On a pseudorandom property of binary sequences

Katalin Gyarmati

Abstract

C. Mauduit and A. Sárközy proposed the use of well-distribution measure and correlation measure as measures of pseudorandomness of finite binary sequences. In this paper we will introduce and study a further measure of pseudorandomness: the symmetry measure. First we will give upper and lower bounds for the symmetry measure. We will also show that there exists a sequence for which each of the well-distribution, correlation and symmetry measures are small. Finally we will compare these measures of pseudorandomness.

2000 AMS Mathematics subject classification number: 11K45.

Key words and phrases: Pseudorandom, symmetry.

1 Introduction

In this paper we will study the symmetry property of finite binary sequences. C. Mauduit and A. Sárközy [2, pp. 367-370] introduced the following measures of pseudorandomness:

For a binary sequence

$$E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N,$$

write

$$U(E_N, t, a, b) = \sum_{j=1}^{t} e_{a+jb}$$

and, for $D = (d_1, \dots, d_k)$ with non-negative integers $0 \le d_1 < \dots < d_k$,

$$V(E_N, M, D) = \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_k}.$$

Then the *well-distribution measure of* E_N is defined as

$$W(E_N) = \max_{a,b,t} |U(E_N, t, a, b)| = \max_{a,b,t} \left| \sum_{j=1}^{t} e_{a+jb} \right|,$$

where the maximum is taken over all a, b, t such that $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ and $1 \le a + b \le a + tb \le N$, while the *correlation measure of order* k *of* E_N is defined as

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1}, \dots e_{n+d_k} \right|,$$

where the maximum is taken over all $D = (d_1, \dots, d_k)$ and M such that $M + d_k \le N$.

A. Sárközy and C. Mauduit [2, p. 372] observed that if a finite sequence contains a relatively large symmetrical subsequence (namely it contains a subsequence of the form $\{e_1, e_2, \dots, e_n, e_n, \dots, e_2, e_1\}$ or of the form $\{e_1, e_2, \dots, e_{n-1},\ e_n, e_{n-1}, \dots, e_2, e_1\}$), then this sequence certainly cannot be a "typical" random sequence, and this symmetric structure may lead difficulties in certain applications. This observation inspired us to propose a new measure of pseudorandomness.

We will define the *symmetry measure of E_N* by

$$S(E_N) = \max_{a<b} \left| \sum_{j=0}^{[(b-a)/2]-1} e_{a+j} e_{b-j} \right| = \max_{a<b} \left| H(E_N, a, b) \right|,$$

where

$$H(E_N, a, b) = \sum_{j=0}^{[(b-a)/2]-1} e_{a+j} e_{b-j}$$

is defined for all $1 \le a < b \le N$. Considering the sequence $E_N = \{1, 1, \dots, 1\}$ we see that $\max_{E_N} S(E_N) = \left[\frac{N}{2}\right]$. We expect that for a truly random sequence E_N, the symmetry measure is small. First we will prove that the symmetry measure of E_N is around $\sqrt{N}$ for almost all $E_N \in \{-1, +1\}^N$.

Theorem 1 *There is an integer N_0 such that for $N > N_0$ we have*

$$S(E_N) > \frac{7}{20}\sqrt{N}.$$

While for large N, $S(E_N)$ is always greater than a constant times $\sqrt{N}$, the upper bound holds for only the majority of the sequences $E_N \in \{-1, +1\}^N$.

Theorem 2 *For all* $\varepsilon > 0$ *there are numbers* $N_0 = N_0(\varepsilon)$ *such that for* $N > N_0$ *we have*

$$P\left(S(E_N) < 4.25\ (N \log N)^{1/2}\right) > 1 - \varepsilon.$$

We need the following measures of pseudorandomness introduced in [2, p. 371-372]. *Combined* (well-distribution-correlation) *PR-measure of order* k:

$$Q_k(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t} e_{a+jb+d_1} e_{a+jb+d_2} \dots e_{a+jb+d_k} \right|,$$

where a, b, t, $D = (d_1, d_2, \dots, d_k)$ are such that all the subscripts $a + jb + d_l$ belong to $\{1, \dots, N\}$. *Combined PR-measure*:

$$Q(E_N) = \max_{k \le (\log N / \log 2)} Q_k(E_N).$$

C. Mauduit and A. Sárközy [2, p. 373] proved that there is a number p_0 such that if $p > p_0$ is a prime number, $k \in \mathbb{N}$, $k < p$ and if we write

$$E_{p-1} = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \dots, \left(\frac{p-1}{p}\right) \right),$$

then

$$Q_k(E_{p-1}) \le 9kp^{1/2} \log p$$

so that, writing $N = p - 1$, we have

$$Q(E_N) \le 27 N^{1/2} (\log N)^2$$

It follows that for the Legendre symbol both the well-distribution measure and the correlation measure of order 2 are smaller than

$18N^{1/2}\log N$, while the combined PR-measure is smaller than $27N^{1/2}(\log N)^2$. As for all $1 \le k \le \frac{p-1}{2}$ we have $\left(\frac{p-k}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{k}{p}\right)$, the symmetry measure of the Legendre symbol E_{p-1} is $(p-1)/2$. We will show that the symmetry measure of the half of the sequence E_{p-1} is small.

Theorem 3 *If p is an odd prime, and we write*

$$E_{(p-1)/2} = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \dots, \left(\frac{(p-1)/2}{p}\right)\right)$$

then we have

$$S(E_{(p-1)/2}) \le 18p^{1/2}\log p.$$

Finally, we will compare the correlation measure of order 2 with well-distribution and symmetry measures. We expect that these measures of pseudorandomness are relatively independent. In order to show this we will give constructions where one measure is large while the others are small. The following two examples are variants of the ones in [2, p. 371-372].

EXAMPLE 1. Consider a sequence $E_N = (e_1, \dots, e_n) \in \{-1,+1\}^N$ such that each of the symmetry, correlation and well-distribution measure of it are possibly small (by Theorem 1 and 2 in [1] and our Theorem 2, all these measures can be $O\left((N\log N)^{1/2}\right)$ simultaneously) and define $E'_{2N} = (e'_1, e'_2, \dots, e'_{2N}) \in \{-1,+1\}^{2N}$ by

$$e'_n = \begin{cases} e_n & \text{for } 1 \le n \le N, \\ e_{n-N} & \text{for } N < n \le 2N. \end{cases}$$

Then it easy to see that the well-distribution measure of E'_{2N} are less than a constant times the corresponding measure of E_N and $S(E'_{2N}) \leq S(E_N) + C_2(E_N)$, but

$$C_2(E'_N) \geq \left| \sum_{n=1}^{N} e'_n e'_{n+N} \right| = N.$$

EXAMPLE 2. Consider a sequence $E_N = (e_1, \dots, e_n) \in \{-1, +1\}^N$ such that each of the correlation measure of order 2, well-distribution measure and symmetry measure of it are possibly small and define $E'_{2N} = (e'_1, e'_2, \dots, e'_{2N}) \in \{-1, +1\}^{2N}$ by

$$e'_n = \begin{cases} e_n & \text{for } 1 \leq n \leq N, \\ e_{2N-n} & \text{for } N < n \leq 2N. \end{cases}$$

Then the correlation measure of order 2 is less than a constant times $S(E_N) + C_2(E_N)$, while $W(E'_{2N}) \leq 2W(E_N)$. But

$$S(E'_{2N}) \geq \left| \sum_{n=1}^{N} e'_n e'_{2N-n} \right| = N.$$

C. Mauduit and A. Sárközy in [3] expressed the connection between the well-distribution measure and the correlation measure of order 2 in a quantitative form. Accordingly, in the following two theorems we will give a similar quantitative form of the connection between the well-distribution measure and the symmetry measure.

Theorem 4 *For all $N \in E_N$, and $E_N \in \{-1, +1\}^N$ we have*

$$W(E_N) \leq 3(NS(E_N))^{1/2}. \tag{1}$$

Finally, we will show that this result is sharp; there exists a sequence whose well-distribution measure is large and both the correlation measure and the symmetry measure are possibly small. Since the proof of the next theorem is nearly the same as the one in [3] (indeed we have to write $S(E_N)$ in place of $C_2(E_N)$), thus we will only sketch the proof.

Theorem 5 *If* $k, N \in \mathbb{N},\ \ N > N_0$ *and*

$$N^{3/4} \le k \le N \tag{2}$$

then there is a sequence $E_N \in \{-1,+1\}^N$ *with*

$$W(E_N) \ge k \tag{3}$$

and

$$\max\{C_2(E_N), S(E_N)\} < 120 \max\{\frac{k^2}{N}, (N \log N)^{1/2}\}. \tag{4}$$

From Theorem 5 we get that if $k \ge N^{3/4}(\log N)^{1/4}$, then

$$\begin{aligned} W(E_N) \ge k = \frac{N^{1/2}}{11}\left(121\frac{k^2}{N}\right)^{1/2} > \\ > \frac{1}{11}\left(N \max\{C_2(E_N), S(E_N)\}\right)^{1/2}. \end{aligned} \tag{5}$$

This means that (1) is the best possible apart from a constant factor.

One might like to study the generalizations of these measures of pseudorandomness. One possibility is to define the following measure:

$$\max_{\substack{M_1, M_2 \\ f_1(n), f_2(n), \ldots, f_j(n)}} \sum_{\substack{M_1 \le f_i(n) \le M_2 \\ (i=1,\ldots,j)}} e_{f_1(n)} e_{f_2(n)} \cdots e_{f_j(n)}, \tag{6}$$

where the maximum is taken over all $1 \le M_1 < M_2 \le N$ integers and $f_1(n), f_2(n), \ldots, f_j(n)$ polynomials with integer coefficients such that $M_1 \le f_i(n) \le M_2$ holds for all $1 \le n \le N$, $1 \le i \le j$. Of course this generalization also covers certain pathological cases (e.g., $f_1(n) = f_2(n) = \cdots = f_j(n)$), thus to introduce a pseudorandom measure of this type one has to pose certain restrictions on the polynomials $f_1, \ldots, f_j$ involved; we do not go into the details of this here.

When $j = 1$ or 2, for the special values of the polynomials $f_i(n)$, (6) can give the well-distribution measure, the correlation measure of order 2 and the symmetry measure.

Throughout the paper we write $e(x) = e^{2\pi i x}$.

2 Proofs

Proof of Theorem 1

Let $E_N = \{e_1, \ldots, e_n\}$, $f(z) = \sum_{n=1}^{N} e_n z^n$. Using the Cauchy-Schwarz inequality and Parseval formula we obtain:

$$\mathcal{J} \stackrel{\text{def}}{=} \int_0^1 |f(e(\alpha))|^4 \, \mathrm{d}\alpha \ge \left(\int_0^1 |f(e(\alpha))|^2 \, \mathrm{d}\alpha \right)^2 = N^2. \qquad (7)$$

Using the Parseval formula again we get:

$$\mathcal{J} = \int_0^1 \left|f^2\left(e(\alpha)\right)\right|^2 \mathrm{d}\alpha = \int_0^1 \left|\sum_{n=1}^{N}\sum_{m=1}^{N} e_n e_m e((n+m)\alpha)\right|^2 \mathrm{d}\alpha =$$

$$= \int_0^1 \left|\sum_{k=2}^{2N}\left(\sum_{\max\{1,\ k-N\}\le n\le \min\{N,k-1\}} e_n e_{k-n}\right) e(k\alpha)\right|^2 \mathrm{d}\alpha =$$

$$= \sum_{k=2}^{2N}\left|\sum_{\max\{1,\ k-N\}\le n\le \min\{N,k-1\}} e_n e_{k-n}\right|^2$$

By the definition of symmetry measure we have

$$\left|\sum_{\max\{1,\ k-N\}\le n\le \min\{N,k-1\}} e_n e_{k-n}\right| \le 2S(E_N)+1.$$

Therefore

$$\mathcal{J} \le (2N-1)\left(2S(E_N)+1\right)^2. \tag{8}$$

So that, in view of (7) and (8), and since clearly $S(E_N)\ge 1$, for large N we have

$$\frac{7}{20}\sqrt{N} \le S(E_N).$$

Proof of Theorem 2

Write $L = 4.25\ (N\log N)^{1/2}$, then we have:

$$P(S(E_N) > L) = P(\max_{a<b}|H(E_N,a,b)| > L) \le$$

$$\le \sum_{a<b} P(|H(E_N,a,b)| > L)) \le \binom{N}{2}\max_{a<b} P(|H(E_N,a,b)| > L),$$

where both the maximum and the summation are taken over all $a, b \in \mathbb{N}$ such that $1 \le a < b \le N$. Thus in order to prove the theorem, it suffices to show that for all $1 \leq a < b \leq N$ we have:

$$P(|H(E_N,a,b)| > L) = P\left(\left|\sum_{j=0}^{[(b-a)/2]-1} e_{a+j}e_{b-j}\right| > L\right) < \frac{2\varepsilon}{N^2}. \quad (9)$$

Let $t = [(b-a)/2]$, if $t \leq L$ then the probability in (9) is trivially 0 so that we may assume:

$$t = [(b-a)/2] > L = 4.25\ (N\log N)^{1/2}. \quad (10)$$

Write

$$M = 6(t\log t)^{1/2}$$

and

$$|\{j:\ 0 \leq j \leq t-1,\ e_{a+j}e_{b-j} = -1\}| = h. \quad (11)$$

Then we have:

$$\begin{aligned}\sum_{j=0}^{t-1} e_{a+j}e_{b-j} &= |\{j:\ 0 \leq j \leq t-1,\ e_{a+j}e_{b-j} = 1\}| - \\ &- |\{j:\ 0 \leq j \leq t-1,\ e_{a+j}e_{b-j} = -1\}| = (t-h) - h = \\ &= t - 2h.\end{aligned}$$

(11) holds with probability $\frac{1}{2^t}\binom{t}{h}$ so that

$$\begin{aligned}P\left(\left|\sum_{j=1}^{t-1} e_{a+j}e_{b-j}\right| > M\right) &= \sum_{h:\ |t-2h|>M} \frac{1}{2^t}\binom{t}{h} = \\ &= \frac{1}{2^t}\sum_{h:\ |h-t/2|>M/2}\binom{t}{h}. \quad (12)\end{aligned}$$

An easy computation shows that if $t \to \infty$ and $k \le t^{2/3}$, then we have

$$\binom{t}{[t/2]-k} = \binom{t}{[t/2]} \exp\left(-\frac{2k^2}{t} + O(\frac{k^3}{t^2})\right).$$

Using also the fact that $\binom{t}{i}$ is increasing in i for $0 \le i \le t/2$, it follows easily that for N large enough (so that $t = [(b-a)/2]$ is also large by (10)),

$$\begin{aligned}
\sum_{h:\ |h-t/2|>M/2} \binom{t}{h} &= \sum_{h:\ |h-t/2|>3(t\log t)^{1/2}} \binom{t}{h} < \\
&< t\binom{t}{[t/2]+[3(t\log t)^{1/2}]} < \\
&< t\binom{t}{[t/2]} \exp\left(-2\left(3(t\log t)^{1/2}\right)\frac{1}{t} + o(1)\right) = \\
&= t\binom{t}{[t/2]} \exp\left(-18\log t + o(1)\right) < \frac{2^t}{t^{16}}. \qquad (13)
\end{aligned}$$

Since $M \le L$, it follows from (10), (12) and (13) that

$$\begin{aligned}
P\left(\left|\sum_{j=0}^{t-1} e_{a+j}e_{b-j}\right| > L\right) &\le P\left(\left|\sum_{j=0}^{t-1} e_{a+j}e_{b-j}\right| > M\right) < \\
&< \frac{1}{2^t}\frac{2^t}{t^{16}} = \frac{1}{t^{16}} < \frac{1}{L^{16}} = o(\frac{1}{N^8}) < \frac{2\varepsilon}{N^2}
\end{aligned}$$

which proves (9) and this completes the proof of Theorem 2.

Proof of Theorem 3

We shall need the following lemma:

Lemma 1 *If p is a prime number, $f(x) \in F_p[x]$ is a polynomial of degree k such that it is not of the form $f(x) \in b(g(x))^2$ with $b \in F_p$, $g(x) \in F_p[x]$, and X, Y are real numbers with $0 < Y \le p$, then writing*

$$\chi_p^*(n) = \begin{cases} \left(\frac{n}{p}\right) & \textit{for } (n,p) = 1, \\ 0 & \textit{for } p \mid n, \end{cases}$$

we have

$$\left|\sum_{X<n\le X+Y}\chi_p^*(f(n))\right| < 9kp^{1/2}\log p.$$

Proof of Lemma 1

See [2, p. 373]. (Indeed, there this result is deduced from Weil's theorem [4].)

By the definition of H we have:

$$\begin{aligned} H(E_{(p-1)/2},a,b) &= \sum_{j=0}^{[(b-a)/2]-1}\left(\frac{a+j}{p}\right)\left(\frac{b-j}{p}\right) = \\ &= \sum_{j=0}^{[(b-a)/2]-1}\left(\frac{-j^2+(b-a)j+ab}{p}\right). \qquad (14) \end{aligned}$$

Let $f(x) = -x^2+(b-a)x+ab \in F_p[x]$. It is easy to see that $f(x)$ is the form of $b(g(x))^2$ if and only if $a+b\equiv 0\ (p)$. In the present case this is impossible as $1\le a<b\le (p-1)/2$. Applying Lemma 1 with 0 and $(b-a)/2$ in place of X and Y we get:

$$\begin{aligned} \left|\sum_{X\le n\le X+Y}\chi_p^*(f(n))\right| = \sum_{j=0}^{[(b-a)/2]-1}\left(\frac{-j^2+(b-a)j+ab}{p}\right) < \\ < 18p^{1/2}\log p. \qquad (15) \end{aligned}$$

From (14) and (15) we obtain $S(E_{(p-1)/2}) \le 18p^{1/2}\log p$, which proves the theorem.

Proof of Theorem 4

There exist a, b and t natural numbers such that:

$$W(E_N) = |U(E_N,t,a,b)| = \left|\sum_{j=0}^{t-1} e_{a+jb}\right| = \left|\sum_{\substack{a\le n\le a+(t-1)b \\ n\equiv a\quad (b)}} e_n.\right|$$

For all $n \in \mathbb{N}$ let $r(n)$ be the smallest natural number with $r(n) \equiv n \pmod{b}$. Let

$$f(\alpha) \stackrel{\text{def}}{=} \sum_{n=a}^{a+(t-1)b} e_n e\left(r(n)\alpha\right) = \sum_{k=0}^{b-1} \left(\sum_{\substack{a \le n \le a+(t-1)b \\ n \equiv k \quad (b)}} e_n \right) e(k\alpha).$$

The following lemma is well known and very simple.

Lemma 2 *If* $T(\alpha) = \sum_{k=0}^{b-1} c_k e(k\alpha)$ *then*

$$\sum_{h=0}^{b-1} \left| T\left(\frac{h}{b}\right) \right|^2 = b \sum_{k=0}^{b-1} |c_k|^2 .$$

By Lemma 2 we have:

$$\sum_{h=0}^{b-1} \left| f\left(\frac{h}{b}\right) \right|^2 = b \sum_{k=0}^{b-1} \left| \sum_{\substack{a \le n \le a+(t-1)b \\ n \equiv k \quad (b)}} e_n \right|^2 \ge b \left| \sum_{\substack{a \le n \le a+(t-1)b \\ n \equiv a \quad (b)}} e_n \right|^2$$
$$= bW^2(E_N). \tag{16}$$

On the other hand:

$$\begin{aligned} f^2\left(\frac{h}{b}\right) &= \sum_{n=a}^{a+(t-1)b} \sum_{m=a}^{a+(t-1)b} e_n e_m e\left(\frac{r(n)+r(m)}{b} h\right) = \\ &= \sum_{n=a}^{a+(t-1)b} \sum_{m=a}^{a+(t-1)b} e_n e_m e\left(\frac{r(n+m)}{b} h\right) = \\ &= \sum_{k=0}^{b-1} \left(\sum_{n=a}^{a+(t-1)b} \sum_{\substack{m=a \\ n+m \equiv k \quad (b)}}^{a+(t-1)b} e_n e_m \right) e\left(k\frac{h}{b}\right) = \sum_{k=0}^{b-1} c_k e(k\frac{h}{b}), \end{aligned} \tag{17}$$

where $c_k = \sum_{n=a}^{a+(t-1)b} \sum_{\substack{m=a \\ n+m \equiv k \ (b)}}^{a+(t-1)b} e_n e_m$. Replacing $n+m = jb+k$ we get:

$$
\begin{aligned}
|c_k| &= \left| \sum_{n=a}^{a+(t-1)b} \sum_{\substack{m=a \\ n+m\equiv k \quad (b)}}^{a+(t-1)b} e_n e_m \right| \\
&= \left| \sum_{j=\left\lceil \frac{2a-k}{b} \right\rceil}^{\left[\frac{2N-k}{b}\right]} \sum_{n=\max\{a, jb+k-(a+(t-1)b)\}}^{\min\{jb+k-a, a+(t-1)b\}} e_n e_{jb+k-n} \right| \\
&\le \sum_{j=\left\lceil \frac{2a-k}{b} \right\rceil}^{\left[\frac{2N-k}{b}\right]} \left| \sum_{n=\max\{a, jb+k-(a+(t-1)b)\}}^{\min\{jb+k-a, a+(t-1)b\}} e_n e_{jb+k-n} \right| \\
&\le \sum_{j=\left\lceil \frac{2a-k}{b} \right\rceil}^{\left[\frac{2N-k}{b}\right]} (2S(E_N)+1) \le \left(\frac{2N}{b}+1\right)(2S(E_N)+1) \\
&\le 9\frac{N}{b}S(E_N).
\end{aligned}
$$

Using (17) and Lemma 2 with the function $\sum_{k=0}^{b-1} c_k e(k\alpha)$, where c_k has defined above, we get:

$$
\sum_{h=0}^{b-1} \left| f^2\left(\frac{h}{b}\right) \right|^2 = b\sum_{k=0}^{b-1} |c_k|^2 \le 81\frac{N^2}{b}S^2(E_N).
$$

Thus from the Cauchy-Schwarz inequality and (16) we have:

$$
\begin{aligned}
81\frac{N^2}{b}S^2(E_N) \ge \sum_{h=0}^{b-1} \left| f^2\left(\frac{h}{b}\right) \right|^2 \ge \frac{1}{b}\left(\sum_{h=0}^{b-1} \left| f^2\left(\frac{h}{b}\right) \right| \right)^2 \ge \\
\ge \frac{1}{b}\left(bW^2(E_N)\right)^2 = bW^4(E_N),
\end{aligned}
$$

whence:

$$
W(E_N) \le 3\left(\frac{N}{b}S(E_N)\right)^{1/2} \le 3(NS(E_N))^{1/2}
$$

which was to be proved.

Proof of Theorem 5

If $k \geq \frac{N}{10}$ then (4) holds trivially for all E_N satisfying (3), thus we may assume that

$$k \leq \frac{N}{10} \tag{18}$$

Write

$$\Delta = 30 \max \left\{ \frac{k^2}{N}, (N \log N)^{1/2} \right\}$$

so that (4) can be written as

$$\max\{C_2(E_N), S(E_N)\} \leq 4\Delta.$$

If $\mathcal{A}$ is a finite set of positive integers, and $d \in \mathbb{N}$, then denote the number of solutions of

$$a - a' = d, \;\; a \in \mathcal{A}, \;\; a' \in \mathcal{A}, \tag{19}$$

by $f(\mathcal{A}, d)$, and denote the number of solutions of

$$a + a' = d, \;\; a \in \mathcal{A}, \;\; a' \in \mathcal{A}, \tag{20}$$

by $g(\mathcal{A}, d)$.

Lemma 3 *Assume that k satisfies (2) and N is large enough. Then there is an $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ such that*

$$|\mathcal{A}| = k \tag{21}$$

and

$$\max\{f(\mathcal{A}, d), g(\mathcal{A}, d)\} < 30\frac{k^2}{N} \stackrel{\text{def}}{=} M \;\; for \; aall \;\; 1 \leq d \leq 2N - 1. \tag{22}$$

Proof of Lemma 3

Write

$$\mathcal{F} = \{\mathcal{A} :\ \mathcal{A} \subseteq \{1, 2, \ldots, N\}, |\mathcal{A}| = k\},$$

$$\mathcal{F}_d = \{\mathcal{A} :\ \mathcal{A} \in \mathcal{F},\ f(\mathcal{A}, d) \geq M\},$$

$$\mathcal{G}_d = \{\mathcal{A} :\ \mathcal{A} \in \mathcal{F},\ g(\mathcal{A}, d) \geq M\}.$$

Then, clearly, any set $\mathcal{A}$ belonging to

$$\mathcal{F} \setminus \left(\cup_{d=1}^{N-1} \mathcal{F}_d \cup \cup_{d=3}^{2N-1} \mathcal{G}_d\right) \tag{23}$$

satisfies (21) and (22). Thus it suffices to show that the set in (23) is non-empty. To prove this we have to give upper bounds for $|\mathcal{F}_d|$ and $|\mathcal{G}_d|$. In [3] it was proved that

$$|\mathcal{F}_d| \leq \left(\frac{11}{12}\right)^{10N^{1/2}} \binom{N}{k}.$$

From this we get

$$|\mathcal{F}_d| \leq \frac{1}{4N} \binom{N}{k} \tag{24}$$

We will obtain by similar but easier calculations than in [3] that:

$$|\mathcal{G}_d| \leq \left(\frac{1}{3}\right)^{10N^{1/2}} \binom{N}{k} \leq \frac{1}{4N} \binom{N}{k}. \tag{25}$$

Consider a set

$$\mathcal{A} \in \mathcal{G}_d. \tag{26}$$

It follows from $g(\mathcal{A}, d) \geq M$ that there exist $\lceil M \rceil$ different numbers a_i $(i = 1, 2, \ldots \lceil M \rceil)$ with the property that

$$a_i \in \mathcal{A}, \quad d - a_i \in \mathcal{A}.$$

Let

$$\mathcal{A}_0 = \mathcal{A} \setminus \left(\cup_{i=1}^{\lceil M \rceil} \{a_i, d - a_i\} \right).$$

Then $\mathcal{A}$ is the disjoint union of the sets

$$\{a_1, d - a_1\}, \{a_2, d - a_2\}, \dots, \{a_{\lceil M \rceil}, d - a_{\lceil M \rceil}\}, \mathcal{A}_0.$$

Here we may choose $a_1, a_2, \dots, a_{\lceil M \rceil}$ from $\{1, 2, \dots, N\}$ in at most $\binom{N}{\lceil M \rceil}$ ways, these numbers determine $d - a_1, d - a_2, \dots, d - a_{\lceil M \rceil}$ uniquely, and since $|\mathcal{A}_0| = k - 2\lceil M \rceil$, the elements of A_0 can be chosen from the remaining $N - 2\lceil M \rceil$ numbers in at most $\binom{N-2\lceil M \rceil}{k-2\lceil M \rceil}$ ways. It follows that

$$|\mathcal{G}_d| \leq \binom{N}{\lceil M \rceil} \binom{N - 2\lceil M \rceil}{k - 2\lceil M \rceil}.$$

Carrying out similar calculations as in [3], we get:

$$|\mathcal{G}_d| \leq \frac{k^{2\lceil M \rceil}}{\lceil M \rceil!(N - 2\lceil M \rceil)^{\lceil M \rceil}} \binom{N}{k}.$$

By Stirling formula and (18) we get:

$$\lceil M \rceil! \geq \left(\frac{M}{3} \right)^{\lceil M \rceil} = \left(10 \frac{k^2}{N} \right)^{\lceil M \rceil},$$

$$N - 2\lceil M \rceil \geq N - 70 \frac{k^2}{N} \geq \frac{3}{10} N.$$

So we have:

$$|\mathcal{G}_d| \leq \left(\frac{1}{3} \right)^{\lceil M \rceil} \binom{N}{k} < \left(\frac{1}{3} \right)^{30N^{1/2}} \binom{N}{k} < \frac{1}{4N} \binom{N}{k}.$$

Using this, (24) and the fact that $|\mathcal{F}| = \binom{N}{k}$ we get that the set in (23) is non-empty, and this completes the proof of Lemma 3.

Now we fix a set $\mathcal{A} \subseteq \{1, 2, \dots, N\}$ satisfying (21) and (22) in Lemma 3, and let ε denote the set of the binary sequences $E_N \in \{-1, +1\}^N$ with

$$e_n = +1 \;\text{ for } n \in \mathcal{A}$$

so that

$$|\varepsilon| = 2^{N-|\mathcal{A}|} = 2^{N-k}.$$

We consider a "random" element E_N of ε, i.e., we choose each $E_N \in \varepsilon$ with probability $1/2^{N-k}$. In other words, we consider the binary sequence $E_N = \{e_1, e_2, \dots, e_N\}$ where for $n \in \mathcal{A}$ we have $e_n = +1$ while for n values with $n \notin \mathcal{A}$ the e_n's are chosen independently with

$$P(E_n = +1) = P(E_n = -1) = \frac{1}{2} \text{ (for } n \notin \mathcal{A}).$$

C. Mauduit and A. Sárközy [3] proved that

$$\begin{aligned} P(W(E_N) \geq k) &\geq \frac{1}{2}, \\ P(C_2(E_N) > 4\Delta) &\leq \frac{3}{N}. \end{aligned} \tag{27}$$

By the definition of $S(E_N)$ we have

$$\begin{aligned} P(S(E_N) > 4\Delta) &= P(\max_{a,b} |H(E_N, a, b)| > 4\Delta) \\ &\leq \sum_{a,b} P(|H(E_N, a, b)| > 4\Delta). \end{aligned} \tag{28}$$

For all $E_N \in \varepsilon$ we have

$$\begin{aligned} H(E_N, a, b) = & \sum_{\substack{0 \le j \le [(b-a)/2]-1 \\ a+j \in \mathcal{A},\ b-j \in \mathcal{A}}} e_{a+j} e_{b-j} + \sum_{\substack{0 \le j \le [(b-a)/2]-1 \\ a+j \in \mathcal{A},\ b-j \notin \mathcal{A}}} e_{a+j} e_{b-j} \\ & + \sum_{\substack{0 \le j \le [(b-a)/2]-1 \\ a+j \notin \mathcal{A},\ b-j \in \mathcal{A}}} e_{a+j} e_{b-j} + \sum_{\substack{0 \le j \le [(b-a)/2]-1 \\ a+j \notin \mathcal{A},\ b-j \notin \mathcal{A}}} e_{a+j} e_{b-j} \\ & = \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3 + \sum\nolimits_4 . \end{aligned} \tag{29}$$

It can be proved in the same way as in [3] with the change that we write $e_{a+j}e_{b-j}$ in the place of $e_{n+d_1}e_{n+d_2}$ and estimating $P(|\sum_4| \ge \Delta)$ we use Lemma 3 in place of [3, Lemma 1] that

$$P\left(\left|\sum\nolimits_1\right| > \Delta\right) = 0,$$

and for $i = 2, 3, 4$ we have

$$P\left(\left|\sum\nolimits_i\right| > \Delta\right) < \frac{1}{N^4}.$$

From this and (29) we get:

$$P(|H(E_N, a, b)| > 4\Delta) \le \sum_{i=1}^{4} P\left(\left|\sum\nolimits_i\right| > \Delta\right) \le \frac{3}{N^4} \tag{30}$$

Using (28) and (30) we have:

$$P(|S(E_N)| > 4\Delta) \le \frac{3}{N}. \tag{31}$$

It follows from (27) and (31) that (3) and (4) hold simultaneously with probability

$$> \frac{1}{2} - \frac{6}{N} > \frac{1}{3}$$

for N large enough, so that there is at least one $E_N \in \{-1, +1\}^N$ satisfying both (3) and (4), and this completes the proof of Theorem 5.

I would like to thank Professor András Sárközy for the valuable discussions.

References

[1] J. Cassaigne, C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequences VII: The measures of pseudorandomness*, Acta Arithmetica, to appear.

[2] C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequences I: Measures of pseudorandomness, the Legendre symbol*, Acta Arithmetica 82 (1997).

[3] C. Mauduit, A. Sárközy, *On the measures of pseudorandomness of binary sequences*, to appear.

[4] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Act. Sci. Ind. 1041, Hermann, Paris, 1948.