Annales Univ. Sci. Budapest., Sect. Comp. 41 (2013) 347–354

COLLISION AND AVALANCHE EFFECT IN PSEUDORANDOM SEQUENCES

Viktória Tóth (Budapest, Hungary)

Dedicated to Professors Zoltán Daróczy and Imre Kátai on their 75th anniversary

Communicated by Bui Minh Phong

(Received May 31, 2013; accepted July 13, 2013)

Abstract. Pseudorandom binary sequences have many applications. In particular they play a crucial role in cryptography, for example as keystream in the well-known Vernam cipher. The notion of pseudorandomness can be defined in different ways. Recently a constructive theory of pseudorandomness of finite sequences has been developed and many constructions for binary sequences with strong pseudorandom properties have been given. Motivated by applications, Mauduit and Sárközy in [10] generalized and extended this theory from the binary case to k-ary sequences, i.e., to k symbols. They constructed large families of binary and k-ary sequences with strong pseudorandom properties. I adapted the notions of collision and avalanche effect in order to study these pseudorandom properties of two of the most important families in [12] and [13]. Later, the study of the pseudorandom properties mentioned above was extended to k-ary sequences in [14]. This is a survey of my results.

1. Introduction

Mauduit and Sárközy introduced the following notations and definitions in [9]:

Key words and phrases: Pseudorandom, binary sequence, k-ary sequence, collision, avalanche effect.
2010 Mathematics Subject Classification: 11K45.

Consider a binary sequence

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

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

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

where the maximum is taken over all a, b, t with $a, b, t \in \mathbb{N}, 1 \leq a+b \leq a+tb \leq N$. The *correlation measure of order* k of E_N is:

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

where the maximum is taken over all $D = (d_1, ..., d_k)$ (with $d_1 < ... < d_k$ are non-negative integers) and $M \in \mathbb{N}$ with $M + d_k \leq N$.

Then E_N is considered as a "good" pseudorandom sequence if both $W(E_N)$ and $C_k(E_N)$ (at least for "small" k) are "small" in terms of N. Indeed, later, Cassaigne, Mauduit and Sárközy [4] showed that this terminology is justified since for almost all $E_N \in \{-1, 1\}^N$, both $W(E_N)$ and $C_k(E_N)$ are less than $N^{1/2}(\log N)^c$. (See also [2].)

First the existence of collisions in a given family will be studied. This notion appears, e.g., in [11] (see also in [3]) and it can be adapted to our approach in the following way:

Assume that $N \in \mathbb{N}$, $\mathcal{S}$ is a given set and to each $s \in \mathcal{S}$ we assign a unique binary sequence

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

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

$$\mathcal{F} = \mathcal{F}(\mathcal{S}) = \{E_N(s) : s \in \mathcal{S}\}. \tag{1.1}$$

Definition 1.1. If $s \in \mathcal{S}, s' \in \mathcal{S}, s \neq s'$ and

$$E_N(s) = E_N(s'), \tag{1.2}$$

then (1.2) is said to be a *collision* in $\mathcal{F} = \mathcal{F}(\mathcal{S})$. If there is no collision in $\mathcal{F} = \mathcal{F}(\mathcal{S})$, then $\mathcal{F}$ is said to be *collision free*.

In other words, $\mathcal{F} = \mathcal{F}(\mathcal{S})$ is collision free if we have $|\mathcal{F}| = |\mathcal{S}|$. An ideally good family of pseudorandom binary sequences is collision free. If $\mathcal{F}$ is not collision free but the number of collisions is small, then they do not cause many problems. A good measure of the number of collisions is the following:

Definition 1.2. The *collision maximum* $M = M(\mathcal{F}, \mathcal{S})$ is defined by

$$M = M(\mathcal{F}, \mathcal{S}) = \max_{E_N \in \mathcal{F}} |\{s : s \in \mathcal{S}, E_N(s) = E_N\}| .$$

There is another related notion appearing in the literature, namely, the notion of avalanche effect (see, e.g., [3], [5] and [7]). In [12] I introduced the following related definitions:

Definition 1.3. If in (1.1) we have $S = \{-1, +1\}^l$, and for any $s \in S$, changing any element of s changes "many" elements of $E_N(s)$ (i.e., for $s \neq s'$ many elements of the sequences $E_N(s)$ and $E_N(s')$ are different), then we speak about *avalanche effect*, and we say that $\mathcal{F} = \mathcal{F}(\mathcal{S})$ possesses the *avalanche property*. If for any $s \in S, s' \in S, s \neq s'$ at least $(\frac{1}{2} - o(1))N$ elements of $E_N(s)$ and $E_N(s')$ are different then $\mathcal{F}$ is said to possess the *strict avalanche property*.

To study the avalanche property, I introduced the following measure:

Definition 1.4. If

$$N \in \mathbb{N}, E_N = (e_1, ..., e_N) \in \{-1, 1\}^N \quad \text{and} \quad E'_N = (e'_1, ..., e'_N) \in \{-1, 1\}^N,$$

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 (1.1), then the *distance minimum* $m(\mathcal{F})$ of $\mathcal{F}$ is defined by

$$m(\mathcal{F}) = \min_{\substack{s,s' \in \mathcal{S} \\ s \neq s'}} d(E_N(s), E_N(s')).$$

Applying this notion we may say that the family $\mathcal{F}$ in (1.1) is collision free if and only if $m(\mathcal{F}) > 0$, and $\mathcal{F}$ possesses the strict avalanche property if

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

Definitions in the case of k symbols:

Let $k \in \mathbb{N}, k \geq 2$, and let $\mathcal{A} = \{a_1, ..., a_k\}$ be a finite set ("alphabet") of k symbols ("letters"), and consider a sequence $E_N = (e_1, ..., e_N) \in \mathcal{A}^N$ of these symbols. In [10] Mauduit and Sárközy defined the analogues of correlation measure and well-distribution measure to k symbols.

To introduce the new measures of pseudorandomness, we may start in two directions. First, motivated by the definition of normality, we may think in terms of any fixed l-tuple ("word") $(a_1, ..., a_l) \in \mathcal{A}^l$ occurring with the expected

frequency in certain position in E_N. This approach leads to the following definitions: write

$$x(E_N, a, M, u, v) = |\{j : 0 \le j \le M-1, e_{u+jv} = a\}|$$

and for $w = (a_1, ..., a_l) \in \mathcal{A}^l$ and $D = (d_1, ..., d_l)$ with non-negative integers $d_1 < ... < d_l$,

$$g(E_N, w, M, D) = |\{n : 1 \le n \le M, (e_{n+d_1}, ..., e_{n+d_l}) = w\}|$$

Then the *f-well-distribution* ("f" for "frequency") *measure* of E_N is defined as

$$\delta(E_N) = \max_{a,M,u,v} \left| x(E_N, a, M, u, v) - \frac{M}{k} \right|$$

where the maximum is taken over all $a \in \mathcal{A}$ and u, v, M with $1 \le u+(M-1)v \le \le N$, while the *f-correlation measure* of order l of E_N is defined as

$$\gamma_l(E_N) = \max_{w,M,D} \left| g(E_N, w, M, D) - \frac{M}{k^l} \right|$$

where the maximum is taken over all $w \in \mathcal{A}^l$, and $D = (d_1, .., d_l)$ (with $d_1 < < \cdots < d_k$ and M such that $M + d_l \le N$.

The definitions of *collision*, *collision maximum*, *avalanche effect*, *distance* and *distance maximum* introduced for familes of binary sequences can be adapted to the k-ary case without any change. There is a change only in the definition of strict avalanche effect:

Definition 1.5. If for any $s \in S, s' \in S, s \ne s'$ at least $(\frac{k-1}{k} - o(1)) \cdot N$ elements of $E_N(s)$ and $E_N(s')$ are different then $\mathcal{F}$ is said to possess the *strict avalanche property.*

2. Results in the binary case

In [12] I tested the following construction ($f(x) \in \mathbb{F}_p[x]$):

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right), & \text{if } (f(n), p) = 1 \\ +1, & \text{if } p | f(n), \end{cases} \tag{2.1}$$

which is the extended Legendre-symbol construction, studied by Goubin, Mauduit and Sárközy [6]. They proved that under not very strong conditions on $f(x)$ both the well-distribution measure and the correlation measure of small order are small. I proved in [12] that it has further strong pseudorandom properties: a variant of the family described in (2.1) is collision free and it possesses the strong form of the avalanche property:

Theorem 2.1. *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 $E_p(f) = (e_1, ..., e_p)$ by* (2.1) *and $\mathcal{F} = \mathcal{F}(\mathcal{S})$ by* (1.1)*. Then we have*

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

Corollary 2.1. *If $\mathcal{S}$ and $\mathcal{F}$ are defined as in Theorem* 2.1 *and we also have $D < \frac{p^{1/2}}{2}$, then $\mathcal{F}$ is collision free.*

Corollary 2.2. *If $\mathcal{S}$ and $\mathcal{F}$ are defined as in Theorem* 2.1 *and we have $p \to \infty$ and $D = o(p^{1/2})$, then $\mathcal{F}$ possesses the strong avalanche property.*

The other construction I have studied in [12] was presented by Mauduit, Rivat and Sárközy in [8]: let p be an odd prime number, $f(x) \in \mathbb{F}_p[X]$ and define $E_p = (e_1, ..., e_p)$ by

$$e_n = \begin{cases} +1, & \text{if } 0 \leq r_p(f(n)) < p/2 \\ -1, & \text{if } p/2 \leq r_p(f(n)) < p, \end{cases} \tag{2.2}$$

where $r_m(n)$ denotes the least non-negative residue of n modulo m.

They proved that the well-distribution measure and the correlation measure of small order are small, but the correlation measure of large order can be large. I proved in [12] that this family has further weak pseudorandom property: there are "many" collisions in it, even for $k = 2$:

Theorem 2.2. *Now fix a prime p, and for $k \in \mathbb{N}$ write*
$\mathcal{S}_k = \{f(x) : f(x) \in \mathbb{F}_p[x], deg f(x) = k\}$ *and*
$\mathcal{F}_k = \{E_p(f) = (e_1, ...e_p) : f \in \mathcal{S}_k\}$,
then we have

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

($[x]$ denotes the integer part of x.)

Later in [13] I showed that if we replace the given family by a subfamily of it (which is just slightly smaller), 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:

Theorem 2.3. *If $f(x) \in \mathcal{P}_d$, then the family of binary sequences constructed by* (2.2) *is collision free.*

Theorem 2.4. *If $f(x) \in \mathcal{P}_d$, then the family of binary sequences obtained by* (2.2) *possesses the strict avalanche property.*

3. Results in the k-ary case

Mauduit and Sárközy in [10] generalized the Legendre-symbol construction to the case of $k \geq 2$ symbols:

Let p be a prime with $p \equiv 1 \pmod{k}$; by Dirichlet's theorem there are infinitely many primes with this property. Write $N = p - 1$, and let $\mathcal{A}$ denote the set of the k-th roots of unity:

$$\mathcal{A} = \left\{ e\left(\frac{j}{k}\right) : j = 0, 1, ..., k-1 \right\}$$

(where $e(\alpha)$ is the standard notation $e(\alpha) = e^{2\pi i \alpha}$). Let g be a primitive root modulo p, and consider the (multiplicative) character χ_1 modulo p with

$$\chi_1(g) = e\left(\frac{1}{k}\right). \tag{3.1}$$

Clearly, (3.1) determines this character χ_1 uniquely. Moreover, χ_1 is of order k (so that $\chi_1 \neq \chi_0$ by $k \geq 2$), and for all $1 \leq n \leq N = p - 1$ we have

$$\chi_1(n) \in \left\{ e\left(\frac{j}{k}\right) : j = 0, 1, ..., k-1 \right\} = \mathcal{A}.$$

Now define $E_N = (e_1, ..., e_N)$ by

$$e_n = \chi_1(n) \text{ for } n = 1, 2, ..., N. \tag{3.2}$$

It turned out that this sequence is a "good" pseudorandom sequence, i.e. both $\delta(E_N)$ and $\gamma_l(E_N)$ are "small".

With this construction only "a few" k-ary sequences with strong pseudorandom properties can be obtained, but for applications one needs a large family of them. This construction can be extended by inserting a polynomial $f(x)$ in the following way:

Define $E_N = (e_1, ..., e_N)$ by

$$e_n = \begin{cases} \chi_1(f(n)), & \text{if } (f(n), p) = 1 \\ +1, & \text{if } p | f(n), \end{cases} \tag{3.3}$$

for $n = 1, ..., N$, and let $\mathcal{F}$ denote the family of the k-ary sequences obtained in this way.

Ahlswede, Mauduit and Sárközy showed in [1] that this family $\mathcal{F}$ of sequences possesses strong pseudorandom properties. I studied the case of collisions in this family.

Let $\mathcal{H}_D$ be the set of polynomials of degree D which do not have multiple zeros.

Theorem 3.1. *If $f(x) \in \mathcal{H}_D$, then in the family $\mathcal{F}$ of k-ary sequences constructed above we have:*

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

Corollary 3.1. *If $\mathcal{H}_D$ and $\mathcal{F}$ are defined as above, and we also have $16D^2 < p$, then $\mathcal{F}$ is collision free.*

Corollary 3.2. *If $\mathcal{H}_D$ and $\mathcal{F}$ are defined as above, and we have $p \to \infty$ and $D = o(p^{1/2})$, then $\mathcal{F}$ possesses the strong avalanche property.*

The proofs are based on A. Weil's theorem [15].

4. Conclusion

Two of the most important binary constructions were tested in [12] for the new pseudorandom properties, collisions and avalanche effect and it turned out that one of them is ideal from this point of view as well, while the other construction does not possess these pseudorandom properties.

In another paper [13] it is shown that this weakness of the second construction can be corrected: one can take a subfamily of the given family which is just slightly smaller and collision free.

If a large family of binary sequences with strong pseudorandom properties is given, and it turns out that there are many collisions in it, then this negative fact does not mean that the construction must be discarded immediately.

The situation is similar in the k-ary case: there is a family of sequences with small pseudorandom measures and it turned out that it is worth studying this family from a new point of view as well: it is proved in [14] that this family of k-ary sequences is collision free and it possesses the strict avalanche property.

References

[1] **Ahlswede, R., C. Mauduit and A. Sárközy,** Large families of pseudorandom sequences of k symbols and their complexity. I., In: *General theory of information transfer and combinatorics*, Lecture Notes in Comput. Sci., **4123**, Springer, Berlin, 2006, pp. 293–307.

[2] **Alon, N., Y. Kohayakawa, C. Mauduit, C.G. Moreira and V. Rödl,** Measures of pseudorandomness for finite sequences: minimal values, *Combin. Pob. Comput.*, **15** (2005), 1–29.

[3] **Bérczes, A., J. Ködmön and A. Pethő,** A one-way function based on norm form equations, *Period. Math. Hungar.*, **49** (2004), 1–13.

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

[5] **Feistel, H., W.A. Notz and J.L. Smith,** Some cryptographic techniques for machine-to-machine data communications, *Proceedings of the IEEE*, **63** (1975), 1545–1554.

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

[7] **Kam, J. and G. Davida,** Structured design of substitution-permutation encryption networks, *IEEE Transactions on Computers*, **28** (1979), 747–753.

[8] **Mauduit, C., J. Rivat and A. Sárközy,** Construction of pseudorandom binary sequences using additive characters, *Monatshefte Math.*, **141** (2004), 197–208.

[9] **Mauduit, C. and A. Sárközy,** On finite pseudorandom binary sequences I: The measures of pseudorandomness, the Legendre symbol, *Acta Arith.*, **82** (1997), 365–377.

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

[11] **Menezes, A., P.C. van Oorshot and S. Vanstone,** *Handbook of Applied Criptography*, CRS Press, Boca Raton, 1997.

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

[13] **Tóth, V.,** The study of collision and avalanche effect in a family of pseudorandom binary sequences, *Periodica Math. Hungar.*, **59** (2009), 1–8.

[14] **Tóth, V.,** Extension of the notion of collision and avalanche effect to sequences of k symbols, *Periodica Math. Hungar.*, **65** (2012), 229–238.

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

V. Tóth
Eötvös Loránd University
Budapest
Hungary
viktoria@compalg.inf.elte.hu