leq 2^{24} e^{-t^2/2}$$

and consequently

$$\mathbb{P}\left(\max_{\hat{\mathcal{P}} \subset \mathcal{P}} \left\{ \left| \sum_{n \in \hat{\mathcal{P}}} e_n \right| \right\} > 1.4t\sqrt{N/(B+1)} \right) \leq 2^{24} e^{-t^2 k/(2(B+1))}.$$

Thus, again for $t > 2$ and sufficiently large N, we have

$$\begin{aligned} \mathbb{P}\left(W^{(>B)}(E_N) > 1.4t\sqrt{N/(B+1)} \right) &\leq \sum_{k=B+1}^{N} 2^{24} k e^{-t^2 k/(2(B+1))} \\ &\leq 2^{24} \sum_{l=1}^{\infty} 4(B+1)^2 l^2 e^{-t^2/2} 4^{-l+1} \\ &\leq 2^{28} (B+1)^2 e^{-t^2/2}, \end{aligned}$$

which proves the lemma. □

Proof of Lemma 2.6: Let $B \geq 1$ be given. Denote by Q the least common multiple of all the numbers $\{1, \dots, B\}$. Set

$$\mathcal{Q}_k = \{1 \leq n \leq N : \ n \equiv k \mod Q\}, \qquad 1 \leq k \leq Q.$$

Write $\mathcal{A}$ for the class of those *maximal* arithmetic progressions in $\{1, \dots, Q\}$ which have a step size in $\{1, \dots, B\}$. By Donsker's theorem (Lemma 2.2) each of the processes

$$S_k(s) = \frac{\sqrt{Q}}{\sqrt{N}} \sum_{\substack{1 \le n \le sN, \\ n \in \mathcal{Q}_k}} e_n, \qquad 0 \le s \le 1, \qquad 1 \le k \le Q,$$

converges weakly to a standard Wiener process $Z_k(s)$. Since the random variables e_n, $n \ge 1$ are *independent*, we can assume that the Wiener processes $Z_k(s)$ are also independent, for $1 \le k \le Q$. Observe that

$$W^{(\le B)}(E_N) = \frac{\sqrt{N}}{\sqrt{Q}} \sup_{0 \le s_1 \le s_2 \le 1} \max_{A \in \mathcal{A}} \left| \sum_{k \in A} S_k(s_2) - S_k(s_1) \right|.$$

Thus by $S_k \Rightarrow Z_k$ we have for $t \ge 0$

$$\text{(2.5)} \quad \lim_{N \to \infty} \mathbb{P}\left(\frac{W^{(\le B)}(E_N)}{\sqrt{N}} \le t \right) = \mathbb{P}\left(\sup_{0 \le s_1 \le s_2 \le 1} \max_{A \in \mathcal{A}} \left| \sum_{k \in A} (Z_k(s_2) - Z_k(s_1)) \right| \le t\sqrt{Q} \right),$$

where $Z_1, \dots, Z_Q$ are independent Wiener processes. Thus a limit distribution $F_W^{(\le B)}(t)$ of $W^{(\le B)}(E_N)/\sqrt{N}$ exists, which proves the lemma. □

3. Proof of Theorem 1.1

The proof of Theorem 1.1 is split into several parts. Lemma 3.1 shows that the limit distribution function of the normalized well-distribution measure for the arithmetic progressions with short step size $W^{(\le B)}$ is Lipschitz-continuous. Together with the fact that the contribution of the arithmetic progressions with large step size is small (Lemma 2.6), this proves the existence of a limit distribution of the normalized well-distribution measure W_N (Lemma 3.2 and Corollary 3.1). Finally, in Lemmas 3.3 and 3.4 we prove the continuity of the limit distribution and the tail estimate (1.2) in Theorem 1.1.

Lemma 3.1. *For every fixed $t_0 > 0$ there exists a constant $c = c(t_0)$ such that for any $B \ge 1$, $\delta > 0$ and $t \ge t_0$*

$$F_W^{(\le B)}(t + \delta) - F_W^{(\le B)}(t) \le c(t_0)\delta.$$

Lemma 3.2. *Let $\varepsilon > 0$ be given. Then for every $t \in \mathbb{R}$ there exists an $N_0 = N_0(\varepsilon)$ such that for $N_1, N_2 \ge N_0$*

$$\left| \mathbb{P}\left(W(E_{N_1}) N_1^{-1/2} \le t \right) - \mathbb{P}\left(W(E_{N_2}) N_2^{-1/2} \le t \right) \right| \le \varepsilon.$$

Corollary 3.1. *For every $t \in \mathbb{R}$ the limit*

$$F_W(t) = \lim_{N\to\infty} \mathbb{P}\left(W(E_N)N^{-1/2} \le t\right)$$

exists.

Lemma 3.3. *The function $F_W(t)$ (which is defined in Corollary 3.1) is continuous in every point $t \in \mathbb{R}$.*

Lemma 3.4.

$$\lim_{t\to\infty} \frac{t(1-F_W(t))}{e^{-t^2/2}} = \frac{8}{\sqrt{2\pi}}.$$

Proof of Lemma 3.1: Let $t_0 > 0$ be fixed. We use the notation from the previous proof, and formulas (2.1) and (2.5). For $\delta > 0$ we want to estimate

$$F_W^{(\le B)}(t+\delta) - F_W^{(\le B)}(t),$$

which by (2.5) is bounded by

$$\sum_{A\in\mathcal{A}} \mathbb{P}\left(\sup_{0\le s_1\le s_2\le 1}\left|\sum_{k\in A}(Z_k(s_2)-Z_k(s_1))\right| \in \left(t\sqrt{Q},(t+\delta)\sqrt{Q}\right]\right). \tag{3.1}$$

If $Z_1,\dots,Z_K$ are independent standard Wiener processes (for some $K \ge 1$), then $(Z_1+\dots+Z_K)/\sqrt{K}$ is again a standard Wiener process. Thus the probabilities in (3.1) can be computed precisely: if A contains $|A|$ elements, then, writing $Z(t)$ for a standard Wiener process and $d(s)$ for the density function in (2.1), we have

$$\begin{aligned} &\mathbb{P}\left(\sup_{0\le s_1\le s_2\le 1}\left|\sum_{k\in A}(Z_k(s_2)-Z_k(s_1))\right| \in \left(t\sqrt{Q},(t+\delta)\sqrt{Q}\right]\right) \\ &= \mathbb{P}\left(\sup_{0\le s_1\le s_2\le 1}|Z(s_2)-Z(s_1)| \in \left(\frac{t\sqrt{Q}}{\sqrt{|A|}},\frac{(t+\delta)\sqrt{Q}}{\sqrt{|A|}}\right]\right) \\ &= \int_{t\sqrt{Q}/\sqrt{|A|}}^{(t+\delta)\sqrt{Q}/\sqrt{|A|}} d(s)\ ds. \end{aligned} \tag{3.2}$$

It is easily seen that for $k \ge 1$ and $s \ge 2$

$$k^2 e^{-k^2s^2/2} \le e^{-ks^2/2}.$$

Thus for $s \ge 2$ we have

$$d(s) \le \frac{8}{\sqrt{2\pi}}\sum_{k=1}^{\infty} k^2 e^{-k^2s^2/2} \le 4\sum_{k=1}^{\infty} e^{-ks^2/2} \le 5e^{-s^2/2}. \tag{3.3}$$

Clearly for every $k \in \{1,\dots,B\}$ the class $\mathcal{A}$ contains exactly k arithmetic progressions with step size k, and each of them contains Q/k elements.

Thus, by (3.1), (3.2) and (3.3), we have for every $t \geq t_0$

$$\begin{aligned} F_W^{(\leq B)}(t+\delta) - F_W^{(\leq B)}(t) &\leq \sum_{k=1}^{B} k \int_{t\sqrt{k}}^{(t+\delta)\sqrt{k}} d(s)ds \\ &\leq c(t_0)\delta, \end{aligned}$$

where the constant c depends on t_0, but *not* on B. □

Proof of Lemma 3.2: Let $\varepsilon > 0$ be given. Choose $B = B(\varepsilon)$ "large". We have

$$\mathbb{P}\left(W(E_{N_1}) N_1^{-1/2} \leq t\right) \leq \mathbb{P}\left(W^{(\leq B)}(E_{N_1}) N_1^{-1/2} \leq t\right),$$

and

$$\begin{aligned} &\mathbb{P}\left(W(E_{N_2}) N_2^{-1/2} \leq t\right) \\ &\qquad \geq \mathbb{P}\left(W^{(\leq B)}(E_{N_2}) N_2^{-1/2} \leq t\right) - \mathbb{P}\left(W^{(> B)}(E_{N_2}) N_2^{-1/2} > t\right). \end{aligned}$$

By Lemma 2.6 the sequence

$$\mathbb{P}\left(W^{(\leq B)}(E_N) N^{-1/2} \leq t\right)$$

converges as $N \to \infty$, and thus

$$\mathbb{P}\left(W^{(\leq B)}(E_{N_1}) N_1^{-1/2} \leq t\right) - \mathbb{P}\left(W^{(\leq B)}(E_{N_2}) N_2^{-1/2} \leq t\right) \leq \varepsilon/2$$

for sufficiently large N_1, N_2. By Lemma 2.5 for sufficiently large B and $N_2 = N_2(B)$

$$\mathbb{P}\left(W^{(> B)}(E_{N_2}) N_2^{-1/2} > t\right) \leq \underbrace{2^{28}(B+1)^2 e^{-t^2 B/8}}_{\leq \varepsilon/2 \text{ for sufficiently large } B}.$$

Thus

$$\mathbb{P}\left(W(E_{N_1}) N_1^{-1/2} \leq t\right) - \mathbb{P}\left(W(E_{N_2}) N_2^{-1/2} \leq t\right) \leq \varepsilon$$

for sufficiently large B, N_1, N_2, which proves Lemma 3.2. □

Proof of Lemma 3.3: Obviously $F_W(t) = 0$ for $t < 0$. The continuity of $F_W(t)$ at $t = 0$ follows from Theorem A of Alon *et.al.*, see (1.3). Now assume that $t > 0$ is fixed. Let $\delta > 0$ and $B \geq 1$, and assume that δ is "small" and

B is "large". We have

$$\begin{aligned}
&F_W(t+\delta) - F_W(t) \\
&= \lim_{N\to\infty} \mathbb{P}\left(W(E_N)N^{-1/2} \le t+\delta\right) - \lim_{N\to\infty} \mathbb{P}\left(W(E_N)N^{-1/2} \le t\right) \\
&\le \lim_{N\to\infty} \mathbb{P}\left(W^{(\le B)}(E_N)N^{-1/2} \le t+\delta\right) \\
&\quad - \lim_{N\to\infty} \mathbb{P}\left(W^{(\le B)}(E_N)N^{-1/2} \le t\right) \\
&\quad + \limsup_{N\to\infty} \mathbb{P}\left(W^{(>B)}(E_N)N^{-1/2} > t\right) \\
&= \lim_{N\to\infty} \mathbb{P}\left(W^{(\le B)}(E_N)N^{-1/2} \in (t, t+\delta]\right) \\
&\quad + \limsup_{N\to\infty} \mathbb{P}\left(W^{(>B)}(E_N)N^{-1/2} > t\right).
\end{aligned}$$

By Lemma 3.1

$$\lim_{N\to\infty} \mathbb{P}\left(W^{(\le B)}(E_N)N^{-1/2} \in (t, t+\delta]\right) \le \underbrace{c(t)\delta,}_{\le\varepsilon/2 \text{ for sufficiently small } \delta}$$

and by Lemma 2.5 for sufficiently large B and N

$$\limsup_{N\to\infty} \mathbb{P}\left(W^{(>B)}(E_N)N^{-1/2} > t\right) \le \underbrace{2^{28}(B+1)^2 e^{-t^2B/8}.}_{\le\varepsilon/2 \text{ for sufficiently large } B}$$

This proves

$$F_W(t+\delta) - F_W(t) \le \varepsilon$$

for sufficiently small δ. In the same way we can show a similar bound for $F_W(t) - F_W(t-\delta)$. This proves the lemma. □

Proof of Lemma 3.4: For any $t \in \mathbb{R}$

$$1 - F_W(t) \ge 1 - F_W^{(\le 1)}(t) = \int_t^\infty d(s)\ ds.$$

Using the standard estimate

$$\frac{t}{1+t^2}\frac{1}{\sqrt{2\pi}}e^{-t^2/2} < 1 - \Phi(t) < \frac{1}{t}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}, \qquad t > 0,$$

where $\Phi(t) = (2\pi)^{-1/2}\int_{-\infty}^t \phi(s)\ ds$ is the standard normal distribution function, we can easily show

$$\lim_{t\to\infty} \frac{t\left(1 - F_W^{(\le 1)}(t)\right)}{e^{-t^2/2}} = \lim_{t\to\infty} \frac{t\int_t^\infty d(s)\ ds}{e^{-t^2/2}} = \frac{8}{\sqrt{2\pi}},$$

which implies

$$\lim_{t\to\infty} \frac{t(1 - F_W(t))}{e^{-t^2/2}} \ge \frac{8}{\sqrt{2\pi}}. \tag{3.4}$$

On the other hand it is clear that

$$1 - F_W(t) \le 1 - F_W^{(\le 1)}(t) + \limsup_{N\to\infty} \mathbb{P}\left(W^{(>1)}(E_N)N^{-1/2} > t\right).$$

By Lemma 2.5, for sufficiently large t,

$$\limsup_{N\to\infty} \mathbb{P}\left(W^{(>1)}(E_N)N^{-1/2} > t\right) \le 2^{30} e^{-t^2/(1.4)^2},$$

and in particular

$$\lim_{t\to\infty} \frac{t\left(\limsup_{N\to\infty} \mathbb{P}\left(W^{(>1)}(E_N)N^{-1/2} \le t\right)\right)}{e^{-t^2/2}} \le 2^{30} \lim_{t\to\infty} \frac{te^{-t^2/(1.4)^2}}{e^{-t^2/2}} = 0.$$

Thus

$$\lim_{t\to\infty} \frac{t(1-F_W(t))}{e^{-t^2/2}} \le \frac{8}{\sqrt{2\pi}},$$

which together with (3.4) proves the lemma. □

References

[1] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl, *Measures of pseudorandomness for finite sequences: minimal values.* Combin. Probab. Comput. **15(1-2)** (2006), 1–29.

[2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl. *Measures of pseudorandomness for finite sequences: typical values.* Proc. Lond. Math. Soc. (3), **95(3)** (2007), 778–812.

[3] N. Alon, S. Litsyn, and A. Shpunt. *Typical peak sidelobe level of binary sequences.* IEEE Trans. Inform. Theory, **56(1)** (2010), 545–554.

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

[5] I. Berkes, W. Philipp, and R. F. Tichy. *Pseudorandom numbers and entropy conditions.* J. Complexity, **23(4-6)** (2007), 516–527.

[6] P. Billingsley. /it Convergence of probability measures. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition (1999).

[7] J. Cassaigne, C. Mauduit, and A. Sárközy. *On finite pseudorandom binary sequences. VII. The measures of pseudorandomness.* Acta Arith., **103(2)** (2002), 97–118.

[8] W. Feller. *The asymptotic distribution of the range of sums of independent random variables.* Ann. Math. Statistics, **22** (1951), 427–432.

[9] P. Hubert, C. Mauduit, and A. Sárközy. *On pseudorandom binary lattices.* Acta Arith., **125(1)** (2006), 51–62.

[10] Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl. *Measures of pseudorandomness for finite sequences: minimum and typical values.* In Proceedings of WORDS'03, **volume 27** of TUCS Gen. Publ., (2003), 159–169. Turku Cent. Comput. Sci., Turku.

[11] C. Mauduit and A. Sárközy. *On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the Legendre symbol.* Acta Arith., **82(4)** (1997), 365–377.

[12] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, (1996).

Christoph Aistleitner
TU Graz, Department of Analysis and Computational Number Theory (Math A),
Steyrergasse 30/II
8010 Graz, Austria
E-mail: aistleitner@math.tugraz.at