On the complexity of a family related to the Legendre symbol

Katalin Gyarmati

Abstract

Ahlswede, Khachatrian, Mauduit and A. Sárközy introduced the notion family-complexity of families of binary sequences. They estimated the family-complexity of a large family related to Legendre symbol introduced by Goubin, Mauduit and Sárközy. Here their result is improved, and apart from the constant factor the best lower bound is given for the family-complexity.

2000 AMS Mathematics Subject Classification: 11K45.

List of keywords and phrases: pseudorandom, f-complexity.

1 Introduction

In this paper we study large families of finite, binary sequences

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

Research partially supported by Hungarian National Foundation for Scientific Research, Grants No. K67676 and PD72264 and the János Bolyai Research Fellowship.

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

Definition 1 *The complexity $C(\mathcal{F})$ of a family $\mathcal{F}$ of binary sequences $E_N \in \{-1,+1\}^N$ is defined as the greatest integer j so that for any $1 \le i_1 < i_2 < \cdots < i_j \le N$, and for $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_j \in \{-1,+1\}^j$, we have at least one $E_N = \{e_1, \dots, e_N\} \in \mathcal{F}$ for which*

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

In order to get an upper bound for $C(\mathcal{F})$, we take all specifications of the form

$$e_1 = \varepsilon_1,\ e_2 = \varepsilon_2, \dots, e_{C(\mathcal{F})} = \varepsilon_{C(\mathcal{F})}. \tag{1}$$

By the definition of f-complexity, for such a specification, there is a sequence $E \in \mathcal{F}$ for which (1) holds. $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{C(\mathcal{F})}$ may take $2^{C(\mathcal{F})}$ different values, thus,

$$2^{C(\mathcal{F})} \le |\mathcal{F}|\,.$$

So:

Proposition 1

$$C(\mathcal{F}) \le \frac{\log |\mathcal{F}|}{\log 2}.$$

Numerous binary sequences have been tested for pseudorandomness by J. Cassaigne, Z. Chen, X. Du, L. Goubin, K. Gyarmati, S. Ferenczi, S. Li, H. Liu, C. Mauduit, L. Mérai, J. Rivat and A. Sárközy. However, the first constructions produced only "few" pseudorandom sequences, usually for a fixed integer N, the construction provided only one pseudorandom sequence E_N of length N. First L. Goubin, C. Mauduit, A. Sárközy [2] succeeded in constructing large families of pseudorandom binary sequences. Their construction was the following:

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

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{for } (f(n), p) = 1 \\ +1 & \text{for } p \mid f(n). \end{cases} \tag{2}$$

Ahlswede, Khachatrian, Mauduit and Sárközy [1] proved the following:

Theorem B *Let p be a prime. Consider all the polynomials $f(x)$ such that*

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

(where $\deg f(x)$ denotes the degree of $f(x)$) and $f(x)$ has no multiple zero in $\overline{\mathbf{F}}_\mathbf{p}$. For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = (e_1, e_2, \dots, e_p) \in \{-1, +1\}^p$ defined by (2), and let $\mathcal{F}_1$ denote the family of all the binary sequences obtained in this way. Then

$$C(\mathcal{F}_1) \geq K. \tag{3}$$

By Proposition 1 it is clear that

$$|C(\mathcal{F}_1)| \leq \frac{\log |\mathcal{F}_1|}{\log 2} \leq \frac{K}{\log 2} \log p.$$

We will improve on (3) and we will prove the following:

Theorem 1 *Let $p \geq 3$ be a prime. Consider all the polynomials $f(x)$ such that*

$$0 \leq \deg f(x) \leq K$$

(where $\deg f(x)$ denotes the degree of $f(x)$) and $f(x)$ has no multiple zero in $\overline{\mathbf{F}}_{\mathbf{p}}$. For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = (e_1, e_2, \dots, e_p) \in \{-1, +1\}^p$ defined by (2), and let $\mathcal{F}_2$ denote the family of all the binary sequences obtained in this way. Then

$$C(\mathcal{F}_2) \geq \frac{K}{2\log 2}\log p - O(K\log(K\log p)). \tag{4}$$

2 Proof of Theorem 1

In this proof c_1, c_2 will denote absolute constants. For $K \geq p^{1/2}/\log p$ the right-hand side of (4) is negative, so the theorem is trivial. Thus we may suppose that

$$K < p^{1/2}/\log p. \tag{5}$$

Let k be the greatest odd integer with $k \leq K$. Let

$$j \leq \frac{k}{2\log 2}\log p - \frac{c_1 k}{\log 2}\log(k\log p), \tag{6}$$

where we will fix the value of the absolute constant c_1 later. Suppose that we have the specification

$$e_{n_1} = \varepsilon_1,\ e_{n_2} = \varepsilon_2, \dots, e_{n_j} = \varepsilon_j. \tag{7}$$

Let $I = \{n_1, n_2, \dots, n_j\}$. We will consider all polynomials $f(x)$ of the form

$$f_{a_1,a_2,\dots,a_k}(x) = (x-a_1)(x-a_2)\cdots(x-a_k) \tag{8}$$

with $a_i \notin I$, and we will prove by a counting argument that there is at least one k-tuple $a_1, a_2, \dots, a_k$ (where $a_i \notin I$) for which the sequence E_p defined by (2) with $f_{a_1,a_2,\dots,a_k}(x)$ in place of $f(x)$ satisfies (7). Suppose that $\beta_1, \beta_2, \dots, \beta_t$ are the roots of $f(x)$ which have odd multiplicity in the factorization of $f(x)$. Since the degree of $f(x)$ is odd, t the number of these roots are also odd, so $t \geq 1$. Let $f_1(x) = (x-\beta_1)(x-\beta_2)\dots(x-\beta_t)$. Then $f_1(x)$ has no multiple zero and the sequence E'_p defined by (2) with $f_1(x)$ in place of $f(x)$ satisfies (7).

Since this will be true for every $j \leq \frac{k}{2\log 2}\log p - \frac{c_1 k}{\log 2}\log(k\log p)$ from this

$$C(\mathcal{F}) \geq \left[\frac{k}{2\log 2}\log p - \frac{c_1 k}{\log 2}\log(k\log p)\right] \geq \frac{K}{2\log 2}\log p - c_2 K\log(K\log p)$$

follows.

Now consider a k-tuple $a_1, a_2, \dots, a_k$ with $a_i \notin I$, and consider the corresponding polynomial

$$f_{a_1,a_2,\dots,a_k}(x) = (x-a_1)(x-a_2)\cdots(x-a_k).$$

Define the sequence $E_p = \{e_1, e_2, \dots, e_p\}$ by

$$e_n = \begin{cases} \left(\frac{f_{a_1,a_2,\dots,a_k}(n)}{p}\right) & \text{if } (f_{a_1,\dots,a_k}(n), p) = 1, \text{ so } n \neq a_i \text{ for } 1 \leq i \leq k, \\ 1 & \text{if } p \mid f_{a_1,\dots,a_k}(n), \text{ so } n = a_i \text{ for some } 1 \leq i \leq k. \end{cases} \tag{9}$$

Clearly,

$$\frac{1}{2}\left(1 + \varepsilon_i e_{n_i}\right) = \begin{cases} 1 & \text{if } e_{n_i} = \varepsilon_i, \\ 0 & \text{if } e_{n_i} = -\varepsilon_i. \end{cases}$$

If $n_i \neq a_s$ for $1 \leq s \leq l$ then

$$\frac{1}{2}\left(1+\varepsilon_i\left(\frac{(n_i-a_1)(n_i-a_2)\cdots(n_i-a_k)}{p}\right)\right) = \begin{cases} 1 & \text{if } e_{n_i} = \varepsilon_i, \\ 0 & \text{if } e_{n_i} = -\varepsilon_i. \end{cases}$$

Let N be the number of polynomials $f_{a_1,a_2,\dots,a_k}(x) \in \mathbf{F_p}[x]$ with $a_1, a_2, \dots, a_k \in \mathbb{F}_p \setminus I$ such that for the sequence (9) specification (7) holds. Then

$$N = \sum_{\substack{a_1=0 \\ a_1 \notin I}}^{p-1} \sum_{\substack{a_2=0 \\ a_2 \notin I}}^{p-1} \cdots \sum_{\substack{a_k=0 \\ a_k \notin I}}^{p-1} \frac{1}{2^j} \prod_{i=1}^{j} \left(1+\varepsilon_i\left(\frac{(n_i-a_1)(n_i-a_2)\cdots(n_i-a_k)}{p}\right)\right). \tag{10}$$

Here

$$A(a_1,\dots,a_k) \overset{\text{def}}{=} \prod_{i=1}^{j}\left(1+\varepsilon_i\left(\frac{(n_i-a_1)\cdots(n_i-a_k)}{p}\right)\right) = 1 + \sum_{\ell=1}^{j} \sum_{1\leq i_1<i_2<\dots<i_\ell\leq j} \varepsilon_{i_1}\varepsilon_{i_2}\cdots\varepsilon_{i_\ell}\left(\frac{(n_{i_1}-a_1)\cdots(n_{i_1}-a_k)}{p}\right)\left(\frac{(n_{i_2}-a_1)\cdots(n_{i_2}-a_k)}{p}\right)\cdots \left(\frac{(n_{i_\ell}-a_1)\cdots(n_{i_\ell}-a_k)}{p}\right).$$

The Legendre symbol is multiplicative, thus

$$A(a_1,\dots,a_k) = 1 + \sum_{\ell=1}^{j} \sum_{1\leq i_1<i_2<\dots<i_\ell\leq j} \varepsilon_{i_1}\varepsilon_{i_2}\cdots\varepsilon_{i_\ell} \prod_{j=1}^{k}\left(\frac{(n_{i_1}-a_j)(n_{i_2}-a_j)\dots(n_{i_\ell}-a_j)}{p}\right).$$

Writing this in (10) we get

$$N = \sum_{\substack{a_1=0\\ a_1\notin I}}^{p-1} \cdots \sum_{\substack{a_k=0\\ a_k\notin I}}^{p-1} \frac{1}{2^j}\left(1+\sum_{\ell=1}^{j} \sum_{1\le i_1<i_2<\cdots<i_\ell\le j} \varepsilon_{i_1}\cdots\varepsilon_{i_\ell}\right.$$

$$\left.\prod_{t=1}^{k}\left(\frac{(n_{i_1}-a_t)(n_{i_2}-a_t)\dots(n_{i_\ell}-a_t)}{p}\right)\right)$$

$$= \frac{(p-|I|)^k}{2^j} + \frac{1}{2^j}\sum_{\substack{a_1=0\\ a_1\notin I}}^{p-1} \cdots \sum_{\substack{a_k=0\\ a_k\notin I}}^{p-1} \sum_{\ell=1}^{j} \sum_{1\le i_1<i_2<\cdots<i_\ell\le j} \varepsilon_{i_1}\cdots\varepsilon_{i_\ell}$$

$$\prod_{t=1}^{k}\left(\frac{(n_{i_1}-a_t)(n_{i_2}-a_t)\dots(n_{i_\ell}-a_t)}{p}\right)$$

$$= \frac{(p-j)^k}{2^j} + \frac{1}{2^j}\sum_{\ell=1}^{j} \sum_{1\le i_1<i_2<\cdots<i_\ell\le j} \varepsilon_{i_1}\cdots\varepsilon_{i_\ell} \sum_{\substack{a_1=0\\ a_1\notin I}}^{p-1} \cdots \sum_{\substack{a_k=0\\ a_k\notin I}}^{p-1}$$

$$\prod_{t=1}^{k}\left(\frac{(n_{i_1}-a_t)(n_{i_2}-a_t)\dots(n_{i_\ell}-a_t)}{p}\right)$$

$$= \frac{(p-j)^k}{2^j} + \frac{1}{2^j}\sum_{\ell=1}^{j} \sum_{1\le i_1<i_2<\cdots<i_\ell\le j} \varepsilon_{i_1}\cdots\varepsilon_{i_\ell}$$

$$\left(\sum_{\substack{a=0\\ a\notin I}}^{p-1} \frac{(n_{i_1}-a)(n_{i_2}-a)\dots(n_{i_\ell}-a)}{p}\right)^k. \tag{11}$$

Lemma 1 *Suppose that* p *is a prime,* χ *is a non-principal character modulo* p *of order* d*,* $f \in \mathbf{F}_p[x]$ *has* s *distinct roots in* $\overline{\mathbf{F}}_p$*, and it is not the constant multiple of the* d*-th power of a polynomial over* $\mathbf{F}_p$*. Then*

$$\left|\sum_{n\in\mathbf{F_p}} \chi(f(n))\right| \le sp^{1/2}.$$

Poof of Lemma 1

This is Weil's theorem, see [3].

By the triangle-inequality and by Lemma 1:

$$\left|\sum_{\substack{a=0\\ a\notin I}}^{p-1}\left(\frac{(n_{i_1}-a)(n_{i_2}-a)\dots(n_{i_\ell}-a)}{p}\right)\right|\le$$
$$\left|\sum_{a=0}^{p-1}\left(\frac{(n_{i_1}-a)(n_{i_2}-a)\dots(n_{i_\ell}-a)}{p}\right)\right|+j\le \ell p^{1/2}+j\le jp^{1/2}+|I|\,.$$

Thus by (11) and the triangle-inequality

$$N\ge\frac{(p-j)^k}{2^j}-\frac{1}{2^j}\sum_{\ell=1}^{j}\sum_{1\le i_1<i_2<\dots<i_\ell\le j}(jp^{1/2}+j)^k=\frac{(p-j)^k}{2^j}-(jp^{1/2}+j)^k.$$

Thus $N>0$ follows from

$$\frac{p-j}{2^{j/k}}>jp^{1/2}+j$$
$$p>2^{j/k}(jp^{1/2}+j)+j. \tag{12}$$

Thus it remains to prove (12). By (6)

$$2^{j/k}(jp^{1/2}+j)\le 2^{\left(\frac{1}{2\log 2}\log p-\frac{c_1}{\log 2}\log(k\log p)\right)}\left(\frac{k}{2\log 2}p^{1/2}\log p+\frac{k}{2\log 2}\log p\right)$$
$$+\frac{k}{2\log 2}\log p\le\frac{p^{1/2}}{(k\log p)^{c_1}}\left(\frac{k\log p}{2\log 2}p^{1/2}+\frac{kp^{1/2}\log p}{2\cdot 3^{1/2}\log 2}\right)$$
$$+\frac{k}{2\log 2}\log p\le\frac{p^{1/2}}{(k\log p)^{c_1}}1.138(k\log p)p^{1/2}+\frac{k}{2\log 2}\log p.$$

By this and (5)

$$\begin{aligned}2^{j/k}(jp^{1/2}+j)&\le 1.138\frac{p}{(k\log p)^{c_1-1}}+\frac{p^{1/2}}{2\log 2}\\&\le 1.138\frac{p}{(k\log p)^{c_1-1}}+\frac{p}{2\cdot 3^{1/2}\log 2}\\&\le 1.138\frac{p}{(k\log p)^{c_1-1}}+0.414p.\end{aligned}$$

For $c_1=9$ we have

$$2^{j/k}(jp^{1/2}+j)\le 1.138\frac{p}{(\log 3)^8}+0.414p<p$$

which proves (12). Thus for $j \leq \frac{k}{2\log 2}\log p - \frac{9k}{\log 2}\log(k\log p)$ we have that (12) holds. Then $N > 0$. So there is a sequence E_p for which specification (7) holds. Thus we proved

$$C(\mathcal{F}) \geq \left[\frac{k}{2\log 2}\log p - \frac{9k}{\log 2}\log(k\log p)\right] \geq \frac{K}{2\log 2}\log p - O(K\log(K\log p)).$$

References

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

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

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

Eötvös Loránd University
Department of Algebra and Number Theory
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
email: gykati@cs.elte.hu