Construction of large families of pseudo-random binary sequences

by

Louis GOUBIN

Cryptography Research, Schlumberger Smart Cards
36-38 rue de la Princesse, BP45
F - 78430 Louveciennes Cedex, France
E-mail : LGoubin@slb.com

Christian MAUDUIT

Institut de Mathématiques de Luminy, UPR 9016 CNRS
63 Av. de Luminy, Case 907
F - 13288 Marseille Cedex 9, France
E-mail : mauduit@iml.univ-mrs.fr

and

András SÁRKÖZY*
Department of Algebra and Number Theory
Eötvös Loránd University
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
E-mail : sarkozy@cs.elte.hu

∗ Research partially supported by the Hungarian National Foundation for Scientific Research, Grant N° T029759, and MKM fund FKFP - 0139/1997 and "Balaton" French-Hungarian exchange program TéT-18/00.

Abstract

In a series of papers Mauduit and Sárközy (partly with coauthors) studied finite pseudorandom binary sequences. They showed that the Legendre symbol forms a "good" pseudorandom sequence, and they also tested other sequences for pseudorandomness, however, no large family of "good" pseudorandom sequences has been found yet.

In this paper, a large family of this type is constructed by extending the earlier Legendre symbol construction.

1991 Mathematics Subject Classification : 11 K 45.

Key words : pseudorandom, binary sequence, Legendre symbol.

1 Introduction

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

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

In particular, in Part I [5] first they introduced the following measures of pseudorandomness : Write

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

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

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

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

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

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

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

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

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

Moreover, it was shown in [5] that the Legendre symbol forms a "good" pseudorandom sequence. More exactly, let p be an odd prime, and

$$N = p - 1, e_n = \left(\frac{n}{p}\right), E_N = \{e_1, \dots, e_N\}. \tag{1}$$

Then by Theorem 1 in [5] we have

$$W(E_N) \ll p^{1/2} \log p \ll N^{1/2} \log N$$

and

$$C_k(E_N) \ll kp^{1/2} \log p \ll kN^{1/2} \log N.$$

Since then numerous binary sequences have been tested for pseudorandomness but still construction (1) is the best (see also [9] for another construction which is just slightly worse). However, all the constructions given so far for "good" pseudorandom binary sequences produce only a "few" good sequences while in certain applications (e.g., in cryptography) one needs "large" families of "good" pseudorandom binary sequences.

In this paper our goal is to construct large families of this type. We remark that our results could be extended to the more general case of sequences of r symbols (see [7]), however, we prefer to focus here on the slightly simpler case $r = 2$.

2 A further construction related to the Legendre symbol

The pseudorandom properties of the Legendre symbol $\left(\frac{n}{p}\right)$ have been studied in numerous papers, see *e.g.* [8] and [4]. Further references can be found in [5]. Here we consider the more general case $\left(\frac{f(n)}{p}\right)$.

In [6] we extended construction (1) by generalizing the definition of e_n to

$$e_n = \left(\frac{f(n)}{p}\right)$$

where $f(n)$ is a permutation polynomial over $\mathbf{F}_p$ (= the field of the modulo p residue classes).

However, very little is known on permutation polynomials and we know only very few of them.

Now we shall be able to use a much greater family of "good" polynomials. Before describing the family of these polynomials, we have to introduce two definitions.

DEFINITION 1. If $M \in \mathbb{N}$, $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}$ (= 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 $c \in \mathbb{Z}_m$,

$$a + b = c, a \in \mathcal{A}, b \in \mathcal{B} \tag{2}$$

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

Let us denote by $\overline{\mathbf{F}}_p$ the algebraic closure of $\mathbf{F}_p$.

THEOREM 1. If p is a prime number, $f(X) \in \mathbf{F}_p[X]$ ($\mathbf{F}_p$ being the field of the modulo p residue classes) has degree $k(> 0)$, $f(X)$ has no multiple zero in $\overline{\mathbf{F}}_p$, and the binary squence $E_p = \{e_1, \ldots, e_p\}$ is defined by

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

then

(i) we have

$$W(E_p) < 10\ kp^{1/2}\log p; \tag{3}$$

(ii) if $\ell \in \mathbb{N}$ is such that the triple (r,ℓ,p) is admissible for all $r \le k$, then

$$C_\ell(E_p) < 10\ k\ell p^{1/2}\log p. \tag{4}$$

PROOF OF THEOREM 1. The proof of both assertions (i) and (ii) will be based on

LEMMA 1. Suppose that p is a prime number, χ is a non-principal character modulo p of order d (so that $d|p-1$), $f(X) \in \mathbf{F}_p[X]$ has degree k and a factorization $f(X) = b(X-X_1)^{d_1}\dots(X-X_s)^{d_s}$ (where $X_i \neq X_j$ for $i \neq j$) in $\overline{\mathbf{F}}_p$ with

$$(d, d_1, \dots, d_s) = 1. \tag{5}$$

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

Then

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

PROOF OF LEMMA 1. This is Theorem 2 in [5] and, indeed, there we derived it from A. Weil's theorem [11]. We remark that a small correction is needed in our proof there : in the proof we used formula (4.1) which said that under certain assumptions on p, χ, d, and $f(X)$, for all $a \in \mathbb{Z}$ we have

$$\left|\sum_{X\in\mathbf{F}_p} \chi(f(X))\exp\left(\frac{2i\pi aX}{p}\right)\right| \le sp^{1/2}.$$

We wrote that "this is a part of Theorem 2G in [11, p.45]". This reference is correct for $(a,p)=1$, but for $p|a$ it must be replaced by a reference to Theorem 2C or 2C' in [11, p.43].

Moreover, we remark that in Theorem 2 in [5] our goal was to state the inequality in question in a possibly simple form which holds uniformly in p. For large p the constant factor in the upper bound can be improved considerably and, indeed, the same proof also gives that for $p \to \infty$ the constant factor 9 in Theorem 2 in [5] and in Lemma 1 above can be replaced by a factor $1+o(1)$. Using this form of the lemma, we obtain that for $p \to \infty$ the constant factor 10 in both (3) and (4) above can be replaced by a factor $1+o(1)$. (However, one would expect that the truth is $o(kp^{1/2}\log p)$, resp. $o(k\ell p^{1/2}\log p)$; perhaps even $\mathcal{O}(kp^{1/2}\log\log p)$, resp. $\mathcal{O}(k\ell p^{1/2}\log\log p)$ is also true, but certainly not more.)

(i) Assume that $a \in \mathbb{Z}, b, t \in \mathbb{N}$ and

$$1 \le a \le a + (t-1)b \le p, \tag{6}$$

and write $g(X) = f(a+bX)$ so that $g(X) \in \mathbf{F}_p[X]$.

Clearly, $g(X) \equiv 0 \pmod p$ has at most k solutions thus, defining $\left(\frac{a}{p}\right)$ as 0 for $p|a$, we have

$$|u(E_p,t,a,b)| = \left|\sum_{j=0}^{t-1} e_{a+jb}\right| \le \left|\sum_{j=0}^{t-1}\left(\frac{f(a+jb)}{p}\right)\right| + k = \left|\sum_{j=0}^{t-1}\left(\frac{g(j)}{p}\right)\right| + k. \tag{7}$$

Clearly, f and g are of the same degree, and if the factorization of f in $\overline{\mathbf{F}}_p$ is

$$f(X) = c(X-X_1)\dots(X-X_k)$$

where $X_i \neq X_j$ for $i \neq j$, then the factorization of $g(X)$ is

$$g(X) = f(a+bX) = cb^k(X - b^{-1}(X_1 - a)) \dots (X - b^{-1}(X_k - a))$$

so that $g(X)$ does not have multiple zeros either. Thus in order to estimate the sum in (7), we may apply Lemma 1 with $\left(\frac{n}{p}\right), 2$ and $g(n)$ in place of $\chi(n), d$ and $f(n)$, respectively. We obtain that

$$|u(E_p, t, a, b)| = \left|\sum_{j=0}^{t-1} \left(\frac{g(j)}{p}\right)\right| + k < 9kp^{1/2}\log p + k < 10\ kp^{1/2}\log p$$

for all a, b, t satisfying (6) which completes the proof of (3).

(ii) Write $f(X) = bf_1(X)$ with $b \in \mathbb{Z}_p$ where $f_1(X)$ is a unitary polynomial. For any integers $d_1, \dots, d_\ell$ and $M \in \mathbb{N}$ with

$$0 \leq d_1 < \dots < d_\ell, M + d_\ell \leq p, \tag{8}$$

$$f(n + d_i) \equiv 0(\bmod p), 1 \leq n \leq M, 1 \leq i \leq \ell$$

has at most $k\ell$ solutions. Thus writing again $\left(\frac{0}{p}\right) = 0$, we have

$$V(E_p, M, D) = \left|\sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_\ell}\right| \leq \left|\sum_{n=1}^{M} \left(\frac{f(n+d_1)}{p}\right)\left(\frac{f(n+d_2)}{p}\right)\dots\left(\frac{f(n+d_\ell)}{p}\right)\right| + k\ell =$$

$$= \left|\left(\frac{b^\ell}{p}\right)\sum_{n=1}^{M}\left(\frac{f_1(n+d_1)f_1(n+d_2)\dots f_1(n+d_\ell)}{p}\right)\right| + k\ell.$$

Write $h(n) = f_1(n+d_1)f_1(n+d_2)\dots f_1(n+d_\ell)$. It suffices to show :

LEMMA 2. If f, k, ℓ are defined as in Theorem 1, then $h(X)$ has at least one zero in $\overline{\mathbf{F}}_p$ whose multiplicity is odd.

Indeed, assuming that Lemma 2 has been proved, the proof of (4) can be completed in the following way : by Lemma 2, we may apply Lemma 1 with $\left(\frac{n}{p}\right), 2$ and $h(X)$ in place of χ, d and $f(X)$, respectively (since then (5) holds with $d = 2$). The degree of $h(X)$ is clearly $k\ell$, thus applying Lemma 1 we obtain

$$|V(E_p, t, a, b)| \leq \left|\sum_{n=1}^{M}\left(\frac{h(n)}{p}\right)\right| + k\ell < 9k\ell p^{1/2}\log p + k\ell < 10\ k\ell p^{1/2}\log p$$

for all $d_1, \dots, d_\ell, M$ satisfying (8) which proves (4). Thus it remains to prove the lemma :

PROOF OF LEMMA 2. We will say that the polynomials $\varphi(X) \in \mathbf{F}_p[X], \psi(X) \in \mathbf{F}_p[X]$ are equivalent : $\varphi \sim \psi$ if there is an $a \in \mathbf{F}_p$ such that $\psi(X) = \varphi(X + a)$. Clearly, this is an equivalence relation.

Write $f_1(X)$ as the product of irreducible polynomials over $\mathbf{F}_p$. It follows from our assumption on $f(X)$ that these irreducible factors are distinct. Let us group these factors so that in each group the equivalent irreducible factors are collected. Consider a typical group $\varphi(X + a_1), \dots, \varphi(X + a_r)$.

Then writing $h(X)$ as the product of unitary irreducible polynomials over $\mathbf{F}_p$, all the polynomials $\varphi(X + a_i + d_j)$ with $1 \leq i \leq r$, $1 \leq j \leq \ell$ occur amongst the factors. All these polynomials are equivalent, and no other irreducible factor belonging to this equivalence class will occur amongst the irreducible factors of $h(X)$.

Since distinct unitary irreducible polynomials cannot have a common zero, thus the conclusion of Lemma 2 fails, i.e., each of the zeros of h is of even multiplicity, if and only if in each group, formed by equivalent irreducible factors $\varphi(X + a_i + d_j)$ of $h(X)$, every polynomial of form $\varphi(X + c)$ occurs with even multiplicity, i.e., for even number of pairs a_i, d_j. In other words, writing $\mathcal{A} = \{a_1, \ldots, a_r\}, \mathcal{D} = \{d_1, \ldots, d_\ell\}$, for each group $\mathcal{A} + \mathcal{D}$ must possess property P. Now consider any of these groups (by $\deg f > 0$ there is at least one such group).

Since $\mathcal{A} + \mathcal{D}$ possesses property P, thus (r, ℓ, p) (with $r = |\mathcal{A}|$) is **not** an admissible triple. Clearly here we have $r \leq \deg f_1 = \deg f = k$ which contradicts our assumption on ℓ and thus, indeed, the conclusion of Lemma 2 cannot fail, and this completes the proof.

3 Sufficient criteria for admissibility.

To be able to use Theorem 1, one needs criteria for a triple (k, ℓ, p) being admissible.

Here we will present and prove three sufficient criteria of this type :

THEOREM 2.

(i) For every prime p and $k \in \mathbb{N}, k < p$ the triple $(k, 2, p)$ is admissible.

(ii) If p is prime, $k, \ell \in \mathbb{N}$ and

$$(4\ell)^k < p, \tag{9}$$

then $(k; \ell, p)$ is admissible.

(iii) If p is a prime such that 2 is a primitive root modulo p, then for every pair $k, \ell \in \mathbb{N}$ with $k < p, \ell < p$, the triple (k, ℓ, p) is admissible.

PROOF.

(i) Assume that contrary to the assertion, there is a prime p and a $k \in \mathbb{N}$ with

$$k < p \tag{10}$$

such that the triple $(k, 2, p)$ is not admissible, i.e., there are $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_p$ such that $|\mathcal{A}| = k, |\mathcal{B}| = 2$ and (2) has even number of solutions for all $c \in \mathbb{Z}_p$. Write $\mathcal{B} = \{b, b + d\}$ (where $d \neq 0$).

Every element of $\mathcal{A} + b$ must have (at least) 2 representations in form (2) whence it follows easily that $\mathcal{A} + b = \mathcal{A} + b + d$. Therefore, $\mathcal{A} + b = \mathcal{A} + b + rd$ for any r, thus $\mathcal{A} + b = \mathcal{A} + b + s$ for any $s \in \mathbb{Z}_p$, in particular for any $s \in \mathcal{A} + b$. Hence, $\mathcal{A} + b$ is an additive subgroup of $\mathbb{Z}_p$ thus $\mathcal{A} = \mathcal{A} + b = \mathbb{Z}_p$.

(ii) Assume that k, ℓ, p satisfy (9), and we have $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_p, |\mathcal{A}| = k, |\mathcal{B}| = \ell$.

It suffices to show that then there is a $c \in \mathbb{Z}_p$ for which (2) has exactly one solution. Moreover, if $m \in \mathbb{N}, (m, p) = 1$, then (2) and

$$ma + mb = mc, a \in \mathcal{A}, b \in \mathcal{B}$$

have the same solutions, and if c runs over the elements of $\mathbb{Z}_p$, then here $c' = mc$ does the same. Thus it suffices to show that there are $m \in \mathbb{N}, c' \in \mathbb{Z}_p$ such that $(m, p) = 1$, and

$$ma + mb = c', a \in \mathcal{A}, b \in \mathcal{B} \tag{11}$$

has exactly one solution.

For $a \in \mathbb{Z}$, let $r(a)$ denote the absolute least residue of a modulo p, i.e., definie $r(a) \in \mathbb{Z}$ by

$$r(a) \equiv a(\text{mod } p), |r(a)| \leq \frac{p-1}{2}.$$

We need

LEMMA 3. If $k, \ell, p, \mathcal{A}$ are defined as above, and the residue classes in $\mathcal{A}$ are represented by $a_1, \ldots, a_k$, then there is an $m \in \mathbb{N}$ such that $(m, p) = 1$ and

$$|r(ma_i)| \leq \frac{1}{2}\left[\frac{p}{\ell}\right] \text{ for } i = 1, 2, \ldots, k. \tag{12}$$

PROOF OF LEMMA 3. Consider the p k tuples

$$\underline{u}_j = (r(ja_1), \ldots, r(ja_k)), \quad j = 1, 2, \ldots, p \tag{13}$$

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

$$r(ja_i) \in \left\{-\frac{p-1}{2} + t_i D, -\frac{p-1}{2} + t_i D + 1, \ldots, -\frac{p-1}{2} + (t_i + 1)D - 1\right\} \text{ for } i = 1, 2, \ldots, k,$$

and for these integers t_i clearly we have

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

The number of the possible k tuples $t_1, \ldots, t_k$ with (14) is, by (9),

$$Z^k = \left(\left[\frac{p}{D}\right] + 1\right)^k < \left(2\frac{p}{D}\right)^k < \left(2\frac{p}{p/2\ell}\right)^k = (4\ell)^k < p,$$

thus there is at least one k tuple $t_1, \ldots, t_k$ which is assigned to at least two distinct j values j_1, j_2 :

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

Then we have

$$-\frac{p-1}{2} + t_i D \leq r(j_1 a_i), r(j_2 a_i) < -\frac{p-1}{2} + (t_i + 1)D$$

whence

$$|r(j_1 a_i) - r(j_2 a_i)| < D \text{ for } i = 1, 2, \ldots, k. \tag{16}$$

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

Then it follows from (16) that

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

which completes the proof of Lemma 3.

In order to complete the proof of assertion (ii), consider an integer m satisfying $(m, p) = 1$ and (12) in Lemma 3, and denote the representatives of the residue classes in $\mathcal{B}$ by $b_1, \ldots, b_\ell$. Now we represent the elements of $\mathbb{Z}_p$ on a circle, more exactly, assign the consecutive vertices $Q(1), \ldots, Q(p)$ of a regular p-gon to the consedutive residue classes $1, 2, \ldots, p$.

Consider the vertices $Q(mb_1), \ldots, Q(mb_\ell)$ and consider a pair $Q(mb_i), Q(mb_j)$ of vertices such that, moving on the circle in positive direction, there is no further vertex $Q(mb_X)$ between them, moreover, this is the pair of consecutive vertices with the maximal distance between them. We may

assume that moving in the positive direction $Q(mb_i)$ comes first, followed by $Q(mb_j)$ (with no $Q(mb_x)$ between). In other words,

$$mb_x \notin \{mb_i + 1, mb_i + 2, \ldots, mb_j - 1\} \text{ for } x = 1, 2, \ldots, \ell$$

(in $\mathbb{Z}_p$ sense, i.e., modulo p). A straightforward application of the pigeon hole principle gives that the maximal modulo p distance between two consecutive $mb'_x s$, i.e., the distance between mb_i and mb_j is at least $[p/\ell] + 1$ (note that p/ℓ is not integer).

Consider the numbers $r(ma_1), \ldots, r(ma_k)$, and reorder them according to their size, denote the numbers obtained in this way by $r_1, \ldots, r_k$ so that, by (12),

$$-\frac{1}{2}\ \left[\frac{p}{\ell}\right] \leq r_1 < \ldots < r_k \leq \frac{1}{2}\ \left[\frac{p}{\ell}\right].$$

By (12) we have

$$(mb_j + r_1) - (mb_i + r_k) = (mb_j - mb_i) + r_1 - r_k \geq \left(\left[\frac{p}{\ell}\right] + 1\right) - \frac{1}{2}\ \left[\frac{p}{\ell}\right] - \frac{1}{2}\ \left[\frac{p}{\ell}\right] = 1 > 0. \quad (17)$$

If u, v are defined by $r(ma_u) = r_1$ and $r(ma_v) = r_k$, then it follows easily from (12) and (17) that the numbers

$$mb_i + r_k = mb_i + ma_v$$

and

$$mb_j + r_1 = mb_j + ma_u$$

do not have any further representations in form (11) which completes the proof.

(iii) From practical point of view this is, perhaps, the most important of the three criteria. Namely, this criterion enables us to control even correlations of very high order provided that there are "many" primes p such that 2 is a primitive root modulo p.

Unfortunately, it is not known yet that there are infinitely many primes with this property: this is Artin's conjecture for primitive roots [1]. However, it is conjectured [2] that a positive proportion of the primes is of this type. Partly because of the importance of this criterion, partly in order to help to understand the notion of admissibility and the related difficulties better, we will give a detailed discussion of this case in the next section. This discussion will lead not only to the proof of criterion (iii), but it will also provide negative examples.

4 Admissibility, "good" primes. Negative examples.

DEFINITION 3. A positive integer m is said to be **good** if for any pair $k, \ell \in \mathbb{N}$ with $k < m, \ell < m$, the triple (k, ℓ, m) is admissible.

THEOREM 3. An odd prime p is good if and only if 2 is a primitive root modulo p.

PROOF OF THEOREM 3. For any $\mathcal{C} \subset \mathbb{Z}_p$ let us consider the polynomial $P_{\mathcal{C}}(X) \in F_2[X]$ defined by $P_{\mathcal{C}}(X) = \sum_{c \in \mathcal{C}} X^{s(c)}$ where $s(c)$ denotes the least **non negative** element of the residue class c modulo p.

We remark that for any $u \in \mathbb{Z}_p$, the polynomial $P_{u+\mathcal{C}}(X)$ is equal to the residue of $X^u.P_{\mathcal{C}}(X)$ modulo $(1 + X^p)$ in $F_2[X]$.

It follows from this remark that for any $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_p$, the sum $\mathcal{A} + \mathcal{B}$ has property P if and only if $(1 + X^p)$ divides $P_{\mathcal{A}}(X)P_{\mathcal{B}}(X)$ in $F_2[X]$.

If $1 + X + \ldots + X^{p-1}$ is reducible in $F_2[X]$, let us write $1 + X + \ldots + X^{p-1} = P_1(X)P_2(X)$ with $2 \leq \deg P_i \leq p-3$ for $i \in \{1, 2\}$ (polynomials of degree 1 do not divide $1 + X + \ldots + X^{p-1}$ in $F_2[X]$).

If we define $\mathcal{A}$ and $\mathcal{B}$ by $P_1(X) = \sum_{a \in \mathcal{A}} X^{s(a)}$ and $(1 + X)P_2(X) = \sum_{b \in \mathcal{B}} X^{s(b)}$, we see that the sum $\mathcal{A} + \mathcal{B}$ has property P, so that p is not a good prime.

Conversely if $1 + X + \ldots + X^{p-1}$ is irreducible in $F_2[X]$, for any $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_p$ such that $\mathcal{A} + \mathcal{B}$ has property P, the polynomial $1 + X + \ldots + X^{p-1}$ must divide $P_{\mathcal{A}}(X)$ or $P_{\mathcal{B}}(X)$ in $F_2[X]$, which implies $\mathcal{A} = \mathbb{Z}_p$ or $\mathcal{B} = \mathbb{Z}_p$, so that p is a good prime.

Thus we have proved that a prime p is good if and only if the polynomial $1 + X + \ldots + X^{p-1}$ is irreducible in $F_2[X]$.

It follows from a well known result concerning cyclotomic polynomials (see for example [6, Theorem 2.47, page 65]) that the polynomial $1+X+\ldots+X^{p-1}$ factors into $\dfrac{p-1}{d}$ distinct irreducible polynomials of the same degree d in $F_2[X]$, where d is the least positive integer such that $2^d \equiv 1 mod p$. In particular, this shows that the polynomial $1+X+\ldots+X^{p-1}$ is irreducible in $F_2[X]$ if and only if 2 is a primitive root modulo p, which completes the proof of theorem 3.

REMARK. The same method shows that an integer m is good if only if $m = 4, p^k$ or $2p^k$, where p is an odd prime, $k \geq 0$ and 2 is a primitive root modulo m (see [6, page 7]).

If p is not a good prime, then this method provides many examples of $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_p$ such that $\mathcal{A} + \mathcal{B}$ has property P, i.e., of pairs $k, \ell \in \mathbb{N}$ such that (k, ℓ, p) is not admissible.

EXAMPLE 1. If $p = 7$, then the factorization of $1 + X^7$ in $F_2[X]$ as $(1 + X + X^3)(1 + X + X^2 + X^4)$ shows that if $\mathcal{A} = \{0, 1, 3\}$ and $\mathcal{B} = \{0, 1, 2, 4\}$, then $\mathcal{A} + \mathcal{B}$ has property $\mathcal{P}$. It follows that (3,4,7) and (4,3,7) are not admissible.

For $p = 7$, it is actually easy to find all $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}_7$ such that $\mathcal{A} + \mathcal{B}$ has property $\mathcal{P}$. The problem is equivalent to find all polynomials $P_{\mathcal{A}}$ and $P_{\mathcal{B}}$ of degree less than 7 in $F_2[X]$ such that $P_{\mathcal{A}}(X)P_{\mathcal{B}}(X)$ is a multiple of the product of the three irreducible polynomials $(1 + X), (1 + X + X^3)$ and $(1 + X^2 + X^3)$ in $F_2[X]$.

For example, the factorization of $(1 + X + X^2 + X^3)(1 + X^7)$ in $F_2[X]$ as $(1 + X^3 + X^5 + X^6)(1 + X + X^2 + X^4)$ shows that if $\mathcal{A} = \{0, 3, 5, 6\}$ and $\mathcal{B} = \{0, 1, 2, 4\}$, then $\mathcal{A} + \mathcal{B}$ has property $\mathcal{P}$. Moreover a little computation shows that (3,4,7), (4,3,7) and (4,4,7) are the only non-admissible triple for $p = 7$.

EXAMPLE 2. If $p = 17$, then the factorization of $1 + X^{17}$ in $F_2[X]$ as $(1 + X + X^3 + X^6 + X^8 + X^9)(1 + X + X^2 + X^4 + X^6 + X^7 + X^8)$ shows that if $\mathcal{A} = \{0, 1, 3, 6, 8, 9\}$ and $\mathcal{B} = \{0, 1, 2, 4, 6, 7, 8\}$, then $\mathcal{A} + \mathcal{B}$ has property $\mathcal{P}$.

In the same way, the factorization of $1 + X^{17}$ in $F_2[X]$ as $(1 + X^3 + X^4 + X^5 + X^8)(1 + X^3 + X^4 + X^5 + X^6 + X^9)$ shows that if $\mathcal{A} = \{0, 3, 4, 5, 8\}$ and $\mathcal{B} = \{0, 3, 4, 5, 6, 9)$, then $\mathcal{A} + \mathcal{B}$ has property $\mathcal{P}$. It follows that (6,7,17), (7,6,17),(5,6,17) and (6,5,17) are not admissible.

EXAMPLE 3. If $p = 31$, then the factorization of $1 + X^{31}$ in $F_2[X]$ as $(1 + X^2 + X^5)(1 + X^2 + X^4 + X^5 + X^6 + X^8 + X^9 + X^{13} + X^{14} + X^{15} + X^{16} + X^{17} + X^{20} + X^{21} + X^{23} + X^{26})$ shows that if $\mathcal{A} = \{0, 2, 5\}$ and $\mathcal{B} = \{0, 2, 4, 5, 6, 8, 9, 13, 14, 15, 16, 17, 20, 21, 23, 26\}$ then $\mathcal{A} + \mathcal{B}$ has property P. It follows that (3,16,31) and (16,3,31) are not admissible.

5 Consequences, the algorithm.

Combining Theorem 1 and (i) in Theorem 2 we obtain.

COROLLARY 1. If p, f, k, E_p are defined as in Theorem 1, then we have

$$W(E_p) < 10kp^{1/2}\log p \tag{18}$$

and

$$C_2(E_p) < 20kp^{1/2}\log p.$$

In other words, the well-distribution measure and the correlation of order 2 are always small. If we also want to control correlations of higher order, this is possible if the order is not "very large" in terms of k and p. Indeed, combining Theorem 1 and (ii) in Theorem 2 we obtain.

COROLLARY 2. If p, f, k, E_p are defined as in Theorem 1, and either
(i) 2 is a primitive root modulo p and $\ell < p$, or
(ii) we have

$$\ell < \frac{p^{1/4}}{4},$$

then both (18) and

$$C_\ell(E_p) < 10k\ell_p^{1/2}\log p \tag{19}$$

hold.

Based on Corollary 2, we propose the following algorithm for constructing pseudorandom binary sequences of a given length p (where p is prime) :

THE ALGORITHM. Suppose a prime p and an integer $L \in \mathbb{N}$ are given with

$$L < \begin{cases} p \text{ for all } p \\ \dfrac{p}{4} \text{ if 2 is not primitive root modulo } p \end{cases} \tag{20}$$

(and typically L is "much smaller", than $p/4$, say, $L < p^{1/4}$). Suppose we want to control the correlation of order ℓ for all $\ell \le L$.

Let $L \in \mathbb{N}$ be a "large" number but such that

$$k < \frac{\log\ p}{\log(4L)} \quad \text{if 2 is not a primitive root modulo } p. \tag{21}$$

Write $t = [k/2]$. Then consider polynomials $g(X) \in \mathbf{F}_p[X]$ of the form

$$g(X) = X^k + \sum_{i=0}^{t} a_i X^i \tag{22}$$

where any coefficients a_i can be chosen with

$$a_i \in \mathbb{Z}_p \text{ for } i = 0, 1, \ldots, t-1 \text{ and } a_t \in \mathbb{Z}_p \backslash 0. \tag{23}$$

Let $d(X)$ denote the greatest common divisor of the polynomials $g(X)$ and $g'(X)$, and compute

$$f(X) = \frac{g(X)}{d(X)} = \frac{g(X)}{(g(X), g'(X))}. \tag{24}$$

Then computing the sequence E_p defined as in Theorem 1, we obtain a "good" pseudorandom sequence E_p.

Indeed, it follows from (24) that $f(X)$ has no multiple zeros. The degree of $f(X)$ is

$$\deg f(X) \le \deg g(X) = k.$$

Moreover, by (20) and (21), for all $\ell \le L$ we have

$$\ell \le L < \begin{cases} p \text{ for all } p \\ \dfrac{1}{4}\, p^{1/4} \text{ if 2 is not a primitive root modulo } p \end{cases}$$

so that the assumptions in Corollary 2 hold. Thus it follows from Corollary 2 that both (18) and (19) (for all $\ell \le L$) holf so that, indeed, every E_p definied in this way is of the desired properties.

Note that $a_0, a_1, \ldots, a_t$ in (22) and (23) can be chosen in $(p-1)p^{t-1} \ge \dfrac{1}{2}p^t = p^{[k/2]}$ ways so that there are "many" polynomials $g(X)$ of form (22). Moreover, for all $g(X)$ we have

$$d(X) | (kg(X) - Xg'(X)) = (k-t)a_t X^t + \ldots$$

(where $(k-t)a_t X^t$ is the highest degree term) whence $\deg d(X) \le t$

so that

$$\deg f(X) = \deg g(X) - \deg d(X) \ge k - t = k - [k/2] \ge k/2.$$

This shows that although different polynomials $g(X)$ may lead to the same polynomial $f(X)$, there is little chance for this, secondly, when reducing the polynomials $g(X)$ to the polynomials $f(X)$, the resulting polynomials in general will not be "too simple", "degenerated" (say, linear) polynomials, in other words, the family of the sequences E_p defined in this way is both "large" and "of high complexity", we will return to this question in a subsequent paper.

REFERENCES

[1] E. Artin, Collected Papers, Ed. S. Lang and J. T. Tate, New York, Springer-Verlag, pp. viii-ix, 1965.

[2] C. Hooley, On Artin's Conjecture, J. reine angew. Math. 225 (1967), 209-220.

[3] R. Lidl and H. Niederreiter, Finite Fields, Encyclopedia of Math. Appl., Vol. 20, Addison-Wesley, Reading, MA, 1983.

[4] Yu. V. Linnik, Ergodic Properties of Algebraic Fields, Ergebnisse der Mathematik und Ihrer Grenzgebiete, Vol. 45, Springer, 1968. In Chapter 10.

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

[6] C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences II : The Champernowne, Rudin-Shapiro, and Thue - Morse sequences. A further construction, J. Number Theory 73 (1998), 256-276.

[7] C. Mauduit and A. Sárközy, On finite pseudorandom sequences of k symbols, Indag. Mathem., 13 (2002), 89-101.

[8] A. Sárközy, Number Theory and Its Applications (in Hungarian), Müszaki Könyvkiadó, Budapest, 1978.

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

[10] W. Schmidt, Equations over Finite Fields. An Elementary Approach, Lecture Notes in Math. 536, Springer, New York, 1976.

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