Eötvös Loránd University
Faculty of Sciences

Richárd Sebők
Master’s thesis

Pseudorandomness of binary sequences and their generalization

Supervisors: András Sárközy and Katalin Gyarmati
Department of Algebra and Number Theory

Budapest, 2014
Contents

Acknowledgments 3

Introduction 4

1 Pseudorandom binary sequences 6
   1.1 Vernam cypher and One-time pad .......................... 6
   1.2 Pseudorandom bits and sequences .......................... 8

2 Measures of pseudorandomness 10
   2.1 Some measures of pseudorandomness .......................... 10
   2.2 Previous constructions ....................................... 12
      2.2.1 Constructions based on the Legendre symbol .......... 12
      2.2.2 A construction based on the multiplicative inverse ... 14
      2.2.3 A construction based on additive characters .......... 15
      2.2.4 A construction based on the discrete logarithm ....... 16
   2.3 Some properties of Pseudorandom Measures .................. 17
      2.3.1 Connections between Pseudorandom Measures .......... 17
      2.3.2 Typical values of Pseudorandom Measures ............ 24
      2.3.3 Minimal values of Pseudorandom Measures ............ 24
   2.4 The construction of Mauduit and Sárközy .................... 26
      2.4.1 Lemmas on multiplicative characters .................. 27
      2.4.2 Completion of Theorem XIX ......................... 28
      2.4.3 Proof of Theorem II .................................. 29

3 A construction based on prime powers 30
   3.1 The construction ............................................. 30
   3.2 The new results ............................................. 33
   3.3 Proof of Theorem 3.2.1 ..................................... 34
   3.4 Further problems ............................................ 40
Family measures

4.1 Basic definitions ..................................................... 41
4.2 The definition of the cross-correlation measure ................. 43
4.3 Construction using Legendre symbol ............................. 46
4.4 Construction using additive characters ......................... 48

Bibliography .................................................................. 51
Acknowledgments

I would like to thank Professors K. Gyarmati and A. Sárközy who introduced me to the number theory, and whose valuable advices helped me a great way to write this thesis.
Introduction

Pseudorandom binary sequences play a crucial role in cryptography. In particular, they are used as “key” in the most important and widely used stream cipher, the Vernam cipher. The most frequently used classical definition of pseudorandomness uses computational complexity (“next bit test”, “linear complexity”). However, this approach is not quite satisfactory, it has certain weak points. Thus Mauduit and Sárközy proposed a new, more constructive and quantitative approach. First in [33] they introduced some measures for studying pseudorandom (briefly PR) properties of finite binary sequences.

It was also shown in [33] that the Legendre symbol sequence

\[ E_{p-1} = \left( \left( \frac{1}{p} \right) ; \left( \frac{2}{p} \right) ; \ldots ; \left( \frac{p-1}{p} \right) \right) \]  

possesses strong PR properties in this sense (see Theorem II in Section 2.2.1 and Section 2.4). Since that, many “good” PR sequences have been constructed by different authors, a survey of these results is presented by Gyarmati in [14]. Apart from the constructions given in [10], [11], [18] and [40], a common feature of these constructions is that they are of modular nature with a prime modulus \( p \) (so that the length of the sequence is \( p \) or \( p - 1 \)), in other words, finite fields \( \mathbb{F}_p \) of prime order are used. Thus one might look for constructions of different nature, in particular, for modular construction with composite moduli \( m \). Rivat and Sárközy [37] extended the most important modular constructions from prime moduli to “RSA moduli” \( m = pq \), where \( p \) and \( q \) are of the same order of magnitude, and they showed that the sequence obtained in this way have weak PR properties: e.g., the correlation of order 4 is large (see also [28] [29] [30] and [7]). In the case when \( m = uv, (u, v) = 1 \) and both \( u \) and \( v \) are large, probably the situation is similar. The only remaining interesting case is when \( m \) is a prime power \( p^r \) with \( r \geq 2 \); indeed, the four exceptional constructions presented in [10], [11], [18] and [40] are of this type. Recently I generalized a result in [18] from \( r = 2 \) to \( r \geq 2 \), and also sharpened the results in [40], for more details, see Section 3.1.

In the first chapter we will present some definitions and we will study the most important tools of the general, qualitative approach of pseudorandomness.

In the second chapter we will discuss the most important measures of pseudorandomness, the connections between them and some of the frequently used constructions.
usually based on fields of prime order. In Section 2.3.1, I present a new result about two important measures (Theorem 2.3.3).

In the third chapter we will present the construction of Sárközy and Winterhof [40], and later I sharpen it in a special case.

In the fourth chapter we will discuss the foundations of the theory of family measures, and see that how “good” are the families of the most important construction.
Chapter 1

Pseudorandom binary sequences

One of the most important applications of number theory is the simulation of randomness. Most of the computer softwares use a table of random numbers and most of these tables are generated by number theoretical methods.

We will not discuss the basic definitions of cryptography in this thesis, like the definition of alphabets, keys, encryption, decryption, etc.. For a great overview see [36].

Frequently cited in the literature are Kerckhoff’s desiderata, a set of requirements for cipher systems. They are presented here similarly (originally stated in French) as Kerckhoff [25] stated them in 1883:

1. the system should be, if not theoretically unbreakable, unbreakable in practice,
2. compromise of the system details should not inconvenience the correspondents,
3. the key should be rememberable without notes and easily changed,
4. the cryptogram should be transmissible by telegraph,
5. the encryption apparatus should be portable and operable by a single person; and
6. the system should be easy, requiring neither the knowledge of a long list of rules nor mental strain.

Point 2 allows that the encryption transformations being used be publicly known and that the security of the system should reside only in the key chosen.

1.1 Vernam cypher and One-time pad

Definition 1.1.1. The Vernam cipher is defined on the alphabet $\mathcal{A} = \{0, 1\}$. A binary message $m_1m_2\ldots m_t$ is operated on by a binary key string $k_1k_2\ldots k_t$ of the same length.
to produce a ciphertext string \( c_1c_2\ldots c_t \) where
\[
c_i \equiv m_i + k_i \pmod{2} \quad 1 \leq i \leq t.
\]

If the key string is randomly chosen and never used again, the Vernam cipher is called a one-time system or a one-time pad.

If the key string is reused there are ways to attack the system. For example, if \( c_1c_2\ldots c_t \) and \( c'_1c'_2\ldots c'_t \) are two ciphertext strings produced by the same keystream \( k_1k_2\ldots k_t \) then
\[
c_i \equiv m_i + k_i \pmod{2} \quad c'_i \equiv m'_i + k_i \quad (\text{mod } 2)
\]
and
\[
c'_i + c'_i \equiv m_i + m'_i \pmod{2}.
\]
The redundancy in the latter may permit cryptanalysis.

The one-time pad can be shown to be theoretically unbreakable (Shannon [42] proved that the one-time pad has a property he termed perfect secrecy; that is, the ciphertext \( c \) gives absolutely no additional information about the plaintext). That is, if a cryptanalyst has a ciphertext string \( c_1c_2\ldots c_t \) encrypted using a random key string which has been used only once, the cryptanalyst can do no better than guess at the plaintext being any binary string of length \( t \) (i.e., \( t \)-bit binary strings are equally likely as plaintext). It has been proven that to realize an unbreakable system requires a random key of the same length as the message. This reduces the practicality of the system in all but a few specialized situations. Reportedly during the Cold war the communication line between Moscow and Washington was secured by a one-time pad. Transport of the key was done by trusted advisor [36].

If the bits of the key stream are generated by a physical device, which is, the usual case, then the bit sequences obtained in this way must be tested by using statistical tests (testing of this kind are called a posteriori, by Knuth [27]).

While it is impossible to give a mathematical proof that a generator is indeed a random bit generator, the tests presented below help detect certain kinds of weaknesses the generator may have. This is accomplished by taking a sample output sequence of the generator and subjecting it to various statistical tests. Each of the statistical test determines whether or not the sequence has a certain attribute that a truly random sequence would be likely to have; the conclusion of each test is not definite, only probabilistic. An example is that the sequence should have roughly the same number of 0’s as 1’s. If the sequence shows that it have failed any one of the statistical tests, the generator may be rejected as being non-random; alternatively, the generator may be subjected to further testing. On the other hand, if the sequence passes all of the statistical tests, the generator is accepted as being random. More precisely, the term “accepted” should be replaced by “not rejected”, since passing the tests only provides probabilistic evidence
that the generator produces sequences which have characteristics of random sequences [36].

A short list of the most widely used tests:

1. Frequency test (monobit test),
2. Serial test (two-bit test),
3. Poker test,
4. Runs test,
5. Autocorrelation test.

For more details, see [36].

1.2 Pseudorandom bits and sequences

As Menezes, van Oorschot and Vanstone wrote it in [36] “An obvious drawback of the one-time pad is that the key should be as long as the plaintext which increases the difficulty of key distribution and key management. This motivates the design of stream ciphers where the key stream is pseudorandomly generated from a smaller secret key, with the intent that the keystream appears random to a computationally bounded adversary. Such stream ciphers do not offer unconditional security..., but the hope is that they are computationally secure” This motivates us to present the following definitions.

**Definition 1.2.1.** A random bit generator is a device or algorithm which outputs a sequence of statistically independent and unbiased binary digits.

**Definition 1.2.2.** A pseudorandom bit generator (=PRBG) is a deterministic algorithm which, given a truly random binary sequence of length \( k \), outputs a binary sequence of length \( l > k \), which appears to be random. The input to the PRBG is called the seed, while the output of the PRBG is called a pseudorandom bit sequence.

It is important to see that the output of a PRBG is not random; in fact, the number of possible output sequences is at most a small fraction, namely \( 2^k / 2^l \), of all possible binary sequences of length \( l \) [36].

**Definition 1.2.3.** A pseudorandom bit generator is said to pass all polynomial-time (the running time of the test is bounded by a polynomial in the length \( l \) of the output sequence) statistical tests if no polynomial-time algorithm can correctly distinguish between an output sequence of the generator and a truly random sequence of the same length with probability significantly greater than \( \frac{1}{2} \) (considered as \( P \ll \frac{1}{2}(1 + o(1)) \)).
Definition 1.2.4. A pseudorandom bit generator is said to pass the next-bit test if there is no polynomial-time algorithm which, on input of the first \( l \) bits of an output sequences, can predict the \((l + 1)\)st bit of the sequence with probability significantly greater than \( \frac{1}{2} \).

Although Definition 1.2.3 appears to impose a more stringent security requirement on pseudorandom bit generators than Definition 1.2.4 does, the following result of Yao [48] shows that they are, in fact, equivalent.

Theorem I (Universality of the next-bit test). A pseudorandom bit generator passes the next-bit test if and only if it passes all polynomial-time statistical tests.

We note that it is hard to use this test on other than recursive series: what if a generator is accepted by the next-bit test but we can calculate every element of the sequences with knowing the first few and the last bit?

As it is written by Sárközy in [39], these definitions have several weaknesses:

1. “In the practice one has to construct a PR sequence of a give length \( N \); in the definitions above nothing is said on the dependence of the degree and coefficients of the polynomial in the “polynomial time algorithm” on \( N \), thus the definitions cannot be adopted to this situation, they are of asymptotic nature only.”

2. “The non-existence of a polynomial time algorithm has never been shown unconditionally yet; thus there is no PRBG whose cryptographical security has been proved unconditionally. On the other hand, “plausible but unproved mathematical assumptions” can be disputed (even they may turn out to be wrong); e.g., the assumption on the “difficulty of factorization” used to be widely accepted but now, after the discovery of polynomial time primality testing, it becomes more and more disputed.”

3. “These definitions measure only the quality of the PRBG’s but not that of the output sequences so that, even in case of “good” PRBG’s one may get, e.g., an all zero output sequence. Thus to avoid the use of sequences which “are not of random type”, one also needs “a posteriori” testing.”

“Based on all these facts, we may conclude that the notion of pseudorandomness needs another, a quantitative approach, which is more constructive than the one based on complexity theory, it does not use unproved assumptions, it studies pseudorandomness of single sequences instead of generators, and which also makes testing of “a priori” type possible.” In the the rest of the thesis, we will follow this approach.
Chapter 2

Measures of pseudorandomness

2.1 Some measures of pseudorandomness

Mauduit and Sárközy in [33] introduced different pseudorandom measures of finite binary sequences in order to study their pseudorandom (often called as PR) properties.

For a binary sequence $E_N = (e_1, \ldots, e_N) \in \{-1, +1\}^N$ of length $N$ and for $k \in \mathbb{N}$, $M \in \mathbb{N}$, $X = (x_1, \ldots, x_k) \in \{-1, +1\}^k$, write

$$T(E_N, M, X) = |\{n : 0 \leq n < M, (e_{n+1}, e_{n+2}, \ldots, e_{n+k}) = X\}|,$$

then the normality measure of order $k$ of $E_N$ is defined as

$$N_k(E_N) = \max_{X, M} |T(E_N, M, X) - M/2^k|,$$

where the maximum is taken over all $X = (x_1, \ldots, x_k) \in \{-1, +1\}^k$, and $M$ such that $0 < M \leq N - k + 1$.

The normality measure is defined as

$$N(E_N) = \max_{k \leq (\log N)/\log 2} N_k(E_N).$$

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 $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ such that $1 \leq a + b \leq a + bt \leq N$.

The correlation measure of order $k$ of $E_N$ is defined as

$$C_k(E_N) = \max_{M, D} |V(E_N, M, D)| = \max_{M, D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \cdots e_{n+d_k} \right|, \quad (2.1.1)$$
where the maximum is taken over all \( D = (d_1, \ldots, d_k) \) with non-negative integers \( d_1 < \cdots < d_k \) and \( M \in \mathbb{N} \) such that \( M + d_k \leq N \).

The correlation measure is defined as

\[
C(E_N) = \max_{k \leq (\log N)/\log 2} C_k(E_N). \tag{2.1.2}
\]

In [33], the authors considered that a finite binary sequence as a good PR-sequence, if both the well-distribution measure and the correlation measure are small. This is justified by the fact (proved in [6]) that for a random binary sequence \( E_N \in \{-1, +1\}^N \) both these measures are small (as we will see in Section 2.3.2). They proved that to ensure that both these measures are small, it is not enough if we have a good upper bound for only one of them. Mauduit and Sarkozy combined these two measures, and this way they introduced the combined PR-measures, which was used by many mathematicians. For more details, see Katalin Gyarmati’s survey paper [14].

The combined (well-distribution-correlation) PR-measure of order \( k \) of \( E_N \) is defined as

\[
Q_k(E_N) = \max_{a, b, t, D} \left| Z(a, b, t, D) \right| \tag{2.1.3}
\]

\[
= \max_{a, b, t, D} \left| \sum_{j=0}^{t} e_{a+jb+d_1}e_{a+jb+d_2}\cdots e_{a+jb+d_k} \right|
\]

where the maximum is taken over all \( a, b, t, D = (d_1, d_2, \ldots, d_k) \) such that all the subscripts \( a + jb + dl \) belongs to \( \{1, \ldots, N\} \).

The combined PR-measure of \( E_N \) is defined as

\[
Q(E_N) = \max_{k \leq (\log N)/\log 2} Q_k(E_N),
\]

and sometimes we also use the following notation

\[
Q^*_N = \sum_{k=1}^{\infty} Q_k(E_N)/2^k.
\]

The symmetry measure of \( E_N \) was defined in [16] as

\[
S(E_N) = \max_{a < b} \left| \sum_{j=0}^{[\frac{b-a}{2}]-1} e_{a+j}e_{b-j} \right| = \max_{a < b} |H(E_N, a, b)|. \tag{2.1.4}
\]

While it is trivial that a “truly random” sequence cannot be “very symmetric”, most authors just rarely use this measure. In Sections 2.2.1 and in 2.2.4 we will state same result about the symmetric measures of different constructions.
In [33], the authors proposed that "one should be able to estimate this PR-measure at least for certain ‘nice’ sequences." By using the Legendre symbol, they constructed sequence of length \( p - 1 \) and proved that it has good PR properties by a consequence of André Weil’s theorem [46].

In the last few years, many new constructions have been given [14], but most of them are modular ones, so that the lengths of the sequences are produced by this way are of length \( p \) (or \( p - 1 \), where \( p \) is a prime number. One would like to extend this construction to composite moduli, in particular, Rivat and Sárközy [37] studied the case of “RSA moduli”, \( p \cdot q \), where \( p \) and \( q \) are of the same order of magnitude, but it turned out that in these constructions the correlation of order 4 is large (see also [28] [29] [30] and [7]).

Thus one might to look for constructions of different nature, and the only remaining interesting case is when \( m \) is a prime power \( p^r \) with \( r \geq 2 \); indeed, the four exceptional constructions presented in [10], [11], [18] and [40] are of this type.

In this thesis, my goal is to show that the construction of Sárközy and Winterhof [40] (which is a construction based on fields of prime power order and has good pseudorandom properties) can be sharpened in a special case, but here I will prove this only in the case when \( q = p^2 \), because there is an easier way with slightly different proof than the general case \( q = p^r, r \geq 2 \).

2.2 Previous constructions

2.2.1 Constructions based on the Legendre symbol

In [33], using the Legendre symbol Mauduit and Sárközy showed that in the case \( r = 1 \) (i.e. \( N = p - 1 \)).

**Theorem II.** There is a number \( p_0 \) such that if \( p > p_0 \) is a prime number, \( k \in \mathbb{N} \) and writing

\[
E_{p-1} = \left( \left( \frac{1}{p} \right), \left( \frac{2}{p} \right), \ldots, \left( \frac{p-1}{p} \right) \right)
\]

then

\[
Q_k(E_{p-1}) \leq 9kp^{1/2} \log p
\]

so that, writing \( N = p - 1 \),

\[
Q(E_N) = \max_{k \leq (\log N)/\log 2} Q_k(E_N) \leq 27N^{1/2}(\log N)^2
\]

\[
Q^*(E_N) = \sum_{k=1}^{\infty} Q_k(E_N)/2^k \leq 33N^{1/2} \log N.
\]
Mauduit and Sárközy observed that their construction contains a large symmetrical subsequence, so the symmetric measure of their Legendre symbol construction is $p^{-1/2}$, due to the fact

$\left(\frac{p-k}{p}\right) = \left(\frac{-1}{p}\right) \left(\frac{k}{p}\right)$.

In [16], Gyarmati showed the following result:

**Theorem III.** If $p$ is an odd prime, then for the sequence

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

we have

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

In [12], Goubin, Mauduit and Sárközy constructed a large family of binary sequences with good pseudorandom properties.

**Theorem IV.** If $p$ is a prime number, $f(x) \in \mathbb{F}_p[x]$ has degree $k(>0)$ and no multiple zero in $\overline{\mathbb{F}}_p$ (the algebraic closure of $\mathbb{F}_p$), and the binary sequence $E_p = (e_1, \ldots, e_p)$ is defined by

$$e_n = \begin{cases} 
\left(\frac{f(n)}{p}\right) & \text{for } (f(n), p) = 1 \\
+1 & \text{for } p|f(n),
\end{cases} \quad (2.2.4)$$

then we have

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

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

(i) $l = 2$;

(ii) $l < p$, and $2$ is a primitive root modulo $p$;

(iii) $(4k)^l < p$,

then we also have

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

They also presented examples showing that if none of these conditions holds, then $C_l(E_p)$ can be large.

The authors got these conditions by solving some additive number theoretical question with the help of the following two definitions.
Definition 2.2.1. [Property P] If \( m \in \mathbb{N}, \mathcal{A}, \mathcal{B} \subset \mathbb{Z}_m \) (=ring of the modulo \( m \) residue classes), and \( \mathcal{A} + \mathcal{B} \) represents every element of \( \mathbb{Z}_m \) with even multiplicity, i.e., for all \( c \in \mathbb{Z}_m; \) 
\[ a + b = c \quad a \in \mathcal{A}, \ b \in \mathcal{B} \]
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 P.

Definition 2.2.2. [Admissible triple] If \( k, l, m \in \mathbb{N} \) and \( k, l \leq m \), then \((k, l, m)\) is said to be an admissible triple if there are no \( \mathcal{A}, \mathcal{B} \subset \mathbb{Z}_m \) such that \( |\mathcal{A}| = k, |\mathcal{B}| = l \) and \( \mathcal{A} + \mathcal{B} \) possesses property P.

In [12], (ii) of Theorem 1 was about the case when our polynomial got degree \( k \), if \( l \in \mathbb{N} \) is such that the triple \((r, l, p)\) is admissible for all \( r \leq k \), then for the construction presented in (2.2.4) we have
\[ C_l(E_p) < 10klp1/2 \log p. \]

In the rest of the paper they were presenting sufficient criteria for admissibility. Admissibility plays a crucial role in others constructions as well (e.g. the construction of Sárközy and Winterhof, see (3.1.3) in Section 3.1).

In Section 4.3, we will return to investigate this construction (2.2.4) from a different angle.

2.2.2 A construction based on the multiplicative inverse

In [32], Mauduit and Sárközy showed a construction based on the multiplicative inverse, which be calculated fast (in polynomial time), and the pseudorandom properties are only “slightly” worse than in the case (2.2.4).

Theorem V. Assume that \( p \) is a prime number, \( f(x) \in \mathbb{F}_p[x] \) has degree \((0 <) k(< p)\) and no multiple zero in \( \mathbb{F}_p \). For \((a, p) = 1\), denote the multiplicative inverse of \( a \) by \( a^{-1}: \)
\[ aa^{-1} \equiv 1 \pmod{p}. \]

Define the binary sequence \( E_p = (e_1, \ldots, e_p) \) by
\[ e_n = \begin{cases} 
1 & \text{if } (f(n), p) = 1, \ r_p(f(n)^{-1}) < \frac{p}{2}, \\
-1 & \text{if either } (f(n), p) = 1, \ r_p(f(n)^{-1}) > \frac{p}{2}, \text{ or } p \mid f(n) 
\end{cases} \tag{2.2.5} \]
where \( r_p(n) \) denotes the unique \( r \in \{0,1, \ldots, p-1\} \) such that \( n \equiv r \pmod{p} \). Then we have
\[ W(E_p) \ll kp^{1/2}(\log p)^2. \]

Assume also that \( l \in \mathbb{N} \) with \( 2 \leq l \leq p \), and one of the following conditions holds:
1. $l = 2$,
2. $(4k)^l < p$.

Then we also have

$$C_l(E_p) \ll klp^{1/2}(\log p)^{l+1}.$$  

Proof. This is Theorem 1 and Theorem 2 in [32].

2.2.3 A construction based on additive characters

The first constructions were based on multiplicative characters. In [35], Mauduit, Rivat and Sárközy gave a simple construction based on additive characters instead of multiplicative ones.

Let $p$ be an odd prime number, $f(x) \in \mathbb{F}_p[x]$, and define $E_p = (e_1, \ldots, e_p)$ by

$$e_n = \begin{cases} +1 & \text{if } 0 \leq r_p(f(n)) < p/2, \\ -1 & \text{if } p/2 \leq r_p(f(n)) < p, \end{cases} \quad (2.2.6)$$

where $r_p(n)$ denotes the unique $r \in \{0, 1, \ldots, p - 1\}$ such that $n \equiv r \pmod{p}$. This sequence can be computed fast. Moreover, the authors showed that for this sequence both $W(E_n)$ and the correlations of small order are small:

**Theorem VI.** For $f(x) \in \mathbb{F}_p[x]$ of degree $d \geq 2$ and $E_p = (e_1, \ldots, e_p)$ defined by (2.2.6), we have

$$W(E_p) \ll dp^{1/2}(\log p)^2,$$

and for $2 \leq k \leq d - 1$, we have

$$C_k(E_p) \ll dp^{1/2}(\log p)^{k+1},$$

on the other hand for any $k = 2^t$, there exists a constant $c = c(k) > 0$ such that if $p$ is a prime number large enough, $f \in \mathbb{F}_p[x]$ is of degree $k$ and $E_p = (e_1, \ldots, e_p)$ is defined by (2.2.6), then

$$C_k(E_p) \gg p.$$  

Proof. This is Theorem 2 and Theorem 3 and Corollary 1 in [35].

If the control of the correlation of large order is an important aspect in an application, then either we need to use polynomials of large order, which makes the computations much slower, or we have to return to the earlier construction based on multiplicative characters [35].

In Section 4.4, we will show some further pseudorandom properties of this construction.
2.2.4 A construction based on the discrete logarithm

In [15], Gyarmati presented a construction of large family of pseudorandom sequences by using the notion of index (discrete logarithm).

If $p$ is a fixed prime and $g$ is a fixed primitive root modulo $p$, and $(a, p) = 1$, then let $\text{ind } a$ denote the (modulo $p$) index of $a$ (to the base $g$), so that

$$g^{\text{ind } a} \equiv a \pmod{p},$$

and to make the value of index unique, we may add the condition

$$1 \leq \text{ind } a \leq p - 1.$$

Write $N = p - 1$ and define the sequence $E_n = (e_1, \ldots, e_N)$ by

$$e_n = \begin{cases} +1 & \text{if } 1 \leq \text{ind } f(n) \leq (p - 1)/2, \\ -1 & \text{if } (p + 1)/2 \leq \text{ind } f(n) < p - 1. \end{cases} \quad (2.2.7)$$

Then we have

$$W(E_N) < 20N^{1/2} (\log N)^2,$$

and for all $k \in \mathbb{N}, k < p$,

$$C_k(E_N) < 27k8^kN^{1/2} (\log N)^{k+1}.$$

In [15], also a large family of pseudorandom binary sequences were presented similarly as in Section 2.2.1.

Let $p$ be an odd prime, $g$ a primitive root modulo $p$. Let $f(x) \in \mathbb{F}_p[x]$ be a polynomial of the degree $k$. Then define the sequence $E_{p-1} = (e_1, \ldots, e_{p-1})$ by

$$e_n = \begin{cases} +1 & \text{if } 1 \leq \text{ind } f(n) \leq (p - 1)/2, \\ -1 & \text{if } (p + 1)/2 \leq \text{ind } f(n) < p - 1. \end{cases} \quad (2.2.8)$$

**Theorem VII.** For all $f(x) \in \mathbb{F}_p[x]$, if $E_{p-1}$ is gained by (2.2.8), then we have

$$W(E_{p-1}) < 38kp^{1/2} (\log p)^2.$$

Suppose that at least one of the following four conditions holds

(i) $f$ is irreducible,

(ii) If $f$ has the factorization $f = \phi_1^{\alpha_1} \phi_2^{\alpha_2} \ldots \phi_u^{\alpha_u}$, where $\alpha_i \in \mathbb{N}$ and $\phi_i$ is irreducible over $\mathbb{F}_p$, then there exists a $\beta$ such that exactly one or two $\phi_i$’s have degree $\beta$,

(iii) $l = 2$
(iv) \((4l)^k < p\) or \((4k)^l < p\).

Then we have

\[ C_l(E_{p-1}) < 10kl^4p^{l/2}(\log p)^{l+1}. \]

**Proof.** This is Theorems 2 and 3 in [15].

**Theorem VIII.** Let \(f(x) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_0, a_k \not\equiv 0 \pmod{p}, k < p, \) and define \(t\) by

\[ ka_k t \equiv -2a_{k-1} \pmod{p}. \]

Let \(E'_{v-u+1} = (e_u, e_{u+1}, \ldots, e_v)\) be a subsequence of \(E_{p-1}\) defined in (2.2.8). If \(t < 2u\) or \(t > 2v\) or \(f(x) \not\equiv \pm f(t-x)\), then

\[ S(E'_{v-u+1}) < 88kp^{1/2}\log p)^3. \]

**Proof.** This is Theorem 4 in [15].

### 2.3 Some properties of Pseudorandom Measures

#### 2.3.1 Connections between Pseudorandom Measures

One might like to study whether there are some relations between the different PR-measures, or even between the same kind of PR-measures with different order. In this section, we will review some of the often used results of this type. For example, Mauduit and Sárközy [33] showed that the normality measure can be bounded by the maximum of correlation measures.

**Theorem IX.**

\[ N_l(E_N) \leq \max_{1 \leq l \leq d} C_l(E_N). \tag{2.3.1} \]

This result explains the fact that most of the papers do not handle the normality measures separately, they just give a non-trivial upper bound for the well-distribution and the correlation measures.

**Proof.** For all \(k, N \in \mathbb{N}, X = (x_1, \ldots, x_k) \in \{-1, 1\}^k\) and \(1 \leq M \leq N + 1 - k\) we have

\[ \left| T(E_N, M, X) - \frac{M}{2^k} \right| = \]

\[ = \left| \{n : 0 \leq n \leq M, (e_{n+1}, e_{n+2}, \ldots, e_{n+k}) = X\} \right| - \frac{M}{2^k} = \]
\[
\sum_{n=0}^{M-1} \frac{x_1 \ldots x_k}{2^k} \prod_{j=1}^{k} (e_{n+j} + x_k) - \frac{M}{2^k} = \\
\prod_{j \in \{1, \ldots, k\}\setminus\{d_1, \ldots, d_t\}} x_j \sum_{n=0}^{M-1} e_{n+d_1} \ldots e_{n+d_t} \leq \\
\leq \frac{1}{2^k} \sum_{D \subset \{1, 2, \ldots, k\}} |V(E_N, M, D)| \leq \frac{1}{2^k} \sum_{t=1}^{k} \binom{k}{t} C_t = (E_N) \leq \\
\leq \max_{1 \leq t \leq l} C_t(E_N),
\]

which proves the theorem. \qed

It is easy to see that \( C_k(E_N) \leq Q_k(E_N) \) and \( W(E_N) = Q_1(E_N) \), that was the motivation to define the combined measure.

In [34] Mauduit and Sárközy found a connection between the well distribution measure of \( E_N \) and the correlation measure of order 2 of \( E_N \), for every \( E_N \in \{-1, +1\}^N \).

**Theorem X.** For all \( E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N \), we have

\[ W(E_N) \leq 3\sqrt{NC_2(E_N)}. \] (2.3.2)

**Proof.** This result will follow from Theorem 2.3.3. \qed

In order to prove our main result, (3.2.4) of Theorem 3.2.1, we will use Theorem X of Mauduit and Sárközy. (In [13], Gyarmati generalized Theorem X).

Nothing can be said in the opposite direction. It may occur that both normality measure and well distribution measure are very small, but the correlation measure is very large [33].

**Example 2.3.1.** Take a sequence \( E_N = (e_1, \ldots, e_N) \in \{-1, +1\}^N \) such that both the normality measure and the well-distribution measure of it are possibly small, and define \( E'_2N = (e'_1, \ldots, e'_2N) \in \{-1, +1\}^{2N} \) by

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

Then the normality measure and the well-distribution measure if \( E'_2N \) are less than a constant times the corresponding measure of \( E_N \), while

\[
C_2(E'_2N) \geq \left| \sum_{n=1}^{N} e'_n e'_{n+N} \right| = N.
\]
Another example [33] for that not even these measures can characterize the “good” pseudorandom sequences.

**Example 2.3.2.** Consider the sequence \( E_N = (e_1, \ldots, e_N) \in \{-1, 1\}^N \) such that both the correlation measure and the well-distribution measure (or even the combined measure) of it are small, and define \( E'_{2N} = (e'_1, \ldots, e'_{2N}) \in \{-1, 1\}^{2N} \) 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 we can see that the correlation measure and the well-distribution measure of \( E'_{2N} \) are less than a constant times the corresponding measure of \( E_N \), so in the terms of our PR-measures the sequence \( E'_{2N} \) must be considered as a PR-sequence, although it is symmetric, and a “truly random” sequence cannot be as symmetric as \( E'_{2N} \).

Cassaigne, Mauduit and Sárközy in [6] Theorem 4 proved some results for connections between correlation measures of different order.

**Theorem XI.** For \( k \in \mathbb{N} \), \( l \in \mathbb{N}, k|l \), \( N \in \mathbb{N} \) \( E_N \in \{-1, 1\}^N \), we have

\[
C_k(E_N) \leq N^{1 - \frac{k}{l}} \left( C_l^\frac{1}{l} \left( \frac{(l!)}{k!} \right) + \left( \frac{l^2}{2} \right) \right).
\]

In [6], the authors also proved the following in a sharper form.

**Theorem XII.** Suppose that \( 2 \leq k, l \) and \( k \nmid l \). Then there is a sequence \( E_N \in \{-1, 1\}^N \) for which

\[
C_k(E_N) > \frac{N}{k} - 1 - 54k^2 \log N,
\]

and

\[
C_l(E_N) < 27k^2N^{1/2} \log N.
\]

Recently I managed to generalize Theorem X in the following form:

**Theorem 2.3.3.** For all \( E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N \), \( k \in \mathbb{N} \), we have

\[
Q_k(E_N) \leq 2 \sqrt{N \max_{1 \leq i \leq k} C_{2i}(E_N)}.
\] (2.3.3)

**Proof.** Assume that

\[
a, b, t \in \mathbb{N} \quad \text{and} \quad 1 \leq a \leq a + (t - 1)b \leq a + (t - 1)b + d_k \leq N
\] (2.3.4)

Write

\[
e_N = 0 \quad \text{for } n > N.
\]
If $t = 1$, then

$$|Z(a, b, t, D)| = |Z(a, b, 1, D)| = 1 \leq \max_{1 \leq i \leq k} C_{2i} \leq (N \max_{1 \leq i \leq k} C_{2i})^{1/2}. \quad (2.3.5)$$

If $t \geq 2$, then it follows (2.3.4) that

$$b < N$$

and

$$t - 1 \leq (t - 1)b \leq N - a \leq N - 1 \quad (2.3.6)$$

whence

$$t \leq N.$$

Let $g_i = d_i - d_1$. Then $g_1 = 0$, but we will write it down sometimes if it makes things more understandable.

\[
\begin{aligned}
&\sum_{i=a}^{a+b-1} \left( \sum_{j=0}^{t-1} e_{i+jb+d_1} \cdots e_{i+jb+d_k} \right)^2 \\
&= \sum_{i=a}^{a+b-1} \left( \sum_{j_1=0}^{t-1} e_{i+j_1b+d_1} \cdots e_{i+j_1b+d_k} \right) \left( \sum_{j_3=0}^{t-1} e_{i+j_2b+d_1} \cdots e_{i+j_2b+d_k} \right) \\
&= \sum_{i=a}^{a+b-1} \left( \sum_{j_1=0}^{t-1} 1 + 2 \sum_{0 \leq j_1 < j_2 \leq t-1} e_{i+j_1b+d_1} \cdots e_{i+j_1b+d_k} e_{i+j_2b+d_1} \cdots e_{i+j_2b+d_k} \right) \\
&= \sum_{i=a}^{a+b-1} \left( \sum_{j_1=0}^{t-1} 1 + 2 \sum_{0 \leq j_1 < j_2 \leq t-1} e_{i+j_1b+g_1} \cdots e_{i+j_1b+g_k} e_{i+j_2b+g_1} \cdots e_{i+j_2b+g_k} \right) \\
&= \sum_{i=a}^{a+b-1} \left( \sum_{j_1=0}^{t-1} 1 + 2 \sum_{0 \leq j_1 < j_2 \leq t-1} e_{i+j_1b+g_1} \cdots e_{i+j_1b+g_k} e_{i+j_1b+fb+g_1} \cdots e_{i+j_1b+fb+g_k} \right) \\
&= \sum_{j_1=0}^{t-1} \sum_{i=a}^{a+b-1} \sum_{f=1}^{t-1} e_{i+j_1b+g_1} \cdots e_{i+j_1b+g_k} e_{i+j_1b+fb+g_1} \cdots e_{i+j_1b+fb+g_k} \\
&= (t - 1)b + b + 2 \sum_{f=1}^{t-1} \sum_{n=a+d_1}^{a+d_1+(t-f)b-1} e_{n+g_1} \cdots e_{n+g_k} e_{n+fb+g_1} \cdots e_{n+fb+g_k} \\
&< N + 2 \sum_{f=1}^{t-1} \sum_{n=a+d_1}^{a+d_1+(t-f)b-1} e_{n+g_1} \cdots e_{n+g_k} e_{n+fb+g_1} \cdots e_{n+fb+g_k} \quad (2.3.7)
\end{aligned}
\]

By investigating the innermost sum, we need to find the cases when

$$n + g_i = n + fb + g_j,$$
so
\[ fb = g_i - g_j = d_i - d_1 - (d_j - d_1) = d_i - d_j \]

for some \( 1 \leq f \leq t - 1 \).

If there is no such \( i \) and \( j \) that \( fb = d_i - d_j \), then we sum the product of \( 2k \) different elements of the sequence \( E_N \), and by the choice of \( D' = (d'_1, \ldots, d'_{2k}) \), with

\[ 0 \leq d'_1 \leq \ldots \leq d'_{2k} \leq a + (t - 1)b + d'_{2k} \leq N \]

and
\[
\{a + d'_1 + g_1, a + d'_1 + g_2, \ldots, a + d'_1 + g_k, \\
a + d'_1 + fb + g_1, a + d'_1 + fb + g_2, \ldots, a + d'_1 + fb + g_k\} = \{d'_1, \ldots, d'_{2k}\},
\]

then we got that
\[
\sum_{n=0}^{(t-f)b-1} e_{n+a+d'_1+g_1} \ldots e_{n+a+d'_1+g_k} e_{n+a+d'_1+fb+g_1} \ldots e_{n+a+d'_1+fb+g_k} = V(E_N, (t - f)b - 1, D'),
\]

thus
\[
\sum_{n=0}^{(t-f)b-1} e_{n+a+d'_1+g_1} \ldots e_{n+a+d'_1+g_k} e_{n+a+d'_1+fb+g_1} \ldots e_{n+a+d'_1+fb+g_k} \leq (2.3.8)
\]

\[
\leq |V(E_N, (t - f)b - 1, D')| \leq C_{2k}(N).
\]

If there exist some \( i \) and \( j \) that \( fb = d_i - d_j \), then we sum the product of \( 2k \) elements of the sequence \( E_N \) in the innermost sum, but some of them are pairwise equal since their indices are identical. Of course the product of these elements are 1, and \( n + a + d_1 + g_0 \) are smaller than all the other indices, and \( n + a + d_1 + fb + g_k \) are greater than all the other indices, thus at least two elements with those indices will remain.

\[
\sum_{n=0}^{(t-f)b-1} e_{n+a+d'_1+g_1} \ldots e_{n+a+d'_1+g_k} e_{n+a+d'_1+fb} \ldots e_{n+a+d'_1+fb+g_k} =
\]

\[
= \sum_{n=0}^{(t-f)b-1} e_{n+j_1} e_{n+j_2} e_{n+j_3} \ldots e_{n+j_{2t}},
\]

where \( 1 \leq l < k \). With \( D'' = (j_1, j_2, \ldots, j_{2t}) \)

\[
\sum_{n=0}^{(t-f)b-1} e_{n+j_1} e_{n+j_2} \ldots e_{n+j_{2t}} = V(E_N, (t - f)b - 1, D'),
\]

21
thus
\[
\left| \sum_{n=a+1}^{a+d_1+(t-f)b-1} e_{n+j_1}e_{n+j_2} \cdots e_{n+j_{2l}} \right| \leq (2.3.9)
\]
\[
\leq |V(E_N, (t-f)b-1, D'')| \leq C_2l(N).
\]

If we take a look at (2.3.7) again, in
\[
\sum_{f=1}^{t-1} \left| \sum_{n=a+1}^{a+d_1+(t-f)b-1} e_n \cdots e_{n+g_k} e_{n+fb} \cdots e_{n+fb+g_k} \right|
\]
for every \( f \), we can give an upper bound to the innermost sum as
\[
\sum_{n=a+1}^{a+d_1+(t-f)b-1} e_n \cdots e_{n+g_k} e_{n+fb} \cdots e_{n+fb+g_k} \leq C_2l(E_N)
\]
for some \( 1 \leq l \leq k \), which means
\[
\sum_{f=1}^{t-1} \left| \sum_{n=a+1}^{a+d_1+(t-f)b-1} e_n \cdots e_{n+g_k} e_{n+fb} \cdots e_{n+fb+g_k} \right| \leq \sum_{f=1}^{t-1} \max_{1 \leq l \leq k} C_{2l}. \tag{2.3.10}
\]

So by (2.3.10) and (2.3.6), from (2.3.7) we get that
\[
(Z(a, b, t, D))^2 < 2N + 2(t-1) \max_{1 \leq l \leq k} C_{2l}
\]
\[
\leq (2N + 2(t-1)) \max_{1 \leq l \leq k} C_{2l}
\]
\[
< 4N \max_{1 \leq l \leq k} C_{2l}, \tag{2.3.11}
\]
in the case when \( t \geq 2 \), and with (2.3.5), it proves the theorem.

As a consequence of Theorem 2.3.3 and Theorem XI, the following result is also true:

**Corollary 2.3.4.** For all \( E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N \), we have
\[
Q_2(E_N) \leq 5\sqrt{N}C_4(E_N). \tag{2.3.12}
\]

**Proof.** The proof until (2.3.7) is the same with \( k = 2 \), so
\[
\sum_{i=a}^{a+b-1} \left( \sum_{j=0}^{t-1} e_{i+jb+d} e_{i+jb+d} \right)^2 \leq 2N + 2 \sum_{f=1}^{t-1} \left| \sum_{n=a+1}^{a+d_1+(t-f)b-1} e_n e_{n+fb} e_{n+fb+d} \right|,
\]

22
where \(d = d_2 - d_1\) and \(b\) are fixed numbers.

In the case \(b|d\), for a fix \(f\), \(fb = d\), the upper bound for the innermost sum is given by \(C_2(E_N)\), otherwise by \(C_4(E_N)\), and using Theorem XI we got the following result:

\[
C_2(E_N) \leq \sqrt{N} \left( \sqrt{C_4(E_N)} \sqrt{6} + 4 \right) < 7\sqrt{NC_4(E_N)}.
\] (2.3.13)

Thus

\[
(Z(a, b, t, D))^2 = \left( \sum_{j=0}^{t-1} e_{i+jb+d_1} e_{i+jb+d_2} \right)^2 \leq \sum_{i=a}^{a+b-1} \left( \sum_{j=0}^{t-1} e_{i+jb+d_1} e_{i+jb+d_2} \right)^2
\]

\[
< 2N + 2(t - 2)C_4(E_N) + 2C_2(E_N)
\]

\[
\leq 2NC_4(E_N) + 2(N - 2)C_4(E_N) + 2 \cdot 7\sqrt{NC_4(E_N)}
\]

\[
\leq 2NC_4(E_N) + 2(N - 2)C_4(E_N) + 14NC_4(E_N)
\]

\[
< 18NC_4(E_N).
\] (2.3.14)

In the case \(b \nmid d\)

\[
(Z(a, b, t, D))^2 = \left( \sum_{j=0}^{t-1} e_{i+jb+d_1} e_{i+jb+d_2} \right)^2 \leq \sum_{i=a}^{a+b-1} \left( \sum_{j=0}^{t-1} e_{i+jb+d_1} e_{i+jb+d_2} \right)^2
\]

\[
< 2N + 2 \sum_{f=1}^{t-1} C_4(E_N) = 2N + 2(t - 1)C_4(E_N)
\]

\[
\leq 2NC_4(E_N) + 2(N - 1)C_4(E_N)
\]

\[
< 4NC_4(E_N)
\] (2.3.15)

which proves the corollary.

The following corollary will show a connection between a correlation of odd order and some other correlations of even order.

**Corollary 2.3.5.** For all \(E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N\) and for odd \(k\) we have

\[
C_k(E_N) \leq 2 \sqrt{N \max_{1 \leq l \leq k} C_{2l}(E_N)}.
\] (2.3.16)

**Proof.** From (2.3.3) we got that

\[
C_k(E_N) \leq Q_k(E_N) \leq 2 \sqrt{N \max_{1 \leq l \leq k} C_{2l}(E_N)}.
\]
Note that the previous Corollary also true for even $k$, but
\[ C_{2j}(E_N) \leq 2 \sqrt{N \max_{1 \leq l \leq 2j} C_{2l}(E_N)}. \]
holds because
\[ C_{2j}(E_N) \leq 2\sqrt{NC_{2j}(E_N)}. \]
true ($C_{2j}(E_N)$).

2.3.2 Typical values of Pseudorandom Measures

In [6], Cassaigne, Mauduit and Sárközy proved that for the majority of the sequences $E_N \in \{-1, 1\}^N$ the measures $W(E_N)$ and $C_k(E_N)$ are around $N^{1/2}$ (up to some logarithmic factors). Later Alon, Kohayakawa, Mauduit, Moreira and Rödl [3] improved on these bounds.

**Theorem XIII.** Suppose we choose each $E_N \in \{-1, 1\}^N$ with probability $1/2^N$. For all $\varepsilon > 0$ there exist $N_0 = N_0(\varepsilon)$ and $\delta = \delta(\varepsilon)$ such that for $N > N_0$ we have
\[ P \left( \delta\sqrt{N} < W(E_N) < \frac{1}{\delta}\sqrt{N} \right) > 1 - \varepsilon. \]

**Theorem XIV.** Suppose we choose each $E_N \in \{-1, 1\}^N$ with probability $1/2^N$. For all $0 < \varepsilon < 1/16$ there is a constant $N_0 = N_0(\varepsilon)$ such that for $N > N_0$ we have
\[ P \left( \frac{2}{5} \sqrt{N \log \left( \frac{N}{k} \right)} < C_k(E_N) < \frac{7}{4} \sqrt{N \log \left( \frac{N}{k} \right)} \right) > 1 - \varepsilon. \]

In [16], Gyarmati showed that the symmetry measure of $E_N$ can be upper bounded in the sense it was presented in Theorems XIII and XIV.

**Theorem XV.** Suppose we choose each $E_N \in \{-1, 1\}^N$ with probability $1/2^N$. For all $\varepsilon > 0$ there exist $N_0 = N_0(\varepsilon)$ such that for $N > N_0$ we have
\[ P \left( S(E_N) < 4.25\sqrt{N \log N} \right) > 1 - \varepsilon. \]

2.3.3 Minimal values of Pseudorandom Measures

In [3] and [2] the minimal values of pseudorandom measures were also studied. Write
\[ m(N) = \min_{\{-1, 1\}^N} W(E_N), \quad M_k(N) = \min_{\{-1, 1\}^N} C_k(E_N). \quad (2.3.17) \]
The estimate of $m(N)$ is a classical problem. In 1996 Roth [38] proved that $m(n) \gg N^{1/4}$, and upper bounds were given by Sárközy [8] and Beck [5], and finally Matoušek and Spencer [31] proved that $m(N) \ll N^{1/4}$.

The value of $M_k(N)$ depends on the value of the order $k$. Cassaigne, Mauduit and Sárközy [6] showed that

$$M_k(N) \ll \sqrt{kN \log N}.$$  

The results of [3] improved the constant factor (see Theorem XIV).

As the upper bound goes for $M_k(E_N)$, first Cassaigne, Mauduit and Sárközy proved in [6] that

$$M_k(E_N) \gg \log \left(\frac{N}{k}\right), \quad \text{for even } k.$$

Later Alon, Kohayakawa, Mauduit, Moreira and Rödl [2] and [26] improved considerably the lower bound and got

**Theorem XVI.** If $k$ is even then

$$M_k(N) \geq \sqrt{\frac{1}{2} \left[\frac{N}{k+1}\right]}.$$  

The proof of the theorem uses deep linear algebraic tools, and any simplification so far resulted with weaker results (for more details and references, see [14]).

Cassaigne, Mauduit and Sárközy [6] observed that although for the sequence $E_N = (-1, 1, -1, 1, \ldots) \in \{-1, 1\}^N$ we have $C_k(E_N) = 1$ for odd $k$, since

$$e_{n+1+d_1} \ldots e_{n+1+d_k} = (-e_{n+d_1}) \ldots (-e_{n+d_k}) = (-1)^k e_{n+d_1} \ldots e_{n+d_k},$$

thus

$$\left| \sum_{n=1}^{M} e_{n+d_1} \ldots e_{n+d_k} \right| = |1 - 1 + 1 - 1 + \ldots| = \begin{cases} 1 & \text{if } M \text{ is odd}, \\ 0 & \text{if } M \text{ is even}. \end{cases}$$

So $C_k(E_N) = 1$ and thus $M_k(N) = 1$ for odd $k$.

Cassaigne, Mauduit and Sárközy [6] also remarked that although for the sequence $E_N = (-1, 1, -1, 1, \ldots) \in \{-1, 1\}^N$, $C_3(E_N) = 1$, the correlation of order 2 is large: $C_2(E_N) = \left[\frac{N}{2}\right]$.

In [17], Gyarmati proved that $C_2(E_N)C_3(E_N) \gg N^{2/3}$ always holds. Later Anantharam [4] improved this by showing $C_2(E_N)C_3(E_N) \gg N$ also true. By the method of the proofs it became possible to compare correlation measures of odd and even order. Gyarmati and Mauduit proved the following sharp result in [19].

---

25
Theorem XVII. There is a constant $c_{k,l}$ depending only on $k$ and $l$ such that if

$$C_{2k+1}(E_N) < c_{k,l}N^{1/2}$$

then

$$C_{2k+1}(E_N)^2C_{2l}(E_N)^{2k+1} \gg N^{2k+1}$$

where the implied constant factor depends only on $k$ and $l$.

This theorem has the following consequences:

Corollary I. If $C_{2k+1}(E_N) = O(1)$, then $C_{2l}(E_N) \gg N$, where the implied constant factor depends on $k$ and $l$.

Corollary II.

$$C_{2k+1}(E_N)C_{2l}(E_N) \gg N^{c_{k,l}}$$

where the implied constant factor depends only on $k$ and $l$ and where

$$c_{k,l} = \begin{cases} 1 & \text{if } k \geq l, \\ \frac{1}{2} + \frac{2k+1}{4l} & \text{if } k < l. \end{cases}$$

In [16], Gyarmati showed the following result about symmetry measure:

Theorem XVIII. There is an integer $N_0$ such that for $N > N_0$ we have

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

The minimum of the normality measure was studied in [2] and [26] and very recently Aistleither tightened the gap between the upper and the lower bounds.

2.4 The construction of Mauduit and Sárközy

In this section, I will present proof of the result Mauduit and Sárközy [33] presented earlier as Theorem I.

To prove Theorem I, we will need some lemmas and Theorem XIX.

Theorem XIX. Suppose that $p$ is a prime number, $\chi$ is a non-principal character of $\mathbb{F}_p$ of order $d$ (so that $d|p-1$) $f(x) \in \mathbb{F}_p[x]$ ($\mathbb{F}_p$ being the field of modulo $p$ residue classes) has degree $k$ and a factorization $f(x) = b(x-x_1)^{d_1} \cdots (x-x_s)^{d_s}$ in $\mathbb{F}_p$ (the algebraic closure of $\mathbb{F}_p$), where $x_i \neq x_j, (i \neq j)$ with

$$(d,d_1,\ldots,d_s) = 1$$  \hspace{1cm} (2.4.1)

Let $X,Y$ be real numbers with $0 < Y \leq p$. Then

$$\left| \sum_{X < n \leq X+Y} \chi(f(n)) \right| \leq 9kp^{1/2}\log p.$$  \hspace{1cm} (2.4.2)
Corollary 2.4.1. If $p$ is a prime, $f(x) \in \mathbb{F}_p[x]$ is a polynomial of degree $k$ such that it is not of the form $f(x) = b(g(x))^2$, with $b \in \mathbb{F}_p$, $g(x) \in \mathbb{F}_p[x]$ (in other words, in the factorization of $f$ in $\overline{\mathbb{F}}_p$ as in Theorem XIX, there is at least one odd exponent $d_i$), and $X, Y$ are real numbers with $0 < Y \leq q$, then writing

$$\chi^*_{p}(\mathbb{F}_p) = \begin{cases} \left( \frac{n}{p} \right) & \text{for } (n, p) = 1, \\ 0 & \text{for } p|n. \end{cases}$$

we have

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

2.4.1 Lemmas on multiplicative characters

Lemma I. Let $\chi$ be a multiplicative character $\neq \chi_0$ of order $d$ with $d|q - 1$ and $\psi$ an additive character $\neq \psi_0$ of $\mathbb{F}_q$. Let $f(x) \in \mathbb{F}_q[x]$ have precisely $m$ distinct ones among his roots, and let $g(x) \in \mathbb{F}_q[x]$ have degree $n$. Suppose that either

1. $(d, \deg f) = (n, q) = 1$, or
2. $y^d - f(x)$ and $z^q - z - g(x)$ are absolutely irreducible.

Then

$$\left| \sum_{x \in \mathbb{F}_q} \chi(f(x))\psi(g(x)) \right| \leq (m + n - 1)q^{1/2}$$

(2.4.3)

Proof. This is Theorem 2G in Chapter 2 of [41], and a consequence of Andre Weil's famous theorem on curves over finite fields [46].

Lemma II. Suppose the polynomial $y^d - f(x)$ has coefficients in a field $K$. Then the following conditions are equivalent:

(i) $y^d - f(x)$ is absolutely irreducible

(ii) if $f(x) = a(x-x_1)^{d_1} \cdots (x-x_s)^{d_s}$ is the factorization of $f$ in $\overline{K}$, with $x_i \neq x_j, (i \neq j)$, then $(d, d_1, \ldots, d_s) = 1.$

Proof. This is Lemma 2C in Chapter 1 of [41].

Lemma III. If $p$, $\chi$, $d$, $f(x)$ and $k$ is defined as in Theorem XIX and $a \in \mathbb{Z}$, then

$$\left| \sum_{x \in \mathbb{F}_q} \chi(f(x))e(ax/q) \right| \leq kq^{1/2}.$$
Proof. In Theorem XIX, (2.4.1) is assumed, (ii) of Lemma II holds with $K = \mathbb{F}_p$, so by Lemma II, $y^d - f(x)$ is absolutely irreducible.

Next, we apply Lemma II with $\mathbb{F}_p$, $x$, and $z^p - z$ in place of $K$, $y^d$ and $f(x)$, respectively. Since now $d = 1$, (ii) of Lemma II holds so that by Lemma II the polynomial $x - (z^p - z)$, and thus also it’s negative, $z^p - z - x$ is absolutely irreducible.

Now we can apply Lemma I. with $g(x) = x$, so $g(x)$ has a degree of 1. Since $s \leq k$, (2.4.4) follows from (2.4.3).

We will use Lemma 4 from [33] without any modification, which is based on the Vinogradov principle generalized by Erdős and Turán. Here we state this result in the form presented by Friedlander and Iwaniec [9], where the authors wrote “In this form (2.4.5) follows for instance from two application of (3.4) of [24].

**Lemma IV.** If $m \in \mathbb{N}$, the function $g(x) : \mathbb{Z} \to \mathbb{C}$ is periodic with period $m$, and $X, Y$ are real number with $Y > 0$ then

$$\left| \sum_{X < n \leq X+Y} g(n) \right| \leq \frac{Y + 1}{m} \left| \sum_{n=1}^{m} g(n) \right| + \sum_{1 \leq |h| \leq m/2} |h|^{-1} \left| \sum_{n=1}^{m} g(n) e \left( \frac{hn}{m} \right) \right|. \quad (2.4.5)$$

### 2.4.2 Completion of Theorem XIX

First, we will use Lemma IV with $p$ and $\chi(f(n))$ in place of $m$ and $g(n)$, and then we apply Lemma III, to get

$$\left| \sum_{X < n \leq X+Y} \chi(f(n)) \right| \leq \frac{Y + 1}{p} \left| \sum_{n=1}^{p} \chi(f(n)) \right| + \sum_{1 \leq |h| \leq p/2} |h|^{-1} \left| \sum_{n=1}^{p} \chi(f(n)) e \left( \frac{hn}{p} \right) \right| \leq 2kp^{1/2} + 2 \sum_{1 \leq k \leq p/2} h^{-1} kp^{1/2} \leq 2kp^{1/2}(1 + (1 + \log(p/2))) < 2kp^{1/2}(2 + \log p) \leq 2kp^{1/2} \left( \frac{2 \log p}{\log 2} + \log p \right) < 9kp^{1/2} \log p.$$

Which proves Theorem XIX.

Now we proceed with the proof of Corollary 2.4.1.

With the choice of $\chi(a) = \chi^*_{\mathbb{F}_p}(a)$ in Theorem XIX, we have $d = 2$, so that (2.4.1) holds if and only if one of the exponents $d_1, \ldots, d_s$ is odd in the factorization of $f$ in $\mathbb{F}_p$, so $f(x)$ is not of the form $f(x) = b(g(x))^2$. 28
2.4.3 Proof of Theorem II

Defining $Z(a, b, t, D)$ as in (2.1.3) with $e_n = \binom{n}{p}$, for $k < p$, we have

$$|Z(a, b, t, D)| = \left| \sum_{n=0}^{t} \left( \frac{a + nb + d_1}{p} \right) \left( \frac{a + nb + d_2}{p} \right) \cdots \left( \frac{a + nb + d_k}{p} \right) \right|$$  \hspace{1cm} (2.4.6)

for all $a, b, t, D = (d_1, d_2, \ldots, d_k)$ such that

$$a + nb + d_l \in \{1, \ldots, q - 1\} \text{ for } n = 0, 1, \ldots, t \text{ and } l = 1, \ldots, k.$$  \hspace{1cm} (2.4.7)

We may assume that $(b, p) = 1$, otherwise

$$|Z(a, b, t, D)| = t \cdot \left| \prod_{i=1}^{k} \left( \frac{a + d_i}{p} \right) \right| = t.$$

Then let be an integer with $\bar{b}b \equiv 1 \pmod{p}$, and for $j = 1, \ldots, k$, let $h_j$ denote an integer with

$$h_j \equiv (a + d_j)\bar{b} \pmod{p}$$

so that

$$h_i \not\equiv h_j \pmod{p} \text{ for } 1 \leq i < j \leq k.$$  \hspace{1cm} (2.4.8)

Take $f(n) = (n + h_1)(n + h_2) \ldots (n + h_k)$. Then it follows from (2.4.6)

$$|Z(a, b, t, D)| = \left| \sum_{n=0}^{t} \left( \frac{\bar{a}b + n + d_1\bar{b}}{p} \right) \left( \frac{\bar{a}b + n + d_2\bar{b}}{p} \right) \cdots \left( \frac{\bar{a}b + n + d_k\bar{b}}{p} \right) \right|$$

$$= \left| \sum_{n=0}^{t} \left( \frac{n + h_1}{p} \right) \left( \frac{n + h_2}{p} \right) \cdots \left( \frac{n + h_k}{p} \right) \right|$$

$$= \left| \sum_{n=0}^{t} \left( \frac{f(n)}{p} \right) \right| = \left| \sum_{n=0}^{t} \chi_p^*(f(n)) \right|$$

with the character $\chi_p^*$ defined in Corollary 2.4.1.

Writing $X = -1, Y = t+1$, clearly we may assume that $0 < Y = t+1 \leq N+1 = p$. Moreover, since $f(x)$ has no multiple zero by (3.3.1), Corollary 2.4.1 can be applied. We obtain

$$|Z(a, b, t, D)| < 9kp^{1/2} \log q$$

which proves (2.2.1) of Theorem 1. (2.2.2) and (2.2.3) follows with more computation.
Chapter 3

A construction based on prime powers

3.1 The construction

To present the construction of Sárközy and Winterhof based on prime power fields, we will also need the multidimensional extension of the theory of pseudorandomness of binary sequences. In [23] Hubert, Mauduit and Sárközy introduced the following definitions:

Denote by $I^n_N$ the set of $n$-dimensional vectors whose coordinates are integers between 0 and $N - 1$:

$$I^n_N = \{x = (x_1, \ldots, x_n) : x_i \in \{0, 1, \ldots, N - 1\}\}.$$  

This set is called an $n$-dimensional $N$-lattice or briefly an $N$-lattice. They extended the definition of binary sequences to more dimensions by considering functions of type $\eta : I^n_N \to \{-1, +1\}$, called binary lattice.

If $x = (x_1, \ldots, x_n)$ so that $\eta(x) = \eta((x_1, \ldots, x_n))$ then we will simplify the notation by writing $\eta(x) = \eta(x_1, \ldots, x_n)$.

Let $u_1, u_2, \ldots, u_n$ be $n$ linearly independent $n$-dimensional vectors over the fields of the real numbers such that the $i$-th coordinate of $u_i$ is a positive integer and the other coordinates of $u_i$ are 0, so that, writing $z_i = |u_i|$, $u_i$ is of the form $(0, \ldots, 0, z_i, 0, \ldots, 0)$. Let $t_1, t_2, \ldots, t_n$ be integers with $0 \leq t_1, t_2, \ldots, t_n < N$. Then we call the set

$$B^*_N = \{x = x_1 u_1 + \cdots + x_n u_n : 0 \leq x_i z_i \leq t_i (< N) \text{ for } i = 1, \ldots, n\}$$

$n$-dimensional box $N$-lattice or briefly a box $N$-lattice.

Later Gyarmati, Mauduit and Sárközy [20] introduced the following measure of pseudorandomness of binary lattices: The correlation measure of order $l$ of the lattice
\[ \eta : I_N^n \to \{-1, +1\} \text{ is defined by} \]
\[ C_k(\eta) = \max_{B', d_1, \ldots, d_k} \left| \sum_{x \in B'} \eta(x + d_1) \ldots \eta(x + d_k) \right|, \]
where the maximum is taken over all distinct \( d_1, \ldots, d_k \in I_N^n \) and all box lattices \( B' \) of the special form
\[ B' = \{ x = (x_1, \ldots, x_n) : 0 \leq x_1 \leq t_1(<N), \ldots, 0 \leq x_n \leq t_n(<N) \} \]
such that \( B' + d_1, \ldots, B' + d_k \in I_N^n \).

In [23], in Theorem 2 the authors presented a construction based on the quadratic character of \( \mathbb{F}_q^\times \), so that for \( a \in \mathbb{F}_q^\times \) we have \( \gamma(a) = +1 \) if \( a \) is a quadratic residue and \( \gamma(a) = -1 \) otherwise.

Considered \( \mathbb{F}_q \) as an \( r \) dimensional vector space over \( \mathbb{F}_p \), and let \( v_1, v_2, \ldots, v_r \) be a basis of this vector space. Then \( \eta(x) : I_p^r \to \{-1, +1\} \) is defined by
\[ \eta(x) = \begin{cases} \gamma(x_1v_1 + \cdots + x_nv_n) & \text{for } (x_1, \ldots, x_n) \neq (0, \ldots, 0) \\ 1 & \text{for } (x_1, \ldots, x_n) = (0, \ldots, 0) \end{cases} \tag{3.1.1} \]
for any \( x_1, \ldots, x_n \in \mathbb{F}_p \), and \( \eta(x) = \eta(x_1, \ldots, x_n) \).

**Theorem XX (Mauduit and Sárközy).** If \( p \) is a prime, \( n \in \mathbb{N} \), \( k \in \mathbb{N} \), and the \( n \)-dimensional binary \( p \)-lattice is defined by (3.1.1), then we have
\[ C_k(\eta) < kq^{1/2}(1 + \log p)^n. \tag{3.1.2} \]

By using finite fields \( \mathbb{F}_q \) of order \( q = p^r \) (with any \( r \in \mathbb{N} \)), in [40] Sárközy and Winterhof constructed binary sequences of length \( p^r \) with strong pseudorandom properties. Take the following ordering of the elements of \( \mathbb{F}_q \): as \( \mathbb{F}_q \) is an \( r \)-dimensional vector space over \( \mathbb{F}_p \), consider any basis \( \theta_0, \theta_1, \ldots, \theta_{r-1} \) of this vector space. Take the unique representation of any integer \( n \in \{0, 1, \ldots, q-1\} \) in the number system of base \( p \):
\[ n = \varepsilon_0 + \varepsilon_1p + \varepsilon_2p^2 + \cdots + \varepsilon_{r-1}p^{r-1}, \]
with \( 0 \leq \varepsilon_i < p \) for \( 0 \leq i \leq r-1 \). We assign the element
\[ \xi_n = \varepsilon_0\theta_0 + \varepsilon_1\theta_1 + \varepsilon_2\theta_2 + \cdots + \varepsilon_{r-1}\theta_{r-1} \]
of \( \mathbb{F}_q \) to the integer \( n \). Then for a polynomial \( f_q(x) \in \mathbb{F}_q[x] \) of degree \( l > 0 \) with no multiple zero in \( \mathbb{F}_q^\times \) define the binary sequence \( L_q = (l_0, l_1, \ldots, l_{q-1}) \in \{-1, +1\}^q \) by
\[ l_n = \begin{cases} \gamma(f(\xi_n)) & \text{if } f(\xi_n) \neq 0, \\ 1 & \text{if } f(\xi_n) = 0 \end{cases} \tag{3.1.3} \]
for \( n = 0, 1, \ldots, q-1 \).
Theorem XXI (Sárközy and Winterhof). For $r \geq 2$ we have

(i) $W(L_q) < c_13^{r-1}rl^{1/r}p^{r-1/2}$

and if $r \geq 3$ and $l < p$ then this can be improved to

$W(L_q) < c_22^{3r/2}l^{1/2}q^{3/4}(\log p)^{r/2}$, and

(ii) if either $k = 2$ and $l < p$, or $4^{r(k+1)} < p$, then

$C_k(L_q) < c_32^{(r-1)k}r2^klq^{1/2}(\log p)^r$.

where $c_1, c_2, c_3$ are positive absolute constants.

Note that the original proof of this result uses the definition of admissible triples (see Definition 2.2.2), and that is the reason we have a slight restriction ($4^{r(k+1)} < p$) on the cases when we can give good upper bound on the correlation measure.

Observe that in the $q = p^2$ (i.e., $r = 2$) special case Theorem XXI gives in the most important special case $f(x) = x$ that

Corollary 3.1.1. If $q = p^2$ and the sequence $L_q = L_{p^2} = (l_0, l_1, \ldots, l_{p^2-1}) \in \{-1, +1\}^{p^2}$ is defined by $l_n = \gamma(\xi_n)$ for $\xi_n \neq 0$ and $l_0 = 1$, then we have

(i) $W(L_{p^2}) < 6c_1p^{3/2} = 6c_1q^{3/4}$ for $r \geq 3$, and

(ii) if either $k = 2$, or $4^{r(k+1)} < p$, then

$C_k(L_{p^2}) < 8c_3k2^klq^{1/2}(\log p)^2$.

In [18] Gyarmati studied the $q = p^2$ special case again. First she proved that if $\eta$ is a two dimensional $N$-lattice with small correlation measures, then one can prepare a binary sequence of length $N^2$ from it which is also has small correlation measure. Indeed, define $E_{N^2}(\eta) \in \{-1, +1\}^{N^2}$ so that we take the first (from the bottom) row of the lattice then the second row of the lattice, etc.:

$e_{i,N+j} = \eta((j - 1, i))$ for $i = 0, 1, \ldots, N - 1$, $j = 1, 2, \ldots, N$.

She proved:

Theorem XXII (Gyarmati). For any two dimensional binary $N$-lattice $\eta$ for $1 < l < N$ we have

$C_l(E_{N^2}(\eta)) \leq (l + 2)C_l(\eta)$.  \hfill (3.1.4)
Now we apply this theorem in the special case when $N = p$ is a prime, and $\eta$ is the binary $p$-lattice defined in (3.1.1) with $n = 2$, and $\theta_0, \theta_1$ in place of $v_1, v_2$. Then $E_{N^2}$ coincides with the sequence $L_{\mu^2}$ occurring in Corollary 3.1.1, thus combining Theorems XX and XXI we get

$$C_k(L_{\mu^2}) = C_k(E_{N^2}(\eta)) \leq (k + 2)C_k(\eta) <$$

$$(k + 2)kq^{1/2}(1 + \log p)^2 < c_4k^2q^{1/2}(\log p)^2.$$  

Observe that this improves on the upper bound in Corollary 3.1.1, (ii) considerably: apart from an absolute constant factor this bound is smaller in terms of $k$ by an exponential factor $2^k$.

At the end of this subsection we remark that the generalisation of Theorem XXII is still an interesting open question, one would like to know the relations between the correlations of higher dimensional lattices and the correlation of the corresponding sequences. In [21], Gyarmati, Mauduit and Sárközy gave the answer to the opposite question in 2 dimension, they showed that it may happen that the pseudorandom measures of the sequence $E_{N^2}(\eta)$ are small, however, the corresponding pseudorandom measures of the lattice $\eta$ are large.

### 3.2 The new results

In the rest of the paper we will restrict ourselves to the most important special case $f(x) = x$ of the construction (3.1.3), so that now $q = p^r$ and the sequence $L_q = (l_0, l_1, \ldots, l_{q-1}) \in \{-1, 1\}^q$ is defined by

$$l_n = \begin{cases} 
\gamma(\xi_n) & \text{for } n = 1, 2, \ldots, q - 1, \\
1 & \text{for } n = 0. 
\end{cases}$$  

**Theorem 3.2.1.** If $p$ is a prime number and $q = p^r$, $k \in \mathbb{N}$, $k < q$, then

$$C_k(L_q) \leq (k + 1)^{r-2}(r + k)kq^{1/2}(1 + \log p)^r,$$  

and

$$Q_k(L_q) \leq 2(2k + 1)^{(r-2)/2} \sqrt{(r + 2k)(2k)} q^{3/4}(1 + \log p)^{r/2}.$$  

As an immediate consequence of this theorem, by definition

$$W(L_q) \leq 2 \cdot 3^{(r-2)/2} \sqrt{(r + 2)2} q^{3/4}(1 + \log p)^{r/2}$$  

holds.

Thus for any choice of the basis $\theta_0, \theta_1, \ldots, \theta_{r-1}$, the sequence $L_q$ generated by this construction possesses good PR properties. In this way we may generate many “good”
binary sequences of length $q$. Two different bases may generate the same sequence $L_q$, however, it is easy to see that counting only the distinct sequences, still we got many (more than $cq^r$) sequences.

Note that the proof of Theorem 3.2.1 does not use the definition of admissible triples (see Definition 2.2.2) and therefore we do not have a kind of restriction like the conditions of (ii) in Theorem XXI.

Also note that in the special case $f(x) = x$ Theorem XXI gives

$$W(L_q) \leq c_2 2^{3r/2} r^{3/2} q^{3/4} (\log p)^{r/2} \quad \text{(for } r \geq 3)$$

(3.2.5)

and if either $k = 2$ or $4^{r(k+1)} < p$, then

$$C_k(L_q) \leq c_3 2^{(r-1)k} 2^r k q^{1/2} (\log p)^r.$$  

(3.2.6)

Comparing Theorem 3.2.1 with these results we find that, apart from a constant factor, in (3.2.4) the factor $2 \cdot 3^{(r-2)/2}(1 + \log p)^{r/2}$ replaces the greater factor $2^{3r/2}(\log p)^{r/2}$ and in (3.2.2) the factor $(k + 1)^{r-2}(r + k)(1 + \log p)^r$ replaces the much greater factor $2^{(r-1)k} 2^r (\log p)^r$. Thus, our Theorem 3.2.1 improves considerably on the $f(x) = x$ special case of Theorem XXI.

Moreover, in the special case $r = 2$ it follows from (3.2.2) in our Theorem 3.2.1 that

$$C_k(L_q) < c_3 k^2 q^{1/2} (\log p)^2$$

which, apart from the constant factor, is the same upper bound as the one in (3.1.5). Thus, indeed, our Theorem 3.2.1 generalizes the upper bound (3.1.5) which can be obtained by Gyarmati's approach in [18].

### 3.3 Proof of Theorem 3.2.1

First, we will prove (3.2.3), with the assumption that (3.2.2) holds. Then, by Theorem 2.3.3, we got

$$Q_k(L_q) \leq 2 \sqrt{NC_{2k}(L_q)} \leq 2 \sqrt{(2k + 1)^{r-2}(r + 2k)2k q^{1/2}(1 + \log p)^r q^{1/2} =}$$

$$= 2(2k + 1)^{(r-2)/2} \sqrt{(r + 2k)(2k)} q^{3/4}(1 + \log p)^{r/2},$$

which proves (3.2.3)

Next we prove (3.2.2), here only in the case of $q = p^2$.  

34
Now, with $L_q = (l_0, l_1, l_2, \ldots, l_{q-1})$, where $l_n = \gamma(\xi_n)$ for $1 \leq n \leq p^r - 1$ and $l_0 = 1$ and by the definition of the correlation measure in (2.1.1), we get

$$|V(L_q, M, D)| = \left| \sum_{n=0}^{M} l_{n+d_1}l_{n+d_2} \ldots l_{n+d_k} \right|$$

$$= \left| \sum_{n=0}^{M} \gamma(\xi_{n+d_1})\gamma(\xi_{n+d_2}) \ldots \gamma(\xi_{n+d_k}) \right|$$

**Lemma 3.3.1.** For $j = 1, \ldots, k$ let $\omega_j(n)$ denote the function

$$\omega_j(n) = \xi_{n+d_j} - \xi_n$$

where $n + d_j \leq M + d_k \leq N$. Then

$$\omega_i(n) \neq \omega_j(n) \quad \text{for } 1 \leq i < j \leq k \quad (3.3.1)$$

for all $n$, where $0 \leq n \leq M$.

**Proof.** Let $n = a_0 + a_1 \cdot p$ and $d_j = d_{j,0} + d_{j,1} \cdot p$, where

$$a_0, a_1, d_{j,0}, d_{j,1} \in \{0, 1, 2, \ldots, p - 1\}.$$

If $a_0 + d_{j,0} < p$, then $n + d_j = a_0 + d_{j,0} + (a_1 + d_{j,1}) \cdot p$, thus

$$\xi_{n+d_j} = (a_0 + d_{j,0}) \cdot \theta_0 + (a_1 + d_{j,1}) \cdot \theta_1,$$

on the other hand, if $n_0 + d_{j,0} \geq p$, then $n + d_j = a_0 + d_{j,0} - p + (a_1 + d_{j,1} + 1) \cdot p$, thus

$$\xi_{n+d_j} = (a_0 + d_{j,0} - p)\theta_0 + (a_1 + d_{j,1} + 1) \cdot \theta_1,$$

where $0 \leq a_0 + d_{j,0} - p < p$, henceforth

$$\omega_i(n) = \xi_{n+d_j} - \xi_n = \begin{cases} 
  d_{j,0} \cdot \theta_0 + d_{j,1} \cdot \theta_1, & \text{if } a_0 + d_{j,0} < p \\
  d_{j,0} \cdot \theta_0 + (d_{j,1} + 1) \cdot \theta_1, & \text{if } a_0 + d_{j,0} \geq p.
\end{cases} \quad (3.3.2)$$

We need to consider that in the case $n_0 + d_{j,0} \geq p$, $0 \leq a_0 + d_{j,0} - p < p$ true, but the first coordinate of $\xi_{n+d_j} - \xi_n$ is

$$a_0 + d_{j,0} - p - a_0 = d_{j,0} - p \equiv d_{j,0} \pmod{p},$$

because the characteristic of the field is $p$.

From now on, the proof is by contradiction, we assume that there exist an $n$, such that $\omega_i(n) = \omega_j(n)$ for some $i < j$ indices.
If \( d_{i,0} \cdot \theta_0 + d_{i,1} \cdot \theta_1 = d_{j,0} \cdot \theta_0 + d_{j,1} \cdot \theta_1 \), or \( d_{i,0} \cdot \theta_0 + (d_{i,1} + 1) \cdot \theta_1 = d_{j,0} \cdot \theta_0 + (d_{j,1} + 1) \cdot \theta_1 \), then we will get that \( d_i = d_j \) which implies \( i = j \) which contradicts our assumption.

So let us suppose \( d_{i,0} \cdot \theta_0 + d_{i,1} \cdot \theta_1 = d_{j,0} \cdot \theta_0 + (d_{j,1} + 1) \cdot \theta_1 \). This implies \( d_{i,0} = d_{j,0} \) and \( d_{i,1} = d_{j,1} + 1 \), but these two implies \( d_i > d_j \), which contradicts that \( i < j \).

The only case left is when \( d_{i,0} \cdot \theta_0 + (d_{i,1} + 1) \cdot \theta_1 = d_{j,0} \cdot \theta_0 + (d_{j,1}) \cdot \theta_1 \), so \( d_{i,0} = d_{j,0} \) and \( d_{i,1} + 1 = d_{j,1} \), which gives us

\[
d_i + p = d_{i,0} + (d_{i,1} + 1) \cdot p = d_{j,0} + d_{j,1} \cdot p = d_j,
\]

but then, \( p \leq n_0 + d_{i,0} = n_0 + d_{j,0} < p \), which is a contradiction, and that proves the lemma.

Take \( f(\xi_n) = (\xi_n + \omega_1(n))(\xi_n + \omega_2(n)) \ldots (\xi_n + \omega_k(n)) \). Then

\[
|V(L_q, M, D)| = \left| \sum_{n=0}^{M} \gamma(\xi_{n+d_1}) \gamma(\xi_{n+d_2}) \ldots \gamma(\xi_{n+d_k}) \right| = \sum_{n=0}^{M} \gamma(\xi_n + \omega_1(n)) \gamma(\xi_n + \omega_2(n)) \ldots \gamma(\xi_n + \omega_k(n)) = \sum_{n=0}^{M} \gamma((\xi_n + \omega_1(n))(\xi_n + \omega_2(n)) \ldots (\xi_n + \omega_k(n))).
\]

We should take a partition of \( \bigcup_i I_i = I = \{\xi_1, \xi_2, \ldots, \xi_M\} \), such that for all \( i \), all \( \xi_n \in I_i \), \( \omega_1(n), \omega_2(n), \ldots, \omega_k(n) \) will be constant. We saw that in (3.3.2), that \( \omega_j(n) \) depends only on \( a_0 + d_{j,0} \). So for all \( j \in \{1, 2, \ldots, k\} \), \( n, n' \) such that \( n = a_0 + a_1 \cdot p \), \( n' = a_0' + a_1' \cdot p \), and \( a_0 = a_0' \), then \( \omega_j(n) = \omega_j(n') \). Thus, if \( \xi_n \in I_i \), then \( \xi_{n+t_1} \in I_i \), if \( n + l \cdot p \leq M \).

The partition of \( I \) only depends on \( d_1, d_2, \ldots, d_k \), more precisely on \( d_{1,0}, d_{2,0}, \ldots, d_{k,0} \). Note that we do not know the relations between them. By switching the indices, we get

\[
d_{i_1,0} \leq d_{i_2,0} \leq \cdots \leq d_{i_k,0}.
\]

We only need to pay attention to the different ones, so assume there is \( u \) different ones of them, \( 1 \leq u \leq k \), and from now on, we will work with

\[
d_{i_1,0} < d_{i_2,0} < \cdots < d_{i_u,0}.
\]

Now we define precisely \( \bigcup_{j=0}^{u} I_j = I = \{\xi_1, \xi_2, \ldots, \xi_M\} \), such that for \( j = \{1, \ldots, u-1\} \)

\[
I_j = \{\xi_n : n = a_0 + a_1 \cdot p, a_0 + d_{i_j,0} < p \text{ and } a_0 + d_{i_{j+1},0} \geq p\}
\]

and

\[
I_0 = \{\xi_n : n = a_0 + a_1 \cdot p, a_0 + d_{i_1,0} \geq p\},
\]

36
\[ I_u = \{ \xi_n : n = a_0 + a_1 \cdot p, a_0 + d_n, 0 < p \}. \]

Now for every \( j \), every \( \omega_i(n) \) is a constant, if \( \xi_n \in I_j \) so \((\xi_n + \omega_1(n))(\xi_n + \omega_2(n)) \ldots (\xi_n + \omega_k(n))\) is a one variable polynomial of degree \( k \), for \( I_j \) we will denote it as \( f_j \).

\[
\left| \sum_{n=0}^{M} \gamma((\xi_n + \omega_1(n))(\xi_n + \omega_2(n)) \ldots (\xi_n + \omega_k(n))) \right| = \left| \sum_{j=0}^{u} \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right|. \tag{3.3.4}
\]

To give a good estimate for \( \left| \sum_{j=0}^{u} \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right| \), we will use a result of A. Winterhof:

**Theorem XXIII (Winterhof).** Let \( p \) be an odd prime, \( n \in \mathbb{N}, q = p^n \) and consider the linear vector space formed by the elements of \( \mathbb{F}_q \) over \( \mathbb{F}_p \), and let \( v_1, \ldots, v_n \) be a basis of this vector space. Denote \( \chi \) a multiplicative character of \( \mathbb{F}_q \) of order \( d > 1 \), \( f \in \mathbb{F}_q[x] \) is a non-constant polynomial which is not a \( d \)-th power and which has \( m \) distinct zeros in its splitting field over \( \mathbb{F}_q \), and \( k_1, \ldots, k_n \) are non-negative integers with \( k_1 \leq p, \ldots, k_n \leq p \), then writing

\[ B = \left\{ \sum_{i=1}^{n} j_i v_i : 0 \leq j_i < k_i \right\}, \]

we have

\[
\left| \sum_{z \in B} \chi(f(z)) \right| < mq^{1/2}(1 + \log p)^n. \tag{3.3.5}
\]

**Proof.** This is a part of Theorem 2 in [47] (where its proof was based on A. Weil’s theorem [46]).

A set \( B \subset \mathbb{F}_q \) of the form appearing in Theorem XXIII will be called a box lattice. As

\[
\left| \sum_{j=0}^{u} \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right| \leq \sum_{j=0}^{u} \left| \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right|,
\]

thus it suffices to give an upper bound on each \( \left| \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right| \). We can translate nearly all \( I_j \), such that \( I_j \) will become a box lattice, and by this we mean there is at most one piece with an “excess”.

37
Lemma 3.3.2. In the case \( q = p^2 \), there is at most one \( I_j \) which cannot be translated into a box lattice. If \( M = m_0 + m_1 \cdot p \) such that \( 0 \leq m_k < p \) for \( k = 0, 1 \), then for the \( I_j \), where \( \xi_M \in I_j \), it may occur that the greatest box lattice inside of \( I_j \) does not contain all the elements of \( I_j \). Then we will call those elements as the excess of \( I_j \), and denote them as

\[
S_j = I_j - B_j
\]

where \( B_j \) is the greatest box lattice inside of \( I_j \). If \( S_j \neq \emptyset \), then \( \xi_M \in S_j \). These elements of the excess got the same last coordinate, thus they are all of the form

\[
\{ \xi_n : n = n_0 + m_1 \cdot p, 0 \leq n_0 \leq p - 1 \},
\]

therefore they form a translated box lattice.

Proof. If \( j \) is such that \( M \in I_j \) and \( S_j \neq \emptyset \), then we will see that \( \xi_M \in S_j \). By the definition of \( I_j \), \( m_0 + d_{j,0} < p \) and \( m_0 + d_{j+1,0} \geq p \). There is an \( \xi_n \), which in \( \xi_n \in S_j \), and by contradiction, if \( \xi_M \notin S_j \), then \( \xi_M \in I_j \setminus S_j = B_j \), thus \( \xi_M \) is in the greatest translated box lattice inside \( I_j \). But then

\[
B_j = \{ \xi_n : n = a_0 + a_1 \cdot p, a_0 + d_{j,0} < p \text{ and } a_0 + d_{j+1,0} \geq p, 0 \leq a_1 \leq m_1 \}
\]

so every element of \( I_j \) is in \( B_j \), therefore \( S_j = \emptyset \). Contradiction.

If \( \xi_n \in S_j \), for \( n < n' \leq M \) we have \( \xi_{n'} \in \bigcup_{j=0}^n S_j \). By contradiction, if there is an \( \xi_{n''} \notin \bigcup_{j=0}^n S_j \), then for a \( j \) index, \( \xi_{n''} \in B_j \). We also know that for \( n'' < n' \) \( \xi_{n''} \in \bigcup_{j=0}^n I_j \). If \( \xi_{n''} \in I_k \), then

\[
A'_k := \{ \xi_n : n = a_0 + a_1 \cdot p, a_0 + d_{k,0} < p \text{ and } a_0 + d_{k+1,0} \geq p, 0 \leq a_1 = n_1'' \} \subseteq I_k
\]

holds, and \( A'_k \) is a translated box lattice inside \( I_k \), so \( A'_k \) also inside the greatest translated box lattice in \( I_k \), so \( A'_k \subseteq B_k \) holds, which implies that for all \( n'' < n' \), \( \xi_{n''} \in \bigcup_{j=0}^n B_j \), even for \( n \), which is a contradiction.

It is also easy to see that all the elements which are in \( S_j \) got the same last coordinate, otherwise there would be an \( \xi_n \in S_j \) such that \( n = n_0 + n_1 \cdot p, n_1 < m_1 \). But \( \xi_M \in S_j \), so there are no such elements \( \xi_{n''} \in B_j \), with \( n < n' < M \). But it also implies two things. First, that all the elements

\[
A_j := \{ \xi_n : n = a_0 + a_1 \cdot p, a_0 + d_{j,0} < p \text{ and } a_0 + d_{j+1,0} \geq p, a_1 = n_1 \} \subseteq I_j
\]

and secondly, that for all \( \xi_{n''} \in B_j \), \( n_1'' < n_1 \). But then \( A_j \cup B_j \) also a translated box lattice, and

\[
B_j \subseteq A_j \cup B_j \subseteq I_j
\]

which contradicts the choice of \( B_j \) as the greatest translated box lattice in \( I_j \).
Now all that is left to check is that there is at most one \( I_j \) with an excess. If there would be more than one, an \( I_j \) and an \( I_k \), \( i \neq k \), and we can assume that \( \xi_M \in S_j \). All the elements in the excess got the same last coordinate, \( m_1 \), for \( k \), this means that

\[
A_k := \{ \xi_n : n = a_0 + a_1 \cdot p, n_0 + d_{i_k,0} < p \text{ and } n_0 + d_{i_k+1,0} \geq p : 0 \leq a_1 = m_1 \} \subseteq I_k
\]

so all the elements of \( I_k \) with last coordinate \( m_1 \) are in \( I_k \), so they are also in \( B_k \), so \( I_k = B_k \) which implies \( S_k = \emptyset \). This contradiction proves the lemma. □

\[
I_j = \{ \xi_n : n = a_0 + a_1 \cdot p, a_0 + d_{j,0} < p \text{ and } a_0 + d_{j+1,0} \geq p \}
\]

so by computing the smallest \( a_0(j) \) of an \( n : \xi_n \in I_j \), such that \( a_0 + d_{j+1,0} \geq p \) holds, \( a_0 + d_{j+1,0} = p \), and therefore

\[
a_0(j) = p - d_{j+1,0}.
\]

In \( I_j - a_0(j) \cdot \theta_0 = \{ \xi_n - a_0(j) \cdot \theta_0 : \xi_n \in I_j \} \) (this \( \theta_0 \) is from the basis of the linear vector space formed by the elements of \( \mathbb{F}_q \) over \( \mathbb{F}_p \) take the greatest box lattice, \( B_j := I_j - a_0(j) \cdot \theta = \{ \xi_n - a_0(j) \cdot \theta : \xi_n \in I_j \} = B' + \{ \xi_{n_1}, \xi_{n_2}, \ldots, \xi_{n_s} \} \), such that \( B' \) is a translated box lattice, and \( n_1, n_2, \ldots, n_s \) are consecutive integers, and \( \xi_{n_s} \) is the translated of \( \xi_M \), so \( \xi_{n_s} = \xi_M - a_0(j) \cdot 1 \). Therefore \( n_s = M - a_0(j) = M - p - d_{i_{j+1},0} \) We cannot take \( \xi_{n_1}, \xi_{n_2}, \ldots, \xi_{n_s} \) into \( B' \) to form a box lattice, but we can use them to form a box lattice in a smaller dimension, in an affine subspace. The elements \( \xi_{n_1}, \xi_{n_2}, \ldots, \xi_{n_s} \) form a 1 dimensional box lattice by Lemma 3.3.2. Also in the general case, this will be the method to give a treatable upper bound for the \( I_j \)’s with an excess.

Only left to deal with the polynomials \( f_j \)’s, when we change form \( I_j \) to \( B_j \). Note that for an element \( \xi_n \in I_j \) \( f_j(\xi_n) = f_j((\xi_n - a_0(j) \cdot \theta_0) + a_0(j) \cdot \theta_0) \), where \( \xi_n - a_0(j) \cdot \theta_0 \in B_j \).

So we can change our polynomials \( f_j(\xi_n) \) to \( f_j(\xi_n + a_0(j) \cdot \theta_0) \), which are also polynomials with the same degrees, and also have \( k \) distinct roots.

Now we can finish the estimation of \( (3.3.4) \) by Theorem XXIII and by remembering that \( u + 1 \leq k + 1 \). If \( \xi_M \in I_j \), then by Lemma 3.3.2 only \( S_j \) is not empty:

\[
\left| \sum_{j=0}^u \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right| \leq \sum_{j=0}^u \left| \sum_{\xi_n \in I_j} \gamma(f_j(\xi_n)) \right| \leq \sum_{j=0}^u \left| \sum_{\xi_n \in B_j} \gamma(f_j(\xi_n + a_0(j) \cdot \theta_0)) \right| + \left| \sum_{\xi_n \in S_j} \gamma(f_j(\xi_n + a_0(j) \cdot \theta_0)) \right| < (k + 1)kq^{1/2}(1 + \log p)^2 + kq^{1/2}(1 + \log p)^2 < (k + 2)kq^{1/2}(1 + \log p)^2,
\]

which proves the theorem. □
3.4 Further problems

There are a few related problems which have not been settled yet.

**Problem 1.** For a random binary sequence $E_q$ of length $q$ we have $W(E_q) \ll q^{1/2+\varepsilon}$ by [6], and for the sequence $L_q$ studied here we have $W(L_q) \ll q^{3/4+\varepsilon}$ by Theorem XXI. What is the smallest $c$ with $W(L_q) \ll q^{c+\varepsilon}$?

**Problem 2.** Can we extend Gyarmati’s Theorem XXII so that under a suitable condition on $\eta$, how small must $W(E_{N^2}(\eta))$ be?

**Problem 3.** Can one extend Gyarmati’s Theorem XXII from two dimensional lattices to $r$ dimensional lattices for any $r \geq 2$?

**Problem 4.** Can one extend our Theorem 3.2.1 on the sharpening of Theorem XXI from $f(x) = x$ to a large family of polynomials $f(x)$?

I hope to return to some of these problems in subsequent papers.
Chapter 4
Family measures

In many applications, one needs not just a single “good” sequence: we need a large family of them. However, it is not enough if our family $F$ is large. For example, if all the sequences in $F$ are identical except for bits in “a few” fixed positions, then one cannot use more than one sequence from that family. It is very important to know that $F$ got a “complex”, “rich” structure in the sense there are many “independent” sequences in it which differ in many bits. For these purposes one needs quantitative measures to study the structural properties of families of binary sequences.

In this chapter I would like to give a brief introduction to the most commonly used family measures, based on the work of V. Tóth [43] [44], K. Gyarmati, A. Sárközy and C. Mauduit [22].

We will study the existence of collisions in the given family. This notion appears in many paper, for example in [36], and it can be adapted to our approach.

4.1 Basic definitions

Assume that $N \in \mathbb{N}$, $S$ is a given set (e.g. a set of certain polynomials or the set of all the binary sequences of a given length much less than $N$), to each $s \in S$, we assign a unique binary sequence

$$E_N(s) = (e_1, \ldots, e_N) \in \{-1, 1\}^N,$$

and let $F = F(S)$ denote the family of the binary sequences obtained in this way:

$$F = F(S) = \{E_N(s) : s \in S\}. \quad (4.1.1)$$

Example 4.1.1. If we look at the family generated by (3.1.3), in the case when $f(x) = x$, then $S$ will be the bases of $\mathbb{F}_p$ over $\mathbb{F}_p$. 

41
Definition 4.1.2. If \( s \in S, s' \in S, s \neq s' \) and

\[
E_N(s) = E_N(s')
\]

(4.1.2)

then (4.1.2) is said to be a collision in \( \mathcal{F} = \mathcal{F}(S) \). If there is no collision in \( \mathcal{F} = \mathcal{F}(S) \), then \( \mathcal{F} \) is said to be collision free.

So \( \mathcal{F} = \mathcal{F}(S) \) is collision free if we have \( |\mathcal{F}| = |S| \). We wonder whether or not our pseudorandom family is collision free. It is easy to see that our family need not to be collision free: if the number of collisions is limited, they do not cause much trouble. Let us define a measure for the number of collisions:

Definition 4.1.3. The collision maximum \( M = \mathcal{F}(S) \) is defined by

\[
M = \mathcal{F}(S) = \max \{|s : s \in S, E_N(s) = E_N\}.
\]

thus \( M \) is the maximal number of elements of \( S \) generating the same binary sequence \( E_N \).

We will need some other notions, namely the avalanche property and a variant of the Hemming distance to handle the avalanche property.

Definition 4.1.4. If in (4.1.1) we have \( S = \{-1,1\}^l \), and for any \( s \in S \), changing any element of \( s \) changes “many” elements of \( E_N(s) \), in other words, for \( s \neq s' \) many elements of the sequences \( E_N(s) \) and \( E_N(s') \) are different, then we speak about the avalanche effect, and we say that \( \mathcal{F} = \mathcal{F}(S) \) possesses the avalanche property. If for any \( s \in S, s' \in S, s \neq s' \) at least \( \left( \frac{1}{2} - o(1) \right) N \) elements of \( E_N(s) \) and \( E_N(s') \) are different, then \( \mathcal{F} \) is said to possess the strict avalanche property.

As we mentioned above, to study the avalanche property, one may introduce the following measure:

Definition 4.1.5. If \( N \in \mathbb{N}, E_N(e_1, \ldots, e_N) \in \{-1,1\}^N \) and \( E'_N = (e'_1, \ldots, e'_N) \in \{-1,1\}^N \), then the distance \( d(E_N, E'_N) \) between \( E_N \) and \( E'_N \) is defined by

\[
d(E_N, E'_N) = |\{n : 1 \leq n \leq N, e_n \neq e'_n\}|
\]

Moreover, if \( \mathcal{F} \) is a family of form (4.1.1), then the distance minimum \( m(\mathcal{F}) \) of \( \mathcal{F} \) is defined by

\[
m(\mathcal{F}) = \min_{s,s' \in S} d(E_N(s), E_N(s')).
\]

Applying this notion we may say that the family \( \mathcal{F} \) of form (4.1.1) is collision free if and only if \( m(\mathcal{F}) > 0 \), and \( \mathcal{F} \) possesses the strict avalanche property if

\[
m(\mathcal{F}) \geq \left( \frac{1}{2} - o(1) \right) N.
\]

(4.1.3)
The notions introduced in Definitions 4.1.4 and 4.1.5 can also be used when there is no parameter set $S$ is given: we may say that $F$ possesses the avalanche property if for any $E_N \in F$, $E'_N \in F$, $E_N \neq E'_N$ the sequences $E_N$ and $E'_N$ have many different elements, and the distance minimum can be defined as

$$m(F) = \min_{E_N \in F, E_N \neq E'_N} d(E_N, E'_N).$$

Another approach was presented by Ahlswede, Khachatrian, Mauduit and Sarkozy in [1] when they introduced the notion of family complexity or shortly $f$-complexity (which is a really good measure for cryptographic applications):

**Definition 4.1.6.** The $f$-complexity $\Gamma(F)$ of a family $F$ of binary sequences $E_N \in \{-1, 1\}^N$ is defined as the greatest integer $j$ so that for any specification

$$e_i = e_1, \ldots, e_j = e_j \quad (1 \leq i_1 < \cdots < i_j \leq N)$$

which $e_1, \ldots, e_j \in \{-1, 1\}$, there is at least one $E_N = (e_1, \ldots, e_N) \in F$ with satisfies it. If there is no $j \in \mathbb{N}$ with the property above then we define the $f$-complexity of $F$ as $\Gamma(F) = 0$.

### 4.2 The definition of the cross-correlation measure

In this section, we will present the definition of the cross-correlation measure of a family of binary sequences and we will investigate the connection between this new measure and the other related notions listed in Section 4.1.

We already showed that $C_k(E_N)$, the correlation measure of order $k$ of the binary sequence $E_N$ is one of the most important measure of pseudorandomness of a single binary sequence. In the definition of this measure we only considered a fixed sequence and we compared different elements of it (so this is an autocorrelation type quantity). [22] A logical move for us in the generalization of the correlation measure, instead of a single binary sequence, we want to characterize a family of sequences, so we will compare elements of different sequences taken from the family.

**Definition 4.2.1.** Let $N \in \mathbb{N}, k \in \mathbb{N}$, and for any $k$ binary sequences $E_N^{(1)}, \ldots, E_N^{(k)}$ with

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

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

$$0 \leq d_1 \leq \ldots \leq d_k < M + d_k \leq N,$$

(4.2.1)
write
\[
V_k \left( E_N^{(1)}, \ldots, E_N^{(k)}, M, D \right) = \sum_{n=1}^{M} e_n^{(1)} \ldots e_n^{(k)}
\] (4.2.2)

Let
\[
\hat{C}_k \left( E_N^{(1)}, \ldots, E_N^{(k)} \right) = \max_{M,D} V_k \left( E_N^{(1)}, \ldots, E_N^{(k)}, M, D \right),
\] (4.2.3)

where the maximum is taken over all \( D = (d_1, \ldots, d_k) \) and \( M \in \mathbb{N} \) satisfying (4.2.1) 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 \( \mathcal{F} \) of binary sequences \( E_N \in \{-1,1\}^N \) is defined as
\[
\Phi_k(\mathcal{F}) = \max \hat{C}_k \left( E_N^{(1)}, \ldots, E_N^{(k)} \right)
\] (4.2.4)

where the maximum is taken over all \( k \)-tuples of binary sequences \( \left( E_N^{(1)}, \ldots, E_N^{(k)} \right) \) with
\[
E_N^{(i)} \in \mathcal{F} \quad \text{(for } i = 1, 2, \ldots, k).\] (4.2.5)

For a better understanding of the definition of \( \hat{C}_k \left( E_N^{(1)}, \ldots, E_N^{(k)} \right) \), for every \( E_N \in \{-1,1\}^N \) we have
\[
\hat{C}_k \left( E_N, \ldots, E_N \right) = C_k(E_N)
\]
thus it follows from (4.2.4) that

**Corollary III.** We have
\[
\Phi_k(\mathcal{F}) \geq \max_{E_N \in \mathcal{F}} C_k(E_N).
\] (4.2.6)

This is actually an important result. An upper bound for the cross-correlation of order \( k \) of the family \( \mathcal{F} \) is also an upper bound for correlation of order \( k \) of every sequence \( E_N \in \mathcal{F} \). Thus it is enough to estimate \( \Phi_k(\mathcal{F}) \); if we have a “good” upper bound for \( \Phi_k(\mathcal{F}) \), then we know that all the elements of \( \mathcal{F} \) possess strong pseudorandom properties.

The definition of the cross-correlation measure seems to be a complicated definition, and looking at it for the first time it is hard to see, what “bonus” can it add to our “tool-kit”.

**Proposition I.** If \( N \in \mathbb{N} \) and \( E_N = (e_1, \ldots, e_N) \in \mathcal{F} \), \( E'_N = (e'_1, \ldots, e'_N) \in \mathcal{F} \), \( \mathcal{F} \subset \{-1,1\}^N \), then we have
\[
\left| d(E_N, E'_N) - \frac{N}{2} \right| \leq \frac{1}{2} \hat{C}_2(E_N, E'_N) \leq \frac{1}{2} \Phi_2(\mathcal{F}).
\] (4.2.7)
Proof. We have
\[
\frac{(e_n - e'_n)^2}{2} = \begin{cases} 
1 & \text{if } e_n = e'_n \\
0 & \text{if } e_n \neq e'_n
\end{cases}
\]
therefore
\[
d(E_N, E'_N) = \sum_{n=1}^{N} \frac{(e_n - e'_{n})^2}{2} = \frac{N}{2} - \frac{1}{2} \sum_{n=1}^{N} e_ne'_n
\]
whence by (4.2.2) (4.2.3) (4.2.4),
\[
\left| d(E_N, E'_N) - \frac{N}{2} \right| = \frac{1}{2} \left| \sum_{n=1}^{N} e_ne'_n \right| \leq \frac{1}{2} \tilde{c_2}(E_N, E'_N) \leq \frac{1}{2} \Phi_2(\mathcal{F})
\]
which proves (4.2.7).

If the cross-correlation of order 2 of the family \(\mathcal{F} \subseteq \{-1, 1\}^N\) is \(o(N)\):
\[
\Phi_2(\mathcal{F}) = o(N)
\]
then it follows from Definition 4.1.5, (4.2.7) and (4.2.8) that
\[
m(\mathcal{F}) = \min_{E_N, E'_N \in \mathcal{F}} d(E_N(s), E'_N(s')) \geq \frac{N}{2} - \frac{1}{2} \Phi_2(\mathcal{F}) = \frac{N}{2} - o(N),
\]
so that (4.1.3) holds. This proves

Proposition II. If \(N \in \mathbb{N}\), \(\mathcal{F} \subseteq \{-1, 1\}^N\) and (4.2.8) holds, then the family \(\mathcal{F}\) possesses the strict avalanche property.

So far we have seen that there is a close connection between collision, distance minimum and avalanche property in a family of binary sequences on one hand and its cross-correlation on the other hand. With two examples, we will show that the family complexity and the cross-correlation are "independent" in the following sense: it may happen that a family \(\mathcal{F}\) is "good" considering its family complexity, i.e. \(\Gamma(\mathcal{F})\) is large, but it is "bad" concerning its cross-correlation, i.e. \(\Phi_k(\mathcal{F})\) is also large for every small \(k\), and it is also possible that \(\mathcal{F}\) is considered "good" since \(\Phi_k(\mathcal{F})\) is small, however, \(\mathcal{F}\) is "bad" concerning its small family complexity. In other words, it is not enough to study only one of \(\Gamma(\mathcal{F})\) and \(\Phi_k(\mathcal{F})\), we have to estimate both of them.

Example 4.2.2. Let \(N \in \mathbb{N}\) and let \(\mathcal{F}\) be the set of all the binary sequences of length \(N\): \(\mathcal{F} = \{-1, 1\}^N\). Then clearly \(\Gamma(N) = N\) maximal. On the other hand, \(E_N = (1,\ldots, 1) \in \mathcal{F}\) thus by (4.2.6), for \(k \in \mathbb{N}\), \(k \leq N\) we have
\[
\Phi_k(\mathcal{F}) \geq C_k(E_N) = \sum_{n=1}^{N-k+1} e_ne_{n+1} \cdots e_{n+k-1} = \sum_{n=1}^{N-k+1} 1 = N - k + 1
\]
which is also large.
Example 4.2.3. Consider any family $\mathcal{F}$ of binary sequences of length $N$ with small cross-correlation of order $k$ for any small $k$; e.g., we may take $N = p$ = prime and the family $\mathcal{F}_1$ which construction will be in the next section, and which satisfies the inequality

$$\Phi_k(\mathcal{F}_1) < 10kd p^{1/2} \log p$$

(for any $1 < k < p$), Then for at least half of the sequences $E_p = (e_1, \ldots, e_p) \in \mathcal{F}_1$ either $e_1 = 1$ or $e_1 = -1$ holds: we may assume that the first equality does. Then let

$$\mathcal{F}'_1 = \{E_p = (e_1, \ldots, e_p) : e_1 = 1\}$$

so that $|\mathcal{F}'_1| \geq \frac{|\mathcal{F}_1|}{2}$, we have

$$\Phi_k(\mathcal{F}'_1) < 10kd p^{1/2} \log p$$

(which is small), and

$$\Gamma(\mathcal{F}'_1) = 0$$

(which is also small) since there is no $E_p = (e_1, \ldots, e_p) \in \mathcal{F}'_1$ satisfying the specification

$$e_1 = -1.$$

For more details and results on the cross-correlation measure, see [22].

4.3 Construction using Legendre symbol

In [12], the authors extended the Legendre symbol construction presented in Theorem II in the way showed in (2.2.4). In Theorem IV, Goubin, Mauduit and Sárközy proved that under not very strong conditions on $f(x)$, both the correlation measure of small order and the well distribution measure are small.

As Theorem IV and the discussion in [1] show, this construction in (2.2.4) is one of the best constructions for large families of binary sequences with strong pseudorandom properties. V. Tóth [43] showed that a variant of the family described in Theorem IV is collision free and possesses a strong form of the avalanche property.

**Theorem XXIV.** Let $S$ be the set of polynomials $f(x) \in \mathbb{F}_p[x]$ of degree $D \geq 2$ which do not have multiple zeros. Define $E_p = (e_1, \ldots, e_p)$ by (2.2.4) and $\mathcal{F} = \mathcal{F}(S)$ by (4.1.1). Then we have

$$m(\mathcal{F}) \geq \frac{1}{2}(p - (2D - 1)p^{1/2} - 2D).$$

**Corollary IV.** If $S$, $\mathcal{F}$ are defined as in Theorem XXIV and we have $D < \frac{\sqrt{p}}{2}$, then $\mathcal{F}$ is collision free.
Proof. If $D < \frac{\sqrt{p}}{2}$, then it follows from Theorem XXIV that

$$m(\mathcal{F}) \geq \frac{1}{2}(p - (2D - 1)p^{1/2} - p^{1/2}) = \frac{1}{2}(p - 2Dp^{1/2}) > 0$$

and therefore $\mathcal{F}$ is collision free. \hfill \Box

**Corollary V.** If $S$, $\mathcal{F}$ are defined as in Theorem XXIV and we have $p \to \infty$, $D = o(p^{1/2})$, then $\mathcal{F}$ possesses the strong avalanche property.

**Proof.** If $p \to \infty$ and $D = o(p^{1/2})$, then Theorem XXIV gives

$$m(\mathcal{F}) \geq \left(\frac{1}{2} - o(1)\right)$$

which proves the corollary. \hfill \Box

In [22], the authors showed two different ways to modify slightly the definition of the family given by Goubin, Mauduit and Sárközy to have a ‘reasonable control” over the cross-correlation measure.

**Theorem XXV.** Let $d \in \mathbb{N}$, $p$ a prime number, $d < p$, consider all the irreducible polynomials $f(x) \in \mathbb{F}_p[x]$ of the form

$$f(x) = x^d + a_2x^{d-2} + a_3x^{d-3} + \ldots + a_d$$

and let $\mathcal{F}_1$ denote the family of the binary sequences $E_p = E_p(f)$ assigned to these polynomials $f$ by the formula (2.2.4). Then we have

1. $\Phi_k(\mathcal{F}_1) < 10kd p^{1/2} \log p$

   for all $k \in \mathbb{N}$, $1 < k < p$, and

2. if $d < \frac{p^{1/2}}{20 \log p}$, then

   $$|\mathcal{F}_1| \geq p^{\left\lfloor d/3 \right\rfloor - 1}.$$ 

Theorem XXV gives a good upper bound for the cross-correlation, and the size of the family is also large. The only weakness of the theorem is the following: since no good algorithm is known for constructing “many” irreducible polynomials over $\mathbb{F}_p$, thus Theorem XXV proves only existence but it does not provide an explicit construction. Thus in [22], the authors presented another construction which will be more explicit, but we will be able to control the cross-correlation of order $k$ only if $k = 2$ or $k$ is odd.
Theorem XXVI. Let $d \in \mathbb{N}$, $p$ a prime number, $d < p$, and consider all the polynomials $f(x) \in \mathbb{F}_p[x]$ of the form

$$f(x) = (x - x_1)(x - x_2) \ldots (x - x_d)$$

where

$$x_1, x_2, \ldots, x_d \text{ are distinct elements of } \mathbb{F}_p$$

and

$$x_1 + x_2 + \ldots + x_d = 0.$$ 

Let $\mathcal{F}_2$ denote the family of the binary sequences $E_p = E_p(f)$ assigned to these polynomials $f$ by the formula (2.2.4). Then we have

1. $$\Phi_k(\mathcal{F}_2) < 10kd^{1/2}\log p$$
   if $k = 2$ or $k$ is odd

2. $$|\mathcal{F}_2| = \frac{1}{d}(p - 1).$$

4.4 Construction using additive characters

In this section, we will present the result of Tóth [43] and [44], where she investigated “how good” are some family measures of the construction (2.2.6).

Proposition III. For any family $\mathcal{F} = \mathcal{F}(S)$ of type (2.2.6) the collision maximum $M = (\mathcal{F}, S)$ satisfies

$$M \geq \frac{|S|}{2^N}.$$ 

Proof. This is Proposition 1 in [43]. We have

$$|S| = \sum_{E_N \in \{-1,1\}^N} |\{s : s \in SE_N(s) = E_N\}|$$

$$\leq \sum_{E_N \in \{-1,1\}^N} M = M \sum_{E_N \in \{-1,1\}^N} 1 = M \cdot 2^N.$$

\[\square\]
If $|S| > 2^N$, then it is trivial that there are collisions in $\mathcal{F}$.
Now fix a prime $p$, and for $k \in \mathbb{N}$, write

$$S_k = \{ f(x) : f(x) \in \mathbb{F}_p[x], \deg f(x) = k \}$$

and

$$\mathcal{F}_k = \{ E_p(f) = (e_1, \ldots, e_p), f \in S_k \},$$

where $E_p = E_p(f)$ is defined by (2.2.6). Then by Proposition III if we have

$$\frac{|S|}{2^p} \geq p^k = \exp(k \log p - p \log 2) \to \infty,$$

or equivalently,

$$\frac{k}{p(\log p)^{-1}} \to \infty,$$  \hspace{1cm} (4.4.1)

then we have

$$M = (\mathcal{F}_k, S_k) \to \infty$$

so that for $k$ satisfying (4.4.1), there are many collisions in $\mathcal{F}_k$.

In [43], Theorem 2 presented the fact that there are many collisions in $\mathcal{F}_k$ also for small $k$, even for $k = 2$. If we denote the integer part of $x$ by $[x]$, then we have:

**Theorem XXVI.** If $p$ is a fixed prime and $\mathcal{F}_2$, $S_2$ are defined as above then we have

$$M = (\mathcal{F}_2, S_2) \geq \left[ \frac{1}{6} \log p \right].$$

**Proof.** This is Theorem 2 in [43]. \hfill $\square$

In [44], Tóth constructed a collision free version of the construction (2.2.6), so she showed that just because there are many collisions in a construction, it does not mean that it has to be discarded immediately, it may occur that the situation can be saved by replacing the given family by a subfamily of it which is just slightly smaller. In this case, it can be generated easily, it is collision free, and it possesses even the strict avalanche property.

Let $\mathcal{P}_d$ be the set of monic polynomials of degree $d$ whose constant term is 0:

$$\mathcal{P}_d = \{ f(x) \in \mathbb{F}_p[x] : f(x) = \sum_{i=0}^{d} a_i x^i \text{ with } a_0 = 0, a_d = 1 \}.$$

**Theorem XXVIII.** If $f(x) \in \mathcal{P}_d$, then the family of binary sequences obtained by (2.2.6) is collision free and possesses the strict avalanche property.
If \( f(x) \in \bigcup_{d=2}^{D} \mathcal{P}_d \) (the degrees of the polynomials are not the same but their constant terms are zero), then the family of binary sequences obtained by (2.2.6) is collision free and possesses the strict avalanche property.

Proof. This is a combination of Theorems 1, 2, 3 and 4 in [44].
Bibliography


[22] K. Gyarmati, C. Mauduit, A. Sárközy, The cross-correlation measure for families of binary sequences, Proceeding to Applications of Algebra and Number Theory (Lectures on the occasion of Harald Niederreiter’s 70th Birthday),


