Generation of further pseudorandom binary sequences, I

(Blowing up a single sequence)

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

Christian Mauduit

Université Aix-Marseille

Institut de Mathématiques de Luminy

CNRS, FRE3529,

163 avenue de Luminy, 13288 Marseille cedex 9, France

E-mail: mauduit@iml.univ-mrs.fr

and

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

2010 Mathematics Subject Classification: Primary 11K45.

Keywords and phrases: pseudorandom, families of binary sequences, well-distribution, correlation, family complexity.

Research partially supported by Hungarian National Foundation for Scientific Research, Grants No. K100291 and NK104183, the János Bolyai Research Fellowship, the Agence

Abstract

Assume that a binary sequence is given with strong pseudorandom properties. An algorithm is presented and studied which prepares many further binary sequences from the given one. It is shown that if certain conditions hold then each of the sequences obtained in this way also possesses strong pseudorandom properties. Moreover, it is proved that certain large families of these sequences also posses strong pseudorandom properties.

1 Introduction

Finite binary sequences with strong pseudorandom properties (briefly "PR sequences") play a crucial role in cryptography. The PR sequences are usually generated by algorithms called pseudorandom bit generators ("PRBG"). There are hundreds of known PRBG's of various quality. In spite of this, in cryptography one always needs further and further constructions for "good" PR sequences. In this series we will present and study constructions of the type that a "good" PR sequence or a family of "good" PR sequences is given, and then we construct further "good" PR sequences from the given sequence(s). In particular, here in Part I we will study the case when we start out from a single PR sequence.

Nationale de la Recherche grant ANR-10-BLAN 0103 called MUNUM and the MTA-ELTE Geometric and Algebraic Combinatorics Research Group.

2 The measures of pseudorandomness of binary sequences

In order to characterize the quality of a PR sequence we will need *measures for the pseudorandomness of binary sequences.* We will start out from the PR measures introduced by the second and third author in [11]:

Consider the finite binary sequence

$$E_N = (e_1, e_2, \dots, e_N) \in \{-1, +1\}^N. \tag{2.1}$$

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

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

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

$$C_\ell(E_N) = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \cdots e_{n+d_\ell} \right|$$

where the maximum is taken over all $D = (d_1, d_2, \dots, d_\ell)$ and M such that $0 \le d_1 < \cdots < d_\ell \le N - M$. The *combined PR-measure of order ℓ of E_N* is defined as

$$Q_\ell(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t-1} e_{a+jb+d_1} e_{a+jb+d_2} \cdots e_{a+jb+d_\ell} \right|$$

where the maximum is taken over all $a, b, t \in \mathbb{N}$ and $D = (d_1, d_2, \dots, d_\ell)$ such that all the subscripts $a + jb + d_i$ belong to $\{1, 2, \dots, N\}$.

Here we also have to introduce *the cyclic* versions of these measures. Consider again the binary sequence E_N in (2.1), and extend it to an infinite sequence $\overset{\circ}{E}_N$ in the following way:

Definition 1 *If E_N is the binary sequence given in* (2.1)*, then the infinite binary sequence*

$$\overset{\circ}{E}_N = (\dots, e_{-2}, e_{-1}, e_0, e_1, e_2, \dots) \tag{2.2}$$

(infinite in both directions) is defined so that for $i \in \mathbb{Z}$ let $r(i)$ be the integer with $r(i) \equiv i \pmod{N}$, $1 \le r(i) \le N$, and then $e_i = e_{r(i)}$.

(In other words, $\overset{\circ}{E}_N$ is the periodic extension of E_N with period length N.)

Definition 2 *The* cyclic well-distribution measure *of the sequence E_N in* (2.1) *is defined by*

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

where the maximum is taken over all $a \in \mathbb{Z}$ and $b, t \in \mathbb{N}$ such that $(0 \le)$ $(t-1)b < N$ *(and the terms e_{a+jb} are defined by* (2.2)*).*

Definition 3 *The* cyclic correlation measure of order ℓ *of the sequence E_N in* (2.1) *is defined by*

$$\overset{\circ}{C}_\ell(E_N) = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_\ell} \right| \tag{2.3}$$

where the maximum is taken over all $D = (d_1, d_2, \dots, d_\ell)$ and M such that *the d_i's are integers with* $0 \le d_1 < d_2 < \dots < d_\ell < N$ *and* $M \in \mathbb{N}$, $M \le N$ *(and the terms* e_{n+d_i} *are defined by* (2.2)*).*

Definition 4 *The* cyclic combined PR-measure of order ℓ *of the sequence* E_N *in* (2.1) *is defined by*

$$\overset{\circ}{Q}_\ell(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t-1} e_{a+jb+d_1} e_{a+jb+d_2} \dots e_{a+jb+d_\ell} \right|$$

where the maximum is taken over all $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ *and* $D = (d_1, d_2, \dots, d_\ell)$ *such that* $(0 \le)$ $(t-1)b < N$ *and the d_i's are integers with* $0 \le d_1 < d_2 < \dots < d_\ell < N$ *(and the terms* e_{a+jb+d_i} *are defined by* (2.2)*).*

We remark that there are also other options to define the cyclic well-distribution measure (and thus also the cyclic combined PR measure). We will return to these different definitions and their analysis and comparison in a subsequent paper.

Then for every sequence E_N of form (2.1) we have

Proposition 1

$$W(E_N) \le \overset{\circ}{W}(E_N) \le 2W(E_N), \tag{2.4}$$

$$C_\ell(E_N) \le \overset{\circ}{C}_\ell(E_N) \le (\ell+1)C_\ell(E_N), \tag{2.5}$$

$$Q_\ell(E_N) \le \overset{\circ}{Q}_\ell(E_N) \le (\ell+1)Q_\ell(E_N). \tag{2.6}$$

Proof of Proposition 1 (2.4) and the first inequalities in (2.5) and (2.6) are trivial. In order to prove the second inequality in (2.5) consider the sum

$\sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_\ell}$ in (2.3) for which the maximum is attained so that

$$\overset{\circ}{C}(E_N) = \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_\ell} \right| .$$

Clearly we may assume $0 \le d_1 < d_2 < \dots < d_\ell \le N-1$. Let $I_\ell = [1, N-d_\ell]$, $I_{\ell-1} = [N+1-d_\ell, N-d_{\ell-1}], \dots,$ $I_i = [N+1-d_{i+1}, N-d_i], \dots,$ $I_1 = [N+1-d_2, N-d_1]$, $I_0 = [N+1-d_1, N]$.

If $n \in I_i$ then $1 \le n+d_1, n+d_2, \dots, n+d_i \le N$ and $N+1 \le n+d_{i+1}, n+d_{i+2}, \dots, n+d_\ell \le 2N$.

Thus

$$\left| \sum_{n \in I_i} e_{n+d_1} \dots e_{n+d_\ell} \right| = \left| \sum_{n \in I_i} e_{n+d_1} \dots e_{n+d_i} e_{n+d_{i+1}-N} \dots e_{n+d_\ell - N} \right| \le C_\ell(E_N).$$

Since $I_0 \cup I_1 \cup \dots \cup I_\ell = [1, 2, \dots, N]$ thus there exists a j such that $M \in I_j$, let $I_j^* = [N+1-d_{j+1}, M]$. Then by the triangle-inequality

$$\begin{aligned}
\overset{\circ}{C}_\ell(E_N) &= \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_\ell} \right| \\
&= \left| \sum_{n \in I_j^*} e_{n+d_1} \dots e_{n+d_\ell} + \sum_{i=j+1}^{\ell} \sum_{n \in I_i} e_{n+d_1} \dots e_{n+d_\ell} \right| \\
&\le \left| \sum_{n \in I_j^*} e_{n+d_1} \dots e_{n+d_\ell} \right| + \left| \sum_{i=j+1}^{\ell} \sum_{n \in I_i} e_{n+d_1} \dots e_{n+d_\ell} \right| \\
&\le (\ell+1) C_\ell(E_N).
\end{aligned}$$

It is easy to see that (2.4) and the first inequality in both (2.5) and (2.6) is sharp. We can also show that the constant factor $\ell+1$ in the upper bounds in (2.5) and (2.6) cannot be replaced by a number less than ℓ:

Example 1 *Let $N \in \mathbb{N}$, $\ell \in \mathbb{N}$, $\ell \mid N$, $N = \ell K$, and consider all the binary sequences $E_N = (e_1, e_2, \ldots, e_N) \in \{-1,+1\}^N$ such that*

$$e_{i+(\ell-1)K} = e_i e_{i+K} e_{i+2K} \ldots e_{i+(\ell-2)K} \text{ for } i = 1, 2, \ldots, K. \tag{2.7}$$

For each of the $2^{(\ell-1)K}$ sequences of this form we have

$$\begin{aligned} C_\ell(E_N) &\geq \left| \sum_{n=1}^{K} e_n e_{n+K} \ldots e_{n+(\ell-2)K} e_{n+(\ell-1)K} \right| \\ &\geq \left| \sum_{n=1}^{K} \left(e_n e_{n+K} \ldots e_{n+(\ell-2)K} \right) e_{n+(\ell-1)K} \right| \\ &= \left| \sum_{n=1}^{K} \left(e_{n+(\ell-1)K} \right)^2 \right| = \sum_{n=1}^{K} 1 = K = \frac{N}{\ell}, \end{aligned}$$

and with a little work it could be also shown that for almost all of these sequences E_N we also have

$$C_\ell(E_N) \leq (1 + o(1)) \frac{N}{\ell} \tag{2.8}$$

(we leave the details to the reader). Thus for almost all of these sequences we have

$$C_\ell(E_N) = (1 + o(1)) \frac{N}{\ell}. \tag{2.9}$$

On the other hand, for each of the sequences E_N satisfying (2.7) we have

$$\overset{\circ}{C}_\ell(E_N) \geq \left| \sum_{n=1}^{N} e_n e_{n+K} \ldots e_{n+(\ell-1)K} \right| = \sum_{n=1}^{N} 1 = N$$

whence

$$\overset{\circ}{C}_\ell(E_N) = N. \tag{2.10}$$

It follows from (2.8) and (2.10) that for almost all the sequences satisfying (2.7) we have

$$(\ell - o(1))C_\ell(E_N) \le \overset{\circ}{C}_\ell(E_N)$$

which proves that, indeed, the constant factor $\ell+1$ in (2.5) and (2.6) cannot be replaced by a number less than ℓ.

Thus the best constant in the upper bounds (2.5) and (2.6) is between ℓ and $\ell+1$; we have not been able to determine this constant.

We remark that there are many papers in the litterature in which the quantitative PR measures W, C_ℓ of certain special binary sequences E_N have been estimated. In many of these cases these (see, e.g., [2], [3], [4], [7], [8], [9], [10], [13], [15], [16], [17])

estimates also go through for $\overset{\circ}{W}$ and $\overset{\circ}{C}_\ell$ without any change, so that exactly the same estimates can be given for the cyclic measures.

3 The construction

Fix a binary sequence E_N of form (2.1) and a positive integer K with $K \le N$. Let k be positive integer with $1 \le k \le K$, and let $t_1, t_2, \dots, t_k$ be integers with $0 \le t_1 < t_2 < \dots < t_k < N$. Then

Definition 5 *Let*

$$F_N(t_1, t_2, \dots, t_k) = (f_1, f_2, \dots, f_N) \in \{-1, +1\}^N \tag{3.1}$$

denote the sequence whose i-th element is

$$f_i = e_{i+t_1} e_{i+t_2} \dots e_{i+t_k} \tag{3.2}$$

where the e_i's are defined as in Definition 1.

In this paper our goal is to study the pseudorandom properties of these sequences $F_N(t_1, t_2, \dots, t_k)$. We will also study the pseudorandom properties of the following *families* of these sequences:

Definition 6 *Let L be a positive integer with $L < N$. Then define the family $\mathcal{F}(K, L)$ by*

$$\mathcal{F}(K, L) = \{F_N(t_1, t_2, \dots, t_k) : \ 1 \le k \le K, \ 0 \le t_1 < t_2 < \dots < t_k \le L\},$$

and write also

$$\mathcal{F}(K) = \mathcal{F}(K, N-1).$$

4 An example

First we will show by an example that the strong PR properties of E_N do not guarantee without further assumption that $F_N(t_1, t_2, \dots, t_k)$ also possesses good PR properties:

Example 2 *Let $M \in \mathbb{N}$, $N = 2M$ and*

$$E_N = E_{2M} = (e_1, e_2, \dots, e_{2M}) \in \{-1, +1\}^N$$

any binary sequence with strong pseudorandom properties, and consider

$$F_N = F_N(0, M) = (e_1 e_{M+1}, e_2 e_{M+2}, \dots, e_M e_{2M}, e_{M+1} e_1, e_{M+2} e_2, \dots, e_{2M} e_M).$$

This sequence is periodic with period M, thus the correlation measure of order 2 is large:

$$C_2(F_N) \geq \left| \sum_{n=1}^{M} f_N f_{N+M} \right| = \left| \sum_{n=1}^{M} f_N^2 \right| = M = \frac{N}{2}.$$

This example shows that we need further assumption on N or the integers t_i's to ensure that the pseudorandom measures of $F_N(t_1, t_2, \dots, t_k)$ should be small. First we will study the case when N is a prime. (In Example 2 $N = 2M$ was a composite number).

5 The case of prime modulus

Let $N = p$ be a prime, $E_p = (e_1, e_2, \dots, e_p)$ and consider the sequence $F_p(t_1, t_2, \dots, t_k) = (f_1, f_2, \dots, f_p)$ defined by (3.2) where $t_1, t_2, \dots, t_k$ are integers with $0 \leq t_1 < t_2 < \dots < t_k \leq p - 1$. Then we have

$$\begin{aligned}
\overset{\circ}{W}(F_p) &= \max_{a,b,t} \left| \sum_{j=0}^{t-1} f_{a+jb} \right| \\
&= \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+jb+t_1} e_{a+jb+t_2} \cdots e_{a+jb+t_k} \right| \\
&\leq \overset{\circ}{Q}_k(E_p) \\
&\leq (k+1) Q_k(E_p)
\end{aligned}$$

so that the *cyclic* well-distribution measure of F_p can be estimated in terms of the combined measure of order k of E_p.

In order to estimate the correlation measure $\overset{\circ}{C}_\ell(F_p)$ we have to estimate $\left|\sum_{n=1}^{M} f_{n+d_1} f_{n+d_2} \dots f_{n+d_\ell}\right|$ where $0 \le d_1 < d_2 < \dots < d_\ell \le p-1$, $1 \le M \le p$.

Let $\mathcal{A} = \{t_1, t_2, \dots, t_k\}$, $\mathcal{B} = \{d_1, d_2, \dots, d_\ell\}$. Let $\mathcal{C}$ be the set of those integers c for which the equation

$$a + b \equiv c \pmod{p}, \ a \in \mathcal{A}, \ b \in \mathcal{B}$$

has an *odd* number of solutions.

Then

$$f_{n+d_1} \cdots f_{n+d_\ell} = \prod_{t_i \in \mathcal{A}} \prod_{d_j \in \mathcal{B}} e_{n+t_i+d_j} = \prod_{c \in \mathcal{C}} e_{n+c}.$$

Let $|\mathcal{C}| = s$. Clearly, $s \le k\ell$. If $\mathcal{C}$ is non empty $(s \ge 1)$ then

$$\begin{aligned} \left| \sum_{n=1}^{M} f_{n+d_1} \cdots f_{n+d_\ell} \right| &= \left| \sum_{n=1}^{M} \prod_{c \in \mathcal{C}} e_{n+c} \right| \\ &\le \max_{1 \le s \le k\ell} \overset{\circ}{C}_s(E_p) \\ &\le \max_{1 \le s \le k\ell} (s+1) C_s(E_p) \\ &\le (k\ell + 1) \max_{1 \le s \le k\ell} C_s(E_p). \end{aligned}$$

(Here $C_1(E_p) \le W(E_p)$.) Thus in order to estimate $\overset{\circ}{C}_\ell(E_p)$ we have to answer the following question: how can one ensure that for every $\mathcal{A} \subseteq \{0, 1, 2, \dots, p-1\}$, $1 \le |\mathcal{A}| \le k$ and $\mathcal{B} \subseteq \{0, 1, 2, \dots, p-1\}$, $1 \le |\mathcal{B}| \le \ell$ there is a $c \in \mathbb{Z}_p$

such that

$$a + b \equiv c \pmod{p}, \ a \in \mathcal{A}, \ b \in \mathcal{B}$$

has an *odd* number of solutions? Goubin, Mauduit and Sárközy introduced the following notion:

Definition 7 *If $M \in \mathbb{N}$, $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}_M$ and $\mathcal{A} + \mathcal{B}$ represents every element of $\mathbb{Z}_M$ with even multiplicity, i.e., for all $c \in \mathbb{Z}_M$, the equation*

$$a + b = c, \quad a \in \mathcal{A}, \ b \in \mathcal{B}$$

has even number of solutions (including the case when there are no solutions) then the sum $\mathcal{A} + \mathcal{B}$ is said to have property $\mathcal{P}$.

Definition 8 *If $k, \ell, M \in \mathbb{N}$ and $k, \ell \leq M$ then (k, ℓ, M) is said to be an* admissible triple *if there are no $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}_M$ such that $|\mathcal{A}| = k$, $|\mathcal{B}| = \ell$, and $\mathcal{A} + \mathcal{B}$ possesses property $\mathcal{P}$.*

Goubin, Mauduit and Sárközy proved that if one of the following three conditions holds, then (k, ℓ, p) is admissible:

a) $\ell = 2$,

b) $4^{k+\ell} < p$,

c) 2 is a primitive root modulo p.

(a) and c) appears in [4] while b) is the $n = 1$ special case of Lemma 4 in [14]).

Using this result we get the following

Theorem 1 *Let p be a prime, $E_p = (e_1, e_2, \dots, e_p) \in \{-1, +1\}^p$ and define $F_p(t_1, t_2, \dots, t_k)$ as in Definition 5. Then*

$$W(F_p) \leq \overset{\circ}{W}(F_p) \leq \overset{\circ}{Q}_k(E_p) \leq (k+1)Q_k(E_p). \tag{5.1}$$

Moreover, if one of the following three conditions holds:

a) $\ell = 2$,

b) $4^{k+\ell} < p$,

c) 2 is a primitive root modulo p,

then we have

$$C_\ell(F_p) \leq \overset{\circ}{C}_\ell(F_p) \leq \max_{1 \leq s \leq k\ell} \overset{\circ}{C}_s(E_p) \leq (k\ell + 1) \max_{1 \leq s \leq k\ell} C_s(E_p). \tag{5.2}$$

We remark that (5.1) involves the combined measure Q_k. If we have a good upper bound only for $C_k(E_p)$ but not for $Q_k(E_p)$, then by using (5.2) and Theorem 1 in [12], we may estimate $W(F_p)$ in the following way:

$$W(F_p) \leq (NC_2(F_p))^{1/2} \leq N^{1/2}(2k+1)^{1/2} \left(\max_{1 \leq s \leq 2k} C_s(E_p) \right)^{1/2}.$$

6 The case of composite modulus

In Section 4 we showed that if N is composite, then there are $t_1, t_2, \dots, t_k$ such that $F_N(t_1, t_2, \dots, t_k)$ has weak pseudorandom properties, thus $t_1, \dots, t_k$ can not be chosen arbitrarily. Here we give a sufficient condition for choosing

$t_1, \dots, t_k$ so that $F_N(t_1, \dots, t_k)$ has strong pseudorandom properties.

As in the case of prime modulus, we have

$$W(F_N) \le \overset{\circ}{W}(F_N) \le \overset{\circ}{Q}_k(F_N) \le (k+1)Q_k(F_N)$$

without any assumption. Now we will estimate the correlation measure: Let $\mathcal{A} = \{t_1, t_2, \dots, t_k\}$. We say $\mathcal{A}$ is L-good if for any $\mathcal{B} \subseteq \{0, 1, 2, 3, \dots, N-1\}$, $1 \le |\mathcal{B}| \le L$ there exists a $c \in \{0, 1, 2, 3, \dots, N-1\}$ such that

$$a + b \equiv c \pmod{N}, \quad a \in \mathcal{A},\ b \in \mathcal{B}$$

has an odd number of solutions. For $x \in \mathbb{N}$ let $r_N(x)$ denote the unique integer for which

$$x \equiv r_N(x) \pmod{N} \text{ and } 0 \le r_N(x) \le N-1.$$

If $\mathcal{A}$ is L-good, then as in Section 5 we have

$$C_\ell(F_p) \le \overset{\circ}{C}_\ell(F_p) \le \max_{1 \le s \le k\ell} \overset{\circ}{C}_s(E_p) \le (k\ell+1) \max_{1 \le s \le k\ell} C_s(E_p)$$

for $1 \le \ell \le L$.

Lemma 1 *If $\mathcal{A} = \{t_1, t_2, \dots, t_k\}$ is such that*

$$0 \le t_1 < t_2 < \cdots < t_k < \frac{N}{L},$$

then $\mathcal{A}$ is L-good.

Proof of Lemma 1. Let $\mathcal{B} = \{d_1, d_2, \ldots, d_\ell\}$, where $0 \le d_1 < d_2 < \cdots < d_\ell < N$ and $1 \le \ell < L$.

Let H denote the maximum of

$$r_N(d_2 - d_1),\ r_N(d_3 - d_2), \ldots, r_N(d_\ell - d_{\ell-1}),\ r_N(d_1 - d_\ell).$$

Clearly,

$$\begin{aligned}\ell H &\ge r_N(d_2 - d_1) + r_N(d_3 - d_2) + \cdots + r_N(d_\ell - d_{\ell-1}) + r_N(d_1 - d_\ell)\\ &= (d_2 - d_1) + (d_3 - d_2) + \cdots + (d_\ell - d_{\ell-1}) + (N + d_1 - d_\ell) = N,\end{aligned}$$

whence

$$H \ge \frac{N}{\ell}.$$

Thus $0 \le t_1 < t_2 < \cdots < t_k < \frac{N}{L} \le \frac{N}{\ell} \le H$.

First suppose that H is attained for

$$H = r_N(d_m - d_{m-1}),\ 2 \le m \le \ell.$$

(The other case $H = r_N(d_1 - d_\ell)$ can be handled in the same way.)

Then writing $d_{m-1} + t_k = c$, we will prove that the equation

$$a + b \equiv c \pmod{N},\ a \in \{t_1, t_2, \ldots, t_k\},\ b \in \{d_1, d_2, \ldots, d_\ell\} \tag{6.1}$$

has only one solution. First we remark that

$$0 \le d_{m-1} \le c = d_{m-1} + t_k < d_{m-1} + H = d_m < N.$$

Case 1: $b = d_i$ where $1 \le i \le m-2$. Then

$$\begin{aligned} 0 &\le b + a = d_i + a \le d_i + t_k \le d_{m-2} + t_k < d_{m-1} + t_k = c \\ &< d_{m-1} + H = d_m < N. \end{aligned}$$

So

$$0 \le b + a < c < N.$$

Thus

$$b + a \not\equiv c \pmod{N}.$$

Case 2: $b = d_{m-1}$ Then

$$d_{m-1} + a \equiv a + b \equiv c \equiv d_{m-1} + t_k \pmod{N}$$

holds only for $a = t_k$. In this case there is exactly one solution of (6.1).

Case 3: $b = d_i$ where $m \le i \le \ell$. In this case we distinguish two cases according $r_N(a+b) = a + b$ or $a + b - N$.

Case 3A: $r_N(a+b) = a + b$. Then $r_N(a+b) = a + b \ge d_m + a \ge d_m > c$. Since $0 \le c < N$ and $0 \le r_N(a+b) < N$ in this case

$$a + b \equiv c \pmod{N}$$

is not possible.

Case 3B: $r_N(a+b) = a+b-N$. Then

$$0 \leq r_N(a+b) = (a+b-N) = (d_m - N) + a < a \leq t_k \leq d_{m-1} + t_k = c.$$

Again by $0 \leq c < N$ and $0 \leq r_N(a+b) < N$ we can not have

$$a + b \equiv c \pmod{N}.$$

Using Lemma 1 we obtain

Theorem 2 *Let $N \in \mathbb{N}$, $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N$, $0 \leq t_1 < t_2 < \cdots < t_k < \frac{N}{L}$ and define $F_N(t_1, t_2, \ldots, t_k)$ as in Definition 5. Then for $\ell \leq L$ we have*

$$C_\ell(F_N) \leq \overset{\circ}{C}_\ell(F_N) \leq \max_{1 \leq s \leq k\ell} \overset{\circ}{C}_s(E_N) \leq (k\ell + 1) \max_{1 \leq s \leq k\ell} C_s(E_N).$$

7 The non-cyclic case

So far we studied the cyclic case. One might like to also study the case when the length of the output sequence may vary. Fix the sequence E_N in (2.1) and let $t_1, t_2, \ldots, t_k$ be k integers with $0 \leq t_1 < t_2 < \cdots < t_k < N$.

Definition 9 *Let*

$$F^*_{N-t_k}(t_1, \ldots, t_k) = (f_1, f_2, \ldots, f_{N-t_k})$$

where each f_i is defined as

$$f_i = e_{i+t_1} e_{i+t_2} \cdots e_{i+t_k}.$$

Then $F^*_{N-t_k}$ is the truncated version of F_N (their lengths are different).

Let $0 \le d_1 < d_2 < d_3 < \cdots < d_\ell < M + d_\ell \le N - t_k$. Then

$$\left| \sum_{n=1}^{M} f_{n+d_1} \cdots f_{n+d_\ell} \right| = \left| \sum_{n=1}^{M} e_{n+d_1+t_1} \cdots e_{n+d_\ell+t_k} \right|.$$

Here every subscript satisfies $1 \le n + d_i + t_j \le n + t_k + d_\ell \le N$.

The maximal elements of the set $\{t_i + d_j : \ 1 \le i \le k, \ 1 \le j \le \ell\}$ is $t_k + d_\ell$ and it occurs exactly once. Let $\mathcal{C}$ be the set of those elements $1 \le c \le N$ for which

$$t_i + d_j = c$$

has an odd number of solutions. Since $t_k + d_\ell \in \mathcal{C}$, we have $1 \le |\mathcal{C}| \le k\ell$. Thus

$$\left| \sum_{n=1}^{M} f_{n+d_1} \cdots f_{n+d_\ell} \right| = \left| \sum_{n=1}^{M} \prod_{c \in \mathcal{C}} e_{n+c} \right| \le \max_{1 \le s \le k\ell} C_s(E_N).$$

In this way we obtain

Theorem 3 *We have*

$$W(F^*_{N-t_k}) \le C_k(E_N)$$

and

$$C_\ell(F^*_{N-t_k}) \le \max_{1\le s\le k\ell} C_s(E_N).$$

In Sections 8 and 9 we will also study the following family of sequences defined in Definition 9:

Definition 10 *For $K, T \in \mathbb{N}$, $T < N$ we write*

$$\mathcal{F}^*(K,T) = \{F^*_{N-T}(t_1,t_2,\dots,t_{K-1},T) : \ 1 \le k \le K,$$
$$0 \le t_1 < t_2 < \cdots < t_{K-1} < T\}.$$

8 The complexity of the families constructed

So far we have studied the individual sequences constructed, and we have shown that under suitable assumptions these sequences have strong PR properties. However, in the applications it is usually not enough to construct a single (or a few) "good" sequences, one needs large "good" families of "good" sequences. But when can one say that a family is "good", it possesses strong PR properties? There are different measures for the PR quality of families of binary sequences. The most important of these measures is, perhaps, the *family complexity*, which was introduced by Ahlswede, Khachatrian, Mauduit and Sárközy [1] (which plays an especially important role in cryptography):

Definition 11 *If $N \in \mathbb{N}$, $j \in \mathbb{N}$, $j < N$, $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_j) \in \{-1,+1\}^j$, $i_1, i_2, \dots, i_j$ are integers with $1 \le i_1 < i_2 < \cdots < i_j \le N$ and $E_N =$*

$(e_1, e_2, \dots, e_N) \in \{-1, +1\}^N$ *is a binary sequence such that*

$$e_{i_1} = \varepsilon_1,\ e_{i_2} = \varepsilon_2, \dots, e_{i_j} = \varepsilon_j, \tag{8.1}$$

then we say that the sequence E_N satisfies the specification (8.1).

Definition 12 *The* f-complexity *of a family $\mathcal{F}$ of binary sequences $E_N \in \{-1, +1\}^N$ is defined as the greatest integer j so that for any specification* (8.1) *there is at least one $E_N \in \mathcal{F}$ which satisfies it. The f-complexity of $\mathcal{F}$ is denoted by $\Gamma(F)$. (If there is no $j \in \mathbb{N}$ with the property above, then we set $\Gamma(\mathcal{F}) = 0$.)*

We quote [1]: "... if we can construct a family $\mathcal{F}$ of high f-complexity and of "good" pseudorandom binary sequences, then the cryptosystem based on it ... has good security properties." Thus our next goal is to show that the families $\mathcal{F}(K)$, $\mathcal{F}(K, L)$ and $\mathcal{F}^*(K, T)$ defined above (which consist of binary sequences of strong PR properties by Theorems 1, 2 and 3) also have large f-complexity. Indeed, we will prove the following theorems:

Theorem 4 *Let $E_N \in \{-1, +1\}^N$ and $\mathcal{F}(K, L)$ be the family generated from E_N (as described in Definition 5 and 6). Let*

$$K^* = \begin{cases} K & \text{if } K \text{ is odd,} \\ K - 1 & \text{if } K \text{ is even.} \end{cases}$$

If for some $R \in \mathbb{N}$ we have

$$2^{R/K^*} \max_{1 \le t \le R} \overset{\circ}{C}_t(E_N) < L + 1, \tag{8.2}$$

then

$$\Gamma(\mathcal{F}(K,L)) \geq R.$$

Note that this theorem also covers $\mathcal{F}(K)$ since we have $\mathcal{F}(K) = \mathcal{F}(K, N-1)$. Moreover, observe that in (8.2) we also use $C_1(E_N)$ $(\leq W(E_N))$.

By (2.5) the assumption (8.2) can be replaced by

$$2^{R/K^*} \max_{1\leq t\leq R} (t+1) C_t(E_N) < L+1.$$

Theorem 5 *Let $E_N \in \{-1,+1\}^N$ and $\mathcal{F}^*(K,T)$ be the family generated from E_N (as described in Definitions 9 and 10). If for some $R \in \mathbb{N}$ we have*

$$2^{R/(K-1)} \max_{1\leq t\leq R} C_t(E_N) < T, \tag{8.3}$$

then

$$\Gamma(\mathcal{F}^*(K,T)) \geq R.$$

Since these two theorems are rather complicated, thus here first we present two more transparent corollaries and two examples, and the proofs of these theorems and corollaries will be presented only in the next section.

Corollary 1 *Let $E_N \in \{-1,+1\}^N$ and $\mathcal{F}(K,L)$ be the family generated from E_N, and ε be a positive number such that*

$$\max_{1\leq t\leq \frac{\varepsilon}{\log 2}K^*\log(L+1)} \overset{\circ}{C}_t(E_N) < (L+1)^{1-\varepsilon}. \tag{8.4}$$

Then we have

$$\Gamma(\mathcal{F}(K,L)) \geq \left[\frac{\varepsilon}{\log 2} K^* \log(L+1)\right].$$

Corollary 2 *Let $E_N \in \{-1,+1\}^N$ and $\mathcal{F}^*(K,T)$ be the family generated from E_N, and ε be a positive number such that*

$$\max_{1\le t\le \frac{\varepsilon}{\log 2}(K-1)\log T} \overset{\circ}{C}_t(E_N) < T^{1-\varepsilon}.$$

Then we have

$$\Gamma(\mathcal{F}^*(K,T)) \ge \left[\frac{\varepsilon}{\log 2}(K-1)\log T\right].$$

Example 3 *Let p be a prime and let E_p denote the modulo p Legendre symbol sequence completed by a $+1$ at the end: $E_p = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \dots, \left(\frac{p-1}{p}\right), 1\right)$. Let $\mathcal{F}(K) = \mathcal{F}(K, p-1)$ be the family generated from this sequence E_p (in the manner of Definitions 5 and 6), and consider a sequence $F_p(t_1, t_2, \dots, t_k) = (f_1, f_2, \dots, f_{p-1}) \in \mathcal{F}(K)$ with $1 \le k \le K$, $0 \le t_1 < t_2 < \cdots < t_k \le p-1$. Then every element f_i of this sequence is of form* (3.2)*. If we write $h(x) = (x+t_1)(x+t_2)\cdots(x+t_k)$ and $p \nmid h(i)$ then* (3.2) *can be rewritten as*

$$f_i = \left(\frac{i+t_1}{p}\right)\left(\frac{i+t_2}{p}\right)\cdots\left(\frac{i+t_k}{p}\right) = \left(\frac{h(i)}{p}\right) \quad \textit{for } p \nmid h(i). \qquad (8.5)$$

Since $1 \le i \le p$ and $p \mid h(i)$ holds for k values of i, thus f_i is of the form (8.5) *for all but k values of i (while for the k exceptional values of i all we can say is $f_i \in \{-1,+1\}$). The family $\mathcal{F}(K)$ obtained this way is very similar to a subfamily of $\mathcal{F}_0$ using polynomials of degree at most K and Legendre symbol, defined in Theorem 1 in [4]. Indeed, by* (8.5) *for every $F_p \in \mathcal{F}(K)$ there exists an $F'_p \in \mathcal{F}_0$ such that F_p and F'_p differ at most k elements.*

Let $\varepsilon = \frac{1}{3}$, $p > p_0$. *It follows from Theorem 1 in [11] that*

$$C_t(E_p) \le 9tp^{1/2}\log p + 1 < 10tp^{1/2}\log p$$

whence

$$\max_{1\le t\le \frac{1}{3\log 2}\log p} C_t(E_p) \le 10\left(\frac{1}{3\log 2}\log p\right)p^{1/2}\log p < p^{2/3}.$$

Thus (8.3) *holds. Using Corollary 1 we get*

$$\Gamma(\mathcal{F}(K)) \ge \left[\frac{1}{3\log 2}K^*\log p\right].$$

This result is not completely new: a variant of it was proved in [5]. Here we have obtained it as a special case of Corollary 1.

Example 4 *Let* p *be a prime,* g *be a primitive root modulo* p *and let* $E_{p-1} = (e_1, e_2, \ldots, e_{p-1})$ *be defined by*

$$e_n = \left(\frac{g^n - 1}{p}\right) \quad \text{for } 1 \le n \le p-2 \quad \text{and} \quad e_{p-1} = 1$$

Let $\mathcal{F}(K) = \mathcal{F}(K, p-2)$ *be the family generated from this sequence* E_p *(in the manner of Definition 5 and 6), and consider a sequence* $F_p(t_1, t_2, \ldots, t_k) = (f_1, f_2, \ldots, f_{p-1}) \in \mathcal{F}(K)$ *with* $1 \le k \le K$, $0 \le t_1 < t_2 < \cdots < t_k \le p-2$. *Then every element* f_i *of this sequence is of form* (3.2). *If we write* $h(x) =$

$(g^{t_1}x-1)(g^{t_2}x-1)\cdots(g^{t_k}x-1)$ and $p \nmid h(i)$ then (3.2) *can be rewritten as*

$$f_i = \left(\frac{g^{i+t_1}-1}{p}\right)\left(\frac{g^{i+t_2}-1}{p}\right)\cdots\left(\frac{g^{i+t_k}-1}{p}\right) = \left(\frac{h(g^i)}{p}\right) \quad \text{for } p \nmid h(g^i). \tag{8.6}$$

Since $1 \le i \le p-1$ and $p \mid h(g^i)$ holds for k values of i, thus f_i is of the form (8.6) *for all but k values of i (while for the k exceptional values of i all we can say is $f_i \in \{-1,+1\}$). The family $\mathcal{F}(K)$ obtained in this way is very similar to a subfamily of $\mathcal{F}_0$ using polynomials of degree at most K and the Legendre symbol, defined in (3) in [8]. Indeed, by* (8.6) *for every $F_p \in \mathcal{F}(K)$ there exists an $F'_p \in \mathcal{F}_0$ such that F_p and F'_p differ in at most k elements.*

Let $\varepsilon = \frac{1}{3}$, $p > p_0$. From Theorem 2 in [8] follows that

$$C_t(E_p) \le 5tp^{1/2}\log p + 1 < 6tp^{1/2}\log p$$

whence

$$\max_{1 \le t \le \frac{1}{3\log 2}\log p} C_t(E_p) \le 6\left(\frac{1}{3\log 2}\log p\right)p^{1/2}\log p < (p-1)^{2/3}.$$

Thus (8.3) *holds. Using Corollary 1 we get*

$$\Gamma(\mathcal{F}(K)) \ge \left[\frac{1}{3\log 2}K^*\log p\right].$$

9 Proofs of Theorems 4 and 5 and their corollaries

We will prove the two theorems simultaneously. We will refer to the problem considered in Theorem 4 (the study of $\mathcal{F}(K,L)$) as Case 1, and the problem in Theorem 5 (the study of $\mathcal{F}^*(K,T)$) will be called Case 2. We define the integer M as $M = N$ in Case 1 and $M = N - T$ in Case 2, moreover $\mathcal{F}$ is defined as $\mathcal{F} = \mathcal{F}(K,L)$ in Case 1 and $\mathcal{F} = \mathcal{F}^*(K,T)$ in Case 2.

In order to prove Theorems 4 and 5 it suffices to prove that for every choice of

$$(1 \le \ ) \ s_1 < s_2 < \cdots < s_R \ (\le M) \tag{9.1}$$

and

$$(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_R) \in \{-1, +1\}^R \tag{9.2}$$

there is a sequence $F_M = (f_1, f_2, \ldots, f_M) \in \mathcal{F}$ such that

$$f_{s_1} = \varepsilon_1, \ f_{s_2} = \varepsilon_2, \ldots, f_{s_R} = \varepsilon_R. \tag{9.3}$$

We will use the following notation: if $a_1, a_2, \ldots, a_k$ are (not necessarily distinct) integers with

$$0 \le a_1, a_2, \ldots, a_k \le L \quad \text{in Case 1}$$

and

$$0 \le a_1, a_2, \dots, a_k \le T \quad \text{in Case 2}$$

then write

$$F(a_1, a_2, \dots, a_k) = (f_1^{(a_1,\dots,a_k)}, f_2^{(a_1,\dots,a_k)}, \dots, f_M^{(a_1,\dots,a_k)}) \tag{9.4}$$

where

$$f_i^{(a_1,\dots,a_k)} = e_{i+a_1} e_{i+a_2} \cdots e_{i+a_k} \quad \text{for } i = 1, 2, \dots, M. \tag{9.5}$$

Observe that if for some $1 \le u < v \le k$ we have $a_u = a_v$, then the product $e_{i+a_u} e_{i+a_v}$ appearing on the right hand side of (9.5) is

$$e_{i+a_u} e_{i+a_v} = (e_{i+a_u})^2 = 1,$$

thus this product can be dropped (if there is at least one further e_{i+a_W}). If k is odd, then we may simplify the right hand side of (9.5) by dropping all these products, and then in the remaining factors e_{i+a_j} the a_j's will be distinct, and there will be at least one of them. It follows that for every $F(a_1, a_2, \dots, a_k)$ of form (9.4) we have

$$F(a_1, a_2, \dots, a_k) \in \mathcal{F} \quad \text{if } k \text{ is odd and } k \le K. \tag{9.6}$$

(We will need (9.6) in the proof of Theorem 4.)

We will also use the following notations: $R, N, Z, k \in \mathbb{N}$, $Z = L + 1$ in Case 1 and $Z = T$ in Case 2, $S = (s_1, s_2, \dots, s_R)$ is a sequence of integers of form (9.1) and $X = (x_1, x_2, \dots, x_R) \in \{-1, +1\}^R$. Let $V(E_N, Z, S, X, k)$

denote the number of the k-tuples $(a_1, a_2, \dots, a_k)$ of integers with $0 \le a_i < Z$ (for $i = 1, 2, \dots, k$) such that for the sequence $F_M = F_M(a_1, a_2, \dots, a_k) = (f_1^{(a_1,\dots,a_k)}, f_2^{(a_1,\dots,a_k)}, \dots, f_M^{(a_1,\dots,a_k)})$ (defined in the way described in Definition 5) we have

$$f_{s_i}^{(a_1,\dots,a_k)} = x_i \quad \text{for } i = 1, 2, \dots, R.$$

Then if we can prove that under the assumptions of Theorem 4 for every S and X we have

$$V(E_N, L+1, S, X, K^*) = V(E_N, Z, S, X, K^*) > 0 \tag{9.7}$$

then using this with $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_R)$ in place of X we obtain there is an $F_M = F_M(a_1, a_2, \dots, a_{K^*})$ which satisfies (9.3), and since $F_M \in \mathcal{F}(K, L)$ by (9.6) this proves Theorem 4.

The situation is more complicated in Case 2. Then we have to prove

$$V(E_N, T, S, X, K-1) = V(E_N, Z, S, X, K-1) > 0. \tag{9.8}$$

In order to derive the solvability of (9.3) from this, we use (9.8) with

$$X = (\varepsilon_1 e_{s_1+T}, \varepsilon_2 e_{s_2+T}, \dots, \varepsilon_R e_{s_R+T}).$$

By (9.8) there exists an $F_M = F_M(a_1, a_2, \dots, a_{K-1}) = \left(f_1^{(a_1,\dots,a_{K-1})}, f_2^{(a_1,\dots,a_{K-1})}, \dots, f_M^{(a_1,\dots,a_{K-1})}\right)$ such that

$$0 \le a_i \le T-1 \quad \text{for } 1 \le i \le K-1 \tag{9.9}$$

and

$$f_{s_1}^{(a_1,\dots,a_{K-1})} = \varepsilon_1 e_{s_1+T},\ f_{s_2}^{(a_1,\dots,a_{K-1})} = \varepsilon_2 e_{s_2+T},\ \dots, f_{s_R}^{(a_1,\dots,a_{K-1})} = \varepsilon_R e_{s_R+T}. \tag{9.10}$$

Then considering the sequence

$$F_M^* = F_M(a_1, \dots, a_{K-1}, T) = (f_1^{(a_1,\dots,a_{K-1},T)}, f_2^{(a_1,\dots,a_{K-1},T)}, \dots, f_M^{(a_1,\dots,a_{K-1},T)})$$

(defined as in Definition 5) by (9.10) we have

$$\begin{aligned} f_{s_i}^{(a_1,\dots,a_{K-1},T)} &= e_{s_i+a_1} e_{s_i+a_2} \cdots e_{s_i+a_{K-1}} e_{s_i+T} = f_{s_i}^{(a_1,\dots,a_{K-1})} e_{s_i+T} \\ &= (\varepsilon_i e_{s_i+T}) e_{s_i+T} = \varepsilon_i \quad \text{for } i = 1, 2, \dots, R \end{aligned}$$

which proves that this sequence F_M^* satisfies (9.3). Moreover, we have

$$f_{s_i}^{(a_1,\dots,a_{K-1},T)} = e_{s_i+a_1} e_{s_i+a_2} \cdots e_{s_i+a_{K-1}} e_{s_i+T}$$

and in the products on the right hand side all the subscripts $i + a_j$ are less than the last subscript $i + T$. Thus simplifying this product in the way described after formula (9.5), the factor e_{i+T} will not be cancelled out. This guarantees that we have

$$F_M^* = F_M(a_1, \dots, a_{K-1}, T) \in \mathcal{F}^*(K, T)$$

which completes the proof of Theorem 5.

It remains to prove (9.7) and (9.8). We need the following lemma:

Lemma 2 *Using the notations above we have*

$$\left|V(E_N, Z, S, X, k) - \frac{Z^k}{2^R}\right| \le \begin{cases} \left(\max_{1\le t\le R} \overset{\circ}{C}_t(E_N)\right)^k & \textit{in Case 1,} \\ \left(\max_{1\le t\le R} C_t(E_N)\right)^k & \textit{in Case 2.} \end{cases} \tag{9.11}$$

Proof of Lemma 2. We have

$$\begin{aligned}
&\left|V(E_N,Z,S,X,k) - \frac{Z^k}{2^R}\right| \\
&\quad = \left|\sum_{a_1=0}^{Z-1}\sum_{a_2=0}^{Z-1}\cdots\sum_{a_k=0}^{Z-1}\frac{1}{2^R x_1 x_2\cdots x_R}\prod_{j=1}^{R}\left(f_{s_j}^{(a_j,\dots,a_k)} + x_j\right)\right| \\
&\quad = \frac{1}{2^R}\left|\sum_{a_1=0}^{Z-1}\sum_{a_2=0}^{Z-1}\cdots\sum_{a_k=0}^{Z-1}\sum_{\substack{\mathcal{J}=(j_1,\dots,j_t)\ne\emptyset\\ 1\le j_1<\dots<j_t\le R}}\prod_{i=1}^{t} f_{s_{j_i}}^{(a_1,\dots,a_k)}\prod_{h\in\{1,2,\dots,R\}\setminus\mathcal{J}} x_h\right| \\
&\quad = \frac{1}{2^R}\left|\sum_{\substack{\mathcal{J}=(j_1,\dots,j_t)\ne\emptyset\\ 1\le j_1<\dots<j_t\le R}}\prod_{h\in\{1,2,\dots,R\}\setminus\mathcal{J}} x_h\sum_{a_1=0}^{Z-1}\sum_{a_2=0}^{Z-1}\cdots\sum_{a_k=0}^{Z-1}\prod_{i=1}^{t} f_{s_{j_i}}^{(a_1,\dots,a_k)}\right| \\
&\quad \le \frac{1}{2^R}\sum_{\substack{\mathcal{J}=(j_1,\dots,j_t)\ne\emptyset\\ 1\le j_1<\dots<j_t\le R}}\left|\sum_{a_1=0}^{Z-1}\sum_{a_2=0}^{Z-1}\cdots\sum_{a_k=0}^{Z-1}\prod_{i=1}^{t}\prod_{r=1}^{k} e_{s_{j_i}+a_r}\right| \\
&\quad = \frac{1}{2^R}\sum_{\substack{\mathcal{J}=(j_1,\dots,j_t)\ne\emptyset\\ 1\le j_1<\dots<j_t\le R}}\left|\sum_{a_1=0}^{Z-1}\sum_{a_2=0}^{Z-1}\cdots\sum_{a_k=0}^{Z-1}\prod_{r=1}^{k}\left(\prod_{i=1}^{t} e_{s_{j_i}+a_r}\right)\right| \\
&\quad = \frac{1}{2^R}\sum_{\substack{\mathcal{J}=(j_1,\dots,j_t)\ne\emptyset\\ 1\le j_1<\dots<j_t\le R}}\left|\left(\sum_{a=0}^{Z-1}\prod_{i=1}^{t} e_{s_{j_i}+a}\right)^k\right|.
\end{aligned} \tag{9.12}$$

By the definitions of $C_t(E_N)$ and $\overset{\circ}{C}_t(E_N)$ the inner sum can be estimated in the following way:

$$\left|\sum_{a=0}^{Z-1}\prod_{i=1}^{t} e_{s_{j_i}+a_r}\right| \le \begin{cases} \overset{\circ}{C}_t(E_N) & \text{in Case 1,} \\ C_t(E_N) & \text{in Case 2.} \end{cases} \tag{9.13}$$

The first sum $\sum_{\mathcal{J}}$ in the last line in (9.12) has less than 2^R terms thus (9.11) follows from (9.12) and (9.13) which completes the proof of the lemma.

By using Lemma 2 we may complete the proof of Theorem 4 in the following way: applying Lemma 2 with $L+1$ and K^* in place of Z and k we get from (9.11) that

$$\left|V(E_N, L+1, S, X, K^*) - \frac{(L+1)^{K^*}}{2^R}\right| \le \left(\max_{1\le t\le R} \overset{\circ}{C}_t(E_N)\right)^{K^*}$$

whence

$$V(E_N, L+1, S, X, K^*) \ge \frac{(L+1)^{K^*}}{2^R} - \left(\max_{1\le t\le R} \overset{\circ}{C}_t(E_N)\right)^{K^*}.$$

By our assumption (8.2) the right hand side of this inequality is positive which proves (9.7).

It can be proved similarly (by using Lemma 2 and (8.3)) that (9.8) also holds and this completes the proof of Theorems 4 and 5.

Proof of Corollary 1. Let $R = \left[\frac{\varepsilon}{\log 2} K^* \log(L+1)\right]$. Then by (8.4) we have

$$\begin{aligned} 2^{R/K^*} \max_{1\le t\le R} \overset{\circ}{C}_t(E_N) &\le 2^{(\varepsilon/\log 2)\log(L+1)} \max_{1\le t\le R} \overset{\circ}{C}_t(E_N) = (L+1)^{\varepsilon} \max_{1\le t\le R} \overset{\circ}{C}_t(E_N) \\ &< L+1 \end{aligned}$$

so that (8.2) holds. Thus by Theorem 4 we have

$$\Gamma(\mathcal{F}(K,L)) \geq R = \left[\frac{\varepsilon}{\log 2} K^* \log(L+1)\right]$$

which was to be proved.

Proof of Corollary 2. This can be derived from Theorem 5 similarly to the proof of Corollary 1.

10 Other measures of pseudorandomness of families

In Sections 8 and 9 we studied the *family complexity* of the families $\mathcal{F}(K,L)$ and $\mathcal{F}^*(K,T)$. There are also other measures of pseudorandomness of families of binary sequences. In particular, in [4] we introduced the following measure:

Definition 13 *Let $N \in \mathbb{N}$, $k \in \mathbb{N}$, and for any k binary sequences $E_N^{(1)}, \dots, E_N^{(k)}$ with*

$$E_N^{(i)} = (e_1^{(i)}, \dots, e_N^{(i)}) \in \{-1,+1\}^N \quad \textit{(for } i = 1,2,\dots,k\textit{)}$$

and any $M \in \mathbb{N}$ and k-tuple $D = (d_1, \dots, d_k)$ of non-negative integers with

$$0 \leq d_1 \leq \dots \leq d_k < M + d_k \leq N, \tag{10.1}$$

write

$$\widetilde{C}_k\left(E_N^{(1)},\ldots,E_N^{(k)}\right)=\max_{M,D}\left|\sum_{n=1}^{M} e_{n+d_1}^{(1)}\cdots e_{n+d_k}^{(k)}\right|$$

where the maximum is taken over all $D=(d_1,\ldots,d_k)$ *and* $M\in\mathbb{N}$ *satisfying* (10.1) *with the additional restriction that if* $E_N^{(i)}=E_N^{(j)}$ *for some* $i\neq j$*, then we must not have* $d_i=d_j$*. Then the* cross-correlation measure of order k *of the family of binary sequences* $E_N\in\{-1,+1\}^N$ *is defined as*

$$\Phi_k(\mathcal{F})=\max\widetilde{C}_k\left(E_N^{(1)},\ldots,E_N^{(k)}\right)$$

where the maximum is taken over all k*-tuples of binary sequences* $\left(E_N^{(1)},\ldots,E_N^{(k)}\right)$ *with* $E_N^{(i)}\in\mathcal{F}$ *for* $i=1,2,\ldots,k$*.*

Then for a "good" family $\Phi_k(\mathcal{F})$ must be small (at least for "small" k): ideally, it must be as small as $O(N^{1/2+\varepsilon})$.

One might like to estimate the cross-correlation measures of the families $\mathcal{F}(K)$, $\mathcal{F}(K,L)$ and $\mathcal{F}^*(K,T)$. The following examples show that, e.g., the cross-correlation of order 3 of these families can be large:

Example 5 *Again we start out from the binary sequence* E_N *of form* (2.1) *and, using the notations of Definitions 1, 5 and 6, consider the binary sequences*

$$F_N(0,1)=(f_1^{(1)},f_2^{(1)},f_3^{(1)},\ldots,f_N^{(1)})=(e_1e_2,e_2e_3,e_3e_4,\ldots,e_Ne_1),$$
$$F_N(0,2)=(f_1^{(2)},f_2^{(2)},f_3^{(2)},\ldots,f_N^{(2)})=(e_1e_3,e_2e_4,e_3e_5,\ldots,e_Ne_2),$$
$$F_N(1,2)=(f_1^{(3)},f_2^{(3)},f_3^{(3)},\ldots,f_N^{(3)})=(e_2e_3,e_3e_4,e_4e_5,\ldots,e_1e_2).$$

Then for $N \geq 3$, $2 \leq L \leq N-1$ clearly we have

$$F_N(0,1), F_N(0,2), F_N(1,2) \in \mathcal{F}(2,L) \subseteq \mathcal{F}(2,N-1) = \mathcal{F}(2)$$

whence

$$\begin{aligned}\Phi_3(\mathcal{F}(2)) \geq \Phi_3(\mathcal{F}(L)) \geq \widetilde{C}_3\left(F_N(0,1), F_N(0,2), F_N(1,2)\right) \geq \left|\sum_{n=1}^{N} f_n^{(1)} f_n^{(2)} f_n^{(3)}\right| \\ = \left|\sum_{n=1}^{N} (e_n e_{n+1})(e_n e_{n+2})(e_{n+1} e_{n+2})\right| = \left|\sum_{n=1}^{N} (e_n e_{n+1} e_{n+2})^2\right| = \sum_{n=1}^{N} 1 \\ = N.\end{aligned}$$

Example 6 *Using the same notations as in Example 5 and also the notations in Definitions 9 and 10, for* $T, N \in \mathbb{N}$, $5 \leq T < N$ *consider all the* $\binom{5}{2} = 10$ *pairs* (u,v) *of integers with* $0 \leq u < v \leq 4$, *and denote these pairs by* $(u_1,v_1),(u_2,v_2),\ldots,(u_{10},v_{10})$. *For* $i = 1,2,\ldots,10$ *consider the binary sequence*

$$\begin{aligned}&F_{N-T}(u_i,v_i,T) = (f_1^i, f_2^i, f_3^i, \ldots, f_{N-T}^i) \\ &= (e_{1+u_i} e_{1+v_i} e_{1+T}, e_{2+u_i} e_{2+v_i} e_{2+T}, e_{3+u_i} e_{3+v_i} e_{3+T}, \ldots, e_{N-T+u_i} e_{N-T+v_i} e_N).\end{aligned}$$

Then clearly we have

$$F^*_{N-T}(u_i,v_i,T) \in \mathcal{F}^*(3,T)$$

whence

$$\begin{aligned}
\Phi_{10}(\mathcal{F}^*(3,T)) &\geq \widetilde{C}_{10}\left(F_{N-T}(u_1,v_1,T),\dots,F_{N-T}(u_{10},v_{10},T)\right)\\
&\geq \left|\sum_{n=1}^{N-T} f_n^{(1)}\cdots f_n^{(10)}\right| = \left|\sum_{n=1}^{N-T}\prod_{i=1}^{10} e_{n+u_i}e_{n+v_i}e_{n+T}\right|\\
&= \left|\sum_{n=1}^{N-T}(e_n e_{n+1}e_{n+2}e_{n+3}e_{n+4})^4 e_{n+T}^{10}\right|\\
&= \sum_{n=1}^{N-T} 1 = N-T.
\end{aligned}$$

The large cross-correlation does not mean that this construction is useless and these families must be discarded. Indeed, one may expect that we may achieve by dropping "not too many" sequences belonging these families that the remaining subfamily has small cross-correlation and it still has large family complexity. So the problem to settle is:

Problem 1 *Show that the families $\mathcal{F}(K)$, $\mathcal{F}(K,L)$, $\mathcal{F}^*(K,T)$ have possibly large subfamilies which have small cross-correlation and large family complexity.*

There is another related problem to study:

Problem 2 *What can one say about collisions, distance minimum, avalanche property in the families studied by us?*

(See [6] for the definitions of these notions and related references.)

References

[1] R. Ahlswede, L. H. Khachatrian, C. Mauduit and A. Sárközy, *A complexity measure for families of binary sequences*, Periodica Math. Hungar. 46 (2003), 107-118.

[2] Z. X. Chen, X. N. Du and G. Z. Xiao, *Sequences related to Legendre/Jacobi sequences*, Inform. Sci. 177 (2007), 4820-4831.

[3] J. Folláth, *Construction of pseudorandom binary sequences using additive characters over $GF(2^k)$*, Period. Math. Hungar. 57 (2008), 73-81.

[4] L. Goubin, C. Mauduit and A. Sárközy, *Construction of large families of pseudorandom binary sequences*, J. Number Theory 106 (2004), 56-69.

[5] K. Gyarmati, *On the complexity of a family related to the Legendre symbol*, Period. Math. Hung. 58 (2009), 209-215.

[6] K. Gyarmati, C. Mauduit and A. Sárközy, *The cross-correlation measure for families of binary sequences*, in: Applications of Algebra and Number Theory (Lectures on the Occasion of Harald Niederreiter's 70th Birthday).

[7] K. Gyarmati, *On a family of pseudorandom binary sequences*, Period. Math. Hungar. 49 (2004), 45-63.

[8] K. Gyarmati, A. Pethő and A. Sárközy, *On linear recursion and pseudorandomness*, Acta Arith. 118 (2005), 359-374.

[9] H. N. Liu, *A family of pseudorandom binary sequences constructed by the multiplicative inverse*, Acta Arith. 130 (2007), 167-180.

[10] C. Mauduit, J. Rivat and A. Sárközy, *Construction of pseudorandom binary sequences using additive characters*, Monatsh. Math. 141 (2004), 197-208.

[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] C. Mauduit and A. Sárközy, *On the measures of pseudorandomness of binary sequences*, Discrete Math. 271 (2003), 195-207.

[13] C. Mauduit and A. Sárközy, *Construction of pseudorandom binary sequences by using the multiplicative inverse*, Acta Math. Hungar. 108 (2005), 239-252.

[14] C. Mauduit and A. Sárközy, *On large families of pseudorandom binary lattices*, Unif. Distrib. Theory 2 (2007), 23-37.

[15] L. Mérai, *A construction of pseudorandom binary sequences using both additive and multiplicative characters*, Acta Arith. 139 (2009), 241-252.

[16] L. Mérai, *Construction of large families of pseudorandom binary sequences*, Ramanujan J. 18 (2009), 341-349.

[17] A. Sárközy, *A finite pseudorandom binary sequence*, Studia Sci. Math. Hungar. 38 (2001), 377-384.