A Complexity Measure for Families of Binary Sequences

Rudolf Ahlswede and Levon H. Khachatrian
Fakultät für Mathematik, Universität Bielefeld
Postfach 100131, D–33501 Bielefeld, Germany,
e-mail: ahlswede@mathematik.uni-bielefeld.de
C. Mauduit
Institut de Mathématiques de Luminy
CNRS–UPR 9016, 163 Avenue de Luminy, Case 907
F–13288 Marseille Cedex 9,France
e-mail: mauduit@iml.univ-mrs.fr
and
A. Sárközy *
Eötvös Loránd University
Department of Algebra and Number Theory
H–1117 Budapest, Pázmány Péter Sétány 1/c, Hungary
e-mail: sarkozy@cs.elte.hu

Abstract

In earlier papers finite pseudorandom binary sequences were studied, quantitative measures of pseudorandomness of them were introduced and studied, and large families of "good" pseudorandom sequences were constructed. In certain applications (cryptography) it is not enough to know that a family of "good" pseudorandom binary sequences is large, it is a more important property if it has a "rich", "complex" structure. Correspondingly, the notion of "f–complexity" of a family of binary sequences is introduced. It is shown that the family of "good" pseudorandom binary sequences constructed earlier is also of high f–complexity. Finally, the cardinality of the smallest family achieving a prescibed f–complexity and multiplicity is estimated.

* Research partially supported by Hungarian National Foundation for Scientific Research, Grant No. T029759 and MKM fund FKFP–0139/1997.

1991 AMS Mathematics Subject Classification number: primary 11K45, secondary 94A60, 68Q25
List of key words and phrases: pseudorandom, complexity, binary sequence, specification, Legendre symbol, covering hypergraphs.

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 [4] 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

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

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

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

where the maximum is taken over all $D = (d_1, \dots, d_k)$ and M such that $M + d_k \leq N$. Then the sequence E_N is considered as a "good" pseudorandom sequence, if both these measures $W(E_N)$ and $C_k(E_N)$ (at least for small k) are "small" in terms of N (in particular, both are $o(N)$ as $N \to \infty$). Indeed, it is shown in [2] that for a "truely random" $E_N \in \{-1, +1\}^N$ both, $W(E_N)$ and, for fixed k, $C_k(E_N)$, are around $N^{1/2}$ with "near 1" probability.

Moreover, it was shown in [4] that the Legendre symbol forms a "good" pseudorandom sequence. More exactly, let p be an odd prime, and

$$N = p - 1, e_n = \left(\frac{n}{p} \right), E_N = (e_1, \dots, e_N). \tag{1.1}$$

Then by Theorem 1 in [4] we have

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

and

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

(Here $\ll$ is Vinogradov's notation, i.e., $f(x) \ll g(x)$ means $f(x) = O\big(g(x)\big)$.)

Later this construction was extended [5], and another modular construction of a "good" pseudorandom sequence was given in [7]. Numerous other binary sequences have been tested for pseudorandomness by Cassaigne, Ferenczi, Mauduit, Rivat and Sárközy, but none of them proved to be nearly as "good" as the ones mentioned above.

However, these "good" constructions produce only a "few" good sequences while in many applications, e.g., in cryptography, one needs "large" families of "good" pseudorandom binary sequences. Note that here we speak of *families* of binary sequences instead of *sets* of them. The reason is that there is a natural and often used bijection between binary sequences $E_N = (e_1, \dots, e_N) \in \{-1, +1\}^N$ and subsets of $\{1, 2, \dots, N\}$:

$$E_N = (e_1, \dots, e_N) \leftrightarrow \{n : 1 \le n \le N, e_n = +1\}.$$

This bijection maps *sets* of binary sequences onto *families* of subsets which would force us to use the terminologies "set" and "family" alternately. This may cause a confusion; to avoid this, it is simpler and safer to use the terminology "family" in both cases.

Very recently, Goubin, Mauduit and Sárközy [3] succeeded in constructing **large families** of pseudorandom binary sequences. Their most important results can be summarized as follows:

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

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{for } (f(n), p) = 1 \\ +1 & \text{for } p \mid f(n), \end{cases} \tag{1.2}$$

then we have

$$W(E_p) < 10kp^{1/2} \log p. \tag{1.3}$$

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

(i) $\ell = 2$;

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

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

Then we also have

$$C_\ell(E_p) < 10k\ell p^{1/2} \log p. \tag{1.4}$$

(Note that the crucial tool in the proofs of (1.3) and (1.4) is an estimate for incomplete character sums of the form $\sum\limits_{A<x<B} \chi\big(f(x)\big)$, where $\chi \ne \chi_o$ is a character modulo p and $f(x) \in F_p[x]$. This estimate was deduced in [4] from a theorem of Weil [9]. Examples

show that, perhaps, conditions (i) – (iii) can be relaxed, but that they cannot be omitted completely.)

It is easy to see that this theorem generates "large" families of "good" pseudorandom binary sequences. However, in many applications it is not enough to know that our family $\mathcal{F}$ of "good" binary sequences is large; it can be much more important to know that $\mathcal{F}$ has a "rich", "complex" structure, there are many "independent" sequences in it. Thus one might like to introduce a quantitative measure of the complexity of the structure of families of binary sequences. Before defining such a measure, consider the following well–known model.

Suppose a family $\mathcal{F}$ of binary sequences $E_N \in \{-1,+1\}^N$ is given, that we want to use in cryptography as key space.

First a sequence $E_N = (e_1, \dots, e_N) \in \mathcal{F}$ is choosen as the *key*. If the sequences in $\mathcal{F}$ are expressed in terms of certain parameters, then picking E_N means to fix the values of these parameters, and the sender and receiver of the messages may let each other know about the choice of the parameters (by using public key cryptosystem for instance). Using this key $E_N = (e_1, \dots, e_N)$, the message to be sent can be coded in the following way: first the text to be sent should be expressed in terms of binary sequences U_N belonging to $\{-1,+1\}^N$, and then each of these sequences U_N is coded by

$$U_N = (u_1, \dots, u_N) \xrightarrow{E_N} V_N = (v_1, \dots, v_n) = (e_1 u_1, \dots, e_N u_N);$$

then the message consists of these binary sequences $V_N = E_N(U_N)$. To decode the message, one has to repeat this operation, so that

$$U_N = E_N(V_N) = (e_1 v_1, \dots, e_N v_N) = (u_1, \dots, u_N).$$

(This procedure is clearly equivalent to using 0–1 sequences and adding them modulo 2.)

Suppose now that there is an eavesdropper, who wants to break the code. He hopes to do this in two steps:

Step 1. He hopes that the coding operation $U_N \xrightarrow{E_N} V_N$ does not destroy the structural properties (certain regularities, repetitions, etc.) of the message to be coded completely, and by using the remaining structure, he is able to determine many elements e_i of the key $E_N = (e_1, \dots, e_N)$, say, that is he finds a **specification** of E_n of length j

$$e_{i_1} = \varepsilon_1, \dots, e_{i_j} = \varepsilon_j \quad (i_1 < \cdots < i_j) \tag{1.5}$$

for a possibly large j.

Step 2. He hopes that the specification (1.5) uniquely determines the rest of the key E_N (or, in the worst case, there are only a few E_N's with specification (1.5) so that it suffices to check a few possibilities only).

Thus our cryptosystem has reasonable security properties if both, Step 1 and 2, are made possibly, ideally hopelessly, difficult. In case of Step 1, this goal can be achieved by taking "good" pseudorandom sequences E_N; this is the problem studied by Mauduit and Sárközy, and satisfactorily settled in form of Theorem A above. In case of Step 2, the requirement motivates the following definition of f–complexity of families of binary sequences, which is also inspired by the concept of **local randomness** introduced in [8] and further studied in [6]. It is also related to the concept of **unicity** introduced by Shannon. "Complexity" is an often used word but it usually concerns a property of single sequences and not that of families of sequences. Moreover, it expresses the nature of the property to be studied by us very well. Thus in spite of the other uses of the word we propose the use of it in this context as well, but to avoid confusion we will speak about f–complexity ("f" for family).

Definition. The *f–complexity* $\Gamma(\mathcal{F})$ 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 specification (1.5) there is at least one $E_N \in \mathcal{F}$ which satisfies it. The f–complexity of $\mathcal{F}$ is denoted by $\Gamma(\mathcal{F})$. (If there is no $j \in \mathbb{N}$ with the property above, we set $\Gamma(\mathcal{F}) = 0$.)

Now indeed, for $j < \Gamma(\mathcal{F})$ and any specification (1.5) of length j, clearly there are at least $2^{\Gamma(\mathcal{F})-j}$ sequences $E_N \in \mathcal{F}$ with the given specification. This means that the eavesdropper knowing a specification of length j has many options for the true key and therefore a difficult task, if $\Gamma(\mathcal{F})$ is **large**.

We conclude that if we can construct *a family $\mathcal{F}$ of high f–complexity and of "good" pseudorandom binary sequences*, then *the cryptosystem based on it* (as described above) *has good security properties.*

Therefore we are looking for families $\mathcal{F}$ which have as their elements "good" pseudorandom sequences E_N and **at the same time** have large f–complexity $\Gamma(\mathcal{F})$. Our finding is Theorem 1 in Section 2, which shows that the **construction** in Theorem A does not only give a set with "good" pseudorandom sequences, but under condition (iii) also with "high" f–complecity.

Next we adress the question how large a family (key space) $\mathcal{F}$ is needed to achieve a prescribed f–complexity.

A simple observation illustrates the difference between large size of $\mathcal{F}$ and f–complexity $\Gamma(\mathcal{F})$.

Let

$$\mathcal{F} = \left\{ E_N = (e_1, \dots, e_N) : E_N \in \{-1,+1\}^N, e_N = +1 \right\}.$$

This family contains 2^{N-1} sequences E_N, which is half of the total number of binary sequences of length N, so that this is a "very large" family. On the other hand, there is no $E_N = (e_1, \dots, e_N) \in \mathcal{F}$ with specification

$$e_N = -1$$

so that $\Gamma(\mathcal{F}) = 0$.

On the other hand, we will see that high f–complexity enforces large size.

Theorem 2 in Section 3 gives rather sharp estimates on the values of these two quantities. In addition it incorporates multiplicity of $\mathcal{F}$, a parameter measuring the ambiguity for the eavesdropper in selecting a key confining to the specification known to him.

2 Construction of a family of pseudorandom sequences with high f–complexity

The following result significantly improves Theorem A, because it establishes an additional complexity property of the construction, expressed in (2.6).

Theorem 1. *Let p be a prime number, and $K \in \mathbb{N}$, $L \in \mathbb{N}$,*

$$(4K)^L < p. \tag{2.1}$$

Consider all the polynomials $f(x) \in F_p[x]$ with the properties that

$$0 < \deg f(x) \le K \tag{2.2}$$

(where $\deg f(x)$ *denotes the degree of $f(x)$) and*

$$f(x) \textit{ has no multiple zero in } \overline{F}_p. \tag{2.3}$$

For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = (e_1, \dots, e_p) \in \{-1, +1\}^p$ defined by (1.2), and let $\mathcal{F}$ denote the family of all the binary sequences obtained in this way. Then for all $E_p \in \mathcal{F}$ we have

$$W(E_p) < 10Kp^{1/2} \log p \tag{2.4}$$

and

$$C_\ell(E_p) < 10KLp^{1/2} \log p \textit{ for all } \ell \in \mathbb{N}, 1 < \ell \le L. \tag{2.5}$$

Moreover, we have

$$\Gamma(\mathcal{F}) \ge K. \tag{2.6}$$

Note that his result is based on the use of assumption (iii) in Theorem A. If we are satisfied with estimating $C_2(E_p)$ and we do not insist on estimating correlations of higher order as well, then we may use assumption (i) instead of (iii), and then the strong assumption (2.1) in Theorem 1 can be dropped.

Moreover, if we assume that p is a prime such that 2 is a primitive root modulo p, then we may use (ii) in Theorem A instead of (iii), and then assumption (2.1) in Theorem 1 can be replaced by the much weaker $L < p$, which is natural anyway. However, the problem with this approach is that we do not know whether 2 is a primitive root for infinitely many primes p. By a well–known conjecture of Artin, there are infinitely many primes p with this property. Unfortunately, there is no hope for proving this conjecture in the near future, and even if it gets proved, we would also need information on the distribution of these primes and a good algorithm for finding such a p.

Proof of Theorem 1. We will need the following simple result:

Lemma 1. *If T is a field and $g(x) \in T[x]$ is a non–zero polynomial, then it can be written in the form*

$$g(x) = \big(h(x)\big)^2 g^*(x) \tag{2.7}$$

where $h(x) \in T[x]$, $g^*(x) \in T[x]$ and $g^*(x)$ has no multiple zero in $\overline{T}$ (the algebraic closure of T).

Note that by using the notion and properties of the derivative of a polynomial (which can be defined over any field) a simple and fast algorithm can be given for determining these polynomials $h(x)$ and $g^*(x)$.

Proof. By the theorem on unique factorization into irreducible factors, $g(x)$ can be written as

$$g(x) = a\big(p_1(x)\big)^{2\alpha_1+1} \dots \big(p_r(x)\big)^{2\alpha_r+1}\big(q_1(x)\big)^{2\beta_1} \dots \big(q_s(x)\big)^{2\beta_s}$$

where $a \in T$, the p_i's and q_i's are irreducible polynomials over T which are pairwise essentially distinct (none of them is the constant multiple of another one), the α_i's are non–negative integers and the β_i's are positive integers. Write

$$h(x) = \big(p_1(x)\big)^{\alpha_1} \dots \big(p_r(x)\big)^{\alpha_r}\big(q_1(x)\big)^{\beta_1} \dots \big(q_s(x)\big)^{\beta_s}$$

and

$$g^*(x) = ap_1(x) \dots p_r(x).$$

Then (2.7) holds trivially. Moreover, the polynomials $p_i(x)$, being irreducible, cannot have multiple zero (this follows from $\big(p_i(x), p_i'(x)\big) \mid p_i(x)$), and the zeros of the pairwise essentially distinct irreducible polynomials $p_i(x)$ are distinct. Thus $g^*(x)$ cannot have multiple zero either, and this completes the proof of the lemma.

To prove Theorem 1, we have to show that for any specification of length K:

$$e_{i_1} = \varepsilon_1, \dots, e_{i_K} = \varepsilon_K \quad (i_1 < \dots < i_K), \tag{2.8}$$

there is a polynomial $f(x) \in F_p[x]$ which satisfies (2.2) and (2.3) so that $E_p = E_p(f) \in \mathcal{F}$, and this sequence $E_p = E_p(f)$ satisfies the specification (2.8), and, finally, (2.4) and (2.5) also hold.

(Throughout the proof we will not distinguish between a number $a \in \mathbb{Z}$, the residue class represented by a modulo p, and the corresponding element of F_p. Moreover, if we write, say, $\frac{f(x)}{a}$ with $a \in \mathbb{Z}$, $a \neq 0 \ (\mod p)$ and $f(x) \in F_p(x)$, then we mean $f(x)a^{-1}$ where a^{-1} is the multiplicative inverse of that element of F_p which corresponds to a.)

By (2.1), we have $K < p$, thus there is an integer i_{K+1} with

$$0 < i_{K+1} \leq p,\ i_{K+1} \notin \{i_1, \dots, i_K\}.$$

Let

$$\varepsilon_{K+1} = -\varepsilon_1, \tag{2.9}$$

let q, r be integers with $(q, p) = (r, p) = 1$ and

$$\left(\frac{q}{p}\right) = +1,\ \left(\frac{r}{p}\right) = -1, \tag{2.10}$$

and define $y_1, \dots, y_{K+1}$ by

$$y_i = \begin{cases} q & \text{if } \varepsilon_j = +1 \\ r & \text{if } \varepsilon_j = -1 \end{cases} \quad (\text{for } j = 1, 2, \dots, K+1). \tag{2.11}$$

By the well–known theorem on interpolation, there is a unique polynomial $g(x) \in F_p[x]$ with

$$\deg g(x) \le K \tag{2.12}$$

and

$$g(i_j) = y_j \text{ for } j = 1, 2, \dots, K+1. \tag{2.13}$$

Indeed, this polynomial can be determined by using either Lagrange interpolation or Newton interpolation; e.g., by using Lagrange interpolation, we obtain

$$g(x) = \sum_{j=1}^{K+1} y_j \prod_{\substack{1 \le t \le K+1 \\ t \ne j}} \frac{x - i_t}{i_j - i_t}.$$

(Clearly, this formula can be used for interpolating over any field.)
By Lemma 1, this polynomial $g(x)$ can be written in form (2.7) (where now $T = F_p$). Let

$$f(x) = g^*(x). \tag{2.14}$$

Then by Lemma 1, (2.3) holds. It follows from (2.7), (2.12) and (2.14) that

$$\deg f(x) = \deg g^*(x) \le \deg g(x) \le K. \tag{2.15}$$

By (2.11) and (2.13) we have

$$g(i_j) = y_j = \begin{cases} q & \text{if } \varepsilon_j = +1 \\ r & \text{if } \varepsilon_j = -1 \end{cases} \quad (\text{for } j = 1, 2, \dots, K+1), \tag{2.16}$$

and by (2.10), this implies that

$$\big(g(i_j), p\big) = 1 \quad (\text{for } j = 1, 2, \dots, K+1) \tag{2.17}$$

so that by (2.7), (2.10), (2.14), (2.16) and (2.17) we have

$$\left(\frac{g(i_j)}{p}\right) = \left(\frac{\big(h(i_j)\big)^2}{p}\right)\left(\frac{g^*(i_j)}{p}\right) = \left(\frac{f(i_j)}{p}\right) = \left(\frac{y_i}{p}\right) = \begin{cases} \left(\frac{q}{p}\right) = +1 & \text{if } \varepsilon_j = +1 \\ \left(\frac{r}{p}\right) = -1 & \text{if } \varepsilon_j = -1. \end{cases} \tag{2.18}$$

It follows from (2.9) and (2.18) that

$$\left(\frac{f(i_1)}{p}\right) \ne \left(\frac{f(i_{K+1})}{p}\right)$$

and thus $f(x)$ is not constant, i.e.,

$$\deg f(x) > 0. \tag{2.19}$$

(2.2) follows from (2.15) and (2.19). Moreover, it follows from (1.2) and (2.18) that $E_p(f)$ satisfies the specification (2.8).

It remains to show that (2.4) and (2.5) also hold. By (1.2), (2.2) and (2.3), we may apply Theorem A to estimate $W(E_p)$ and $C_\ell(E_p)$. We obtain that

$$W(E_p) < 10\big(\deg f(x)\big)p^{1/2}\log p \leq 10Kp^{1/2}\log p.$$

Moreover, by (2.1) and (2.2), for all $1 < \ell \leq L$ we have

$$\big(4\deg f(x)\big)^{\ell} \leq (4K)^L < p$$

so that (iii) in Theorem A also holds with $\deg f(x)$ in place of k. Thus by Theorem A and (2.2), for all $1 < \ell \leq L$ we have

$$C_\ell(E_p) < 10\big(\deg f(x)\big)\ell p^{1/2}\log p \leq 10KLp^{1/2}\log p$$

which completes the proof of Theorem 1.

3 On the cardinality of a smallest family achieving a prescribed f–complexity and multiplicity

We consider for positive integers $j \leq K \leq N$ and M

$$\begin{aligned} S(N,j,M) \triangleq \min\{|\mathcal{F}| : \mathcal{F} \subset \{-1,+1\}^N,\ &\text{every} \\ j\text{–specification is covered by } \mathcal{F} &\text{ with multiplicity } \geq M\} \end{aligned} \tag{3.1}$$

and in particular

$$S(N,K) \triangleq S(N,K,1) = \min\big\{|\mathcal{F}| : \mathcal{F} \subset \{-1,+1\}^N, \Gamma(\mathcal{F}) = K\big\}, \tag{3.2}$$

that is, we want to know here how many sequences E_N are needed to cover all K–specifications in $\{-1,+1\}^N$. This can be formulated as a covering problem for the hypergraph $\mathcal{H}(N,K) = \big(\mathcal{V}(N,K), \mathcal{E}(N)\big)$, where $\mathcal{E}(N) = \{-1,+1\}^N$ is the edge set and the vertex set $\mathcal{V}(N,K)$ is defined as the set of K–specifications on $\mathcal{E}(N)$ or, equivalently, as set of $(N-K)$–dimensional subcubes of $\{-1,+1\}^N$ and thus

$$|\mathcal{V}(N,K)| = \binom{N}{K}2^K, |\mathcal{E}(N)| = 2^N. \tag{3.3}$$

$E_N \in \mathcal{E}(N)$ contains specification V iff $E_N \in V$.

We use

Covering Lemma 1 (see [1]). *For any hypergraph* $(\mathcal{V},\mathcal{E})$ *with*

$$\min_{V\in\mathcal{V}} \deg(v) \geq d \tag{3.4}$$

there exists a covering $\mathcal{C}\subset\mathcal{E}$ *with*

$$|\mathcal{C}| \leq \left\lceil \frac{|\mathcal{E}|}{d} \log|\mathcal{V}| \right\rceil .$$

Application to our hypergraph $\mathcal{H}(N,K)$ yields with $d=2^{N-K}$ a family $\mathcal{F}$ with $\Gamma(\mathcal{F})\geq K$,

$$|\mathcal{F}| \leq \left\lceil \frac{2^N}{2^{N-K}} \log \binom{N}{K} 2^K \right\rceil . \tag{3.5}$$

On the other hand one edge E_N covers exactly $\binom{N}{K}$ K–specifications and therefore by (3.3) necessarily

$$|\mathcal{F}| \geq 2^K \tag{3.6}$$

and together with (3.5)

$$2^K \leq S(N,K) \leq 2^K \log \binom{N}{K} 2^K \leq 2^K \cdot K \cdot \log N (K\geq 4). \tag{3.7}$$

As already mentioned in Section 1, to make Step 2 difficult for the eavesdropper, who observes j positions of $E_N\in\mathcal{F}$, we must construct a family $\mathcal{F}$ of high f–complexity $\Gamma(\mathcal{F})$. Then for $j<\Gamma(\mathcal{F})$ the multiplicity $M_j(\mathcal{F})$, that is, the least multiplicity of every j–specification satisfies

$$M_j(\mathcal{F}) \geq 2^{\Gamma(\mathcal{F})-j}, \tag{3.8}$$

because a j–specification can be extended to as many $\Gamma(\mathcal{F})$–specifications with the same support. Therefore

$$\min_{\mathcal{F}:\Gamma(\mathcal{F})\geq K} M_j(\mathcal{F}) \geq 2^{K-j} \tag{3.9}$$

and

$$S(N,j,2^{K-j}) \leq S(N,K) \leq 2^K \cdot K \cdot \log N (K\geq 4). \tag{3.10}$$

On the other hand, since $|\mathcal{V}(N,j)| = \binom{N}{j}2^j$ and an edge E_N covers exactly $\binom{N}{j}$ j–specification, necessarily

$$S(N,j,2^{K-j}) \geq 2^{K-j}\binom{N}{j}2^j\binom{N}{j}^{-1} = 2^K. \tag{3.11}$$

The fact that $S(N,K)$ and thus f–complexity contains almost complete information about the quantity $S(N,j,2^{K-j})$ concerning multiplicity deserves attention. (3.7), (3.10) and (3.11) establish

Theorem 2. *The cardinality* $S(N,K)$ *of a smallest family* $\mathcal{F}\subset\{-1,+1\}^N$ *with* f*–complexity* $\Gamma(\mathcal{F})=K$ *satisfies* *(i)* $2^K \leq S(N,K) \leq 2^K \log \binom{N}{K} 2^K \leq 2^K \cdot K \cdot \log N (K \geq 4)$.

Furthermore, the cardinality $S(N,j,2^{K-j})$ *of a smallest family* $\mathcal{F} \subset \{-1,+1\}^N$ *which covers every* j*–specification with multiplicity* $\geq 2^{K-j}$ $(K \geq j)$ *satisfies*

(ii) $2^K \leq S(N,j,2^{K-j}) \leq S(N,K) \leq 2^K \cdot K \cdot \log N (K \geq 4)$ *for all* $j \leq K \leq N$.

In a statistical context often K is proportional to N, say $K = \kappa N$ and thus we have the

Corollary. *For* $0 < \kappa \leq 1$

(i) $\lim\limits_{N \to \infty} \frac{1}{N} \log_2 S(N, \kappa N) = \kappa$

(ii) $\lim\limits_{N \to \infty} \frac{1}{N} \log_2 S(N, j, 2^{\kappa N - 1}) = \kappa$ *for every* $0 \leq j \leq \kappa N$.

Finally, Theorem 2 shows that the family constructed in Theorem 1 is clearly much larger than $S(4K^L, K)$. This shows that there is space for improvement.

References

[1] R. Ahlswede, Coloring hypergraphs: A new approach to multi–user source coding, Part I, Journ. of Combinatorics, Information and System Sciences, Vol. 4, No. 1, 76–115, 1979; Part II, Journ. of Combinatorics, Information and System Sciences, Vol. 5, No. 3, 220–268, 1980.

[2] J. Cassaigne, C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences VII: The measures of pseudorandomness, Acta Arith., to appear.

[3] L. Goubin, C. Mauduit and A. Sárközy, Construction of large families of pseudorandom binary sequences, J. Number Theory, to appear.

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

[5] C. Mauduit and A. Sárközy, On finite pseudorandom sequences of k symbols, Indagationes Math., to appear.

[6] U. Maurer and J.L. Massey, Local randomness in pseudorandom sequences, J. Cryptology 4, 135–149, 1991.

[7] A. Sárközy, A finite pseudorandom binary sequence, Studia Sci. Math. Hungar. 38, 377–384, 2001.

[8] C.P. Schnorr, On the construction of random number generators and random function generators, Advances in Cryptology – EUROCRYPT '88 (LNCS 330), 225–232, 1988.

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