Concatenation of pseudorandom binary sequences

Katalin Gyarmati

Abstract

In the applications it may occur that our initial pseudorandom binary sequence turns out to be not long enough, thus we have to take the concatenation or merging of it with another pseudorandom binary sequences. Here our goal is study when can we form the concatenation of several pseudorandom binary sequences belonging to a given family? We introduce and study new measures which can be used for answering this question.

2000 AMS Mathematics Subject Classification: 11K45.

List of keywords and phrases: pseudorandom, concatenation, correlation.

1 Introduction

Research partially supported by Hungarian National Foundation for Scientific Research, Grants No. K49693, K67676, K72264 and the János Bolyai Research Fellowship.

In a series of papers C. Mauduit and A. Sárközy (partly with coauthors) studied finite pseudorandom binary sequences

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

In particular, in part I [14] first they introduced the following measures of pseudorandomness:

Write

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

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

$$V(E_N, M, D) = \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2}, \dots e_{n+d_k}. \tag{1}$$

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

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

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

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_k} \right| \tag{2}$$

where the maximum is taken over all $D = (d_1, d_2, \dots, d_k)$ and M such that $1 \leq d_1 < d_2 < \dots < d_k < M + d_k \leq N$.

Then the sequence E_N is considered as a "good" pseudorandom sequence if both these measures $W(E_N)$ and $C_k(E_N)$ (at least for small k) are "small" in terms of N (in particular, both are $o(N)$ as $N \to \infty$).

The goal of this paper is introducing new measures of families of binary sequences. First Anantharam [3] studied correlation measure of a family. Here we will extend his definition. We may expect that if the correlation measure of a family is small, then the sequences in the family are independent in some sense.

Definition 1 *Let $\mathcal{F} \subseteq \{-1,+1\}^N$ be a large family of pseudorandom binary sequences. The f-correlation measure of order k of $\mathcal{F}$ is defined by*

$$C_k(\mathcal{F}) \overset{def}{=} \max_{1 \le \ell \le k,\ E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(\ell)}} C_k(\{E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(\ell)}\}),$$

where the maximum is taken over all $1 \le \ell \le k$, different $E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(\ell)} \in \mathcal{F}$, and where $\{E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(\ell)}\} \in \{-1,+1\}^{\ell N}$ is a binary sequence of length ℓN obtained by writing the elements of $E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(\ell)}$ successively.

Clearly we have

Proposition 1 *If $\mathcal{F}$ and $\mathcal{G}$ are large families of pseudorandom binary sequences with $\mathcal{G} \subseteq \mathcal{F}$ then $C_k(\mathcal{G}) \le C_k(\mathcal{F})$.*

In this paper our goal is to study the importance and applicability of this measure. First I will present a short survey of some related results and facts.

Numerous binary sequences have been tested for pseudorandomness by J. Cassaigne, L. Goubin, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy. In the best constructions we have $W(E_N) \ll N^{1/2}(\log N)^{c_1}$ and $C_k(E_N) \ll$

$N^{1/2}(\log N)^{c_k}$, where $c_1, c_2, \dots$ are positive constants. However, the first constructions produced only a "few" pseudorandom sequences; usually for a fixed integer N, the construction provides only one pseudorandom sequence E_N of length N. First L. Goubin, C. Mauduit and A. Sárközy [7] succeeded in constructing large families of pseudorandom binary sequences. Their construction was the following:

Construction 1 *Suppose that p is a prime number, and $f(x) \in \mathbf{F_p}[x]$ is a polynomial with degree $k > 0$ and no multiple zero in $\overline{\mathbf{F}}_\mathbf{p}$. Define the binary sequence $E_p = \{e_1, \dots, e_p\}$ by*

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{for } (f(n), p) = 1, \\ +1 & \text{for } p \mid f(n) \end{cases} \tag{3}$$

(where $\left(\frac{n}{p}\right)$ denotes the Legendre symbol).

It turns out that under some not too restrictive conditions on p or the degree of the polynomial the pseudorandom measures of E_p are small. Indeed Goubin, Mauduit and Sárközy [7] proved the following

Theorem A *If p is a prime and $f(x)$ is a polynomial as it is described in Construction 1, then for the sequence E_p defined by (3) we have*

$$W(E_p) < 10kp^{1/2}\log p.$$

Moreover, assume that for $\ell \in \mathbb{N}$ one of the following assumptions holds:

(i) $\ell = 2$;

(ii) $\ell < p$ *and 2 is a primitive root modulo p;*

(iii) $(4k)^\ell < p$.

Then we also have

$$C_\ell(E_p) < 10k\ell p^{1/2} \log p.$$

Since then numerous other large families of pseudorandom sequences have been constructed see [8], [9], [10], [11], [12] and [13].

In many applications it is not enough to know that the family contains many binary sequences with strong pseudorandom properties; it is also important that the family has a "rich", "complex" structure, there are many "independent" sequences in it. Ahlswede, Khachatrian, Mauduit and Sárközy [1] introduced the *f-complexity* ("f" for family):

Definition 2 *The complexity* $C(\mathcal{F})$ *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* $1 \le i_1 < i_2 < \cdots < i_j \le N$, *and for any* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_j \in \{-1,+1\}^j$, *we have at least one* $E_N = \{e_1, \ldots, e_N\} \in \mathcal{F}$ *for which*

$$e_{i_1} = \varepsilon_1,\ e_{i_2} = \varepsilon_2, \ldots, e_{i_j} = \varepsilon_j.$$

It is clear from Definition 1 that for $j < C(\mathcal{F})$, there exist at least $2^{C(\mathcal{F})-j}$ sequences $E_N \in \mathcal{F}$ with

$$e_{i_1} = \varepsilon_1,\ e_{i_2} = \varepsilon_2, \ldots, e_{i_j} = \varepsilon_j.$$

However, the high f-complexity ensures only that the family contains many "independent" sequences in this sense, it does not ensure that any pair of sequences in the family are independent. Next we will show an example for a family, where the f-complexity is large, but there are certain connections between almost any pair of sequences.

Example 1 *Let* $3 \mid N$ *and* $E_N = \{e_1, e_2, \dots, e_N\} \in \{-1,+1\}^N$ *be a truly random sequence. Define the family* $\mathcal{F}(E_N)$ *of binary sequences in the following way:*

$$\begin{aligned}\mathcal{F}(E_N) =& \{\{e_1f_1, e_2f_2, \dots, e_Nf_N\} : \ \{f_1, f_2, \dots, f_N\} \in \{-1,+1\}^N \text{ and} \\ & |\{i : \ f_i = 1\}| = N/3\}.\end{aligned}$$

Since $E_N = \{e_1, e_2, \dots, e_N\}$ is a truly random sequence, for arbitrary sequence $\{f_1, f_2, \dots, f_N\}$ the sequence $\{e_1f_1, e_2f_2, \dots, e_Nf_N\}$ is a random type sequence. Thus $\mathcal{F}(E_N)$ consists of random type sequences. Similarly to [5] (however, the proof would be lengthy) it can be seen that almost all sequences from $\mathcal{F}(E_N)$ have strong pseudorandom properties.

The well-known Vernam cipher algorithm uses $\{0,1\}$ sequences. In our example we have $\{-1,+1\}$ sequences. These sequences can be used as a keystream in a variant of the Vernam cipher, where we use multiplication everywhere in place of modulo 2 addition. In other words, we encrypt a message $\{m_1, m_2, \dots, m_N\} \in \{-1,+1\}^N$ by a keystream $\{e_1f_1, e_2f_2, \dots, e_Nf_N\} \in \mathcal{F}(E_N)$ so that the encrypted message is $\{m_1e_1f_1, m_2e_2f_2, \dots, m_Ne_Nf_N\}$.

One obvious drawback of the Vernam-cipher is that it is not recommanded to use the same keystream twice. For similar reasons, no two messages $\{m_1, \ldots, m_N\}$ and $\{m'_1, \ldots, m'_N\}$ can be encrypted by two different keystreams from $\mathcal{F}(E_N)$. Indeed, suppose that $\{m_1, \ldots, m_N\}$ is encrypted by $\{e_1 f_1, e_2 f_2, \ldots, e_N f_N\} \in \mathcal{F}(E_N)$ and $\{m'_1, \ldots, m'_N\}$ is encrypted by $\{e_1 f'_1, e_2 f'_2, \ldots, e_N f'_N\} \in \mathcal{F}(E_N)$. Then the two encrypted messages are $\{m_1 e_1 f_1, m_2 e_2 f_2, \ldots, m_N e_N f_N\}$ and $\{m'_1 e_1 f'_1, m'_2 e_2 f'_2, \ldots, m'_N e_N f'_N\}$. We can take the termwise product of the two encrypted messages which is $\{m_1 e_1 f_1 m'_1 e_1 f'_1, \ldots, m_N e_N f_N m'_N e_N f'_N\} = \{m_1 m'_1 f_1 f'_1, \ldots, m_N m'_N f_N f'_N\}$. Since in both $\{f_1, \ldots, f_N\}$ and $\{f'_1, \ldots, f'_N\}$ the rate of $+1$'s and -1's is $1:2$, by a simple computation we see that in the sequence $\{f_1 f'_1, \ldots, f_N f'_N\}$ the rate of $+1$'s and -1's is usually around $5:4$. This fact may help the eavesdropper to find out the messages $\{m_1, m_2, \ldots, m_N\}$ and $\{m'_1, m'_2, \ldots, m'_N\}$.

Clearly the f-complexity of $\mathcal{F}(E_N)$ is large: $N/3$. We have seen that every sequence in the family is random-type and the f-complexity is large, but in certain applications, for example in the variant of the Vernam-cipher we may use at most one sequence from $\mathcal{F}(E_N)$ as a keystream. This shows that the f-complexity is not enough to guarantee the secure applicability of the family, one also needs the introduction of further measures. In certain applications we need at least a weak independence of all sequences used in the applications. We may expect that the small f-correlation measures assure this weak independence.

I would like to thank Professors László Csirmaz and András Sárközy for the valuable discussions.

2 Theorems

Next we study the f-correlation measure of the large family of pseudo-random binary sequences introduced by Goubin, Mauduit and Sárközy in [7].

Proposition 2 *Let p be a prime number and $R \in \mathbb{N}$. Consider all the polynomials $f(x) \in \mathbf{F_p}[x]$ with leading coefficient 1, which has no multiple roots and*

$$0 < \deg f(x) \leq R,$$

where $\deg f(x)$ denotes the degree of $f(x)$. For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = \{e_1, e_2, \dots, e_p\} \in \{-1, +1\}^p$ defined by (3), and let $\mathcal{F}_1$ denote the family of all binary sequences obtained in this way. Then

$$C_2(\mathcal{F}_1) \geq p - 1.$$

Clearly $\mathcal{F}_1$ contains many independent sequences, but a few sequences from $\mathcal{F}_1$ are not independent. For example $E_p(f(x)) = \{e_1, e_2, \dots, e_p\}$ and $E_p(f(x+1)) = \{e_2, e_3, \dots, e_p, e_1\}$ are both members of the family $\mathcal{F}_1$ and we may get $E_p(f(x))$ by shifting to left by 1 the elements of $E_p(f(x+1))$. Using this property we see that the f-correlation will be large. By using the function V defined in (1)

$$\begin{aligned}C_2(\mathcal{F}) &\geq C_2(\{E_p(f(x)), E_p(f(x+1))\})\\ &\geq |V(\{E_p(f(x)), E_p(f(x+1))\}, p-1, (1,p))| = \left|e_2^2 + e_3^2 + \cdots + e_p^2\right|\\ &= p-1.\end{aligned}$$

Remark 1 Similarly it is easy to prove that for $a \in \mathbf{F_p}$

$$C_2(\{E_p(f(x)), E_p(f(x+a))\}) \geq \lceil p/2 \rceil .$$

This shows that if the f-correlation measure is smaller than $p/2$, then we may use at most one of the polynomial $f(x), f(x+1), \ldots, f(x+p-1)$ in the construction.

$f(x), f(x+1), \ldots, f(x+p-1)$ are polynomials of the same degree r. If this degree $r < p$, then there exists exactly one polynomial $f(x+a)$ with $a \in \mathbf{F_p}$ such that the coefficient of x^{r-1} is 0. Next we restrict our family to such polynomials.

Theorem 1 *Let p be an odd prime number and $R \in \mathbb{N}$, $R < p$. Consider all the polynomials $f(x) \in \mathbf{F_p}[x]$ which have no multiple roots,*

$$0 < \deg f(x) \leq R$$

and $f(x)$ is of the form

$$f(x) = x^r + a_{r-2}x^{r-2} + a_{r-3}x^{r-3} + \cdots + a_1 x + a_0 \text{ with } 1 \leq r \leq R,\ a_i \in \mathbf{F_p},$$

so that the coefficient of the term $x^{\deg f - 1}$ is $a_{r-1} = 0$. For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = \{e_1, e_2, \ldots, e_p\} \in$

$\{-1,+1\}^p$ *defined by* (3)*, and let* $\mathcal{F}_2$ *denote the family of all binary sequences obtained in this way. (Clearly* $\mathcal{F}_2 \subseteq \mathcal{F}_1$*, where* $\mathcal{F}_1$ *is a family defined in Proposition 2.) Then*

$$C_2(\mathcal{F}_2) \le 80Rp^{1/2}\log p.$$

Viktória Tóth [19] also studied the independence of pairs of sequences. She introduced the distance of two sequences. The correlation measure of order 2 gives an upper bound for the difference of $p/2$ and this distance, but the reverse is not true. For example, the distance of $E_p(f(x))$ and $E_p(f(x+1))$ is around $p/2$, but we have seen that $C_2(\{E_p(f(x)), E_p(f(x+1))\})$ is large.

Theorem 1 is very useful when a weak independence of pairs of sequences is required, but we do not need the independence of 3 or more sequences. However, small f-correlation measure of order 2 does not give full security. In the next theorem we give a family which has small f-correlation measure of order 2, but knowing enough elements of a sequence from the family we can compute the other elements of the sequence relatively quickly.

Theorem 2 *Let* p *be a prime number and* $R \in \mathbb{N}$*,* $R < p$*. Consider all the polynomials* $f(x) \in \mathbf{F_p}[x]$*, where*

$$0 < \deg f(x) \le R$$

and $f(x)$ *is of the form*

$$\begin{aligned} &f(x) = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_r) \\ &\textit{with } 1 \le r \le R,\ \alpha_i \in \mathbf{F_p} \textit{ and } \alpha_1+\alpha_2+\cdots+\alpha_r = 0 \end{aligned} \tag{4}$$

(so that $f(x)$ splits into linear factors over $\mathbf{F_p}$ and the coefficient of the term $x^{\deg f - 1} = x^{r-1}$ is $a_{r-1} = 0$). For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = \{e_1, e_2, \dots, e_p\} \in \{-1, +1\}^p$ defined by

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{if } p \nmid f(n), \text{ i.e., } n \neq \alpha_1, \alpha_2, \dots, \alpha_r, \\ \prod\limits_{\substack{i=1 \\ i \neq s}}^{r} \left(\frac{\alpha_s - \alpha_i}{p}\right) & \text{if } n = \alpha_s, \end{cases}$$

and let $\mathcal{F}_3$ denote the family of all binary sequences obtained in this way. Then

$$C_2(\mathcal{F}_3) \leq 80 R p^{1/2} \log p. \tag{5}$$

Assume that somebody knows the values of $e_{n_1}, e_{n_2}, \dots, e_{n_t}$. Let $w = 2\left[2p^{1-1/(R+1)}\right] + 1$. For $|m| < p^{1-1/(R+1)}$, $m \neq 0$ let A_m be a $w \times t$ matrix whose entries are $a_{i,j} = 1$ if $\left(\frac{mn_j - i}{p}\right) = -1$, otherwise $a_{ij} = 0$ for $j = 1, 2, \dots, t$ and $i = -\left[2p^{1-1/(R+1)}\right], -\left[2p^{1-1/(R+1)}\right] + 1, \dots, \left[2p^{1-1/(R+1)}\right]$. Let ρ denote the maximum of the ranks of the matrices A_m. Then knowing the values of $e_{n_1}, e_{n_2}, \dots, e_{n_t}$ one can compute the other elements of the sequence by $O(2^{w-\rho} t^2 w^2)$ bit operations.

Remark 2 I was not able to estimate the rank of the matrices A_m, but for $t \geq w$ we may expect that the rank of the $w \times t$ matrices A_m is $\min\{w, t\} = w$, $\rho = w$, so we can compute the elements of the sequence by $O(t^2 w^2) = o(p^4)$ bit operations. We note that sometimes more sequences may exist with the fixed values $e_{n_1}, e_{n_2}, \dots, e_{n_t}$. Indeed take two polynomials of type (4), denote the associated sequences by $E_p(f) = \{e_1, e_2, \dots, e_p\}$ and $E_p(f') =$

$\{e'_1, e'_2, \dots, e'_p\}$. We fix the places $n_1, n_2, \dots, n_t$ with

$$e_{n_i} = e'_{n_i} = 1. \tag{6}$$

The number of such n_i's is around $p/4$. Clearly then both $E_p(f)$ and $E_p(f')$ satisfy (6).

Theorem 2 is totally novel, usually it is not studied that from how many elements one may find out the other elements of the sequence. On the other hand Theorem 2 is not very useful in the applications since only from p^{1-c} (where $c > 0$ is small) elements one can compute the other elements of the sequence. Moreover the sizes of the matrices A_m's are very large in Theorem 2.

By Theorem 2 if the rank of the matrices A_m is large, one can compute the elements of the sequence relatively quickly. One explanation of this phenomenon can be that the f-correlation measure of higher order than 2 is large. Indeed, we will prove

Theorem 3 *For the family $\mathcal{F}_3$ defined in Theorem 2 and for $k \geq 3$ we have*

$$C_k(\mathcal{F}_3) \geq p.$$

Fortunately, $\mathcal{F}_2$ has a large subfamily for which the f-correlation is always small.

Theorem 4 *Let p be a prime number and $R \in \mathbb{N}$, $R < p$. Consider all the polynomials $f(x) \in \mathbf{F_p}[x]$ which are irreducible,*

$$0 < \deg f(x) \leq R \tag{7}$$

and $f(x)$ is of the form

$$f(x) = x^r + a_{r-2}x^{r-2} + a_{r-3}x^{r-3} + \cdots + a_1 x + a_0 \text{ with } 1 \le r \le R,\ a_i \in \mathbf{F_p}, \quad (8)$$

so the coefficient of the term $x^{\deg f - 1} = x^{r-1}$ is $a_{r-1} = 0$. For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = \{e_1, e_2, \dots, e_p\} \in \{-1, +1\}^p$ defined by (3), and let $\mathcal{F}_4$ denote the family of all binary sequences obtained in this way. (Clearly $\mathcal{F}_4 \subseteq \mathcal{F}_1$, where $\mathcal{F}_1$ is the family defined in Proposition 2.) Then for $k \ge 2$

$$C_k(\mathcal{F}_4) \le 10Rk^2 2^{k-1} p^{1/2} \log p.$$

Theorem 4 gives a non-trivial upper bound if $R = o\left(\frac{\sqrt{p}}{2^k k^2 \log p}\right)$. Then $\mathcal{F}_4$ has strong pseudorandom properties. (Here we prove a strong upper bound, while by computer it is very slow to compute even the f-correlation measures of small orders.)

The construction of irreducible polynomials over $\mathbf{F_p}$ (which we need in Theorem 4) is an important and difficult subject (see for example [4], [6], [15], [16], [18]).

In order to avoid construction of irreducible polynomials we also introduce the weak f-correlation measure. Here we do not consider the correlation measure of all ℓ-tuples $\left(E_N^{(1)}, E_N^{(2)}, \dots, E_N^{(\ell)}\right)$ just certain ℓ-tuples.

Definition 3 *Let $\mathcal{F} \subseteq \{-1, +1\}^N$ be a family of binary sequences. Let $\mathcal{H}$ be a set of ℓ-tuples of different sequences from $\mathcal{F}$ with $1 \le \ell \le k$. Then the* *week f-correlation measure of order k with respect to $\mathcal{H}$ is*

$$W_{k,\mathcal{H}}(\mathcal{F}) \stackrel{def}{=} \max_{1\le \ell \le k,\ \left(E_N^{(1)},E_N^{(2)},\dots,E_N^{(\ell)}\right)\in\mathcal{H}} C_k(\{E_N^{(1)},E_N^{(2)},\dots,E_N^{(\ell)}\}),$$

where the maximum is taken over all $1 \le \ell \le k$, $\left(E_N^{(1)},E_N^{(2)},\dots,E_N^{(\ell)}\right) \in \mathcal{H}$, where the sequences $E_N^{(1)},E_N^{(2)},\dots,E_N^{(\ell)}$ are different and where $\{E_N^{(1)},E_N^{(2)},\dots,E_N^{(\ell)}\} \in \{-1,+1\}^{\ell N}$ is the binary sequence of length ℓN obtained by writing the elements of $E_N^{(1)},E_N^{(2)},\dots,E_N^{(\ell)}$ successively.

Each of Theorems 1,2 and 4 will be derived from Theorem 5 below:

Theorem 5 *Suppose that $\mathcal{F} \subseteq \mathcal{F}_1$ is a family of pseudorandom binary sequences, where $\mathcal{F}_1$ is the family defined in Proposition 2. Let $\mathcal{H}$ be a set of ℓ-tuples from $\mathcal{F}$ with $1 \le \ell \le k$, for which the following holds: If*

$$(E_p(f_1),E_p(f_2),\dots,E_p(f_\ell)) \in \mathcal{H},$$

where $E_p(f_i) \in \mathcal{F}_1$ is defined by (3) with f_i in place of f, then for $1 \le i_1 \le i_2 \le \dots \le i_k \le \ell$, $a_{i_1},a_{i_2},\dots,a_{i_k} \in \mathbf{F_p}$ where $a_{i_t} \ne a_{i_s}$ if $i_t = i_s$,

$$\prod_{j=1}^{k} f_{i_j}(x+a_{i_j}) \tag{9}$$

is never of the form $cg(x)^2$ with $c \in \mathbf{F_p}$, $g(x) \in \mathbf{F_p}[x]$.

Then for the weak f-correlation measure of order k with respect to $\mathcal{H}$ we have

$$W_{k,\mathcal{H}}(\mathcal{F}) \le 10Rk^2 2^{k-1}p^{1/2}\log p.$$

Theorem 5 is also useful when we have only few sequences in the family. Then we need not worry about the irreducibility of the polynomials involved. We need to check that there is no such product of shifted polynomials which is of the form $cg(x)^2$ with $c \in \mathbf{F_p}$, $g(x) \in \mathbf{F_p}[x]$. However the number of such products can be very large since in (9) the a_{i_j}'s may usually take p different values. The next theorem shows that using the factorization of all polynomials we need not check that all products $\prod_{j=1}^{k} f_{i_j}(x + a_{i_j})$ are not of the form $cg(x)^2$ with $c \in \mathbf{F_p}$ and $g(x) \in \mathbf{F_p}[x]$.

Theorem 6 *Let $f_1(x), f_2(x), \ldots, f_k(x) \in \mathbf{F_p}[x]$. Suppose that $f_i(x)$ factors as*

$$f_i(x) = b_i \prod_{j=1}^{r_i} (x - \alpha_j^{(i)})$$

over $\overline{\mathbf{F}}_\mathbf{p}$, so that the degree of $f_i(x)$ is r_i, its leading coefficient is b_i, and its roots are $\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_{r_i}^{(i)}$. Define

$$\tilde{f}_i(x) \stackrel{def}{=} b_i^p \prod_{j=1}^{r_i} \left(x - \left(\alpha_j^{(i)} \right)^p + \alpha_j^{(i)} \right).$$

Then if

$$\prod_{j=1}^{\ell} f_j(x + a_j)$$

is of the form $cg(x)^2$ with $c \in \mathbf{F_p}$, $g(x) \in \mathbf{F_p}[x]$ then

$$\prod_{j=1}^{\ell} \tilde{f}_j(x) \tag{10}$$

is of the form $\tilde{c}\tilde{g}(x)^2$ with $\tilde{c} \in \mathbf{F_p}$, $\tilde{g}(x) \in \mathbf{F_p}[x]$.

In (10) there are no a_j's, so we must check much less products.

Unfortunately the reverse theorem is not true. Consider the polynomials $f_1(x) = x(x+1)$, $f_2(x) = x(x+2)$, $f_3(x) = x(x+3)$. Then $\tilde{f}_1(x) = \tilde{f}_2(x) = \tilde{f}_3(x) = x^2$, thus $\tilde{f}_1(x)\tilde{f}_2(x)\tilde{f}_3(x) = x^6$ is of the form $\tilde{c}\tilde{g}(x)^2$ with $\tilde{c} \in \mathbf{F_p}$, $\tilde{g}(x) \in \mathbf{F_p}[x]$. On the other hand it is easy to check that there is no $a_1, a_2, a_3 \in \mathbf{F_p}$ such that

$$f_1(x+a_1)f_2(x+a_2)f_3(x+a_3)$$

is of the form $cg(x)^2$ with $c \in \mathbf{F_p}$, $g(x) \in \mathbf{F_p}[x]$.

This theorem can be used when we have only few polynomials in the construction. If we have more polynomials or we do not wish to deal with the factorizations of the polynomials, then Theorem 4 guarantees that using irreducible polynomials the family has strong f-correlation measure.

3 Proofs

First we assume that Theorem 5 has been proved, and we prove the other theorems by using Theorem 5.

Proof of Theorem 1

Let $\mathcal{H}$ be a set which contains every sequences from $\mathcal{F}_2$ and which also contains every pairs of different sequences from $\mathcal{F}_2$.

$$\mathcal{H} = \{E_p : \ E_p \in \mathcal{F}_2\} \cup \{(E_p^{(1)}, E_p^{(2)}) : \ E_p^{(1)} \neq E_p^{(2)} \in \mathcal{F}_2\}.$$

Then $W_{2,\mathcal{H}}(\mathcal{F}_2) = C_2(\mathcal{F}_2)$. We would like to apply Theorem 5 for this set $\mathcal{H}$. In order to apply Theorem 5 we have to show that $\mathcal{F}_2$ is a family such that for every $f \in \mathcal{F}_2, a_1 \neq a_2 \in \mathbf{F_p}$ the product

$$f(x + a_1)f(x + a_2)$$

is not of the form $cg(x)^2$ with $c \in \mathbf{F_p}$ and $g(x) \in \mathbf{F_p}[x]$, and for every $f, h \in \mathcal{F}_2$, $f \not\equiv h$, $a_1, a_2 \in \mathbf{F_p}$

$$f(x + a_1)h(x + a_2)$$

is not of the form $cg(x)^2$ with $c \in \mathbf{F_p}$ and $g(x) \in \mathbf{F_p}[x]$. Gauss proved that in $\mathbf{F_p}[x]$ there is a unique factorization (see, for example [17, Theorem 207].) Therefore $f(x + a_1)h(x + a_2)$ is of the form $cg(x)^2$ with $c \in \mathbf{F_p}$ and $g(x) \in \mathbf{F_p}[x]$ if and only if every irreducible factors appear with even multiplicity. Since both $f(x+a_1)$ and $h(x+a_2)$ have no multiple roots, thus it follows that

$$f(x + a_1) = h(x + a_2)$$

or, in equivalent form,

$$f(x) = h(x + a_2 - a_1) \tag{11}$$

In Remark 1 we noted that only one of the polynomials $h(x), h(x + 1), \ldots, h(x + p - 1)$ belongs to $\mathcal{F}_2$, so only $h(x) \in \mathcal{F}_2$, thus in (11) we have $a_2 - a_1 \equiv 0 \pmod{p}$, and so $f(x) \equiv h(x)$. Thus the condition of Theorem 5 holds, and using Theorem 5 we get the statement.

Proof of Theorem 2

The proof of (5) is similar to the proof of Theorem 1, I leave the details to the reader.

In order to prove the last statement of the theorem, we will use the following lemma:

Lemma 1 *Let p be an odd prime, for $a \in \mathbb{Z}$, let $r_p(a)$ denote the absolute least residue of modulo p, i.e., define $r_p(a) \in \mathbb{Z}$ by*

$$r_p(a) \equiv a \pmod{p}, \ |r_p(a)| \le \frac{p-1}{2}.$$

For $k \in \mathbb{N}$, $a_1, a_2, \ldots, a_k \in \mathbf{Z_p}$ there exists an integer $m \neq 0$ such that

$$|r_p(ma_i)| \le 2p^{1-1/k} \ \text{ for } i = 1, 2, \ldots, k.$$

Proof of Lemma 1 A variant of this lemma is proved in [7, Lemma 3]. Here we adapt their proof. The lemma is trivial for $p \le 2p^{1-1/k}$. Therefore we may assume

$$2p^{1-1/k} < p. \tag{12}$$

Consider the k-tuples

$$u_j = (r_p(ja_1), \ldots, r_p(ja_k)), \ j = 1, 2, \ldots, p. \tag{13}$$

Write $D = \left[2p^{1-1/k}\right]$ and $Z = \left[\frac{p}{D}\right] + 1$. Then $DZ = D\left(\left[\frac{p}{D}\right] + 1\right) > p$, thus each of the k-tuples in (13), there are uniquely determined non-negative integers $t_1 = t_1(j), \dots, t_k = t_k(j)$ such that

$$r_p(ja_i) \in \{-\frac{p-1}{2} + t_iD, -\frac{p-1}{2} + t_iD + 1, \dots, -\frac{p-1}{2} + (t_i+1)D - 1\}$$
$$\text{for } i = 1, 2, \dots, k$$

and for these integers t_i clearly we have

$$t_i \in \{0, 1, \dots, Z-1\} \text{ for } i = 1, 2, \dots, k. \tag{14}$$

By (12) we have $1 < \frac{p}{D}$. Thus the number of the possible k-tuples $(t_1, t_2, \dots, t_k)$ with (14) is

$$Z^k = \left(\left[\frac{p}{D}\right] + 1\right)^k < \left(2\frac{p}{D}\right)^k < \left(2\frac{p}{p/(2p^{1-1/k})}\right)^k = p,$$

thus there is at least one k-tuple $(t_1, t_2, \dots, t_k)$ which is assigned at least two distinct j values j_1, j_2:

$$t_1 = t_1(j_1) = t_1(j_2), \dots, t_k = t_k(j_1) = t_k(j_2). \tag{15}$$

Then we have

$$-\frac{p-1}{2} + t_iD \le r_p(j_1a_i),\ r_p(j_2a_i) < -\frac{p-1}{2} + (t_i+1)D,$$

whence

$$|r_p(j_1a_i) - r_p(j_2a_i)| < D \text{ for } i = 1, 2, \dots, k. \tag{16}$$

Now define m by $m = |j_1 - j_2|$ so that, by $1 \le j_1, j_2 \le p$ and $j_1 \neq j_2$, we have $(m, p) = 1$.

Then it follows from (16) that

$$|r_p(ma_i)| = |r_p((j_1 - j_2)a_i)| \leq |r_p(j_1 a_i) - r_p(j_2 a_i)| < D \text{ for } i = 1, 2, \ldots, k$$

which completes the proof of Lemma 1.

Consider the roots of the polynomial $f(x)$: $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{Z}_p$. Here $r \leq R$. Using Lemma 1 for $1, \alpha_1, \alpha_2, \ldots, \alpha_r$ we get that there exists an integer $m \neq 0$, $|m| < 2p^{1-1/(r+1)} < 2p^{1-1/(R+1)}$ such that

$$|r_p(m\alpha_i)| \leq 2p^{1-1/(R+1)} \text{ for } i = 1, 2, \ldots, r. \tag{17}$$

Unfortunately we do not know the value of this m. Thus we check all $m = -\left[2p^{1-1/(R+1)}\right], \ldots, -1, 1, \ldots, \left[2p^{1-1/(R+1)}\right]$. For all m we determine the polynomial $f(m^{-1}x)$ in place of $f(x)$. We know that there is an m for which (17) holds. Therefore our method is the following: for all $m = -\left[2p^{1-1/(R+1)}\right], \ldots, -1, 1, \ldots, \left[2p^{1-1/(R+1)}\right]$, we compute $f(m^{-1}x)$ by assuming that for all roots of $f(m^{-1}x)$, for $m\alpha_1, m\alpha_2, \ldots, m\alpha_r$ we have

$$|r_p(m\alpha_i)| \leq 2p^{1-1/(R+1)}.$$

From $f(m^{-1}x)$ it is easy to determine $f(x)$ and we check whether for $f(x)$ the sequence $E_p = E_p(f) = (e_1, e_2, \ldots, e_p) \in \{-1, +1\}^p$ is such that $e_{n_1}, e_{n_2}, \ldots, e_{n_t}$ takes the required ± 1 values. If for a polynomial $f(x)$ one knows the values $e_{n_1}, e_{n_2}, \ldots, e_{n_t}$, then one knows that the sequence $\tilde{E}_p = E_p(f(m^{-1}x)) = \{\tilde{e_1}, \tilde{e_2}, \ldots, \tilde{e_p}\}$ is such that $\tilde{e}_{mn_1} = e_{n_1}, \ldots, \tilde{e}_{mn_t} = e_{n_t}$, so one knows the values of $\tilde{e}_{mn_1} = e_{n_1}, \ldots, \tilde{e}_{mn_t} = e_{n_t}$.

Let r denote the degree of $f(x)$. Then

$$\prod_{\substack{i \text{ is a root} \\ \text{of } f(m^{-1}x)}} (x - i) = m^r f(m^{-1}x).$$

Thus for $mn_1, \ldots, mn_t$ we have

$$\prod_{\substack{i \text{ is a root} \\ \text{of } f(m^{-1}x)}} (mn_j - i) = m^r f(m^{-1}mn_j) = m^r f(n_j) \text{ for } j = 1, 2, \ldots, t.$$

If $p \nmid f(n_j)$, so $n_j \neq \alpha_s$ for $1 \leq s \leq r$, then by taking the Legendre symbol of both sides we get

$$\prod_{\substack{i \text{ is a root} \\ \text{of } f(m^{-1}x)}} \left(\frac{mn_j - i}{p}\right) = \left(\frac{m^r f(n_j)}{p}\right) = \left(\frac{m^r}{p}\right)\left(\frac{f(n_j)}{p}\right) = \left(\frac{m^r}{p}\right) e_{n_j}$$

for $j = 1, 2, \ldots, t$. Since $n_j \neq \alpha_s$ for $1 \leq s \leq r$, $mn_j \neq m\alpha_s$ for $1 \leq s \leq r$, so mn_j is not a root of $f(m^{-1}x)$. Thus we write

$$\prod_{\substack{i \text{ is a root} \\ \text{of } f(m^{-1}x), \\ i \neq mn_j}} \left(\frac{mn_j - i}{p}\right) = \left(\frac{m^r f(n_j)}{p}\right) = \left(\frac{m^r}{p}\right)\left(\frac{f(n_j)}{p}\right) = \left(\frac{m^r}{p}\right) e_{n_j} \quad (18)$$

for $j = 1, 2, \ldots, t$. We proved that (18) holds if $n_j \neq \alpha_s$ for $1 \leq s \leq r$. Now we will prove that it also holds for $n_j = \alpha_s$. Indeed, since the roots of $f(m^{-1}x)$ are $m\alpha_1, m\alpha_2, \ldots, m\alpha_r$ and $n_j = \alpha_s$ we have

$$\begin{aligned}\prod_{\substack{i \text{ is a root} \\ \text{of } f(m^{-1}x), \\ i \neq mn_j}} \left(\frac{mn_j - i}{p}\right) &= \prod_{\substack{i=1 \\ \alpha_i \neq n_j}}^{r} \left(\frac{mn_j - m\alpha_i}{p}\right) = \left(\frac{m^r}{p}\right) \prod_{\substack{i=1 \\ \alpha_i \neq n_j}}^{r} \left(\frac{n_j - \alpha_i}{p}\right) \\ &= \left(\frac{m^r}{p}\right) \prod_{\substack{i=1 \\ i \neq s}}^{r} \left(\frac{\alpha_s - \alpha_i}{p}\right) = \left(\frac{m^r}{p}\right) e_{\alpha_s} = \left(\frac{m^r}{p}\right) e_{n_j},\end{aligned}$$

which was to be proved.

We know the values of $e_{n_1}, \ldots, e_{n_t}$. We suppose that for every root β of $f(m^{-1}x)$ we have

$$|r_p(\beta)| \leq 2p^{1-1/(R+1)}.$$

We introduce variables $x_{-\left[2p^{1-1/(R+1)}\right]}, x_{-\left[2p^{1-1/(R+1)}\right]+1}, \ldots, x_0, \ldots, x_{\left[2p^{1-1/(R+1)}\right]}$ such that

$$x_i = \begin{cases} 1 & \text{if } i \text{ occurs with odd multiplicity amongst the roots of } f(mx), \\ 0 & \text{if } i \text{ occurs with even multiplicity amongst the roots of } f(mx), \end{cases}$$

Then (18) becomes

$$\begin{aligned} \prod_{\substack{x_i=1,\\ i\neq mn_j}} \left(\frac{mn_j - i}{p}\right) &= \left(\frac{m^r}{p}\right) e_{n_j} \\ \prod_{\substack{x_i=1,\ \left(\frac{mn_j-i}{p}\right)=-1,\\ i\neq mn_j}} (-1) &= \left(\frac{m^r}{p}\right) e_{n_j} \\ (-1)^S &= \left(\frac{m^r}{p}\right) e_{n_j}, \end{aligned}$$

where

$$S = \sum_{x_i=1,\ \left(\frac{mn_j-i}{p}\right)=-1} 1.$$

Define the integers $c_1, c_2, \ldots, c_t \in \{0, 1\}$ by

$$c_j = \begin{cases} 1 & \text{if } \left(\frac{m^r}{p}\right) e_{n_j} = -1, \\ 0 & \text{if } \left(\frac{m^r}{p}\right) e_{n_j} = 1. \end{cases}$$

Then

$$(-1)^S = (-1)^{c_j}$$

$$S = \sum_{x_i=1,\ \left(\frac{mn_j-i}{p}\right)=-1} 1 \equiv c_j \pmod 2$$

$$\sum_{\left(\frac{mn_j-i}{p}\right)=-1} x_i \equiv c_j \pmod 2 \tag{19}$$

for $j = 1, 2, \ldots, t$. The equations in (19) are linear in the variables $x_{-\left[2p^{1-1/(R+1)}\right]}, x_{-\left[2p^{1-1/(R+1)}\right]+1}, \ldots, x_0, \ldots, x_{\left[2p^{1-1/(R+1)}\right]}$. So we have t linear equations in w variables. By Gauss elimination we solve (19) by $O(t^2 w)$ bit operations. We may get 0, 1 or more solutions. The matrix of this linear equation is the matrix A_m defined in Theorem 2. The rank of this matrix is $\leq \rho$, so the number of solutions of the system of linear equations (19) is $\leq 2^{w-\rho}$. Next we check what solutions lead to a polynomial with degree less than R. So far we needed $O(2^{w-\rho} t^2 w)$ bit operations. m may take $O(w)$ different values, so the algorithm uses $O(2^{w-\rho} t^2 w^2)$ bit operations.

Proof of Theorem 3 We define k polynomials as it follows:

$$f_i(x) = x - i \text{ for } 1 \leq i \leq k-1$$

and

$$f_k(x) = \prod_{i=1}^{k-1} f_i(x) = (x-1)(x-2)\ldots(x-k+1).$$

Define the sequence $E_p^{(i)}$ by

$$E_p^{(i)} = E_p(f_i(x)) = \{e_1^{(i)}, \ldots, e_p^{(i)}\},$$

where

$$e_n^{(i)} = \begin{cases} \left(\frac{f_i(n)}{p}\right) & \text{for } (f_i(n), p) = 1, \\ +1 & \text{for } p \mid f_i(n). \end{cases}$$

Then

$$\begin{aligned} C_k(\mathcal{F}_3) &\geq C_k(\{E_p^{(1)}, \dots, E_p^{(k)}\}) \\ &\geq \left|V(\{E_p^{(1)}, \dots, E_p^{(k)}\}, p, (0, p, 2p, \dots, (k-1)p))\right| \\ &= \left|\sum_{n=1}^{p} e_n^{(1)} e_n^{(2)} \dots e_n^{(k)}\right| \end{aligned} \tag{20}$$

Here for $n > k-1$ we have $(p, (n-1)(n-2)\dots, (n-k+1)) = 1$ thus

$$\begin{aligned} e_n^{(1)} e_n^{(2)} \dots e_n^{(k)} &= \prod_{i=1}^{k} \left(\frac{f_i(n)}{p}\right) = \left(\frac{\prod_{i=1}^{n} f_i(n)}{p}\right) \\ &= \left(\frac{((n-1)\dots(n-k+1))^2}{p}\right) = 1. \end{aligned} \tag{21}$$

For $n \leq k-1$

$$\begin{aligned} e_n^{(1)} e_n^{(2)} \dots e_n^{(k)} &= \prod_{\substack{i=1, \\ i \neq n}}^{k-1} e_n^{(i)} e_n^{(n)} e_n^{(k)} \\ &= \prod_{\substack{i=1, \\ i \neq n}}^{k-1} \left(\frac{n-i}{p}\right) \cdot 1 \cdot \prod_{\substack{i=1, \\ i \neq n}}^{k-1} \left(\frac{n-i}{p}\right) = 1 \end{aligned}$$

Thus (21) holds for all $1 \leq n \leq p$. By this and (20)

$$C_k(\mathcal{F}_3) \geq \left|\sum_{n=1}^{p} e_n^{(1)} e_n^{(2)} \dots e_n^{(k)}\right| = p,$$

which was to be proved.

Proof of Theorem 4 Let $\mathcal{H}$ be a set which contains every ℓ-tuple of sequences from $\mathcal{F}_4$ with $0 \le \ell \le k$:

$$\mathcal{H} = \{(E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)}) : \ E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)} \text{ are different and } \in \mathcal{F}_4\}.$$

Then $W_{k,\mathcal{H}}(\mathcal{F}_4) = C_k(\mathcal{F}_4)$. We would like to apply Theorem 5 for this set $\mathcal{H}$. In order to apply Theorem 5 we have to show that if $f_1, f_2, \dots, f_k$ are irreducible polynomials, $a_1, a_2, \dots, a_k \in \mathbf{F_p}$ where $a_t \neq a_s$ if $f_t = f_s$, then the product

$$\prod_{i=1}^{k} f_i(x + a_i)$$

is never of the form $cg(x)^2$ with $c \in \mathbf{F_p}$ and $g(x) \in \mathbf{F_p}[x]$. In Remark 1 we note that amongst the polynomials $f_i(x), f_i(x+1), \dots, f_i(x+p-1)$ only $f_i(x)$ belongs to $\mathcal{F}_4$. Thus the polynomials $f_1(x + a_1), f_2(x + a_2), \dots, f_k(x + a_k)$ are different. Indeed, if

$$f_t(x + a_t) = f_s(x + a_s)$$

$$f_t(x) = f_s(x + a_s - a_t)$$

then $a_s - a_t = 0$, $f_t = f_s$, which is a contradiction. By the unique factorization in $\mathbf{F_p}[x]$, the product of distinct irreducible polynomials is never of the form $cg(x)^2$ with $c \in \mathbf{F_p}$ and $g(x) \in \mathbf{F_p}[x]$. Thus the conditions of Theorem 5 hold. Using Theorem 5 we get the statement.

Proof of Theorem 5 We have

$$W_{k,\mathcal{H}}(\mathcal{F}) = \max_{1 \le \ell \le k, \left(E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)} \right) \in \mathcal{H}} C_k(\{E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)}\}),$$

where the maximum is taken over all $1 \le \ell \le k$ and $\left(E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)}\right) \in \mathcal{H}$, where $E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)}$ are different. Let $f_1(x), f_2(x), \dots, f_\ell(x)$ be the polynomials for which

$$E_p^{(i)} = E_p(f_i(x))$$

is defined by (3) with $f_i(x)$ in place of $f(x)$. $C_\ell(\{E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)}\})$ is defined by the maximum of V's, see (2). Let $\{E_p^{(1)}, E_p^{(2)}, \dots, E_p^{(\ell)}\} = \{e_1, e_2, \dots, e_{\ell p}\}$. We will prove that for $\mathcal{D} = (d_1, d_2, \dots, d_k)$ with non negative integers $d_1 < d_2 < \dots < d_k$, $M \in \mathbb{N}$ we have

$$\left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_k} \right| \le 10Rk^2 2^{k-1} p^{1/2} \log p. \tag{22}$$

We will use the following lemma.

Lemma 2 *Suppose that p is a prime, χ is a non-principal character modulo p of order d, $f \in \mathbf{F}_p[x]$ has s distinct roots in $\overline{\mathbf{F}}_p$, and it is not a constant multiple of a d-th power of a polynomial over $\mathbf{F}_p$. Let y be a real number with $0 < y \le p$. Then for any $x \in \mathbb{R}$:*

$$\left| \sum_{x < n \le x+y} \chi(f(n)) \right| < 9sp^{1/2} \log p.$$

Poof of Lemma 2

This is a trivial consequence of Lemma 1 in [2]. Indeed, there this result is deduced from Weil theorem, see [20].

For each d_i and n define $a_{i,n}$ and $y_{i,n}$ by

$$n + d_i = (y_{i,n} - 1)p + a_{i,n} \tag{23}$$

where $1 \le a_{i,n} \le p$. Then

$$e_{n+d_i} = \begin{cases} \left(\frac{f_{y_{i,n}}(n+d_i)}{p}\right) & \text{for } (f_{y_{i,n}}(n+d_i), p) = 1, \\ +1 & \text{for } p \mid f_{y_{i,n}}(n+d_i). \end{cases} \tag{24}$$

Suppose that we fix any positive integers $j_1 < j_2 < \cdots < j_k$ and we would like to determine the integers $1 \le n \le M$ such that

$$e_{n+d_1} = \begin{cases} \left(\frac{f_{j_1}(n+d_1)}{p}\right) & \text{for } (f_{j_1}(n+d_1), p) = 1, \\ +1 & \text{for } p \mid f_{j_1}(n+d_1), \end{cases}$$

$$e_{n+d_2} = \begin{cases} \left(\frac{f_{j_2}(n+d_2)}{p}\right) & \text{for } (f_{j_2}(n+d_2), p) = 1, \\ +1 & \text{for } p \mid f_{j_2}(n+d_2), \end{cases}$$

$$\vdots$$

$$e_{n+d_\ell} = \begin{cases} \left(\frac{f_{j_\ell}(n+d_\ell)}{p}\right) & \text{for } (f_{j_\ell}(n+d_\ell), p) = 1, \\ +1 & \text{for } p \mid f_{j_\ell}(n+d_\ell). \end{cases}$$

Then by (23) and (24)

$$\begin{aligned} j_1 &= \left[\frac{n+d_1-1}{p}\right] + 1 \\ j_2 &= \left[\frac{n+d_2-1}{p}\right] + 1 \\ &\vdots \\ j_k &= \left[\frac{n+d_k-1}{p}\right] + 1. \end{aligned} \tag{25}$$

Here $1 \le j_i = \left[\frac{n+d_i-1}{p}\right] + 1 \le \left[\frac{\ell p-1}{p}\right] + 1 = \ell \le k$. It is easy to see that the integers n which satisfy (25) is an interval. We will denote this interval by

$I_{j_1,j_2,\ldots,j_k} \subseteq [1,2,\ldots,\ell p]$. In some cases $I_{j_1,j_2,\ldots,j_k}$ is the empty interval. Since $j_1 = \left[\frac{n+d_1}{p}\right] + 1$ we see $|I_{j_1,j_2,\ldots,j_k}| \leq p$. Then by the triangle inequality

$$\begin{aligned}
\left|\sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \cdots e_{n+d_k}\right| \leq & \left|\sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \cdots \sum_{\substack{j_k=1 \\ I_{j_1,j_2,\ldots,j_k} \neq \emptyset}}^{k} \right. \\
& \left. \sum_{\substack{n \in I_{j_1,j_2,\ldots,j_k} \\ p \nmid f_{j_1}(n+d_1) \cdots f_{j_k}(n+d_k)}} \left(\frac{f_{j_1}(n+d_1)}{p}\right) \cdots \left(\frac{f_{j_k}(n+d_k)}{p}\right) \right| \\
& + \left|\sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \cdots \sum_{\substack{j_k=1 \\ I_{j_1,j_2,\ldots,j_k} \neq \emptyset}}^{k} \sum_{\substack{n \in I_{j_1,j_2,\ldots,j_k} \\ p \mid f_{j_1}(n+d_1) \cdots f_{j_k}(n+d_k)}} 1 \right| \\
\leq & \left|\sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \cdots \sum_{\substack{j_k=1 \\ I_{j_1,j_2,\ldots,j_k} \neq \emptyset}}^{k} \right. \\
& \left. \sum_{\substack{n \in I_{j_1,j_2,\ldots,j_k}, \\ p \nmid f_{j_1}(n+d_1) \cdots f_{j_k}(n+d_k)}} \left(\frac{f_{j_1}(n+d_1) \ldots f_{j_k}(n+d_k)}{p}\right) \right| \\
& + \sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \cdots \sum_{\substack{j_k=1 \\ I_{j_1,j_2,\ldots,j_k} \neq \emptyset}}^{k} Rk. \qquad (26)
\end{aligned}$$

By Lemma 2

$$\begin{aligned}
\left|\sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \cdots e_{n+d_k}\right| & \leq \sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \cdots \sum_{\substack{j_k=1 \\ I_{j_1,j_2,\ldots,j_k} \neq \emptyset}}^{k} \left(9Rkp^{1/2}\log p + Rk\right) \\
& \leq \sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \cdots \sum_{\substack{j_k=1 \\ I_{j_1,j_2,\ldots,j_k} \neq \emptyset}}^{k} 10Rkp^{1/2}\log p. \qquad (27)
\end{aligned}$$

It remains to estimate $\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\cdots\sum_{\substack{j_k=1\\ I_{j_1,j_2,\ldots,j_k}\neq\emptyset}}^{k} 1$. It is clear that j_1 may take k different values. Next we study that for fixed j_1 how many different values j_i may assume. For the fixed j_1 we have

$$j_1 = \left[\frac{n+d_1-1}{p}\right]+1.$$

Thus

$$\begin{aligned}
j_1 - 1 \le \frac{n+d_1-1}{p} &< j_1 \\
(j_1-1)p \le n+d_1-1 &< j_1 p \\
(j_1-1)p + d_i - d_1 \le n+d_i-1 &< j_1 p + d_i - d_1 \\
j_1 - 1 + \frac{d_i-d_1}{p} \le \frac{n+d_i-1}{p} &< j_1 + \frac{d_i-d_1}{p} \\
j_1 - 1 + \left[\frac{d_i-d_1}{p}\right] \le \left[\frac{n+d_i-1}{p}\right] &\le j_1 + \left[\frac{d_i-d_1}{p}\right] \\
j_1 + \left[\frac{d_i-d_1}{p}\right] \le \left[\frac{n+d_i-1}{p}\right] + 1 &\le j_1 + 1 + \left[\frac{d_i-d_1}{p}\right] \\
j_1 + \left[\frac{d_i-d_1}{p}\right] \le j_i &\le j_1 + 1 + \left[\frac{d_i-d_1}{p}\right].
\end{aligned}$$

Thus for fixed j_1 each j_i $(2 \le i \le k)$ may assume at most 2 different values. Thus by (27)

$$\left|\sum_{n=1}^{M} e_{n+d_1}e_{n+d_2}\cdots e_{n+d_k}\right| \le 10Rk^2 2^{k-1} p^{1/2}\log p,$$

which completes the proof.

Proof of Theorem 6 Suppose that

$$f_1(x+a_1)f_2(x+a_2)\ldots f_k(x+a_k) = cg(x)^2 = c\prod_{j=1}^{r}(x-\beta_j)^2$$

with $c \in \mathbf{F_p}$ and $g(x) = \prod_{j=1}^{r}(x - \beta_j) \in \mathbf{F_p}[x]$. Then

$$f_1(x - \ell + a_1) \dots f_k(x - \ell + a_k) = cg(x - \ell)^2 \tag{28}$$

for $\ell = 1, 2, \dots, p-1$. By taking the product of equations (28) for $\ell = 0, 1, 2, \dots, p-1$ we get

$$\prod_{\ell=0}^{p-1} f_1(x - \ell + a_1) \cdots \prod_{\ell=0}^{p-1} f_k(x - \ell + a_k) = c^p \prod_{\ell=0}^{p-1} g(x - \ell)^2. \tag{29}$$

Here

$$\prod_{\ell=0}^{p-1} f_i(x - \ell + a_i) = \prod_{\ell=0}^{p-1} f_i(x - \ell) = \prod_{\ell=0}^{p-1} b_i \prod_{j=0}^{r_i} (x - \ell - \alpha_j^{(i)})$$

where b_i is the leading coefficient of $f_i(x)$ and $\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_{r_i}^{(i)}$ denote the roots of $f_i(x)$. By changing the two products we get

$$\begin{aligned}
\prod_{\ell=0}^{p-1} f_i(x - \ell + a_i) &= b_i^p \prod_{j=0}^{r_i} \prod_{\ell=0}^{p-1} (x - \ell - \alpha_j^{(i)}) \\
&= b_i^p \prod_{j=0}^{r_i} \left(x^p - x - \left(\alpha_j^{(i)} \right)^p + \alpha_j^{(i)} \right) \\
&= \tilde{f}_i(x^p - x).
\end{aligned} \tag{30}$$

Let $\beta_1, \beta_2, \dots, \beta_r$ be the roots of $g(x)$, c_r be the leading coefficient of $g(x)$ and let $\tilde{g}(x) = c_r \prod_{k=1}^{r}(x - \beta^p + \beta)$. Similarly to (30) we get

$$\prod_{\ell=0}^{p-1} g(x - \ell) = \tilde{g}(x^p - x). \tag{31}$$

By (29), (30) and (31) we get

$$\prod_{\ell=0}^{p-1} \tilde{f}_\ell(x^p - x) = c^p \tilde{g}(x^p - x)^2. \tag{32}$$

Since (32) also holds in $\overline{\mathbf{F}}_{\mathbf{p}}[x]$, we may substitute $x^p - x = y$ and get

$$\prod_{\ell=0}^{j} \tilde{f}_\ell(y) = c^p \tilde{g}(y)^2,$$

which proves the theorem.

4 Conclusions

In the applications one may need the concatenation or merging of pseudorandom binary sequences. We were looking for criteria to ensure that the concatenation of several sequences belonging to a large family of "good" pseudorandom sequences also possesses strong pseudorandom properties. In Example 1 we showed that the large f-complexity is not enough to ensure this. Thus we introduce a new measure, the *f-correlation* to study the connection between pseudorandom binary sequences. We applied this f-correlation measure to compare Legendre symbol sequences. It turned out that the f-correlation measure can be large even for families of Legendre symbol sequences otherwise possessing very strong pseudorandom properties. However the situation can be saved by selecting suitable smaller subfamily.

References

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

[2] R. Ahlswede, C. Mauduit, A. Sárközy, *Large families of pseudorandom sequences of k symbols and their complexity, Part I, Part II.*, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Verlag, 293-307.

[3] V. Anantharam, *A technique to study the correlation measures of binary sequences*, Discrete Mathematics, to appear.

[4] M. Ben-Or, *Probabilistic algorithms in finite fields*, 22nd Annual Symposium on Foundations of Computer Science (1981), 394-398.

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

[6] E. Galois, *Sur la théorie des nombres*, Écrits et Mémoires Mathématiques d'Évareste Galois, ed. R. Bourgne and J.-P. Arza, 112-128, Gauthier-Villars, 1830.

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

[8] K. Gyarmati, *A note to the paper "On a fast version of a pseudorandom generator"*, Annales Univ. Sci. Budapest. Eötvös, 49 (2006), 143-149.

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

[10] K. Gyarmati, *On a fast version of a pseudorandom generator*, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Springer, Berlin / Heidelberg 2006, 326-342.

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

[12] Joël Rivat, András Sárközy, *Modular construction of pseudorandom binary sequences with composite moduli.* Periodica Mathematica Hungarica 51 (2005), 75-107.

[13] Christian Mauduit, Joël Rivat, András Sárközy, *Construction of pseudorandom binary sequences using additive characters.* Monatshefte für Mathematik 141 (2004), 197-208.

[14] C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequence I: Measures of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997), 365-377.

[15] E. H. Moore, *A doubly-infinite system of simple groups*, Bully. New York Math. Soc. 3 (1893), 73-78.

[16] M. O. Rabin,*Probabilistic algorithms in finite fields*, SIAM J. Comput. 9 (1980), 273-280.

[17] L. Rédei, *Algebra*, Pergamon Press, Oxford-New York-Toronto, Ont. 1967.

[18] V. Shoup, *Fast construction of irreducible polynomials over finite fields*, Journal of Symbolic Computation 17 (1994), 371-391.

[19] V. Tóth, *Collision and avalanche effect in famillies of pseudorandom binary sequences*, Periodica Math. Hungar. 55 (2007), 185-196.

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