Propriétés aléatoires des suites d'entiers

Katalin Gyarmati

FACULTE DES SCIENCES & TECHNIQUES

U.F.R. Sciences & Techniques Mathématiques, Informatique et Automatique
École Doctorale: Informatique, Automatique, Électronique - Électrotechnique, Mathématiques
Département de Formation Doctorale: Mathématiques

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Mathématiques

par **Katalin GYARMATI**

Propriétés aléatoires des suites d'entiers

Soutenance publique prévue le 12 janvier 2005

Membres du jury :

Rapporteurs :	M. Attila PETHŐ	Professeur, Université de Debrecen
	M. Christian MAUDUIT	Professeur, Université Aix-Marseille 2
Examinateurs :	M. Joël RIVAT	Professeur, Université Aix-Marseille 2 (codir. de thèse)
	M. András SÁRKÖZY	Professeur, Université ELTE de Budapest (codir. de thèse)
	M. Péter Pál PÁLFY (président)	Professeur, Université ELTE de Budapest
	M. Gérald TENENBAUM	Professeur, Université Henri Poincaré Nancy 1

Institut Henri Poincaré de Nancy
Faculté des Sciences & Techniques - 54500 Vandœuvre-lès-Nancy

Remerciements

Je voudrais remercier mon directeur de thèse hongrois András Sárközy et mon directeur de thèse français Joël Rivat, pour avoir attiré mon attention sur la génération des nombres pseudo-aléatoires. J'ai appris beaucoup de nos conversations, et sans leurs conseils, problèmes et questions je n'aurais jamais pu écrire cette thèse.

Je souhaite également remercier les membres de l'Institut Élie Cartan de Nancy et de l'Institut de Mathématiques de Luminy pour leurs excellentes compétences mathématiques et pour les discussions mathématiques agréables que nous avons eues. J'ai écrit en partie ma thèse dans le cadre d'une bourse de "Cotutelle de thèse" dans ces instituts à Nancy et à Marseille et je leur suis reconnaissante pour leur hospitalité.

Résumé de la thèse

Propriétés aléatoires des suites d'entiers

1 Introduction

La génération des nombres pseudoaléatoires joue un rôle important dans beaucoup de domaines des mathématiques et de la physique, en particulier dans les problèmes de cryptographie ou d'analyse numérique.

En cryptographie, l'un des algorithmes de chiffrement les plus sûrs est le « one-time pad » : nous convertissons d'abord le message en une suite d'entiers binaires (des « bits ») $A_N = \{a_1, \ldots, a_N\} \in \{0,1\}^N$ (par exemple, à chaque lettre de l'alphabet, nous assignons un nombre représenté en forme dyadique avec le même nombre de chiffres), donc A_N s'appelle le *texte clair*. Après, nous construisons (par exemple par l'observation d'un phénomène physique) une suite binaire aléatoire $E_N = \{e_1, \ldots, e_N\} \in \{0,1\}^N$ (appelée la *clé*) qui a la même longueur que le texte clair. Nous chiffrons le texte clair A_N bit par bit, en ajoutant modulo 2 ses éléments aux éléments de E_N.

Un inconvénient évident du *one-time pad* est que la clé doit être aussi longue que le texte clair, ce qui augmente la difficulté de la construction, de

la gestion et de la distribution de la clé. Ces difficultés motivent la conception d'un système de chiffrement pour lequel la clé est construite de manière pseudoaléatoire à partir d'une petite clé secrète, avec l'intention que la clé semble aléatoire à un adversaire disposant de moyens de calculs limités. De tels systèmes de chiffrement n'offrent pas une sécurité sans condition, mais on peut conserver l'espoir est qu'ils sont suffisamment sûrs.

Jusqu'à récemment, il n'y avait pas de tentative réussie de construction d'une suite *finie* pseudoaléatoire. Par exemple, les définitions du terme « pseudoaléatoire » qui sont basées sur la non-existence d'un algorithme en temps polynômial pourraient être contestées immédiatement dans le cas des suites finies, puisque dans les algorithmes en temps polynômial, il n'y a aucune restriction pour le degré ou les coefficients du polynôme. Comment dépendent-ils de la longueur de la suite ?

Dans le cas fini, la *complexité linéaire* (voir par exemple [13]) est une définition plus intéressante du pseudoaléatoire, qui repose sur la longueur du « linear feedback shift register » le plus court (récurrence linéaire sur $\mathbb{F}_2$) qui construit une suite ayant E_N pour les N premiers éléments. Mais encore le fait que la complexité linéaire soit grande n'est pas suffisant pour considérer que la suite E_N est aléatoire. La complexité linéaire constitue une bonne mesure pseudoaléatoire, mais elle est loin d'être satisfaisante.

Nouvelles mesures du caractère pseudoaléatoire

En 1997, C. Mauduit et A. Sárközy [11] ont introduit de nouvelles mesures du caractère pseudoaléatoire des suites binaires. Considérons une suite finie

$E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N$ (pour des raisons techniques, nous préférons, pour des raisons de symétrie, travailler avec ± 1 plutôt que 0 ou 1).

Définition 1 *La mesure de bonne-distribution de E_N est définie par*

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

où le maximum est pris sur tous les a, b, t tels que $a, b, t \in \mathbb{N}$ et $1 \leq a \leq a + (t-1)b \leq N$.

Définition 2 *La mesure de corrélation d'ordre k de E_N est définie par*

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

où le maximum est pris sur tous les $D = (d_1, d_2, \ldots, d_k)$ et M tels que $1 \leq d_1 < d_2 < \cdots < d_k < M + d_k \leq N$.

Cassaigne, Mauduit et Sárközy ont montré que, pour la majorité des suites $E_N \subseteq \{-1, +1\}^N$, les mesures $W(E_N)$ et $C_k(E_N)$ sont $\ll \sqrt{N}(\log N)^c$. Ainsi une suite E_N peut être considérée comme une bonne suite pseudoaléatoire si $W(E_N)$ et $C_k(E_N)$ (au moins pour petit k) sont petites en termes de N.

Il y a plusieurs manières de définir des mesures pseudoaléatoires comme W et C_k, mais il n'y a aucune mesure universelle parfaite du caractère pseudoaléatoire. On peut imposer d'autres critères (et dans certaines applications, on est forcé de le faire), et également, on peut introduire de nouvelles mesures pseudoaléatoires. Cependant, il serait de plus en plus difficile de manipuler ces mesures. Ainsi nous devons fixer une limite et nous concentrer sur certains critères pseudoaléatoires qui sont les plus importants dans les

applications. Nous choisirons donc à l'intérieur de cette limite la mesure de bonne-distribution et de corrélation, et dans quelques chapitres de cette thèse nous étudierons également une nouvelle mesure pseudoaléatoire : la mesure de symétrie.

Les cinq chapitres suivants de la thèse sont les articles [8], [6], [7], [9] et [5]. Excepté [7] tous les articles sont reproduits dans leur totalité, mais dans le quatrième chapitre j'omets une partie de [7], puisqu'une version améliorée de certains résultats peut être trouvée dans [5], qui est le dernier chapitre de la thèse.

2 Sur une propriété pseudoaléatoire des suites binaires

Dans cette section nous définissons une nouvelle mesure basée sur l'observation suivante de Mauduit et Sárközy : si une suite finie contient une suite symétrique relativement grande, alors cette structure symétrique peut causer des difficultés dans certaines applications.

Définition 3 *La mesure de symétrie de E_N est définie par*

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

D'abord nous montrons que la mesure de symétrie de E_N est autour de $\sqrt{N}$ pour presque tous $E_N \in \{-1, +1\}^N$.

Théorème 1 *Il y a un nombre entier N_0, tel que pour tous $N > N_0$ on a*

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

Tandis que pour N assez grand, $S(E_N)$ est toujours plus grand que $\sqrt{N}$ multiplié par une constante, une majoration comparable (à un facteur log près) n'a lieu seulement que pour la majorité des suites $E_N \in \{-1,+1\}^N$:

Théorème 2 *Pour tout $\varepsilon > 0$ il existe un nombre entier $N_0 = N_0(\varepsilon)$ tel que pour $N > N_0$ on a*

$$P\left(S(E_N) < 4.25\ (N \log N)^{1/2}\right) > 1 - \varepsilon.$$

Puisque pour tout $1 \leq k \leq \frac{p-1}{2}$ nous avons $\left(\frac{p-k}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{k}{p}\right)$, la mesure de symétrie du symbole de Legendre E_{p-1} est $(p-1)/2$. Nous montrons que la mesure de symétrie de la première partie de la suite E_{p-1} est petite.

Théorème 3 *Si p est un nombre premier impair, et nous posons*

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

on a

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

Dans les deux théorèmes suivants, nous donnons une forme quantitative des relations entre la mesure de bonne-distribution et la mesure de symétrie.

Théorème 4 *Pour tout $N \in \mathbb{N}$ et $E_N \in \{-1,+1\}^N$ on a*

$$W(E_N) \leq 3(NS(E_N))^{1/2}.$$

Enfin, nous voyons que ce résultat est précis : il existe une suite dont la mesure de bonne-distribution est grande, mais les mesures de corrélation et de symétrie sont petites.

Théorème 5 *Si* $k, N \in \mathbb{N}$, $N > N_0$ *et*

$$N^{3/4} \leq k \leq N$$

il existe une suite $E_N \in \{-1, +1\}^N$ *avec*

$$W(E_N) \geq k$$

et

$$\max\{C_2(E_N), S(E_N)\} < 120 \max\{\frac{k^2}{N}, (N \log N)^{1/2}\}.$$

3 Sur une famille de suites pseudoaléatoires

Le caractère pseudoaléatoire de nombreuses suites binaires a été étudié par J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat et A. Sárközy. Cependant, ces constructions produisent seulement peu de suites pseudoaléatoires distinctes, alors que parfois on a besoin des grandes familles des suites pseudoaléatoires.

L. Goubin, C. Mauduit et A. Sárközy ont réussi à construire de grandes familles de suites pseudoaléatoires généralisant une construction basée sur le symbole de Legendre.

Ici, en généralisant une construction de Sárközy [14], nous construisons un autre type de grande famille de suites pseudoaléatoires basées sur la notion de l'indice (logarithme discret). Soit p un nombre premier fixé, g une racine primitive modulo p, $(a, p) = 1$, et désignons par $\operatorname{ind} a$ l'indice de a :

$$g^{\operatorname{ind} a} \equiv a \pmod{p}$$

avec $1 \leq \operatorname{ind} a \leq p - 1$. Donc

Construction 1 *Définissons la suite* $E_{p-1} = \{e_1, \ldots, e_{p-1}\}$ *par*

$$e_n = \begin{cases} +1 & \text{si } 1 \leq \operatorname{ind} f(n) \leq (p-1)/2 \\ -1 & \text{si } (p+1)/2 \leq \operatorname{ind} f(n) \leq p-1 \text{ ou } p \mid f(n). \end{cases} \tag{1}$$

Nous donnons d'abord des estimations pour les mesures de bonne-distribution, de corrélation et de symétrie de la suite E_{p-1}.

Théorème 6 *Pour tout* $f(x) \in F_p[x]$ *de degré* $k \geq 1$ *on a*

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

Le cas de la mesure de corrélation est plus difficile :

Théorème 7 *Supposons qu'au moins une des quatre conditions suivantes est vérifiée :*

a) f *est irréductible.*

b) *Si* f *a la factorisation* $f = \varphi_1^{\alpha_1}\varphi_2^{\alpha_2}\ldots,\varphi_u^{\alpha_u}$ *où* $\alpha_i \in \mathbb{N}$ *et* φ_i *est irréductible sur* F_p, *il existe un* β *tels que exactement un ou deux* φ_i *ont le degré* β ;

c) $\ell = 2$;

d) $(4\ell)^k < p$ *ou* $(4k)^\ell < p$.

Alors on a

$$C_\ell(E_{p-1}) < 10k\ell 4^\ell p^{1/2}(\log p)^{\ell+1}.$$

Dans cette section, on étudie également la mesure de symétrie et la complexité des familles (une notion importante introduite par Ahlswede, Mauduit, Khachatrian et Sárközy [1]).

4 Une version rapide d'une générateur pseudoaléatoire

Dans le troisième chapitre, on a construit une grande famille des suites pseudoaléatoires en utilisant le logarithme discret. Les suites dans cette construction ont des propriétés pseudoaléatoires fortes, mais elles peuvent être construites seulement très lentement, puisqu'on ne connaît pas d'algorithme rapide pour calculer $\operatorname{ind} n$. Dans ce chapitre nous améliorons la construction (1) en modifiant la suite de manière à pouvoir la construire plus vite.

Soit p un nombre premier impair, $f \in \mathbb{F}_p[x]$ un polynôme de degré $k \geq 1$, soit m tel que

$$m \mid p-1,$$

et soit x un nombre premier avec m : $(x, m) = 1$. L'idée cruciale de la construction est de réduire $\operatorname{ind} n$ modulo m :

Construction 2 *On définit* $\operatorname{ind}^* n$ *de la manière suivante : pour tout* $1 \leq n \leq p-1$

$$\operatorname{ind} n \equiv x \cdot \operatorname{ind}^* n \pmod{m}$$

($\operatorname{ind}^* n$ *existe car* $(x, m) = 1$*.) Définissons la suite* $E_{p-1} = \{e_1, \ldots, e_{p-1}\}$ *par*

$$e_n = \begin{cases} +1 & si\ 1 \leq \operatorname{ind}^* f(n) \leq \frac{m}{2}, \\ -1 & si\ \frac{m}{2} < \operatorname{ind}^* f(n) \leq m\ ou\ p \mid f(n). \end{cases} \tag{2}$$

Nous montrons que cette construction a presque autant que bonnes propriétés pseudoaléatoires que celle décrite dans le troisième chapitre, cependant elle peut être construite beaucoup plus rapidement, car nous pouvons

calculer $\mathrm{ind}^* f(n)$ par $c\,(\log p)^4 \max_{p \text{ nombre premier, } p|m} p$ opérations. Si les facteurs premiers de m sont plus petits que $\log p$ alors $\mathrm{ind}^* f(n)$ peut être calculé par $c\,(\log p)^6$ opérations.

5 Sur la corrélation des ordres binaires

Alon, Kohayakawa, Mauduit, Moreira et V. Rödl [2] ont montré que si $1 \leq k \leq N$ sont entiers, on a $C_{2k}(E_N) > \sqrt{\frac{1}{2}\left[\frac{N}{(2k+1)}\right]}$ pour $E_N \in \{-1,+1\}^N$.

La mesure de corrélation d'ordre impair peut être petite : pour la suite $E_N = \{-1,+1,-1,+1,\ldots\}$ on a $C_{2k+1}(E_N) = 1$. Pour cette suite E_N on a $C_2(E_N) = \sum_{n=1}^{N-1} e_n e_{n+1} = N-1$. Cassaigne, Mauduit et Sárközy [3] ont posé le problème suivant :

Problème 1 Pour $N \to \infty$, existe-t-il des suites E_N tels que $C_2(E_N) = O(\sqrt{N})$ et $C_3(E_N) = O(1)$ simultanément ?

Récemment, Mauduit [10] a posé une autre question apparentée :

Problème 2 Est-il vrai que pour tout $E_N \in \{-1,+1\}^N$ on a

$$C_2(E_N)C_3(E_N) \gg N$$

ou au moins

$$C_2(E_N)C_3(E_N) \gg N^c \tag{3}$$

avec une constante $\frac{1}{2} < c \leq 1$?

Dans le cinquième chapitre nous réglerons le problème 1 et partiellement le problème 2 en montrant (3) avec la constante $c = 2/3$.

Théorème 8 *Si* $k, \ell \in \mathbb{N}$, $\log N \geq 2k+1 > 2\ell$, $N \in \mathbb{N}$ *et* $N > 67k^4 + 400$, $E_n \in \{-1, +1\}^N$ *et si*

$$C_{2\ell}(E_N) < \frac{1}{40\sqrt{k(2\ell+1)}} N^{1-\ell/(2k+1)},$$

alors on a

$$C_{2k+1} > \frac{1}{7}\left(\frac{2\ell}{17(2k+1)}\right)^{\ell/2} N^{1/2}.$$

En particulier pour $\ell = 1$ *si*

$$C_2(E_N) < \frac{N^{2/3}}{50\sqrt{\log N}},$$

alors on a

$$C_3(E_N), C_5(E_N), \cdots \gg \sqrt{N}.$$

6 Sur la généralisation d'une inégalité entre les mesures pseudoaléatoires

Mauduit et Sárközy [12] ont prouvé l'inégalité suivante impliquant les mesures pseudoaléatoires W et C_2 : Pour toute suite $E_N \in \{-1, +1\}^N$ on a $W(E_N) \leq 3\sqrt{NC_2(E_N)}$. Dans le dernier chapitre (voir également [5]), nous généralisons cette inégalité à la mesure de corrélation d'ordre supérieur :

Théorème 9 *Pour tout* $E_N \in \{-1, +1\}^N$ *on a*

$$W(E_N) \ll N^{1-1/(2\ell)} (C_{2\ell}(E_N))^{1/(2\ell)}.$$

Mauduit et Sárközy [12] ont montré que leur inégalité est précise en employant des arguments probabilistes. Ici, nous présentons une construction explicite (une suite de la famille décrite dans le quatrième chapitre) pour laquelle même l'inégalité généralisée est précise à un facteur constant près.

Références

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

[2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira et V. Rödl, *Measures of pseudorandomness for finite sequences : minimum values*, Combinatorics, Probability and Computation, à paraître.

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

[4] L. Goubin, C. Mauduit, A. Sárközy, *Construction of large families of pseudorandom binary sequences.*

[5] K. Gyarmati, *An inequality between the measures of pseudorandomness*, Annales Univ. Sci. Budapest. Eötvös, à paraître.

[6] K. Gyarmati, *On a family of pseudorandom binary sequences*, Periodica Math. Hungar., 49 (2), 2004 pp. 1-19.

[7] K. Gyarmati, *On a fast version of a pseudorandom generator*, General Theory of Information Transfer et Combinatorics, Conference Proceedings, à paraître.

[8] K. Gyarmati, *On a pseudorandom property of binary sequences*, The Ramanujan Journal, Vol. 8, No. 3 (2004), 289-302.

[9] K. Gyarmati, *On the correlation of binary sequences*, Studia Sci. Math. Hungar., à paraître.

[10] C. Mauduit, *Construction of pseudorandom finite sequences*, notes de conférence "Information Theory and Some Friendly Neighbours – ein Wunschkonzert", non publié, Bielefeld, 2003.

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

[12] C. Mauduit, A. Sárközy, *On the measures of pseudorandomness of binary sequences*, Discrete Math. 271 (2003), 195-207.

[13] A. J. Menezes, P. C. Oorschot, S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996.

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

Contents

Chapter 1

Introduction

The generation of pseudorandom numbers plays an important role in many fields of mathematics and physics, in particular in the problems of cryptography or numerical analysis.

For example, in cryptography one of the most secure encrypting algorithm is the one-time pad: First we convert the message into a bit sequence $A_N = \{a_1, \ldots, a_N\} \in \{0,1\}^N$ (e.g., to every letter of the alphabet we assign a number represented in diadic form with the same number of digits), then A_N is called *plain-text*. Next, we generate a random binary sequence $E_N = \{e_1, \ldots, e_N\} \in \{0,1\}^N$ (called *key-stream*) with the same length of the plain-text. We encrypt the plain-text A_N bitwise, by adding its elements modulo 2 to the elements of E_N. Then we get the encrypted message $F_N = \{f_1, \ldots f_N\}$ which is called *cipher-text*. Indeed, $f_i = a_i \oplus e_i$ where $\oplus$ is the bitwise addition modulo 2 (XOR function)

$$\begin{aligned} A_N :&\ \{a_1, \ldots, a_N\} \\ \oplus\ E_N :&\ \{e_1, \ldots, e_N\} \\ \hline = F_N :&\ \{f_1, \ldots, f_N\}. \end{aligned}$$

Thus without knowing the key-stream E_N, from the cipher-text F_N we can not compute the plain-text A_N, while if we know the key-stream E_N, it is a very easy exercise to compute A_N by using again bitwise addition modulo 2: $a_i = f_i \oplus e_i$.

The one-time pad is *unconditionally* secure regardless of the statistical distribution of the plain-text, and is optimal in the sense that its key is the smallest possible among all symmetric-key encryption schemes having this property.

The one-time pad was widely used in the second world war and cold war. The source of the random bits (which make up the one time pad) was usually a *physical device* with two possible outputs occurring with (hopefully) equal probability and varying in a (hopefully) random way, say, a diode. However, random bit generators which are based on natural sources of randomness are subject to influence by external factors, and also to malfunction. Thus it is advised that these random bit generators must be tested by using statistical tests ("frequency test", "poker test", "autocorrelation test"). Testing of this type called "aposteriori testing" by Knuth [15].

So the use of the one-time pad is: a bit string is produced by a physical device; the string is tested by statistical tests, each of the communicating partners get a copy of the string, after use, they destroy it.

An obvious drawback of the one-time pad is that the key should be as long as the plain-text which increases the difficulty of key-distribution and key management. This motives the design of stream ciphers where the key-stream is *pseudorandomly* generated from a smaller secret key, with the intent that the key-stream appears random to a computationally bounded adversary. Such stream ciphers do not offer unconditional security, but the hope is that they are computationally secure. More exactly:

Definition 1 *A pseudorandom bit generator is said to pass all polynomial time 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* $1/2$.

This definition could be disputed. For example, when we would like to generate a *finite* pseudorandom binary sequence, say, of length N, then this definition can not be used: it says nothing about the polynomial in the "polynomial time algorithm", there is no restriction for the degree or the coefficients of the polynomial. How do they depend on N? Another

problem with this definition is that the criterion measures only the quality of the pseudorandom bit generator, but not that of the sequences constructed. Finally, the non-existence of a polynomial time algorithm has never been shown yet. Consequently no unconditional proof for the cryptographical security of a pseudorandom bit generator has been given yet.

The first not very successfull attempts to define the pseudorandomness of a single sequence was made by Golomb [7], resp. Knuth [15]. Then Kolmogorov and Chaitin introduced the notion of so-called *Turing-Kolmogorov-Chaitin complexity.* While this complexity measure is of theoretical interest, there is no algorithm known for computing it, hence it has no apparent practical significance. More interesting is:

Definition 2 *The linear complexity of a finite binary sequence E_N, denoted by $L(E_N)$, is the length of the shortest linear feedback shift register (linear recursion over $\mathbb{F}_2$) that generates a sequence having E_N as its first N terms.*

For a "truly" random sequence E_N we have $L(E_N) \sim \frac{N}{2}$. The linear complexity measures an important pseudorandom property, but large linear complexity is not enough to consider E_N to be random. It makes a good, but far not satisfactory pseudorandom measure.

New mesures of pseudorandomness

In 1997 Mauduit and Sárközy [21] introduced the following *measures of pseudorandomness* of binary sequences: Consider finite binary sequences $E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N$ (by technical reasons, we switch from bits to ± 1's). Then

Definition 3 *The well-distribution measure of E_N is*

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

where the maximum is taken over all a, b, t such that $a, b, t \in \mathbb{N}$ and $1 \leq a \leq a + (t-1)b \leq N$.

Definition 4 *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 $D = (d_1, d_2, \dots, d_k)$ and M such that $1 \leq d_1 < d_2 < \dots < d_k < M + d_k \leq N$.

Cassaigne, Mauduit and Sárközy [6] proved that for the majority of the sequences $E_N \subseteq \{-1, +1\}^N$ both $W(E_N)$ and $C_k(E_N)$ are $\ll \sqrt{N}(\log N)^c$. Thus a sequence 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.

There are several ways to define pseudorandom measures like W and C_2, there is no perfect universal measure of pseudorandomness; one may pose further and further criteria for pseudorandomness (and in certain applications, one can forced to do this), and correspondingly, one may introduce further pseudorandom measures. However, it would be more and more difficult to handle these measures; besides posing too many pseudorandom requirements, it may occur that there is no pseudorandom sequence of a given size at all. This difficulty is discussed in [15] in details and, indeed, it is well described in terms of the theory of Kolmogorov complexity. Thus one has to draw the limit somewhere and to focus on certain basic pseudorandom criterias playing the most important role in applications and studied most intensively; we mainly drew this limit by restricting ourselves to the well-distribution and correlation measure, but in several chapters of this thesis we will also study a further pseudorandom measure: the symmetry measure.

The necessity of this new measure is based on the following observation of Mauduit and Sárközy: if a finite sequence contains a relatively large symmetrical subsequence (i.e., it contains a subsequence of the form $\{e_1, e_2, \dots, e_n, e_n, \dots, e_2, e_1\}$ or of the form $\{e_1, \dots, e_{n-1}, e_n, e_{n-1}, \dots, e_1\}$), then this sequence certainly cannot be a "typical" random sequence, and this symmetric structure may lead difficulties in certain applications. This observation inspired me to propose a new measure of pseudorandomness in [12]; this is the second chapter of the thesis.

It was shown in [21] that the Legendre symbol forms a "good" pseudorandom sequence. More exactly, let p be an odd prime, and

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

Then by Theorem 1 in [21] we have

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

and

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

Numerous other binary sequences have been tested for pseudorandomness by J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy. However, these constructions produce only "few" pseudorandom sequences, while in many applications, e.g., in cryptography one needs "large" families of "good" pseudorandom sequences. Very recently L. Goubin, C. Mauduit, A. Sárközy [8] succeeded in constructing large families of pseudorandom binary sequences, generalizing the construction (1.1) by replacing n by a polynomial $f(n)$:

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

In this case there are no general upper bounds for $W(E_N)$ and $C_k(E_N)$, however under some, not too restrictive conditions on the polynomial $f(x)$, these measures can be estimated by $\sqrt{N}(\log N)^c$.

In the third chapter of the thesis (see also [10]) I will generate another type of large family of pseudorandom sequences based on 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 ind a denote the index of a:

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

with $1 \leq \text{ind } a \leq p - 1$. Then define the sequence $E_{p-1} = \{e_1, \ldots, e_{p-1}\}$ by

$$e_n = \begin{cases} +1 & \text{if } 1 \leq \text{ ind } f(n) \leq (p-1)/2 \\ -1 & \text{if } (p+1)/2 \leq \text{ ind } f(n) \leq p-1 \text{ or } p \mid f(n). \end{cases} \tag{1.3}$$

The sequences in this family have as strong pseudorandom properties as the sequences in (1.2). However, these sequences can be generated very slowly, since no fast algorithm is known to compute ind n. The purpose of the fourth chapter (see also [11]) is to modify this family slightly so that the members of the new family can be generated much faster, and they have almost as good pseudorandom properties as the sequences in the original family.

It is also a very important question weather these measures of pseudorandomness are independent. Mauduit and Sárközy [21] gave examples where one of the measures C_2 and W is "large" while the other is "small".

Next we will study the minimum value of these measures. Roth [25] proved that $W(E_N) \gg N^{1/4}$ always holds, and much later Matoušek and Spencer [18] proved that this result is sharp; there exists a sequence $E_N \in \{-1,+1\}^N$ for which $W(E_N) \ll N^{1/4}$. It was a long standing problem whether $C_2(E_N) \gg \sqrt{N}$ holds for all sequences $E_N \in \{-1,+1\}^N$. Finally Alon, Kohayakawa, Mauduit, Moreira and V. Rödl [3] by using tricky linear algebra arguments proved the following: If $1 \leq k \leq N$ are integers, then we have $C_{2k}(E_N) > \sqrt{\left[\frac{N}{2(2k+1)}\right]}$ for any $E_N \in \{-1,+1\}^N$. Later in [4] they also studied the typical value of $C_k(E_N)$, by proving that for almost all sequences $E_N \in \{-1,+1\}^N$, $C_k(E_N)$ is between two constant multiples of $\sqrt{N \log \binom{N}{k}}$.

The correlation measure of odd order can be small: for the sequence $E_N = \{-1,+1,-1,+1,\dots\}$ we have $C_{2k+1}(E_N) = 1$. For this special sequence E_N we have $C_2(E_N) = \sum_{n=1}^{N-1} e_n e_{n+1} = N-1$ Cassaigne, Mauduit and Sárközy [6] asked the following:

Problem 1. For $N \to \infty$, are there sequences E_N such that $C_2(E_N) = O(\sqrt{N})$ and $C_3(E_N) = O(1)$ hold simultaneously?

Recently, Mauduit [19] asked another closely related question

Problem 2. Is it true that for every $E_N \in \{-1,+1\}^N$ we have

$$C_2(E_N)C_3(E_N) \gg N$$

or at least

$$C_2(E_N)C_3(E_N) \gg N^c \tag{1.4}$$

with some $\frac{1}{2} < c \leq 1$?

In the fifth chapter (see also [13]) I will settle both Problem 1 and Problem 2 in the weaker form (1.4) with the constant $c = 2/3$.

Returning to the question of measures of pseudorandomness it turns out that there is a weak connections between them. Mauduit and Sárközy [22] proved the following inequality involving the pseudorandom measures W and C_2: For all sequence $E_N \in \{-1, +1\}^N$ we have $W(E_N) \leq 3\sqrt{NC_2(E_N)}$. In the last chapter (see also [9]) I will generalize this inequality to correlation measure of higher even order: For $E_N \in \{-1, +1\}^N$ we have $W(E_N) \ll N^{1-1/(2\ell)}(C_{2\ell}(E_N))^{1/(2\ell)}$.

Mauduit and Sárközy [22] proved that their inequality is sharp by using probabilistic arguments. Here, I will present an explicit construction (a sequence of the family described in the fourth chapter) for which even the generalized inequality is sharp apart from a constant factor.

The next 5 chapters of the thesis are papers [12], [10], [11], [13] and [9] written by me. Except [11] all papers are fully contained, but in the fourth chapter I omit some part of the the paper [11], since the improved version of these results can be found in [9], which is the last chapter the thesis.

Acknowledgements I would like to thank to my supervisor Professor András Sárközy and my French supervisor Professor Joël Rivat, drawing my attention to the generation of pseudorandom numbers. I have learned a great deal from our consultations, without their valuable advice, problems and questions I would never have been able to write this thesis.

Chapter 2

On a pseudorandom property of binary sequences

Abstract

C. Mauduit and A. Sárközy proposed the use of well-distribution measure and correlation measure as measures of pseudorandomness of finite binary sequences. In this paper we will introduce and study a further measure of pseudorandomness: the symmetry measure. First we will give upper and lower bounds for the symmetry measure. We will also show that there exists a sequence for which each of the well-distribution, correlation and symmetry measures are small. Finally we will compare these measures of pseudorandomness.

2000 AMS Mathematics subject classification number: 11K45.

Key words and phrases: Pseudorandom, symmetry.

2.1 Introduction

In this paper we will study the symmetry property of finite binary sequences. C. Mauduit and A. Sárközy [21, pp. 367-370] introduced the following measures of pseudorandomness:

For a binary sequence

$$E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N,$$

write

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

and, for $D = (d_1, \ldots, d_k)$ with non-negative integers $0 \leq d_1 < \cdots < d_k$,

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

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

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

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

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

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

A. Sárközy and C. Mauduit [21, p. 372] observed that if a finite sequence contains a relatively large symmetrical subsequence (namely it contains a subsequence of the form $\{e_1, e_2, \ldots, e_n, e_n, \ldots, e_2, e_1\}$ or of the form $\{e_1, e_2, \ldots, e_{n-1}, e_n, e_{n-1}, \ldots, e_2, e_1\}$), then this sequence certainly cannot be a "typical" random sequence, and this symmetric structure may lead difficulties in certain applications. This observation inspired us to propose a new measure of pseudorandomness.

We will define the *symmetry measure of* E_N by

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

where

$$H(E_N, a, b) = \sum_{j=0}^{[(b-a)/2]-1} e_{a+j} e_{b-j}$$

is defined for all $1 \leq a < b \leq N$. Considering the sequence $E_N = \{1, 1, \ldots, 1\}$ we see that $\max_{E_N} S(E_N) = \left[\frac{N}{2}\right]$. We expect that for a truly random sequence E_N, the symmetry measure is small. First we will prove that the symmetry measure of E_N is around $\sqrt{N}$ for almost all $E_N \in \{-1, +1\}^N$.

Theorem 1 *There is an integer N_0 such that for $N > N_0$ we have*

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

While for large N, $S(E_N)$ is always greater than a constant times $\sqrt{N}$, the upper bound holds for only the majority of the sequences $E_N \in \{-1, +1\}^N$.

Theorem 2 *For all $\varepsilon > 0$ there are numbers $N_0 = N_0(\varepsilon)$ such that for $N > N_0$ we have*

$$P\left(S(E_N) < 4.25\ (N \log N)^{1/2}\right) > 1 - \varepsilon.$$

We need the following measures of pseudorandomness introduced in [21, p. 371-372]. *Combined* (well-distribution-correlation) *PR-measure of order k*:

$$Q_k(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t} e_{a+jb+d_1} e_{a+jb+d_2} \cdots e_{a+jb+d_k} \right|,$$

where $a,\ b,\ t,\ D = (d_1, d_2, \ldots, d_k)$ are such that all the subscripts $a + jb + d_l$ belong to $\{1, \ldots, N\}$. *Combined PR-measure*:

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

C. Mauduit and A. Sárközy [21, p. 373] proved that there is a number p_0 such that if $p > p_0$ is a prime number, $k \in \mathbb{N}$, $k < p$ and if we write

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

then

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

so that, writing $N = p-1$, we have

$$Q(E_N) \leq 27N^{1/2}(\log N)^2$$

It follows that for the Legendre symbol both the well-distribution measure and the correlation measure of order 2 are smaller than $18N^{1/2}\log N$, while the combined PR-measure is smaller than $27N^{1/2}(\log N)^2$. As for all $1 \leq k \leq \frac{p-1}{2}$ we have $\left(\frac{p-k}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{k}{p}\right)$, the symmetry measure of the Legendre symbol E_{p-1} is $(p-1)/2$. We will show that the symmetry measure of the half of the sequence E_{p-1} is small.

Theorem 3 *If p is an odd prime, and we write*

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

then we have

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

Finally, we will compare the correlation measure of order 2 with well-distribution and symmetry measures. We expect that these measures of pseudorandomness are relatively independent. In order to show this we will give constructions where one measure is large while the others are small. The following two examples are variants of the ones in [21, p. 371-372].

EXAMPLE 1. Consider a sequence $E_N = (e_1, \ldots, e_n) \in \{-1, +1\}^N$ such that each of the symmetry, correlation and well-distribution measure of it are possibly small (by Theorem 1 and 2 in [6] and our Theorem 2, all these measures can be $O\left((N\log N)^{1/2}\right)$ simultaneously) and define $E'_{2N} = (e'_1, e'_2, \ldots, e'_{2N}) \in \{-1, +1\}^{2N}$ 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 it easy to see that the well-distribution measure of E'_{2N} are less than a constant times the corresponding measure of E_N and $S(E'_{2N}) \leq S(E_N) +$

$C_2(E_N)$, but

$$C_2(E'_N) \geq \left| \sum_{n=1}^{N} e'_n e'_{n+N} \right| = N.$$

EXAMPLE 2. Consider a sequence $E_N = (e_1, \ldots, e_n) \in \{-1, +1\}^N$ such that each of the correlation measure of order 2, well-distribution measure and symmetry measure of it are possibly small and define $E'_{2N} = (e'_1, e'_2, \ldots, e'_{2N}) \in \{-1, +1\}^{2N}$ 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 the correlation measure of order 2 is less than a constant times $S(E_N) + C_2(E_N)$, while $W(E'_{2N}) \leq 2W(E_N)$. But

$$S(E'_{2N}) \geq \left| \sum_{n=1}^{N} e'_n e'_{2N-n} \right| = N.$$

C. Mauduit and A. Sárközy in [22] expressed the connection between the well-distribution measure and the correlation measure of order 2 in a quantitative form. Accordingly, in the following two theorems we will give a similar quantitative form of the connection between the well-distribution measure and the symmetry measure.

Theorem 4 *For all $N \in E_N$, and $E_N \in \{-1, +1\}^N$ we have*

$$W(E_N) \leq 3(NS(E_N))^{1/2}. \tag{2.1}$$

Finally, we will show that this result is sharp; there exists a sequence whose well-distribution measure is large and both the correlation measure and the symmetry measure are possibly small. Since the proof of the next theorem is nearly the same as the one in [22] (indeed we have to write $S(E_N)$ in place of $C_2(E_N)$), thus we will only sketch the proof.

Theorem 5 *If $k, N \in \mathbb{N}$, $N > N_0$ and*

$$N^{3/4} \leq k \leq N \tag{2.2}$$

then there is a sequence $E_N \in \{-1,+1\}^N$ *with*

$$W(E_N) \geq k \tag{2.3}$$

and

$$\max\{C_2(E_N), S(E_N)\} < 120 \max\{\frac{k^2}{N}, (N \log N)^{1/2}\}. \tag{2.4}$$

From Theorem 5 we get that if $k \geq N^{3/4}(\log N)^{1/4}$, then

$$W(E_N) \geq k = \frac{N^{1/2}}{11}\left(121\frac{k^2}{N}\right)^{1/2} > \\ > \frac{1}{11}\left(N \max\{C_2(E_N), S(E_N)\}\right)^{1/2}. \tag{2.5}$$

This means that (2.1) is the best possible apart from a constant factor.

One might like to study the generalizations of these measures of pseudorandomness. One possibility is to define the following measure:

$$\max_{\substack{M_1,M_2 \\ f_1(n),f_2(n),\ldots,f_j(n)}} \sum_{\substack{M_1 \leq f_i(n) \leq M_2 \\ (i=1,\ldots,j)}} e_{f_1(n)} e_{f_2(n)} \cdots e_{f_j(n)}, \tag{2.6}$$

where the maximum is taken over all $1 \leq M_1 < M_2 \leq N$ integers and $f_1(n), f_2(n), \ldots, f_j(n)$ polynomials with integer coefficients such that $M_1 \leq f_i(n) \leq M_2$ holds for all $1 \leq n \leq N$, $1 \leq i \leq j$. Of course this generalization also covers certain pathological cases (e.g., $f_1(n) = f_2(n) = \cdots = f_j(n)$), thus to introduce a pseudorandom measure of this type one has to pose certain restrictions on the polynomials $f_1, \ldots, f_j$ involved; we do not go into the details of this here.

When $j = 1$ or 2, for the special values of the polynomials $f_i(n)$, (2.6) can give the well-distribution measure, the correlation measure of order 2 and the symmetry measure.

Throughout the paper we write $e(x) = e^{2\pi i x}$.

2.2 Proofs

Proof of Theorem 1

Let $E_N = \{e_1, \ldots, e_n\}$, $f(z) = \sum_{n=1}^{N} e_n z^n$. Using the Cauchy-Schwarz inequality and Parseval formula we obtain:

$$\mathcal{J} \stackrel{\text{def}}{=} \int_0^1 |f(e(\alpha))|^4 \, \mathrm{d}\alpha \geq \left(\int_0^1 |f(e(\alpha))|^2 \, \mathrm{d}\alpha \right)^2 = N^2. \tag{2.7}$$

Using the Parseval formula again we get:

$$\mathcal{J} = \int_0^1 \left| f^2(e(\alpha)) \right|^2 \mathrm{d}\alpha = \int_0^1 \left| \sum_{n=1}^{N} \sum_{m=1}^{N} e_n e_m e((n+m)\alpha) \right|^2 \mathrm{d}\alpha =$$

$$= \int_0^1 \left| \sum_{k=2}^{2N} \left(\sum_{\max\{1,\ k-N\} \leq n \leq \min\{N, k-1\}} e_n e_{k-n} \right) e(k\alpha) \right|^2 \mathrm{d}\alpha =$$

$$= \sum_{k=2}^{2N} \left| \sum_{\max\{1,\ k-N\} \leq n \leq \min\{N, k-1\}} e_n e_{k-n} \right|^2$$

By the definition of symmetry measure we have

$$\left| \sum_{\max\{1,\ k-N\} \leq n \leq \min\{N, k-1\}} e_n e_{k-n} \right| \leq 2S(E_N) + 1.$$

Therefore

$$\mathcal{J} \leq (2N-1)\left(2S(E_N)+1\right)^2. \tag{2.8}$$

So that, in view of (2.7) and (2.8), and since clearly $S(E_N) \geq 1$, for large N we have

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

Proof of Theorem 2

Write $L = 4.25\ (N \log N)^{1/2}$, then we have:

$$P(S(E_N) > L) = P(\max_{a<b} |H(E_N, a, b)| > L) \leq$$

$$\leq \sum_{a<b} P(|H(E_N, a, b)| > L)) \leq \binom{N}{2} \max_{a<b} P(|H(E_N, a, b)| > L),$$

where both the maximum and the summation are taken over all $a, b \in \mathbb{N}$ such that $1 \leq a < b \leq N$. Thus in order to prove the theorem, it suffices to show that for all $1 \leq a < b \leq N$ we have:

$$P(|H(E_N, a, b)| > L) = P\left(\left| \sum\nolimits_{j=0}^{[(b-a)/2]-1} e_{a+j} e_{b-j} \right| > L \right) < \frac{2\varepsilon}{N^2}. \tag{2.9}$$

Let $t = [(b-a)/2]$, if $t \leq L$ then the probability in (2.9) is trivially 0 so that we may assume:

$$t = [(b-a)/2] > L = 4.25\ (N \log N)^{1/2}. \tag{2.10}$$

Write

$$M = 6(t \log t)^{1/2}$$

and

$$|\{j:\ 0 \leq j \leq t-1,\ e_{a+j}e_{b-j} = -1\}| = h. \tag{2.11}$$

Then we have:

$$\sum_{j=0}^{t-1} e_{a+j}e_{b-j} = |\{j:\ 0 \leq j \leq t-1,\ e_{a+j}e_{b-j} = 1\}| -$$

$$- |\{j:\ 0 \leq j \leq t-1,\ e_{a+j}e_{b-j} = -1\}| = (t-h) - h =$$

$$= t - 2h.$$

(2.11) holds with probability $\frac{1}{2^t}\binom{t}{h}$ so that

$$P\left(\left|\sum_{j=1}^{t-1} e_{a+j}e_{b-j}\right| > M\right) = \sum_{h:\ |t-2h|>M} \frac{1}{2^t}\binom{t}{h} =$$

$$= \frac{1}{2^t} \sum_{h:\ |h-t/2|>M/2} \binom{t}{h}. \tag{2.12}$$

An easy computation shows that if $t \to \infty$ and $k \leq t^{2/3}$, then we have

$$\binom{t}{[t/2]-k} = \binom{t}{[t/2]} \exp\left(-\frac{2k^2}{t} + O(\frac{k^3}{t^2})\right).$$

Using also the fact that $\binom{t}{i}$ is increasing in i for $0 \leq i \leq t/2$, it follows easily that for N large enough (so that $t = [(b-a)/2]$ is also large by (2.10)),

$$\sum_{h:\ |h-t/2|>M/2} \binom{t}{h} = \sum_{h:\ |h-t/2|>3(t\log t)^{1/2}} \binom{t}{h} <$$

$$< t\binom{t}{[t/2] + [3(t\log t)^{1/2}]} <$$

$$< t\binom{t}{[t/2]} \exp\left(-2\left(3(t\log t)^{1/2}\right)\frac{1}{t} + o(1)\right) =$$

$$= t\binom{t}{[t/2]} \exp\left(-18\log t + o(1)\right) < \frac{2^t}{t^{16}}. \tag{2.13}$$

Since $M \leq L$, it follows from (2.10), (2.12) and (2.13) that

$$P\left(\left|\sum_{j=0}^{t-1} e_{a+j}e_{b-j}\right| > L\right) \leq P\left(\left|\sum_{j=0}^{t-1} e_{a+j}e_{b-j}\right| > M\right) <$$
$$< \frac{1}{2^t}\frac{2^t}{t^{16}} = \frac{1}{t^{16}} < \frac{1}{L^{16}} = o(\frac{1}{N^8}) < \frac{2\varepsilon}{N^2}$$

which proves (2.9) and this completes the proof of Theorem 2.

Proof of Theorem 3

We shall need the following lemma:

Lemma 1 *If p is a prime number, $f(x) \in F_p[x]$ is a polynomial of degree k such that it is not of the form $f(x) \in b(g(x))^2$ with $b \in F_p$, $g(x) \in F_p[x]$, and X, Y are real numbers with $0 < Y \leq p$, then writing*

$$\chi_p^*(n) = \begin{cases} \left(\frac{n}{p}\right) & \text{for } (n,p) = 1, \\ 0 & \text{for } p \mid n, \end{cases}$$

we have

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

Proof of Lemma 1

See [21, p. 373]. (Indeed, there this result is deduced from Weil's theorem [28].)

By the definition of H we have:

$$H(E_{(p-1)/2}, a, b) = \sum_{j=0}^{[(b-a)/2]-1} \left(\frac{a+j}{p}\right)\left(\frac{b-j}{p}\right) =$$
$$= \sum_{j=0}^{[(b-a)/2]-1} \left(\frac{-j^2+(b-a)j+ab}{p}\right). \qquad (2.14)$$

Let $f(x) = -x^2 + (b-a)x + ab \in F_p[x]$. It is easy to see that $f(x)$ is the form of $b(g(x))^2$ if and only if $a + b \equiv 0 \ (p)$. In the present case this is impossible as $1 \leq a < b \leq (p-1)/2$. Applying Lemma 1 with 0 and $(b-a)/2$ in place of X and Y we get:

$$\left|\sum_{X\leq n\leq X+Y} \chi_p^*(f(n))\right| = \sum_{j=0}^{[(b-a)/2]-1} \left(\frac{-j^2+(b-a)j+ab}{p}\right) <$$
$$< 18p^{1/2}\log p. \qquad (2.15)$$

From (2.14) and (2.15) we obtain $S(E_{(p-1)/2}) \leq 18p^{1/2}\log p$, which proves the theorem.

Proof of Theorem 4

There exist a, b and t natural numbers such that:

$$W(E_N) = |U(E_N, t, a, b)| = \left|\sum_{j=0}^{t-1} e_{a+jb}\right| = \left|\sum_{\substack{a\leq n\leq a+(t-1)b \\ n\equiv a \pmod b}} e_n.\right|$$

For all $n \in \mathbb{N}$ let $r(n)$ be the smallest natural number with $r(n) \equiv n \pmod b$. Let

$$f(\alpha) \stackrel{\text{def}}{=} \sum_{n=a}^{a+(t-1)b} e_n e\left(r(n)\alpha\right) = \sum_{k=0}^{b-1}\left(\sum_{\substack{a\leq n\leq a+(t-1)b \\ n\equiv k \pmod b}} e_n\right) e(k\alpha).$$

The following lemma is well known and very simple.

Lemma 2 *If* $T(\alpha) = \sum_{k=0}^{b-1} c_k e(k\alpha)$ *then*

$$\sum_{h=0}^{b-1}\left|T\left(\frac{h}{b}\right)\right|^2 = b\sum_{k=0}^{b-1}|c_k|^2.$$

By Lemma 2 we have:

$$\begin{aligned}\sum_{h=0}^{b-1}\left|f\left(\frac{h}{b}\right)\right|^2 &= b\sum_{k=0}^{b-1}\left|\sum_{\substack{a\leq n\leq a+(t-1)b \\ n\equiv k \pmod b}} e_n\right|^2 \geq b\left|\sum_{\substack{a\leq n\leq a+(t-1)b \\ n\equiv a \pmod b}} e_n\right|^2 \\ &= bW^2(E_N). \end{aligned} \tag{2.16}$$

On the other hand:

$$\begin{aligned} f^2\left(\frac{h}{b}\right) &= \sum_{n=a}^{a+(t-1)b}\sum_{m=a}^{a+(t-1)b} e_n e_m e\left(\frac{r(n)+r(m)}{b}h\right) = \\ &= \sum_{n=a}^{a+(t-1)b}\sum_{m=a}^{a+(t-1)b} e_n e_m e\left(\frac{r(n+m)}{b}h\right) = \\ &= \sum_{k=0}^{b-1}\left(\sum_{n=a}^{a+(t-1)b}\sum_{\substack{m=a \\ n+m\equiv k \pmod b}}^{a+(t-1)b} e_n e_m\right) e\left(k\frac{h}{b}\right) = \sum_{k=0}^{b-1} c_k e(k\frac{h}{b}), \end{aligned} \tag{2.17}$$

where $c_k = \sum_{n=a}^{a+(t-1)b} \sum_{\substack{m=a \\ n+m\equiv k \ (b)}}^{a+(t-1)b} e_n e_m$. Replacing $n+m = jb+k$ we get:

$$
\begin{aligned}
|c_k| &= \left| \sum_{n=a}^{a+(t-1)b} \sum_{\substack{m=a \\ n+m\equiv k \ (b)}}^{a+(t-1)b} e_n e_m \right| \\
&= \left| \sum_{j=\left\lceil \frac{2a-k}{b} \right\rceil}^{\left[\frac{2N-k}{b} \right]} \sum_{n=\max\{a, jb+k-(a+(t-1)b)\}}^{\min\{jb+k-a, a+(t-1)b\}} e_n e_{jb+k-n} \right| \\
&\leq \sum_{j=\left\lceil \frac{2a-k}{b} \right\rceil}^{\left[\frac{2N-k}{b} \right]} \left| \sum_{n=\max\{a, jb+k-(a+(t-1)b)\}}^{\min\{jb+k-a, a+(t-1)b\}} e_n e_{jb+k-n} \right| \\
&\leq \sum_{j=\left\lceil \frac{2a-k}{b} \right\rceil}^{\left[\frac{2N-k}{b} \right]} (2S(E_N)+1) \leq \left(\frac{2N}{b} + 1 \right) (2S(E_N)+1) \\
&\leq 9\frac{N}{b} S(E_N).
\end{aligned}
$$

Using (2.17) and Lemma 2 with the function $\sum_{k=0}^{b-1} c_k e(k\alpha)$, where c_k has defined above, we get:

$$
\sum_{h=0}^{b-1} \left| f^2\left(\frac{h}{b}\right) \right|^2 = b \sum_{k=0}^{b-1} |c_k|^2 \leq 81 \frac{N^2}{b} S^2(E_N).
$$

Thus from the Cauchy-Schwarz inequality and (2.16) we have:

$$
\begin{aligned}
81\frac{N^2}{b} S^2(E_N) &\geq \sum_{h=0}^{b-1} \left| f^2\left(\frac{h}{b}\right) \right|^2 \geq \frac{1}{b} \left(\sum_{h=0}^{b-1} \left| f^2\left(\frac{h}{b}\right) \right| \right)^2 \geq \\
&\geq \frac{1}{b} \left(bW^2(E_N) \right)^2 = bW^4(E_N),
\end{aligned}
$$

whence:

$$
W(E_N) \leq 3 \left(\frac{N}{b} S(E_N) \right)^{1/2} \leq 3(NS(E_N))^{1/2}
$$

which was to be proved.

Proof of Theorem 5

If $k \geq \frac{N}{10}$ then (2.4) holds trivially for all E_N satisfying (2.3), thus we may assume that

$$
k \leq \frac{N}{10} \tag{2.18}
$$

Write

$$\Delta = 30 \max\left\{\frac{k^2}{N}, (N \log N)^{1/2}\right\}$$

so that (2.4) can be written as

$$\max\{C_2(E_N), S(E_N)\} \leq 4\Delta.$$

If $\mathcal{A}$ is a finite set of positive integers, and $d \in \mathbb{N}$, then denote the number of solutions of

$$a - a' = d, \ a \in \mathcal{A}, \ a' \in \mathcal{A}, \tag{2.19}$$

by $f(\mathcal{A}, d)$, and denote the number of solutions of

$$a + a' = d, \ a \in \mathcal{A}, \ a' \in \mathcal{A}, \tag{2.20}$$

by $g(\mathcal{A}, d)$.

Lemma 3 *Assume that k satisfies (2.2) and N is large enough. Then there is an $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ such that*

$$|\mathcal{A}| = k \tag{2.21}$$

and

$$\max\{f(\mathcal{A}, d), g(\mathcal{A}, d)\} < 30\frac{k^2}{N} \stackrel{\text{def}}{=} M \ \textit{for aall} \ 1 \leq d \leq 2N - 1. \tag{2.22}$$

Proof of Lemma 3

Write

$$\begin{aligned}
\mathcal{F} &= \{\mathcal{A} : \ \mathcal{A} \subseteq \{1, 2, \ldots, N\}, |\mathcal{A}| = k\}, \\
\mathcal{F}_d &= \{\mathcal{A} : \ \mathcal{A} \in \mathcal{F}, \ f(\mathcal{A}, d) \geq M\}, \\
\mathcal{G}_d &= \{\mathcal{A} : \ \mathcal{A} \in \mathcal{F}, \ g(\mathcal{A}, d) \geq M\}.
\end{aligned}$$

Then, clearly, any set $\mathcal{A}$ belonging to

$$\mathcal{F} \setminus \left(\cup_{d=1}^{N-1} \mathcal{F}_d \cup \cup_{d=3}^{2N-1} \mathcal{G}_d\right) \tag{2.23}$$

satisfies (2.21) and (2.22). Thus it suffices to show that the set in (2.23) is non-empty. To prove this we have to give upper bounds for $|\mathcal{F}_d|$ and $|\mathcal{G}_d|$. In [22] it was proved that

$$|\mathcal{F}_d| \leq \left(\frac{11}{12}\right)^{10N^{1/2}} \binom{N}{k}.$$

From this we get

$$|\mathcal{F}_d| \leq \frac{1}{4N}\binom{N}{k} \tag{2.24}$$

We will obtain by similar but easier calculations than in [22] that:

$$|\mathcal{G}_d| \leq \left(\frac{1}{3}\right)^{10N^{1/2}} \binom{N}{k} \leq \frac{1}{4N}\binom{N}{k}. \tag{2.25}$$

Consider a set

$$\mathcal{A} \in \mathcal{G}_d. \tag{2.26}$$

It follows from $g(\mathcal{A}, d) \geq M$ that there exist $\lceil M \rceil$ different numbers a_i $(i = 1, 2, \ldots \lceil M \rceil)$ with the property that

$$a_i \in \mathcal{A}, \quad d - a_i \in \mathcal{A}.$$

Let

$$\mathcal{A}_0 = \mathcal{A} \setminus \left(\cup_{i=1}^{\lceil M \rceil} \{a_i, d - a_i\}\right).$$

Then $\mathcal{A}$ is the disjoint union of the sets

$$\{a_1, d - a_1\}, \{a_2, d - a_2\}, \ldots, \{a_{\lceil M \rceil}, d - a_{\lceil M \rceil}\}, \mathcal{A}_0.$$

Here we may choose $a_1, a_2, \ldots, a_{\lceil M \rceil}$ from $\{1, 2, \ldots, N\}$ in at most $\binom{N}{\lceil M \rceil}$ ways, these numbers determine $d - a_1, d - a_2, \ldots, d - a_{\lceil M \rceil}$ uniquely, and since $|\mathcal{A}_0| = k - 2\lceil M \rceil$, the elements of A_0 can be chosen from the remaining $N - 2\lceil M \rceil$ numbers in at most $\binom{N-2\lceil M \rceil}{k-2\lceil M \rceil}$ ways. It follows that

$$|\mathcal{G}_d| \leq \binom{N}{\lceil M \rceil}\binom{N - 2\lceil M \rceil}{k - 2\lceil M \rceil}.$$

Carrying out similar calculations as in [22], we get:

$$|\mathcal{G}_d| \leq \frac{k^{2\lceil M\rceil}}{\lceil M\rceil!(N-2\lceil M\rceil)^{\lceil M\rceil}}\binom{N}{k}.$$

By Stirling formula and (2.18) we get:

$$\lceil M\rceil! \geq \left(\frac{M}{3}\right)^{\lceil M\rceil} = \left(10\frac{k^2}{N}\right)^{\lceil M\rceil},$$
$$N - 2\lceil M\rceil \geq N - 70\frac{k^2}{N} \geq \frac{3}{10}N.$$

So we have:

$$|\mathcal{G}_d| \leq \left(\frac{1}{3}\right)^{\lceil M\rceil}\binom{N}{k} < \left(\frac{1}{3}\right)^{30N^{1/2}}\binom{N}{k} < \frac{1}{4N}\binom{N}{k}.$$

Using this, (2.24) and the fact that $|\mathcal{F}| = \binom{N}{k}$ we get that the set in (2.23) is non-empty, and this completes the proof of Lemma 3.

Now we fix a set $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ satisfying (2.21) and (2.22) in Lemma 3, and let ε denote the set of the binary sequences $E_N \in \{-1, +1\}^N$ with

$$e_n = +1 \quad \text{for } n \in \mathcal{A}$$

so that

$$|\varepsilon| = 2^{N-|\mathcal{A}|} = 2^{N-k}.$$

We consider a "random" element E_N of ε, i.e., we choose each $E_N \in \varepsilon$ with probability $1/2^{N-k}$. In other words, we consider the binary sequence $E_N = \{e_1, e_2, \ldots, e_N\}$ where for $n \in \mathcal{A}$ we have $e_n = +1$ while for n values with $n \notin \mathcal{A}$ the e_n's are chosen independently with

$$P(E_n = +1) = P(E_n = -1) = \frac{1}{2} \ (\text{ for } n \notin \mathcal{A}).$$

C. Mauduit and A. Sárközy [22] proved that

$$P(W(E_N) \geq k) \geq \frac{1}{2},$$
$$P(C_2(E_N) > 4\Delta) \leq \frac{3}{N}. \tag{2.27}$$

By the definition of $S(E_N)$ we have

$$P(S(E_N) > 4\Delta) = P(\max_{a,b} |H(E_N, a, b)| > 4\Delta)$$
$$\leq \sum_{a,b} P(|H(E_N, a, b)| > 4\Delta). \tag{2.28}$$

For all $E_N \in \varepsilon$ we have

$$H(E_N, a, b) = \sum_{\substack{0 \leq j \leq [(b-a)/2]-1 \\ a+j \in \mathcal{A},\ b-j \in \mathcal{A}}} e_{a+j} e_{b-j} + \sum_{\substack{0 \leq j \leq [(b-a)/2]-1 \\ a+j \in \mathcal{A},\ b-j \notin \mathcal{A}}} e_{a+j} e_{b-j}$$
$$+ \sum_{\substack{0 \leq j \leq [(b-a)/2]-1 \\ a+j \notin \mathcal{A},\ b-j \in \mathcal{A}}} e_{a+j} e_{b-j} + \sum_{\substack{0 \leq j \leq [(b-a)/2]-1 \\ a+j \notin \mathcal{A},\ b-j \notin \mathcal{A}}} e_{a+j} e_{b-j}$$
$$= \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3 + \sum\nolimits_4. \tag{2.29}$$

It can be proved in the same way as in [22] with the change that we write $e_{a+j}e_{b-j}$ in the place of $e_{n+d_1}e_{n+d_2}$ and estimating $P(|\sum_4| \geq \Delta)$ we use Lemma 3 in place of [22, Lemma 1] that

$$P\left(\left|\sum\nolimits_1\right| > \Delta\right) = 0,$$

and for $i = 2, 3, 4$ we have

$$P\left(\left|\sum\nolimits_i\right| > \Delta\right) < \frac{1}{N^4}.$$

From this and (2.29) we get:

$$P(|H(E_N, a, b)| > 4\Delta) \leq \sum_{i=1}^{4} P\left(\left|\sum\nolimits_i\right| > \Delta\right) \leq \frac{3}{N^4} \tag{2.30}$$

Using (2.28) and (2.30) we have:

$$P(|S(E_N)| > 4\Delta) \leq \frac{3}{N}. \tag{2.31}$$

It follows from (2.27) and (2.31) that (2.3) and (2.4) hold simultaneously with probability

$$> \frac{1}{2} - \frac{6}{N} > \frac{1}{3}$$

for N large enough, so that there is at least one $E_N \in \{-1, +1\}^N$ satisfying both (2.3) and (2.4), and this completes the proof of Theorem 5.

I would like to thank Professor András Sárközy for the valuable discussions.

Chapter 3

On a family of pseudorandom binary sequences

Abstract

Recently, numerous constructions have been given for finite pseudorandom binary sequences. However, in many applications, e.g., in cryptography one needs "large" families of "good" pseudorandom sequences. Very Recently L. Goubin, C. Mauduit, A. Sárközy succeeded in constructing large families of pseudorandom binary sequences based on the Legendre symbol. In this paper we will generate another type of large family of pseudorandom sequences by using the notion of index (discrete logarithm).

2000 AMS Mathematics Subject Classification: 11K45

List of keywords and phrases: pseudorandom, index, discrete logarithm, correlation, linear complexity.

3.1 Introduction

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

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

In particular, in part I [21] 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 $D = (d_1, \ldots, d_k,)$ with non-negative integers $d_1 < \cdots < d_k$,

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

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

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

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ 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 $N \to \infty$).

Moreover it was shown in [21] that the Legendre symbol forms a "good" pseudorandom sequence. More exactly, let p be an odd prime, and

$$N = p - 1, \ e_n = \left(\frac{n}{p}\right), \ E_N = \{e_q, \ldots, e_N\}. \tag{3.1}$$

Then by Theorem 1 in [21] we have

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

and

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

In [12] was introduced the *symmetry measure* of E_N:

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

In [12] it was also shown that for the half of the Legendre symbol sequence, i.e., for the sequence

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

where p is an odd prime, we have

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

Numerous other binary sequence have been tested for pseudorandomness by J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy. However, these constructions produce only "few" pseudorandom sequences, while in many applications, e.g., in cryptography one needs "large" families of "good" pseudorandom sequences. Very recently L. Goubin, C. Mauduit, A. Sárközy [8] succeeded in constructing large families of pseudorandom binary sequences, generalizing the construction (3.1). Their most important results can be summarized as follows:

Theorem 1 *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 $E_p = \{e_q, \ldots, e_p\}$ is defined by*

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{for } (f(n), p) = 1 \\ +1 & \text{for } p \mid f(n), \end{cases} \tag{3.2}$$

then we have

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

Moreover, assume that for $\ell \in \mathbb{N}$ one of the following assumption holds:
(i) $\ell = 2$;
(ii) $\ell < p$ and 2 is a primitive root modulo p;

(iii) $(4k)^{\ell} < p$.
Then we also have

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

This theorem generates "large" families of "good" pseudorandom binary sequences. However, in most applications it is also very important that the "large" family $\mathcal{F}$ of "good" pseudorandom sequences had a "complex" structure, there are many "independent" sequences in it. In [1] a quantitative measure for this property of families of binary sequences was introduced.

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

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

It is clear from Definition 1 that for $j < C(\mathcal{F})$, there exist at least $2^{C(\mathcal{F})-j}$ sequence $E_N \in \mathcal{F}$ with

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

In [1] it was also proved that the complexity measure of the family of the sequence E_p defined by (3.2) is large. More precisely: consider all the polynomials $f(x) \in F_p[x]$ with

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

(where $\deg f(x)$ denotes the degree of $f(x)$) and $f(x)$ has no multiple zero in F_p. For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f) = \{e_1 \ldots, e_p\} \in \{-1,+1\}^p$ defined by (3.2), and let $\mathcal{F}$ denote the family of all binary sequences obtained in this way. Then we have

$$C(\mathcal{F}) > K.$$

In this paper, extending a construction given by A. Sárközy in [26], we will generate a large family of pseudorandom sequences based on the notion

of the index (discrete logarithm). The following pseudorandom sequence was introduced and studied in [26].

If p is a fixed prime and g is a fixed primitive root modulo p, and $(a,p)=1$, then let 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 $E_N=\{e,\ldots,e_N\}$ by

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

Then we have

$$W(E_p) < 4p^{1/2}(\log p)^2 < 20N^{1/2}\,(\log N)^2$$

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

$$C_k(E_N) < 9k4^k p^{1/2}(\log p)^{k+1} < 27k8^k N^{1/2}\,(\log N)^{k+1}\,.$$

We will generate a large family of pseudorandom sequences analogously to Theorem 1, i.e. replacing n by $f(n)$.

Definition 2 *Let p be an odd prime, g a primitive root modulo p. Define* ind n, *by* $1 \leq \text{ind } n \leq p-1$ *and* $n \equiv g^{\text{ind } n} \pmod{p}$. *Let $f(x) \in F[p]$ be a polynomial of the degree k. Then define the sequence $E_{p-1}=\{e_1,\ldots,e_{p-1}\}$ by*

$$e_n = \begin{cases} +1 & \text{if } 1 \leq \text{ ind } f(n) \leq (p-1)/2 \\ -1 & \text{if } (p+1)/2 \leq \text{ ind } f(n) \leq p-1 \text{ or } p \mid f(n). \end{cases} \tag{3.3}$$

This paper is devoted to the study of family described in Definition 2. Throughout this paper we will use these notations: the numbers p,k,g the polynomial f and the sequence E_{p-1} will be defined as in Definition 2. First we give estimates for the well-distribution measure, the correlation measure and the symmetry measure of the sequence E_{p-1}.

Theorem 2 *For all $f(x) \in F_p[x]$ we have*

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

Unlike the corresponding part of Theorem 1, here in Theorem 2 there is no condition on the roots of the polynomial $f(x)$. The case of the correlation measure will be slightly more difficult. As in Theorem 1, the upper bound holds under certain assumptions. The last two conditions are very similar to the conditions of Theorem 1, since these theorems are based on a similar addition lemma.

Theorem 3 *Suppose that at least one of the following 4 condition holds:*

a) *f is irreducible.*

b) *If f has the factorization $f = \varphi_1^{\alpha_1}\varphi_2^{\alpha_2}\ldots,\varphi_u^{\alpha_u}$ where $\alpha_i \in \mathbb{N}$ and φ_i is irreducible over F_p, then there exists a β such that exactly one or two φ_i's have the degree β;*

c) *$\ell = 2$;*

d) *$(4\ell)^k < p$ or $(4k)^\ell < p$.*

Then $C_\ell(E_{p-1}) < 10k\ell 4^\ell p^{1/2}(\log p)^{\ell+1}$.

Clearly, condition b) implies condition a); however, the proof in case a) is simpler, and all the other cases will follow from it in several steps.

Next we will study the symmetry measure.

Theorem 4 *Let $f(x) = a_k x^k + a_{k-1}x^{k-1} + \cdots + a_0$, $a_k \not\equiv 0 \pmod p$, $k < p$, and define t by*

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

Let

$$E'_{v-u+1} = \{e_u, e_{u+1}, \ldots, e_v\} \subseteq E_p$$

where e_i was defined in Definition 2. If $t < 2u$ or $t > 2v$ or $f(x) \not\equiv \pm f(t-x)$, then

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

Suppose that $f(x) \equiv \pm f(t'-x)$ for some t'. Considering the coefficients of x^k and x^{k-1} in $f(x)$ and $f(t'-x)$ we get

$$kt'a_k \equiv -2a_{k-1} \pmod{p}.$$

Thus it follows from $f(x) \equiv \pm f(t'-x) \pmod{p}$ that $t' \equiv t \pmod{p}$.

It is trivial from the definition of the e_i-s, that in the case of $f(x) \equiv \pm f(t-x) \pmod{p}$ we have

$$H(E'_{v-u+1}, u, t-u) = \lceil (t-2u)/2 \rceil \text{ if } t \leq u+v,$$

and

$$H(E'_{v-u+1}, t-v, v) = \lceil (2v-t)/2 \rceil \text{ if } t > u+v.$$

Therefore $S(E'_{v-u+1}) \gg \min\{t-2u, 2v-t\}$. So the condition of Theorem 4 is necessary apart from an additional term $O(p^{1/2}(\log p)^3)$, i.e., the conclusion of the theorem fails if the inequalities $t < 2u$, $t > v$ are replaced by $t < 2u + c_1 p^{1/2}(\log p)^3$, $t > 2v - c_1 p^{1/2}(\log p)^3$, where c_1 is a large enough constant.

Remark 1 *All these theorems are trivial if $k \geq p^{1/2}$, thus throughout the paper we will assume that $k < p^{1/2}$.*

Finally we will study the complexity measure of the family of pseudorandom sequences defined by (3.3).

Theorem 5 *Consider all the polynomials $f(x) \in F_p[x]$ with*

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

For each of these polynomials $f(x)$, consider the binary sequence $E_p = E_p(f)$ defined by (3.3), and let $\mathcal{F}$ denote the family of all binary sequences obtained in this way. Then we have

$$C(\mathcal{F}) > K.$$

3.2 Proofs

Proof of Theorem 1.

We will need the following lemmas:

Lemma 1 *Let $f(x)$ be a polynomial in $F_p[x]$, and let d be a divisor of $p-1$. The following 3 conditions are equivalent:*

(i) $f(x) = b\,(z(x))^d$ *with* $b \in F_p$, $z(x) \in F_p[x]$,

(ii) $f(x) = (h(x))^d$ *with* $h(x) \in \overline{F}_p[x]$,

(iii) $f(x) = b(x-x_1)^{\alpha_1}(x-x_2)^{\alpha_2}\ldots(x-x_s)^{\alpha_s}$ *with* $x_i \in \overline{F}_p$ *and* $d \mid (\alpha_1, \alpha_2, \ldots, \alpha_s)$.

Proof of Lemma 1.

See in [27, p. 51].

Lemma 2 *Suppose that p is a prime, χ is a non-principal character modulo p of order d, $f(x) \in F_p[x]$ has s distinct roots in $\overline{F}_p$, and it is not a constant multiple of a d-th power of a polynomial over F_p. Let y be a real number with $0 < y \leq p$. Then for any $x \in \mathbb{R}$:*

$$\left|\sum_{x<n\leq x+y} \chi(f(n))\right| < 9sp^{1/2}\log p.$$

Poof of Lemma 2

This is a trivial consequence of Lemma 1 and Lemma 2 in [2]. Indeed, there this result is deduced from Weil's theorem [28].

Now, we will prove Theorem 1. Let $f(x) = b(x-x_1)^{\alpha_1}\ldots(x-x_s)^{\alpha_s}$ where $x_i \neq x_j$. By Lemma 1, there exists a polynomial $z(x)$ with $f(x) = b(z(x))^d$ where $d = (\alpha_1, \ldots, \alpha_s)$. It is also obvious from Lemma 1 that $f(x)$ is not a constant multiple of any d'-th power for any $d' \mid d$. Assume now that

$$1 \leq a \leq a + (t-1)b \leq p-1.$$

The following computations and inequalities can be obtained in the same way as in [26], replacing $a+jb$ there by $f(a+jb)$.

$$|U(E_N, t, a, b)| = \frac{2}{p-1}\left|\sum_{\chi\neq\chi_0}\left(\sum_{j=0}^{t-1}\overline{\chi}(f(a+jb)\right)\left(\sum_{k=0}^{(p-1)/2}\chi^k(g)\right)\right|.$$

By the triangle inequality we have

$$|U(E_N,t,a,b)| \le \frac{2}{p-1}\left|\sum_{\chi^d\neq 1}\left(\sum_{j=0}^{t-1}\overline{\chi}(f(a+jb)\right)\left(\sum_{k=1}^{(p-1)/2}\chi^k(g)\right)\right| + \frac{2}{p-1}\left|\sum_{\substack{\chi\neq\chi_0\\ \chi^d=1}}\left(\sum_{j=0}^{t-1}\overline{\chi}(f(a+jb))\right)\left(\sum_{k=1}^{(p-1)/2}\chi^k(g)\right)\right| = \sum\nolimits_1 + \sum\nolimits_2. \tag{3.4}$$

Next we give an upper bound for $\sum_1$ in the same way as in [26]. If we consider a typical term in $\sum_1$: $\left(\sum_{j=0}^{t-1}\overline{\chi}(f(a+jb))\right)\left(\sum_{k=0}^{(p-1)/2}\chi^k(g)\right)$ then the order of χ does not divide d because $\chi^d \neq 1$. Let the order of χ be d'. $d' \nmid d$ so $f(x)$ is not a constant multiple of a d'-th power. Thus we may use Lemma 2:

$$\left|\sum_{j=0}^{t-1}\overline{\chi}(f(a+jb))\right| \le 9sp^{1/2}\log p. \tag{3.5}$$

We need an upper bound for $\sum_{\chi^d\neq 1}\left|\sum_{k=1}^{(p-1)/2}\chi^k(g)\right|$.

Lemma 3 *Let $1 \le d \le p-1$ and $d \mid p-1$. Then*

$$\sum_{\substack{\chi\neq\chi_0\\ \chi^d=1}}\left|\sum_{k=1}^{(p-1)/2}\chi^k(g)\right| \le \sum_{\substack{\chi\neq\chi_0\\ \chi^d=1}}\frac{2}{|1-\chi(g)|} < 2d\log(d+1).$$

Proof of Lemma 3.

The proof is nearly the same as in [27, p. 380-381]. The only difference is in [27, p. 381] at (10), where now we have:

$$\sum_{\substack{\chi\neq\chi_0\\ \chi^d=1}}\frac{1}{|1-\chi(g)|} = \sum_{n=1}^{d-1}\frac{1}{|1-e(n/d)|} \le \frac{1}{4}\sum_{n=1}^{d-1}\frac{1}{\|n/d\|} \le \frac{1}{2}\sum_{n=1}^{[d/2]}\frac{1}{\|n/d\|}$$

$$= \frac{1}{2}\sum_{n=1}^{[d/2]}\frac{d}{n} < \frac{1}{2}d(1+\log(d/2)) < d\log(d+1)$$

which completes the proof.

Since χ is a multiplicative character of order $p-1$, thus we have $\chi^{p-1}=1$. Applying Lemma 3 with $d=p-1$ we get:

$$\sum_{\chi\neq\chi_0}\left|\sum_{k=1}^{(p-1)/2}\chi^k(g)\right|\leq\sum_{\chi\neq\chi_0}\frac{2}{|1-\chi(g)|}<2(p-1)\log(p).$$

It follows that

$$\frac{2}{p-1}\sum\nolimits_1\leq 36sp^{1/2}(\log p)^2. \tag{3.6}$$

Finally we give an upper bound for $\sum_2$:

$$\begin{aligned}\frac{2}{p-1}\sum\nolimits_2&=\frac{2}{p-1}\left|\sum_{\substack{\chi\neq\chi_0\\ \chi^d=1}}\left(\sum_{j=0}^{t-1}\overline{\chi}(bg^d(a+jb))\right)\left(\sum_{k=1}^{(p-1)/2}\chi^k(g)\right)\right|\\&=\frac{2}{p-1}\left|\sum_{\substack{\chi\neq\chi_0\\ \chi^d=1}}\left(\sum_{j=0}^{t-1}\overline{\chi}^d(z(a+jb))\right)\left(\sum_{k=1}^{(p-1)/2}\chi^k(g)\right)\right|\\&\leq\frac{2}{p-1}\sum_{\chi\neq\chi_0}t\left|\sum_{k=1}^{(p-1)/2}\chi^k(g)\right|.\end{aligned}$$

Using Lemma 3 we get:

$$\frac{2}{p-1}\sum\nolimits_2\leq\frac{4}{p-1}td\log d<4d\log d.$$

Using that d is less than the degree of f, which is k, and $k<p^{1/2}$ we get

$$\frac{2}{p-1}\sum\nolimits_2\leq 4k\log k<2p^{1/2}\log p. \tag{3.7}$$

From (3.4), (3.6) and (3.7) we get

$$|U(E_N,t,a,b)|\leq 38kp^{1/2}(\log p)^2.$$

Proof of Theorem 2.

We will use addition theorems as in [8]. First we need the following definition:

Definition 3 *Let $\mathcal{A}$ and $\mathcal{B}$ be multi-sets of the elements of $\mathbb{Z}_p$. If $\mathcal{A}+\mathcal{B}$ represents every element of $\mathbb{Z}_p$ with multiplicity divisible by $p-1$, i.e., for all $c \in \mathbb{Z}_p$, the number of solutions of*

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

(the a's and b's are counted with their multiplicities) is divisible by p-1, then the sum $\mathcal{A}+\mathcal{B}$ is said to have property P.

Lemma 4 *Let $\mathcal{A} = \{a_1, \ldots, a_1, \ldots, a_r \ldots, a_r\}$ and $\mathcal{D} = \{d_1, \ldots, d_1, \ldots, d_\ell, \ldots, d_\ell\}$ be multi-sets of the elements of $\mathbb{Z}_p$ where the multiplicity of a_i is α_i in $\mathcal{A}$ and the multiplicity of d_i is δ_i in $\mathcal{D}$. If one of the following two conditions holds:*

(i) $\min\{r, \ell\} \leq 2$ *and* $\max\{r, \ell\} \leq p-1$,

(ii) $(4\ell)^r \leq p$ *or* $(4r)^\ell \leq p$,

then there exists $c \in \mathbb{Z}_p$ such that

$$a+d=c \quad a \in \mathcal{A},\ d \in \mathcal{D}$$

has exactly $\alpha_i \delta_j$ solutions for some $1 \leq i \leq r$, $1 \leq j \leq \ell$ (i.e. the solution is unique apart from the multiplicities).

Proof of Lemma 4.

Consider the simple sets $\mathcal{A}' = \{a_1, a_2, \ldots, a_r\}$, $\mathcal{D}' = \{d_1, d_2, \ldots, d_r\}$. It was proved in [8, Theorem 2] that under any of these conditions there is a $c \in \mathbb{Z}_p$ such that

$$a+d=c \quad a \in \mathcal{A}',\ d \in \mathcal{D}'$$

has exactly one solution. The statement of the lemma follows easily from this.

To prove Theorem 2, consider any $\mathcal{D} = (d_1, d_2, \ldots, d_\ell)$ with non-negative integers $d_1 < d_2 < \cdots < d_\ell$ and positive integers M with $M + d_\ell \leq N$. Then

arguing as in [26, p. 382] with $f(x+d_j)$ in place of $n+d_j$, we have

$$|V(E_N,M,D)| \le \frac{2^\ell}{(p-1)^\ell} \sum_{\chi_1\neq\chi_0} \cdots \sum_{\chi_\ell\neq\chi_0} \left|\sum_{x=1}^{M} \chi_1(f(x+d_1))\cdots\chi_\ell(f(x+d_\ell))\right| \times \prod \left|\sum_{\ell_j=1}^{(p-1)/2} \overline{\chi}_j(g^{\ell_j})\right|. \tag{3.8}$$

Now, let χ be a generator of the group modulo p characters, e.g. χ can be chosen as the character χ uniquely defined by $\chi(g) = e\left(\frac{1}{p-1}\right)$. The order of χ is $p-1$. Let

$$\chi_u = \chi^{\delta_u} \qquad \text{for } u = 1,2,\ldots,\ell$$

where, by $\chi_1 \neq \chi_0, \ldots, \chi_\ell \neq \chi_0$, we may take

$$1 \le \delta_u < p-1 \qquad \text{for } u = 1,2,\ldots,\ell.$$

Thus in (3.8) we have

$$\left|\sum_{x=1}^{M} \chi_1(f(x+d_1)\ldots\chi_\ell f(x+d_\ell)\right| = \left|\sum_{x=1}^{M} \chi^{\delta_1}(f(x+d_1)\ldots\chi^{\delta_\ell} f(x+d_\ell)\right| = \left|\sum_{x=1}^{M} \chi(f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell))\right|$$

If $f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell)$ is not a perfect $p-1$-th power, then this sum can be estimated by Lemma 3, whence

$$\left|\sum_{x=1}^{M} \chi(f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell))\right| \le 9s\ell p^{1/2}\log p.$$

Therefore by (3.8) and the triangle-inequality we get:

$$|V(E_N,M,D)| \le \frac{2^\ell}{(p-1)^\ell} \left|\sum_{\chi_1\neq\chi_0} \cdots \sum_{\chi_\ell\neq\chi_0} 9s\ell p^{1/2}\log p \prod_{j=1}^{\ell} \left(\sum_{l_j=1}^{(p-1)/2} \chi^{\delta_j}(g^{\ell_j})\right)\right| + \frac{2^\ell}{(p-1)^\ell} \left| \sum_{\substack{1\le\delta_1,\ldots,\delta_\ell\le p-2,\\ f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell) \text{ is}\\ \text{a perfect } p-1\text{-th power}}} (p-1) \prod_{j=1}^{\ell} \left(\sum_{l_j=1}^{(p-1)/2} \chi^{\delta_j}(g^{\ell_j})\right)\right|$$

$$= \frac{2^\ell}{(p-1)^\ell} \sum\nolimits_1 + \frac{2^\ell}{(p-1)^\ell} \sum\nolimits_2. \tag{3.9}$$

By [26, p.384] we have

$$\frac{2^\ell}{(p-1)^\ell}\sum\nolimits_1 \leq 9s\ell 4^\ell p^{1/2}(\log p)^{\ell+1}. \tag{3.10}$$

It remains give an upper bound for $\sum_2$. If f is irreducible it is obvious that $f^{\delta_1}(x+d_1)\dots f^{\delta_\ell}(x+d_\ell)$ is a perfect $p-1$-th power if and only if $p-1 \mid \delta_1,\dots,\delta_\ell$. But in $\sum_2$ we have $1 \leq \delta_1,\dots,\delta_\ell \leq p-2$, therefore $\sum_2 = 0$ which proves theorem 2 in case a). In case b), c), d) we will prove that $\sum_2$ is small. We need the following lemma to estimate $\sum_2$.

Lemma 5 *For all $1 \leq \delta_1,\dots,\delta_\ell \leq p-2$ such that $f^{\delta_1}(x+d_1)\dots f^{\delta_\ell}(x+d_\ell)$ is a perfect $p-1$-th power, there is a δ_i $(1 \leq i \leq \ell)$ and an integer $1 \leq \alpha \leq k$ (where k is the degree of the polynomial $f(x)$) such that $p-1 \mid \alpha\delta_i$.*

We will prove lemma 5 later. Now, from this lemma we verify that

$$\frac{2^\ell}{(p-1)^\ell}\sum\nolimits_2 \leq k(k+1)\ell\, 2^{2\ell-1}(\log p)^{\ell-1}.$$

Consider a fixed ℓ-tuples $\{\delta_1,\dots,\delta_\ell\}$ for which $f^{\delta_1}(x+d_1)\dots f^{\delta_\ell}(x+d_\ell)$ is a perfect $p-1$-th power and $1 \leq \delta_1,\dots,\delta_\ell \leq p-2$. By Lemma 5, then there exits a δ_i with

$$p-1 \mid \delta_i\alpha.$$

But $0 < \alpha\delta_i < \alpha(p-1)$ and $\alpha \leq k$:

$$\begin{aligned} p-1 \leq \delta_i\alpha &\leq (\alpha-1)(p-1) \\ \frac{1}{\alpha} \leq \frac{\delta_i}{p-1} &\leq 1-\frac{1}{\alpha} \\ \frac{1}{k} \leq \frac{1}{\alpha} &\leq \left\| \frac{\delta_i}{p-1} \right\|. \end{aligned}$$

By $|1-e(\alpha)| \geq 4 \parallel \alpha \parallel$ we have

$$\frac{2}{|1-\chi^{\delta_i}(g)|} = \frac{2}{|1-e(\delta_i/(p-1))|} \leq \frac{1}{2\,\|\delta_i/(p-1)\|} < \frac{k}{2}. \tag{3.11}$$

By Lemma 5, we have

$$\sum\nolimits_2 \leq (p-1)\sum_{i=1}^{\ell}\sum\nolimits_{2,i} \tag{3.12}$$

where

$$\sum_{2,i} = \sum_{\substack{1\le\delta_1,\dots,\delta_\ell\le p-2,\\ f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell) \text{ is}\\ \text{a perfect } p-1\text{-th power},\\ \exists\ 1\le\alpha\le k,\ p-1\mid\alpha\delta_i}} \prod_{j=1}^{\ell} \left| \sum_{\ell_j=1}^{(p-1)/2} \chi^{\delta_j}(g^{\ell_j}) \right|.$$

By (3.11) and $\left|\sum_{\ell_j=1}^{(p-1)/2} \chi^{\delta_j}(g^{\ell_j})\right| \le \frac{2}{|1-\chi^{\delta_j}(g)|}$ we get:

$$\sum_{2,i} \le \frac{k}{2} \sum_{1\le\delta_1,\dots,\delta_{i-1},\delta_{i+1},\dots,\delta_\ell\le p-2} \prod_{j\ne i} \frac{2}{|1-\chi^{\delta_j}(g)|} \sum_{\substack{1\le\delta_i\le p-2,\\ f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell) \text{ is}\\ \text{a perfect } p-1\text{-th power}}} 1. \quad (3.13)$$

Next we give an upper bound for

$$\sum_{\substack{1\le\delta_i\le p-2,\\ f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell) \text{ is}\\ \text{a perfect } p-1\text{-th power}}} 1 \stackrel{\text{def}}{=} r.$$

For fixed $\delta_1,\dots,\delta_{i-1},\delta_{i+1},\dots,\delta_\ell$ let $1 \le x_1 < x_2 < \cdots < x_r \le p-2$ denote the numbers for which $f^{\delta_1}(x+d_1)\cdots f^{\delta_{i-1}}(x+d_{i-1}) f^{x_j}(x+d_i) f^{\delta_{i+1}}(x+d_{i+1})\cdots f^{\delta_\ell}(x+d_\ell)$ is a perfect $p-1$-th power. It is clear that the quotient of two polynomials of this form is a $p-1$-th power, so

$$f^{x_j-x_{j-1}}(x+d_i) \qquad (\text{for } j=2,3,\dots,r)$$

is a perfect $p-1$-th power. The degree of $f^{x_j-x_{j-1}}(x+d_i)$ is $(x_j-x_{j-1})k$, ant this degree is divisible by $p-1$, therefore

$$x_j - x_{j-1} \ge \frac{p-1}{k}.$$

So

$$p-1 > x_r > \sum_{j=2}^{r} (x_j - x_{j-1}) \ge (r-1)\frac{p-1}{k}.$$

By this:

$$\sum_{\substack{1\le\delta_i\le p-2,\\ f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell) \text{ is}\\ \text{a perfect } p-1\text{-th power}}} 1 = r \le k+1.$$

Thus we get from (3.13):

$$\sum_{2,i} \leq \frac{k(k+1)}{2} \sum_{1\leq\delta_1,\ldots,\delta_{i-1},\delta_{i+1},\ldots,\delta_\ell\leq p-2} \prod_{j\neq i} \frac{2}{|1-\chi^{\delta_j}(g)|}$$
$$= \frac{k(k+1)}{2} 2^{\ell-1} \left(\sum_{\chi\neq\chi_0} \frac{1}{|1-\chi(g)|} \right)^{\ell-1}.$$

By Lemma 3 we have

$$\sum_{2,i} \leq \frac{k(k+1)}{2} 2^{\ell-1}(p-1)^{\ell-1}(\log p)^{\ell-1}.$$

From this and (3.12):

$$\sum_{2} \leq k(k+1)\ell\ 2^{\ell-2}(p-1)^{\ell}(\log p)^{\ell-1}.$$

By this, (3.9), (3.10) and $k < p^{1/2} - 1$, we get the statement of Theorem 2. It remains to prove Lemma 5.

Proof of Lemma 5.

The following equivalence relation was defined in [8]: We will say that the polynomials $\varphi(x), \psi(x) \in F_p[x]$ are equivalent, if there is an $a \in F_p$ such that $\psi(x) = \phi(x+a)$. Clearly, this is an equivalence relation.

Write $f(x)$ as the products of irreducible polynomials over F_p. Let us group these factors so that in each group the equivalent irreducible factors are collected. Consider a typical group $\phi(x+a_1), \ldots, \phi(x+a_r)$. Then $f(x)$ is of the form $f(x) = \varphi^{\alpha_1}(x+a_1)\ldots\varphi^{\alpha_r}(x+a_r)z(x_r)$ where $z(x)$ has no irreducible factors equivalent with any $\varphi(x+a_i)$ $(1 \leq i \leq r)$.

Let $h(x) = f^{\delta_1}(x+d_1)\cdots f^{\delta_\ell}(x+d_\ell)$. be a perfect $p-1$-th power where $1 \leq \delta_1, \ldots \delta_\ell \leq p-2$. Then writing $h(x)$ as the product of irreducible polynomials over F_p, all the polynomials $\varphi(x+a_i+d_j)$ wit $1 \leq i \leq r$, $1 \leq i \leq r$, $1 \leq j \leq \ell$ 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 irreducible polynomials cannot have a common zero in the algebraic closured of F_p, therefore each of the zeros of h is of multiplicity divisible by $p-1$, 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 multiplicity divisible by $p-1$. In other words writing $\mathcal{A} = \{a_1, \ldots, a_1, \ldots, a_r, \ldots, a_r\}$, $\mathcal{D} = \{d_1, \ldots, d_1, \ldots, d_\ell, \ldots, d_\ell\}$ where a_i has the multiplicity α_i in $\mathcal{A}$ (α_i is the exponent of $\varphi(x+a_i)$ in the factorization of $f(x)$) and d_i has the multiplicity δ_i in $\mathcal{D}$, then for each group $\mathcal{A}+\mathcal{D}$ must possess property P.

If condition b) holds in Theorem 2, then considering the degrees of the irreducible factors of $f(x)$, we see that there exists a group for which $r \leq 2$, i.e., $\mathcal{A}$ contains at most two distinct elements. So if one of the conditions of Theorem 2 holds then there exists a group for which the multi-sets $\mathcal{A}$ and $\mathcal{D}$ satisfy the conditions of Lemma 4, we get that there exists a c such that

$$a + d = c \quad a \in \mathcal{A},\ d \in \mathcal{D}$$

has exactly $\alpha_i \delta_j$ solution for some $1 \leq i \leq r$ and $1 \leq j \leq \ell$. But $\mathcal{A}+\mathcal{D}$ possess property $\mathcal{P}$, therefore $p-1 \mid \alpha_i \delta_j$. Because α_i is the exponent of an irreducible factor in $f(x)$, we also have $1 \leq \alpha_i \leq k$. Which completes the proof of Lemma 5.

Proof of Theorem 3.

We will use the following lemma.

Lemma 6 *If* $f(x) \equiv \pm f(t-x) \bmod p$, *then there exists a permutation* $\{x_1 \ldots, x_s\}$ *of the distinct roots of* $f(x)$ *such that*

$$t \equiv x_1 + x_s \equiv x_2 + x_{s-1} \equiv \cdots \equiv x_{\lceil s/2 \rceil} + x_{s+1-\lceil ss/2 \rceil}$$

and denoting the multiplicity of the root x_i *by* α_i $(1 \leq i \leq s)$ *we also have* $\alpha_i = \alpha_{s+1-i}$.

Proof of Lemma 6.

This is a consequence of the fact every polynomial has unique factorization over F_p. Now we can return to the proof of Theorem 3.

In the same way as the estimates of the correlation measure we obtain:

$$H(E_N,a,b) \le \frac{4}{(p-1)^2}\sum_{\chi_1\neq\chi_0}\sum_{\chi_2\neq\chi_0} 18sp^{1/2}\log p\prod_{j=1}^{2}\left|\sum_{l_j=1}^{(p-1)/2}\chi_j(g^{l_j})\right|$$
$$+\frac{4}{(p-1)^2}\sum_{\substack{1\le\delta_1,\delta_2\le p-2,\\ f^{\delta_1}(a+x)f^{\delta_2}(b-x)\text{ is a}\\ \text{perfect } p-1\text{-th power}}}(p-1)\prod_{j=1}^{2}\left|\sum_{l_j=1}^{(p-1)/2}\chi_j(g^{l_j})\right|$$
$$=\frac{4}{(p-1)^2}\sum\nolimits_1+\frac{4}{(p-1)^2}\sum\nolimits_2. \tag{3.14}$$

Again as in [26, p. 384] we have

$$\frac{4}{(p-1)^2}\sum\nolimits_1 \le 72kp^{1/2}(\log p)^3. \tag{3.15}$$

To give an upper bound for $\sum_2$ we have to handle the case when the polynomial $f^{\delta_1}(a+x)f^{\delta_2}(b-x)$ is a perfect $p-1$-th power. Suppose that there is no permutation $\{x_1\ldots,x_s\}$ with

$$t\equiv x_1+x_s\equiv x_2+x_{s-1}\equiv\cdots\equiv x_{\lceil s/2\rceil}+x_{s+1-\lceil ss/2\rceil}.$$

Then there exists a root of $f(a+x)$ which is not the root of $f(b-x)$ (x is the variable). Denote this root by x_i-a and let α_i the multiplicity of the root x_i-a in $f(a+x)$. Then $p-1\mid\alpha_i\delta_1$ because $f^{\delta_1}(a+x)f^{\delta_2}(b-x)$ is a perfect $p-1$-th power. But also $1\le\alpha_i\le k$, so in this special case in the same way as we get the result of Theorem 2 from Lemma 5, we obtain:

$$\frac{4}{(p-1)^2}\sum\nolimits_2 \le 16k(k+1)(\log p)^3.$$

From this, (3.14), (3.15) we get

$$H(E_N,a,b)\le 88kp^{1/2}(\log p)^3.$$

The case when

$$t\equiv x_1+x_s\equiv x_2+x_{s-1}\equiv\cdots\equiv x_{\lceil s/2\rceil}+x_{s+1-\lceil ss/2\rceil}.$$

holds is slightly more difficult. Considering the multiplicity of the roots $x_i - a \equiv b - x_{s+1-i}$, $x_{s+1-i} - a \equiv b - x_i \bmod p$ in the polynomial $f^{\delta_1}(a+x)f^{\delta_2}(b-x)$ we get:

$$p-1 \mid \delta_1\alpha_i + \delta_2\alpha_{s+1-i},$$
$$p-1 \mid \delta_1\alpha_{s+1-i} + \delta_2\alpha_i$$

Taking the sum and the difference we obtain

$$p-1 \mid (\delta_1 - \delta_2)(\alpha_i - \alpha_{s+1-i}),$$
$$p-1 \mid (\delta_1 + \delta_2)(\alpha_i + \alpha_{s+1-i}) \qquad (1 \le i \le s). \tag{3.16}$$

By Lemma 6 we know that there exists an i for which $\alpha_i \neq \alpha_{s+1-i}$. By $1 \le |\alpha_i - \alpha_{s+1-i}| \le k$, $1 \le |\alpha_i + \alpha_{s+1-i}| \le 2k$ and (3.16), we obtain that both $\delta_1 - \delta_2$ and $\delta_1 + \delta_2$ may assume at most $2k$ different values. Therefore at most $(2k)^2$ pairs $\{\delta_1, \delta_2\}$ exist for which $f^{\delta_1}(a+x)f^{\delta_2}(b-x)$ is a perfect $p-1$-th power.

By $|1 - e(\alpha)| \ge 4 \parallel \alpha \parallel$, $\chi_i = \chi^{\delta_i}$, we have:

$$\prod_{j=1}^{2} \left| \sum_{\ell_j=1}^{(p-1)/2} \chi_j(g^{\ell_j}) \right| \le \frac{4}{|1-\chi^{\delta_1}(g)|\,|1-\chi^{\delta_2}(g)|} \le \frac{1}{4 \parallel \frac{\delta_1}{p-1} \parallel \parallel \frac{\delta_2}{p-1} \parallel}. \tag{3.17}$$

Next we will prove:

$$\frac{1}{\parallel \frac{\delta_1}{p-1} \parallel \parallel \frac{\delta_2}{p-1} \parallel} \le 2pk. \tag{3.18}$$

Lemma 7 *If $x, y \in \mathbb{N}$, $y \neq 0$, $p-1 \mid xy$, but $p-1 \nmid x$ then we have $\parallel \frac{x}{p-1} \parallel \ge \frac{1}{y}$.*

Proof of Lemma 7.

Let $x = r(p-1) + q$ where $1 \le q \le p-2$. Then

$$r(p-1)y < xy < (r+1)(p-1)y.$$

But $p-1 \mid xy$, so:

$$(p-1)(ry+1) \le xy \le (p-1)(ry+y-1),$$
$$r + \frac{1}{y} \le \frac{x}{p-1} \le r + 1 - \frac{1}{y},$$
$$\left\| \frac{x}{p-1} \right\| \ge \frac{1}{y}$$

which was to be proved.

If $p-1 \mid \delta_1 - \delta_2$ and $p-1 \mid \delta_1 + \delta_2$ then by $1 \le \delta_1, \delta_2 \le p-2$ we have $\delta_1 = \delta_2 = \frac{p-1}{2}$ and (3.18) is trivial. We may suppose that at least one of $\delta_1 - \delta_2$, $\delta_1 + \delta_2$ is not divisible by $p-1$. If $p-1 \nmid \delta_1 - \delta_2$ then $p-1 \mid (\delta_1 - \delta_2)(\alpha_i - \alpha_{s+1-i})$, and using Lemma 8 we get:

$$\left\| \frac{\delta_1 - \delta_2}{p-1} \right\| \ge \frac{1}{|\alpha_i - \alpha_{s+1-i}|} \ge \frac{1}{k}.$$

If $p-1 \nmid \delta_1 + \delta_2$ then $p-1 \mid (\delta_1 + \delta_2)(\alpha_i + \alpha_{s+1-i})$, and using again Lemma 8 we get:

$$\left\| \frac{\delta_1 + \delta_2}{p-1} \right\| \ge \frac{1}{|\alpha_i + \alpha_{s+1-i}|} \ge \frac{1}{2k}.$$

By the triangle-inequality in both cases we have:

$$\left\| \frac{\delta_1}{p-1} \right\| + \left\| \frac{\delta_2}{p-1} \right\| \ge \left\| \frac{\delta_1 \pm \delta_2}{p-1} \right\| \ge \frac{1}{2k}.$$

But $\frac{1}{p-1} \le \|\frac{\delta_1}{p-1}\|, \|\frac{\delta_2}{p-1}\|$, so trivially

$$\left\| \frac{\delta_1}{p-1} \right\| \left\| \frac{\delta_2}{p-1} \right\| \ge \frac{1}{2k(p-1)} \ge \frac{1}{2kp},$$

from which (3.18) follows. By (3.17), (3.18), and since there are at least $(2k)^2$ pairs for which $f^{\delta_1}(a+x)f^{\delta_2}(b-x)$ is a perfect $p-1$-th power, we have

$$\sum\nolimits_2 \le \sum_{\substack{1 \le \delta_1, \delta_2 \le p-2, \\ f^{\delta_1}(a+x)f^{\delta_2}(b-x) \text{ is a} \\ \text{perfect } p-1\text{-th power}}} (p-1)\frac{1}{2}pk \le \frac{1}{2}(p-1)pk(2k)^2 = 2(p-1)pk^3. \tag{3.19}$$

From (3.14), (3.15) and (3.19) we get

$$H(E_N, a, b) \le 88kp^{1/2}(\log p)^3$$

which proves the theorem.

Proof of Theorem 4. The proof is exactly the same as in [1, Theorem 1], the only difference is in the definitions of q and r: now we choose q, r as integers with $(q,p) = (r,p) = 1$ and $1 \le \operatorname{ind} q \le \frac{p-1}{2}$, $\frac{p-1}{2} < \operatorname{ind} r \le p-1$.

3.3 Numerical Calculations

In this chapter our goal is to carry out numerical calculations. Partly to see how far our theoretical estimates are from the probable truth, partly to gather numerical data in cases when we cannot prove any theoretical estimates (linear complexity and the correlation measure of higher order). In particular, one might like to gather information on the linear complexity (see, e.g. [23]) which is another characteristic closely related to pseudorandomness. The linear complexity is defined as it follows.

Definition 4 *The linear complexity of a finite binary sequence* $\{s_0, \ldots, s_{N-1}\} \in \{0,1\}^N$ *is the smallest integer* L *for which there exist numbers* $c_1, \ldots, c_{L-1} \in \{0,1\}$ *such that*

$$s_n \equiv c_1 s_{n-1} + c_2 s_{n-2} + \cdots + c_{L-1} s_{n-(L-1)} + s_{n-L} \bmod 2 \qquad \text{for all } n \geq L.$$

We construct a sequence $\{s_0, \ldots, s_{p-2}\} \in \{0,1\}^{p-1}$ from our sequence $E_{p-1} = \{e_1, \ldots, e_{p-1}\} \in \{-1,+1\}^{p-1}$ in the following way: $s_i = \frac{1}{2}(1 - e_{i+1})$ for all $0 \leq i \leq p-2$. One might like to study the linear complexity of this sequence. Unfortunately we haven't been able to prove any non-trivial theoretical result. Thus all we can do in this direction is, again, to carry out numerical computations; we will use the Berlekamp-Massey algorithm [5],[17] (L denotes the linear complexity).

p prime	$\sqrt{p}$	polynomial	W	C_2	C_3	C_4	S	L
1009	31.764	x^3+1	38	98	132	152	72	503
		$x^4+511x^2+123x+851$	45	102	138	157	68	504
		$122x^4+1000x^3+22x^2+626x+500$	37	88	126	158	75	505
		$212x^{20}+567x^{13}+333x^8+9x+12$	60	96	130	146	72	504
1013	31.827	x^3+1	38	123	129	151	84	507
		$x^4+511x^2+123x+851$	40	104	136	146	67	508
		$122x^4+1000x^3+22x^2+626x+500$	42	102	128	165	77	506
		$212x^{20}+567x^{13}+333x^8+9x+12$	59	103	144	150	72	508

p prime	$\sqrt{p}$	polynomial	W	C_2	S	L
100069	316.336	x^3+1	623	1284	923	50036
		$x^4+75638x^2+54322x+81512$	689	1348	1150	50034
		$x^4+34879x^3+98537x^2+12378x+68921$	402	1373	861	50034
		$x^{100}+45623x^{89}+98254x^{63}+74563x^{30}+78346x^{17}$	445	1365	963	50033
100237	316.602	x^3+1	885	1392	919	50117
		$x^4+5433x^2+5432x+23789$	383	1297	859	50118
		$x^4+50000x^3+28657x^2+112211x+854$	410	1367	975	50118
		$x^{100}+x^{84}+456789x^{73}+x^{72}+8789x^4+4$	614	1315	970	50119

The data above seem to point to the direction that our condition on k and ℓ can be relaxed considerably, and that the correlation of not very high

order tends to be relatively small also for k, ℓ values not covered by Theorem 2. Perhaps this data also indicate that the dependence on the degree of the polynomial in the upper bounds for the pseudorandom measures need not be as strong as in our theorems. Most of the time the linear complexity seems to be around $p/2$ as it would happen for truly random sequences, so that our sequence also satisfies the requirement of high linear complexity.

3.4 Conclusion

By using the notion of index (discrete logarithm) we have constructed large families of binary sequences with strong pseudorandom properties. However, the weak point of this construction is that the generation of these sequences is very slow (since there is no fast algorithm for computing the discrete logarithm). One might like to improve on this construction by trying to modify the construction so that we should obtain sequences which still have relatively good pseudorandom properties, however, they can be generated much faster. I will return to this problem in a subsequent paper.

I would like to thank Professors Julien Cassaigne, Joël Rivat, András Sárközy for the valuable discussions.

Chapter 4

On a fast version of a pseudorandom generator

Abstract

In an earlier paper I constructed a large family of pseudorandom sequences by using the discrete logarithm. While the sequences in this construction have strong pseudorandom properties, they can be generated very slowly since no fast algorithm is known to compute ind n. The purpose of this paper is to modify this family slightly so that the members of the new family can be generated much faster, and they have almost as good pseudorandom properties as the sequences in the original family.

2000 AMS Mathematics Subject Classification: 11K45

List of keywords and phrases: pseudorandom, index, discrete logarithm, correlation.

4.1 Introduction

In this work I will continue the work initiated in [10]. C. Mauduit and A. Sárközy [21, pp. 367-370] introduced the following measures of pseudorandomness:

For a finite binary sequence $E_N = \{e_1, e_2, \ldots, e_N\} \in \{-1, +1\}^N$ write

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

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

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

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 $1 \leq a \leq a + (t-1)b \leq N$. The *correlation measure of order* k of E_N is defined as

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

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ and M such that $M + d_k \leq N$. In [12] I introduced a further measure: Let

$$H(E_N, a, b) = \sum_{j=0}^{[(b-a)/2]-1} e_{a+j} e_{b-j},$$

and then the *symmetry measure* of E_N is defined as

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

A sequence E_N is considered as a "good" pseudorandom sequence if each of these measures $W(E_N)$, $C_k(E_N)$ (at least for small k) and $S(E_N)$ is "small" in terms of N (in particular all are $o(N)$ as $N \longrightarrow \infty$). Indeed, it was proved in [6, Theorem 1, 2] and in [12, Theorem 1, 2] that for a truly random sequence $E_N \subseteq \{-1, +1\}^N$ each of these measures is $\ll \sqrt{N \log N}$ and $\gg \sqrt{N}$.

Throughout the paper we will use the following notations: $\| x \|$ is the distance of x from the closest integer, $e(\alpha) = e^{2\pi i\alpha}$, $\overline{\mathbb{F}_p}$ is the algebraic closured of the field $\mathbb{F}_p$. Finally, if p is a prime, α and m are natural numbers we say that $p^\alpha \| m$ if $p^\alpha \mid m$ but $p^{\alpha+1} \nmid m$.

Numerous binary sequences have been tested for pseudorandomness by J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy. The sequences with the strongest pseudorandom properties have been constructed in [8], [10], [21], [20] and [26]. As concerning the strength of the pseudorandom properties these constructions are nearly equally good. But in the construction given by A. Sárközy in [26] and extended by me in [10], the generation of the sequences in question is much more slowly than in the other constructions. Indeed Sárközy's construction is the following:

Let p be an odd prime, $N = p - 1$ and define $E_N = \{e_1, \dots, e_N\} \subseteq \{-1, +1\}^N$ by

$$e_n = \begin{cases} +1 & \text{if } 1 \le \text{ind } n \le \frac{p-1}{2}, \\ -1 & \text{if } \frac{p+1}{2} \le \text{ind } n \le p-1. \end{cases} \tag{4.1}$$

Here ind n denotes the index or discrete logarithm of n modulo p, defined as the unique integer with

$$g^{\text{ind } n} \equiv n \pmod{p}, \tag{4.2}$$

and $1 \le \text{ind } n \le p-1$, where g is a fixed primitive root modulo p. In [10] I extended this construction to a large family of binary sequences with strong pseudorandom properties by replacing n by a polynomial $f(n)$ in (4.1) (in the same way as the Legendre symbol construction in [21] was extended in [8].)

Indeed in [10] I proved for the generalized sequence:

Theorem A *For all $f \in \mathbb{F}_p[x]$ with $k = \deg f$ we have*

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

Moreover if one of the following conditions holds:

a) *f is irreducible;*

b) *If f has the factorization $f = \varphi_1^{\alpha_1}\varphi_2^{\alpha_2}\dots\varphi_u^{\alpha_u}$, where $\alpha_i \in \mathbb{N}$ and the φ_i's are irreducible over $\mathbb{F}_p$, then there exists a β such that exactly one or two φ_i's have the degree β;*

c) $\ell = 2$;

d) $(4\ell)^k < p$ *or* $(4k)^\ell < p$.

Then

$$C_\ell(E_{p-1}) < 10k\ell 4^\ell p^{1/2}(\log p)^{\ell+1}.$$

Finally, if $f(x) \not\equiv f(t-x)$ for all $t \in \mathbb{Z}_p$, then

$$S(E_{p-1}) < 88kp^{1/2}(\log p)^3.$$

As we pointed out earlier these constructions are nearly as good as the others, but the problem is that it is slow to compute e_n since no fast algorithm is known to compute ind n. The Diffie-Hellman key-exchange system utilizes the difficulty of computing ind n.

In this paper my goal is to improve on the construction in Theorem A by replacing the sequence

$$e_n = \begin{cases} +1 & \text{if } 1 \le \text{ind } f(n) \le \frac{p-1}{2}, \\ -1 & \text{if } \frac{p+1}{2} \le \text{ind } f(n) \le p-1 \text{ or } p \mid f(n) \end{cases} \tag{4.3}$$

by a sequence which can be generated faster. I will show that this is possible at the price of giving slightly weaker upper bounds for the pseudorandom measures. Throughout this paper we will use the following:

Notation *Let p be an odd prime, g be a primitive root modulo p. Define* ind n *by* (4.2). *Let $f \in \mathbb{F}_p[x]$ be a polynomial of degree $k \ge 1$, and $f = ch^d$ where $c \in \mathbb{F}_p$ and $h \in \mathbb{F}_p[x]$ is not a perfect power of a polynomial over $\mathbb{F}_p[x]$. Moreover let*

$$m \mid p-1$$

with $m, h \in \mathbb{N}$, and let x be relative prime to m: $(x, m) = 1$.

The crucial idea of the construction is to reduce ind n modulo m:

Construction 1 *Let* $\text{ind}^* n$ *denote the following function: For all* $1 \leq n \leq p-1$

$$\text{ind } n \equiv x \cdot \text{ind}^* n \pmod{m}$$

($\text{ind}^* n$ *exists since* $(x, m) = 1$*.) Define the sequence* $E_{p-1} = \{e_1, \ldots, e_{p-1}\}$ *by*

$$e_n = \begin{cases} +1 & \text{if } 1 \leq \text{ind}^* f(n) \leq \frac{m}{2}, \\ -1 & \text{if } \frac{m}{2} < \text{ind}^* f(n) \leq m \text{ or } p \mid f(n). \end{cases} \tag{4.4}$$

Note that this construction also generalizes the Legendre symbol construction described in [8] and [21]. Indeed in the special case $m = 2$, $x = 1$ the sequence e_n defined in (4.4) becomes

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

(In the special case $m = p - 1$, $x = 1$ we obtain the original construction given in (4.3)).

We will show that the construction presented above has good pseudo-random properties, each of the measures $W(E_{p-1})$, $C_k(E_{p-1})$ is small under certain conditions on the polynomial f. In the case of the well-distribution measure we can control the situation completely.

Theorem 1 *If* $m/(m, d)$ *is even we have*

$$W(E_{p-1}) \leq 36kp^{1/2} \log p \log(m+1).$$

While in the other case, when $m/(m, d)$ *is odd we have:*

$$W(E_{p-1}) = \frac{p-1}{m} + O(kp^{1/2} \log p \log(m+1)).$$

In the case of the correlation measures the situation is slightly more difficult. When the order of the correlation measure is odd we have:

Theorem 2 *If* $f \in \mathbb{F}_p$, $k = \deg f$ *and* ℓ *are odd integers while* m *is an even integer, then we have*

$$C_\ell(E_{p-1}) < 9k\ell 4^\ell p^{1/2} (\log p)^{\ell+1}.$$

Otherwise we need the same conditions on the polynomial f as in [10] in the original construction. If the degree of the polynomial is small depending on m, the same upper bound holds as in [10], while in the general case I will prove a slightly weaker result.

Theorem 3 *i) Suppose that m is even or m is odd with $2m \mid p-1$, and at least one of the following 4 conditions holds:*

a) *f is irreducible;*

b) *If f has the factorization $f = \varphi_1^{\alpha_1}\varphi_2^{\alpha_2}\dots\varphi_u^{\alpha_u}$ where $\alpha_i \in \mathbb{N}$ and the φ_i's are irreducible over $\mathbb{F}_p$, then there exists a β such that exactly one or two φ_i's have the degree β;*

c) *$\ell = 2$;*

d) *$(4\ell)^k < p$ or $(4k)^\ell < p$.*

Then

$$C_\ell(E_{p-1}) < 9k\ell 4^\ell p^{1/2}(\log p)^{\ell+1} + \frac{\ell! k^{\ell(\ell+1)}}{m^\ell} p. \tag{4.5}$$

ii) Moreover if we also have $2^\beta \parallel m$ and $k = \deg f < 2^\beta$ then

$$C_\ell(E_{p-1}) < 9k\ell 4^\ell p^{1/2}(\log p)^{\ell+1}.$$

For fixed m by Heath-Brown's work on Linnik's theorem [14] the least prime number p with $m \mid p-1$ is less than $cm^{5.5}$. Thus the condition $\deg f < 2^\beta \parallel m \mid p-1$ is not too restrictive.

If $m^{2\ell} < p$ holds, then the first term majorizes the second term in (4.5), thus the upper bound becomes $O\left(p^{1/2}(\log p)^{\ell+1}\right)$ where the implied constant factor may depend on k and ℓ.

The study of the symmetry measure also considered in [10] would lead to further complications and I could control it only under the further assumption $\deg f \le 2^{\beta+2}$ where β is defined by $2^\beta \parallel m$. Thus, I do not go into the details of this here.

In applications one should balance between the strength of the upper bounds and the speed of the generation of the sequence depending on our

priorities. By the Pohlig-Hellman algorithm [24] we will show in section 3 that the sequence described in (4.4), in particular ind $^*f(n)$, can be computed faster than the original construction. Indeed, if the prime factors of m are smaller than $\log p$ then ind$^*f(n)$ can be computed by $O((\log p)^6)$ bit operations.

In [1] R. Ahlswede, L.H. Khachatrian, C. Mauduit and A. Sárközy introduced the notion of f-complexity of families of binary sequences as a measure of applicability of the constructions in cryptography.

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

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

We will see that the f-complexity of the family constructed in (4.4) is high.

Theorem 4 *Consider all the polynomials $f \in \mathbb{F}_p[x]$ with*

$$0 < \deg f \le K.$$

For each of these polynomials f, consider the binary sequence $E_{p-1} = E_{p-1}(f)$ defined by (4.4), and let $\mathcal{F}$ denote the family of all binary sequences obtained in this way. Then we have

$$C(\mathcal{F}) > K.$$

4.2 Proofs

4.2.1 Proof of Theorem 1.

First we note that the sequence defined in (4.4) by the polynomial $f = h^d$ and the modulus m, remains the same sequence if we replace in Construction 1 the polynomial $f = h^d$ by the polynomial $h^{d/(m,d)}$ and the modulus m by the modulus $m/(m,d)$. Thus in order to prove this theorem it is sufficient to study the case when $(m,d) = 1$.

The proof of the theorem is very similar to the proof of Theorem 1 in [12]. By the formula

$$\frac{1}{m}\sum_{\chi:\chi^m=1}\overline{\chi}^j(a)\chi(b)=\begin{cases}1 & \text{if } m \mid \operatorname{ind} a-\operatorname{ind} b,\\ 0 & \text{if } m \nmid \operatorname{ind} a-\operatorname{ind} b,\end{cases}$$

we obtain

$$e_n=2\sum_{\substack{1\le j\le m/2\\ jx\equiv \operatorname{ind} f(n) \pmod m}}1-1=\frac{2}{m}\sum_{1\le j\le m/2}\sum_{\chi:\chi^m=1}\overline{\chi}(f(n))\chi(g^{jx})-1.$$

Thus

$$e_n=\frac{2}{m}\sum_{1\le j\le m/2}\sum_{\chi\neq\chi_0:\chi^m=1}\overline{\chi}(f(n))\chi(g^{xj})+\frac{(-1)^m-1}{2m}. \tag{4.6}$$

Assume now that $1\le a\le a+(t-1)b\le N$. Then we have

$$|U(E_{p-1},t,a,b)|=\left|\frac{2}{m}\sum_{\chi\neq\chi_0:\chi^m=1}\left(\sum_{i=0}^{t-1}\overline{\chi}(f(a+ib))\right)\left(\sum_{j=1}^{[m/2]}\chi^j(g^x)\right)+\frac{((-1)^m-1)t}{2m}\right|. \tag{4.7}$$

We will prove the following:

$$S\overset{\text{def}}{=}\left|\frac{1}{m}\sum_{\chi\neq\chi_0:\chi^m=1}\left(\sum_{i=0}^{t-1}\overline{\chi}(f(a+ib))\right)\left(\sum_{j=1}^{[m/2]}\chi^j(g^x)\right)\right|$$
$$\le 18kp^{1/2}(\log p)^2. \tag{4.8}$$

If m is even we obtain the statement of Theorem 1 immediately from (4.7) and (4.8). If m is odd using the triangle inequality we get

$$|U(E_{p-1},t,a,b)|=\frac{t}{m}+O(kp^{1/2}(\log p)^2)$$

which completes the proof of Theorem 1. Thus in order to prove Theorem 1, we have to verify (4.8).

We will use the following lemma:

Lemma 1 *Suppose that p is a prime, χ is a non-principal character modulo p of order z, $f \in \mathbb{F}_p[x]$ has s distinct roots in $\overline{F}_p$, and it is not a constant multiple of a z-th power of a polynomial over $\mathbb{F}_p$. Let y be a real number with $0 < y \leq p$. Then for any $x \in \mathbb{R}$:*

$$\left| \sum_{x<n\leq x+y} \chi(f(n)) \right| < 9sp^{1/2} \log p.$$

Poof of Lemma 1

This is a trivial consequence of Lemma 1 in [2]. Indeed, there this result is deduced from Weil theorem, see [28].

Consider $\sum_{i=0}^{t-1} \overline{\chi}(f(a+ib))$ in (4.7), and here, let the order of χ be z. Since $\chi^m = 1$ we have $z \mid m$. On the other hand $f = ch^d$ is not a constant multiple of a z-th power of a polynomial over $\mathbb{F}_p$, since $1 = (m,d) = (z,d)$ (because of $z \mid m$) and h is not a perfect power of any polynomial over $\mathbb{F}_p$.

Using Lemma 1 we have:

$$\left| \sum_{i=0}^{t-1} \overline{\chi}(f(a+ib)) \right| \leq 9kp^{1/2} \log p$$

and thus by (4.8)

$$S \leq \frac{9kp^{1/2} \log p}{m} \sum_{\chi \neq \chi_0 : \chi^m = 1} \left| \sum_{j=1}^{[m/2]} \chi^j(g^x) \right|.$$

Lemma 2

$$\sum_{\chi \neq \chi_0 : \chi^m = 1} \left| \sum_{j=1}^{[m/2]} \chi^j(g^x) \right| \leq \sum_{\chi \neq \chi_0 : \chi^m = 1} \frac{2}{|1 - \chi(g^x)|} < 2m \log(m+1).$$

Proof of Lemma 2 This is Lemma 3 in [10] with m in place of d, $m/2$ in place of $(p-1)/2$ and g^x in place of g, respectively, and it can be proved in the same way.

Using Lemma 2 we obtain

$$S < 18kp^{1/2} \log p \log(m+1)$$

which proves (4.8) and this completes the proof of Theorem 1.

4.2.2 Proof of Theorem 2 and 3

In this section we may suppose that m is even: In Theorem 2 m cannot be odd. If m is odd in Theorem 3, then considering $2m$ in place of m and f^2 in place of f in Construction 1 we generate the same sequence; however in this case we have $(2m, 2d) > 1$.

To prove Theorems 2 and 3, consider any $\mathcal{D} = \{d_1, d_2, \ldots, d_\ell\}$ with non-negative integers $d_1 < d_2 < \cdots < d_\ell$ and positive integers M with $M + d_\ell \leq p - 1$. Then arguing as in [26, p. 382] with $f(n + d_j)$ in place of $n + d_j$, m in place of $p - 1$, and g^x in place of g from (4.6) and since m is even we obtain:

$$|V(E_N, M, D)| \leq \frac{2^\ell}{m^\ell} \sum_{\substack{\chi_1 \neq \chi_0, \\ \chi_1^m = 1}} \cdots \sum_{\substack{\chi_\ell \neq \chi_0 \\ \chi_\ell^m = 1}} \left| \sum_{n=1}^{M} \chi_1(f(n + d_1)) \cdots \chi_\ell(f(n + d_\ell)) \right| \times \prod \left| \sum_{\ell_j = 1}^{m/2} \overline{\chi}_j(g^{x\ell_j}) \right|. \tag{4.9}$$

Now let χ be a modulo p character of order m; for simplicity we will choose χ as the character uniquely defined by $\chi(g) = e\left(\frac{x^*}{m}\right)$ where $xx^* \equiv 1 \pmod{m}$. Then

$$\chi(g^x) = e\left(\frac{1}{m}\right). \tag{4.10}$$

Let $\chi_u = \chi^{\delta_u}$ for $u = 1, 2, \ldots, \ell$, whence by $\chi_1 \neq \chi_0, \ldots, \chi_\ell \neq \chi_0$, we may take

$$1 \leq \delta_u < m.$$

Thus in (4.9) we have

$$\begin{aligned} &\left| \sum_{n=1}^{M} \chi_1(f(n + d_1)) \ldots \chi_\ell(f(n + d_\ell)) \right| \\ &= \left| \sum_{n=1}^{M} \chi^{\delta_1}(f(n + d_1)) \ldots \chi^{\delta_\ell}(f(n + d_\ell)) \right| \\ &= \left| \sum_{n=1}^{M} \chi\left(f^{\delta_1}(n + d_1) \ldots f^{\delta_\ell}(n + d_\ell)\right) \right|. \end{aligned}$$

If $f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell)$ is not a perfect m-th power, then this sum can be estimated by Lemma 1, whence

$$\left|\sum_{n=1}^{M}\chi(f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell))\right| \le 9s\ell p^{1/2}\log p.$$

Therefore by (4.9) and the triangle-inequality we get:

$$\begin{aligned}|V(E_N,M,D)| &\le \frac{2^\ell}{m^\ell}\sum_{\substack{\chi_1\neq\chi_0\\ \chi_1^m=1}}\cdots\sum_{\substack{\chi_\ell\neq\chi_0\\ \chi_\ell^m=1}} 9s\ell p^{1/2}\log p\left|\prod_{j=1}^{\ell}\left(\sum_{l_j=1}^{m/2}\chi^{\delta_j}(g^{x\ell_j})\right)\right| \\ &+\frac{2^\ell}{m^\ell}\sum_{\substack{1\le\delta_1,\dots,\delta_\ell\le m,\\ f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell)\text{ is}\\ \text{a perfect }m\text{-th power}}}(p-1)\left|\prod_{j=1}^{\ell}\left(\sum_{l_j=1}^{m/2}\chi^{\delta_j}(g^{x\ell_j})\right)\right| \\ &=\sum\nolimits_1+\sum\nolimits_2. \end{aligned} \tag{4.11}$$

From Lemma 2 the same way as in [26, p.384] we have

$$\sum\nolimits_1 \le 9k\ell p^{1/2}(\log p)^{\ell+1}. \tag{4.12}$$

It remains to estimate $\sum_2$. First we claim that in Theorem 2 and in Theorem 3 (ii) we have $\sum_2 = 0$.

Indeed in these cases I will show that if $f^{\delta_1}(n+d_1)\dots f^{\delta_\ell}(n+d_\ell)$ is a perfect m-th power, then there exists a δ_i which is even. Then, if δ_i is even, by (4.10) and $m\nmid\delta_i$ $(1\le\delta_i\le m-1)$ we have

$$\sum_{\ell_j=1}^{m/2}\chi^{\delta_i}(g^{x\ell_j}) = \sum_{\ell_j=1}^{m/2} e\left(\frac{\delta_i/2}{m/2}\ell_j\right) = 0,$$

which means that in $\sum_2$ the product is 0, whence $\sum_2 = 0$. From this, (4.11) and (4.12) Theorem 2 and 3 (ii) follows.

Let us see the proof of those cases for which there exists an even δ_i. In the case of Theorem 2 if $f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell)$ is a perfect m-th power, then m divides the degree of $f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell)$ which is $k(\delta_1+\cdots+\delta_\ell)$. Contrary to our statement, suppose that all δ_i are odd. Then using that k and ℓ are also odd we get that $k(\delta_1+\cdots+\delta_\ell)$ is odd, which contradicts $2\mid m\mid k(\delta_1+\cdots+\delta_\ell)$. In the case of Theorem 3 (ii) we will use the following lemma, which is Lemma 5 of [10] with m in place of $p-1$.

Lemma 3 *Suppose that the conditions of Theorem 3 hold. Then if* $1 \leq \delta_1, \ldots, \delta_\ell \leq m-1$, *and* $f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell)$ *is a perfect* m*-th power, then there is a* δ_i $(1 \leq i \leq \ell)$ *and an integer* $1 \leq \alpha \leq k$ *such that* $m \mid \alpha\delta_i$.

By Lemma 3 we have

$$m \mid \alpha\delta_i,$$
$$\frac{m}{(m,\alpha)} \mid \delta_i.$$

By the conditions of Theorem 3 we have $2^\beta \parallel m$ and $k < 2^\beta$. Thus $(m,\alpha) \leq \alpha \leq k < 2^\beta$. Therefore $2 \mid \frac{m}{(m,\alpha)}$, whence δ_i is even. This completes the proof of Theorem 2 and Theorem 3 (ii).

In order to prove Theorem 3 (i) we need a generalization of Lemma 3. This is the following:

Lemma 4 *Suppose that the conditions of Theorem 3 (i) hold. If* $1 \leq \delta_1, \ldots, \delta_\ell \leq m-1$ *and* $f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell)$ *is a perfect* m*-th power, then there is a permutation* $(\rho_1, \ldots, \rho_\ell)$ *of* $(\delta_1, \ldots, \delta_\ell)$ *such that for all* $1 \leq i \leq \ell$ *there exists an* α_i *with* $1 \leq \alpha_i \leq k^i$ *and*

$$m \mid \alpha_i\rho_i.$$

We postpone the proof of Lemma 4. Now, from this lemma we verify that $\sum_2 \leq \frac{\ell! k^{\ell(\ell+1)}}{m^\ell} p$. Consider a fixed ℓ-tuple $(\delta_1, \ldots, \delta_\ell)$ for which $f^{\delta_1}(n+d_1)\ldots f^{\delta_\ell}(n+d_\ell)$ is a perfect m-th power. We will prove that

$$\prod_{j=1}^{\ell} \left| \sum_{\ell_j}^{m/2} \chi^{\delta_j}(g^{x\ell_j}) \right| \leq \frac{k^{\ell(\ell+1)/2}}{2^\ell}. \tag{4.13}$$

Indeed, by Lemma 4 we have a permutation $(\rho_1, \ldots, \rho_\ell)$ of $(\delta_1, \ldots, \delta_\ell)$ such that for all $1 \leq i \leq \ell$ there exists an α_i with $1 \leq \alpha_i \leq k^i$ and

$$m \mid \alpha_i\rho_i.$$

By this, $0 < \alpha_i \rho_i < \alpha_i m$ and $\alpha_i \le k^i$ we get

$$\begin{aligned} m \le \alpha_i \rho_i &\le (\alpha_i - 1)m, \\ \frac{1}{\alpha_i} \le \frac{\rho_i}{m} &\le 1 - \frac{1}{\alpha_i}, \\ \frac{1}{k^i} \le \frac{1}{\alpha_i} &\le \left|\left|\frac{\rho_i}{m}\right|\right|. \end{aligned}$$

By this, (4.10) and $|1 - e(\alpha)| \ge 4\,||\alpha||$ we have

$$\left|\sum_{\ell_j=1}^{m/2} \chi^{\rho_j}(g^{x\ell_j})\right| \le \frac{2}{|1 - \chi^{\rho_j}(g^x)|} = \frac{2}{|1 - e(\rho_j/m)|} \le \frac{1}{2\,||\rho_j/m||} \le \frac{k^j}{2}. \quad (4.14)$$

Taking the term-wise product in (4.14) for $j = 1, \ldots, \ell$ we obtain (4.13). Thus

$$\sum\nolimits_2 \le p\frac{k^{\ell(\ell+1)/2}}{m^\ell} \sum_{\substack{1 \le \delta_1, \ldots, \delta_\ell \le m, \\ f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell) \text{ is} \\ \text{a perfect } m\text{-th power}}} 1. \quad (4.15)$$

Next we give an upper bound for

$$r \stackrel{\text{def}}{=} \sum_{\substack{1 \le \delta_1, \ldots, \delta_\ell \le m, \\ f^{\delta_1}(n+d_1)\cdots f^{\delta_\ell}(n+d_\ell) \text{ is} \\ \text{a perfect } m\text{-th power}}} 1. \quad (4.16)$$

The number of different permutations $(\rho_1, \ldots, \rho_\ell)$ of $(\delta_1, \ldots, \delta_\ell)$ is $\ell!$. Consider a fixed permutation $(\rho_1, \ldots, \rho_\ell)$. Then by Lemma 4 we have $m \mid \alpha_i \rho_i$ where $1 \le \alpha_i \le k^i$. Thus $\frac{m}{(m,\alpha_i)} \mid \rho_i$. Since $1 \le \rho_i \le m$ we have that ρ_i may assume $(m, \alpha_i) \le \alpha_i \le k^i$ values. Therefore

$$r \le \ell! \prod_{i=1}^{\ell} k^i = \ell! k^{\ell(\ell+1)/2}. \quad (4.17)$$

By (4.15), (4.16) and (4.17) we have

$$\sum\nolimits_2 \le \ell! \frac{k^{\ell(\ell+1)}}{m^\ell} p$$

which proves Theorem 3 (i). It remains to prove Lemma 4.

Proof of Lemma 4

We will need the following definition and lemma:

Definition 2 *Let $\mathcal{A}$ and $\mathcal{B}$ be multi-sets of the elements of $\mathbb{Z}_p$. If $\mathcal{A}+\mathcal{B}$ represents every element of $\mathbb{Z}_p$ with multiplicity divisible by m, i.e., for all $c \in \mathbb{Z}_p$, the number of solutions of*

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

(the a's and b's are counted with their multiplicities) is divisible by m, then the sum $\mathcal{A}+\mathcal{B}$ is said to have property P.

Lemma 5 *Let $\mathcal{A}=\{a_1, a_2, \ldots, a_r\}$, $\mathcal{D}=\{d_1, d_2, \ldots, d_\ell\} \subseteq \mathbb{Z}_p$. If one of the following two conditions holds*
(i) $\min\{r, \ell\} \leq 2$ *and* $\max\{r, \ell\} \leq p-1$,
(ii) $(4\ell)^r \leq p$ *or* $(4r)^\ell \leq p$,
then there exist $c_1, \ldots, c_\ell \in \mathbb{Z}_p$ and a permutation $(q_1, \ldots, q_\ell)$ of $(d_1, \ldots, d_\ell)$ such that for all $1 \leq i \leq \ell$

$$a+d=c_i \qquad a \in \mathcal{A},\ d \in \mathcal{D}$$

has at least one solution, and the number of solutions is less than $i+1$. Moreover for all solution $a \in \mathcal{A}$, $d \in \mathcal{D}$ we have $d \in \{q_1, q_2 \ldots, q_i\}$, and $d=q_i$, $a=c_i-q_i$ is always a solution.

Proof of Lemma 5

We will prove Lemma 5 by induction on i. It was proved in [8, Theorem 2] that for all sets $\mathcal{A}$ and $\mathcal{D}$ with the conditions of Lemma 5, we have a $c \in \mathbb{Z}_p$ such that

$$a+d=c \qquad a \in \mathcal{A},\ d \in \mathcal{D}$$

has exactly one solution.

This proves Lemma 5 in the case $i=1$. Suppose that Lemma 5 holds for $i=j$. Then we will prove that it also holds for $i=j+1$. By the induction hypothesis we have $c_1, \ldots, c_j$ and a permutation $(q_1, \ldots, q_j)$ of $(d_1, \ldots, d_j)$ according to Lemma 5. Let $\mathcal{D}'=\mathcal{D} \setminus \{q_1, \ldots q_j\}$. Since Lemma 5 is true for $i=1$ we have that there exists $c_{j+1} \in \mathbb{Z}_p$ such that

$$a+d=c_{j+1} \qquad a \in \mathcal{A},\ d \in \mathcal{D}'$$

has exactly one solution. Let this unique solution be $\alpha = \alpha_{i+1}$ and $d = q_{j+1}$. Then for the solution of

$$a + d = c_{j+1} \qquad a \in \mathcal{A},\ d \in \mathcal{D}$$

we have $d \in \{q_1, q_2, \ldots, q_{j+1}\}$ which completes the proof of Lemma 5.

Now we return to the proof of Lemma 4. The following equivalence relation was defined in [8] and also used in [10]: We will say that the polynomials $\varphi(x), \psi(x) \in \mathbb{F}_p[x]$ are equivalent, $\varphi \sim \psi$, if there is an $a \in \mathbb{F}_p$ such that $\psi(x) = \varphi(x + a)$. Clearly, this is an equivalence relation.

Write f as the product of irreducible polynomials over $\mathbb{F}_p$. Let us group these factors so that in each group the equivalent irreducible factors are collected. Consider a typical group $\varphi(x+a_1), \ldots, \varphi(x+a_r)$. Then f is of the form $f(x) = \varphi^{\alpha_1}(x + a_1) \ldots \varphi^{\alpha_r}(x + a_r)g(x_r)$ where $g(x)$ has no irreducible factors equivalent with any $\varphi(x + a_i)$ $(1 \leq i \leq r)$.

Let $h(n) = f^{\delta_1}(n + d_1) \cdots f^{\delta_\ell}(n + d_\ell)$ be a perfect m-th power where $1 \leq \delta_1, \ldots, \delta_\ell < m$. Then writing $h(x)$ as the product of irreducible polynomials over $\mathbb{F}_p$, 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 irreducible polynomials cannot have a common zero, each of the zeros of h is of multiplicity divisible by m, 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 multiplicity divisible by m. In other words writing $\mathcal{A} = \{a_1, \ldots, a_1, \ldots, a_r, \ldots, a_r\}$, $\mathcal{D} = \{d_1, \ldots, d_1, \ldots, d_\ell, \ldots, d_\ell\}$ where a_i has the multiplicity α_i in $\mathcal{A}$ (α_i is the exponent of $\varphi(x + a_i)$ in the factorization of $f(x)$) and d_i has the multiplicity δ_i in $\mathcal{D}$ (where $h(n) = f^{\delta_1}(n + d_1) \cdots f^{\delta_\ell}(n + d_\ell)$ is a perfect m-th power), then for each group $\mathcal{A} + \mathcal{D}$ must possess property P.

Let $\mathcal{A}'$ and $\mathcal{D}'$ be the simple set version of $\mathcal{A}$ and $\mathcal{D}$, more exactly, let $\mathcal{A}' = \{a_1, \ldots, a_r\}$ and $\mathcal{D}' = \{d_1, \ldots, d_\ell\}$. $\mathcal{A}'$ and $\mathcal{D}'$ satisfy the conditions of Lemma 5. So by Lemma 5 for the multi-sets $\mathcal{A}$ and $\mathcal{D}$ we have the following: There exist $c_1, \ldots, c_\ell \in \mathbb{Z}_p$ and a permutation $(q_1, \ldots, q_\ell) = (d_{j_1}, \ldots, d_{j_\ell})$ of

$(d_1, \ldots, d_\ell)$ such that if

$$a + d = c_i \qquad a \in \mathcal{A}',\ d \in \mathcal{D}',$$

then we have

$$d \in \{q_1, \ldots, q_i\} = \{d_{j_1}, \ldots, d_{j_i}\}$$

and $d = q_i$, $a = c_i - q_i$ is a solution. Here $(j_1, \ldots, j_\ell)$ is a permutation of $(1, \ldots, \ell)$. Define ρ_i's by $\rho_i = \delta_{j_i}$ (so $(\rho_1, \ldots, \rho_\ell) = (\delta_{j_1}, \ldots, \delta_{j_\ell})$ is the same permutation of $(\delta_1, \ldots, \delta_\ell)$ as the permutation $(q_1, \ldots, q_\ell) = (d_{j_1}, \ldots, d_{j_\ell})$ of $(d_1, \ldots, d_\ell)$). Returning to the multi-set case, using these notation we get that the number of the solutions

$$a + d = c_i \qquad a \in \mathcal{A},\ d \in \mathcal{D}$$

is of the form

$$\epsilon_{i,1}\alpha_{i,1}\rho_1 + \epsilon_{i,2}\alpha_{i,2}\rho_2 + \cdots + \epsilon_{i,i}\alpha_{i,i}\rho_i$$

where $\epsilon_{i,j} \in \{0, 1\}$, $\alpha_{i,j} \in \{\alpha_1, \ldots, \alpha_r\}$ for $1 \leq j \leq i$ and $\epsilon_{i,i} = 1$. (We study the number of the solutions by multiplicity since $\mathcal{A}$ and $\mathcal{D}$ are multi-sets).

Since $\mathcal{A} + \mathcal{D}$ posses property $\mathcal{P}$ we have that for all $1 \leq i \leq \ell$

$$m \mid \epsilon_{i,1}\alpha_{i,1}\rho_1 + \epsilon_{i,2}\alpha_{i,2}\rho_2 + \cdots + \epsilon_{i,i}\alpha_{i,i}\rho_i. \tag{4.18}$$

By induction on i we will prove that

$$m \mid \alpha_{1,1}\alpha_{2,2}, \ldots, \alpha_{i,i}\rho_i. \tag{4.19}$$

Indeed, for $i = 1$ by (4.18) and $\epsilon_{1,1} = 1$ we get $m \mid \alpha_{1,1}\rho_1$. We will prove that if (4.19) holds for $i \leq j - 1$, then it also holds for $i = j$.

By the induction hypothesis we have

$$m \mid \alpha_{1,1}\rho_1,\ m \mid \alpha_{1,1}\alpha_{2,2}\rho_2,\ \ldots,\ m \mid \alpha_{1,1}\alpha_{2,2}\ldots, \alpha_{j-1,j-1}\rho_{j-1}. \tag{4.20}$$

Multiplying (4.18) for $i = j$ by $\alpha_{1,1}\ldots\alpha_{j-1,j-1}$ we get:

$$\begin{aligned} m \mid & \epsilon_{j,1}\alpha_{j,1}\alpha_{1,1}\ldots\alpha_{j-1,j-1}\rho_1 + \epsilon_{j,2}\alpha_{j,2}\alpha_{1,1}\ldots\alpha_{j-1,j-1}\rho_2 + \ldots \\ & + \epsilon_{j,j}\alpha_{j,j}\alpha_{1,1}\ldots\alpha_{j-1,j-1}\rho_i. \end{aligned}$$

From this using (4.20) and $\epsilon_{j,j} = 1$ we get

$$m \mid \alpha_{1,1} \dots \alpha_{j,j} \rho_j$$

which was to be proved.

$\alpha_{1,1}, \dots, \alpha_{i,i} \in \{\alpha_1, \dots, \alpha_r\}$ where α_i's are exponents of irreducible factors of f, thus $1 \leq \alpha_{i,i} \leq \deg f = k$. Therefore $\alpha_{1,1}\alpha_{2,2} \dots \alpha_{i,i} \leq k^i$ and by (4.19) this completes the proof of Lemma 4.

4.2.3 Proof of Theorem 4

The proof is exactly the same as in [1, Theorem 1], the only difference is in the definitions of q and r: now we choose q, r as integers with $(q, p) = (r, p) = 1$ and $1 \leq \text{ind}^* q \leq \frac{m}{2}$, $\frac{m}{2} < \text{ind}^* r \leq m$.

4.3 Time analysis

Construction 1 depends on the key g^x where g is a primitive root and $(x, m) = 1$. We only need g^x, it is not necessary to know the value of g or x. First we prove that it is easy to find a key g^x.

Suppose that the factorization of m is known: $m = p_1^{\alpha_1} \dots p_r^{\alpha_r}$ where $p_1, \dots, p_r$ are primes. The condition $(x, m) = 1$ is equivalent with that $y = g^x$ is not a perfect p_i-th power for any $1 \leq i \leq r$ in $\mathbb{F}_p$. In other words, using Fermat's theorem we have that

$$y^{(p-1)/p_i} \equiv 1 \pmod{p} \tag{4.21}$$

does not hold for all $1 \leq i \leq r$. By using the iterated squaring method to check (4.21), it takes $O\left((\log p)^3\right)$ bit operations (see e.g. in [16]).

We will choose a random $y \in \mathbb{Z}_p$, and by (4.21) we check that $y = g^x$ weather satisfies $(x, m) = 1$ or not. For a fix primitive root g, the number of x's with this property is $\varphi(m)\frac{p-1}{m} \gg \frac{p}{\log \log p}$. Thus after $c \log \log p$ attempts we will find a suitable key g^x with high probability.

By the Pohlig-Hellman [24] algorithm $\text{ind}^* n$ can be computed fast. Here, we present this algorithm: First we determine $\text{ind}^* n$ modulo prime power

divisor q^α of m by $O(\alpha q(\log p)^3)$ bit operations. If we know $\mathrm{ind}^* n$ modulo $p_i^{\alpha_i}$ for all $1 \le \alpha_i \le r$ where $m = p_1^{\alpha_1} \dots p_r^{\alpha_r}$, then using the Chinese Remainder theorem we have determined the value $\mathrm{ind}^* n$ modulo m, which gives $\mathrm{ind}^* n$ because of $1 \le \mathrm{ind}^* n \le m$. Thus to compute $\mathrm{ind}^* n$ we use $O((\log m)^4 + (\log p)^3(\alpha_1 p_1 + \dots + \alpha_r p_r)) \le O((\log m)^4 + (\log p)^3(\alpha_1 + \dots + \alpha_r) \max_{1 \le i \le r} p_i) \le O((\log p)^4 \max_{1 \le i \le r} p_i)$ bit operations.

Let us see the proof of that $\mathrm{ind}^* n$ can be computed modulo prime power divisors q^α of m by $O(\alpha q(\log p)^3)$ bit operations. We will prove this by induction on α. When $\alpha = 0$ the statement is trivial. Suppose that we already know $\mathrm{ind}^* n$ modulo q^i:

$$\mathrm{ind}^* n \equiv s \pmod{q^i}.$$

From this we compute $\mathrm{ind}^* n$ modulo q^{i+1} by $O(q(\log p)^3)$ bit operations if $q^{i+1} \mid m$. In order to prove this statement we will use the following lemma, which is a trivial consequence of the properties of the primitive roots and Fermat's theorem.

Lemma 6 *$q^\alpha \mid m$. Then*

$$\mathrm{ind}^* n \equiv s \pmod{q^\alpha}$$

holds if and only if

$$n/g^{sx} \text{ is a perfect } q^\alpha\text{-th power modulo } p$$

which is equivalent with

$$(n/g^{sx})^{(p-1)/q^\alpha} \equiv 1 \pmod{p}. \tag{4.22}$$

By Lemma 6 we have that n/g^{sx} is a perfect q^i-th power. By Lemma 6, using (4.22), we check that which of the numbers

$$n/g^{sx},\ n/g^{(s+q^i)x},\ n/g^{(s+2q^i)x}, \dots,\ n/g^{s+(q-1)q^i x}$$

is a perfect q^{i+1}-th power. This takes $O(q(\log p)^3)$ bit operations. There is surely one which is a perfect q^{i+1}-th power, because $s,\ s+q^i, \dots,\ s+(q-1)q^i$

run over the residue classes modulo q^{i+1} which are congruent to s modulo q^i. By Lemma 6, n/g^{s+jp^ix} is a perfect p^{i+1}-th power if and only if $\mathrm{ind}^* n \equiv s + jq^ix \pmod{q^{i+1}}$. This completes the proof of the statement.

I would like to thank to Professor András Sárközy for the valuable discussions and to the referee Christian Elsholtz for his careful reading and constructive comments.

Chapter 5

On the correlation of binary sequences

Abstract

C. Mauduit conjectured that $C_2(E_N)C_3(E_N) \gg N^c$ always holds with some constant $1/2 < c \leq 1$. This will be proved for $c = 2/3$, more exactly if for a sequence $E_N \subseteq \{-1.+1\}^N$ we have $C_2(E_N) \ll N^{2/3}$ then $C_3(E_N) \gg N^{1/2}$. Indeed, a more general theorem is proved, involving correlation measures.

2000 AMS Mathematics subject classification number: 11K45.

Key words and phrases: Pseudorandom, correlation measure.

5.1 Introduction

In 1997 Mauduit and Sárközy [21] initiated the systematic study of finite binary sequences $E_N = (e_1, e_2, \ldots, e_N)$ with $e_1, e_2, \ldots, e_N \in \{+1, -1\}$. They proposed to use the following measures of pseudorandomness:

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 $a, b, t \in \mathbb{N}$ with $1 \le a \le a+(t-1)b \le N$, while for $k \in \mathbb{N}, k \ge 2$ the *correlation measure of order k* of E_N is defined as

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

where the maximum is taken over all $M \in \mathbb{N}$ and non-negative integers $d_1 < d_2 < \dots < d_k$ such that $M + d_k \le N$.

Since 1997 about 20 papers have been written on this subject. In the majority of these papers special sequences are constructed and/or tested for pseudorandomness, while in [6], [10], [3] and [22] the measures of pseudorandomness are studied. In particular in [6] Cassaigne, Mauduit and Sárközy compared correlations of different order. They asked the following related question:

Problem 1. For $N \to \infty$, are there sequences E_N such that $C_2(E_N) = O(\sqrt{N})$ and $C_3(E_N) = O(1)$ simultaneously?

Recently, Mauduit [19] asked another closely related question

Problem 2. Is it true that for every $E_N \in \{-1, +1\}^N$ we have

$$C_2(E_N)C_3(E_N) \gg N$$

or at least

$$C_2(E_N)C_3(E_N) \gg N^c \tag{5.1}$$

with some $\frac{1}{2} \le c \le 1$?

In this paper I will settle both Problem 1 and Problem 2 in the weaker form (5.1). The answers will follow from the main result of this paper:

Theorem 1 *If $k, \ell \in \mathbb{N}$, $2k+1 > 2\ell$, $N \in \mathbb{N}$ and $N > 67k^4 + 400$, then for all $E_n \in \{-1, +1\}^N$ we have*

$$\left(11\sqrt{k(2\ell+1)}\, C_{2\ell}\right)^{2k+1} + \left(17\, \frac{2k+1}{2\ell}\right)^{\ell} N^{2k-\ell} C_{2k+1}^2 \ge \frac{1}{23} N^{2k-\ell+1}. \tag{5.2}$$

If follows trivially that

Corollary 1 *If $k, \ell \in \mathbb{N}$, $\log N \ge 2k+1 > 2\ell$, $N \in \mathbb{N}$ and $N > 67k^4 + 400$, $E_n \in \{-1, +1\}^N$ and*

$$C_{2\ell}(E_N) < \frac{1}{40\sqrt{k(2\ell+1)}} N^{1-\ell/(2k+1)}$$

then we have

$$C_{2k+1} > \frac{1}{7}\left(\frac{2\ell}{17(2k+1)}\right)^{\ell/2} N^{1/2}.$$

In particular, for $\ell = 1, 2$ and 3 we obtain:

(i) if

$$C_2(E_N) < \frac{N^{2/3}}{50\sqrt{\log N}},$$

then

$$C_3(E_N), C_5(E_N), \cdots \gg \sqrt{N};$$

(ii) if

$$C_4(E_N) < \frac{N^{3/5}}{64\sqrt{\log N}},$$

then

$$C_5(E_N), C_7(E_N), \cdots \gg \sqrt{N};$$

(iii) if

$$C_6(E_N) < \frac{N^{4/7}}{75\sqrt{\log N}},$$

then

$$C_7(E_N), C_9(E_N), \cdots \gg \sqrt{N};$$

where the implicit constant may depend on the order of the correlation measure.

From the first statement of Corollary 1 (which is an immediate consequence of Theorem 1), follows the parts (i), (ii) and (iii) by using the inequalities $N^{1-\ell/(2k+1)} \geq N^{1-\ell/(2\ell+1)}$ and $\frac{1}{\sqrt{k}} \geq \frac{1}{\sqrt{\log N/2}}$.

Clearly, (i) in the Corollary answers the question in Problem 1. Moreover, since we have

$$C_k(E_N) \geq 1$$

for all $N \geq k$, thus Problem 2 also follows from (i) with $c = 2/3$.

By Theorem 1 for $N > 467$ we have

$$267C_2^3 + NC_3^2 > \frac{1}{700}N^2. \tag{5.3}$$

For a "truly random sequence" $E_N \in \{-1,+1\}^N$ the left hand side of (5.3) is $\ll N^{3/2}+N^2$ which shows that the second term is the best possible apart from the constant factor. On the other hand I do not know whether the exponent 3 in the first term is the best possible. In other words, I have not been able to settle the following problem.

Problem 3. Does there exist a sequence $E_N \in \{-1,+1\}^N$ with $C_2(E_N) = O(N^{2/3})$, $C_3(E_N) = o(N^{1/2})$?

Alon, Kohayakawa, Mauduit, Moreira and V. Rödl proved the following for the correlation measure of even order in [3]:

Theorem 2 *If k and N are natural numbers with even k and $2 \le k \le N$, then*

$$C_k(E_N) > \sqrt{\left[\frac{N}{2(k+1)}\right]}$$

for any $E_N \in \{-1,+1\}^N$.

5.2 Proof of Theorem 1

We may suppose that

$$C_{2k+1}(E_N) \le \sqrt{N} \tag{5.4}$$

otherwise the theorem is trivial. The crucial idea of the proof is the following identity:

Lemma 1 *Let*

$$S_1 \stackrel{def}{=} \sum_{1\le d_1<\cdots<d_{2\ell-1}\le N-(2k+1)} \sum_{\substack{1\le n_1<\cdots<n_{2k+1}\\ \le N-d_{2\ell-1}}} e_{n_1}e_{n_1+d_1}\dots e_{n_1+d_{2\ell-1}}e_{n_2}\dots e_{n_2+d_{2\ell-1}}e_{n_{2k+1}}e_{n_{2k+1}+d_1}\dots e_{n_{2k+1}+d_{2\ell-1}},$$

$$S_2 \stackrel{def}{=} \sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} \sum_{\substack{1\le n_1<\cdots<n_{2\ell}\\ \le N-d_{2k}}} e_{n_1}e_{n_1+d_1}\dots e_{n_1+d_{2k}}e_{n_2}e_{n_2+d_1}\dots e_{n_2+d_{2k}}e_{n_{2\ell}}e_{n_{2\ell}+d_1}\dots e_{n_{2\ell}+d_{2k}},$$

Then

$$S_1 - S_2 = 0 \tag{5.5}$$

We will give an upper bound for $S_1 - S_2$ involving $C_{2\ell}$ and C_{2k+1}. But before this we prove Lemma 1.

Proof of Lemma 1. If a product $e_{n_1}\dots e_{n_{2k+1}+d_{2\ell-1}}$ occurs in S_1, then it also occurs in S_2 and vice-versa, because for all terms $e_{n_1}\dots e_{n_{2k+1}+d_{2\ell-1}}$ in S_1 we have

$$e_{n_1}e_{n_1+d_1}\dots e_{n_1+d_{2\ell-1}}e_{n_2}\dots e_{n_2+d_{2\ell-1}}e_{n_{2k+1}}e_{n_{2k+1}+d_1}\dots e_{n_{2k+1}+d_{2\ell-1}} =$$
$$e_{n_1}e_{n_2}\dots e_{n_{2k+1}}e_{n_1+d_1}e_{n_2+d_1}\dots e_{n_{2k+1}+d_1}\dots e_{n_1+d_{2\ell-1}}e_{n_2+d_{2\ell-1}}\dots e_{n_{2k+1}+d_{2\ell-1}}.$$

Here

$$\begin{aligned} n_{i+1} - n_i &= (n_{i+1}+d_1) - (n_i+d_1) = (n_{i+1}+d_2) - (n_i+d_2) = \dots \\ &= (n_{i+1}+d_{2\ell-1}) - (n_i+d_{2\ell-1}) \end{aligned}$$

for all $1 \le i \le 2k$, which proves that this product also occurs in S_2. Changing the role of S_1 and S_2 we get the inverse statement. Thus indeed $S_1 - S_2 = 0$.

Considering $\sum_{\substack{1\le n_1<\cdots<n_{2k+1}\\ \le N-d_{2\ell-1}}} e_{n_1}\dots e_{n_{2k+1}+d_{2\ell-1}}$ in S_1 we see that this is the sum of all possible products containing $2k+1$ terms from the set $e_1e_{1+d_1}\dots e_{1+d_{2\ell-1}}$, $e_2e_{2+d_1}\dots e_{2+d_{2\ell-1}}, \dots, e_{N-d_{2\ell-1}}e_{N-d_{2\ell-1}+d_1}\dots e_N$. A similar situation holds in the case of S_2. We will use the following lemma.

Lemma 2 *For all $j, M \in \mathbb{N}$, $j \le M$ there is a polynomial $p_{j,M}(x) \in \mathbb{Q}[x]$ with the degree j such that if $x_1, x_2, \dots, x_M \in \{-1, +1\}$ then*

$$p_{j,M}(x_1 + \dots + x_M) = \sum_{1\le i_1<i_2\cdots<i_j\le M} x_{i_1}x_{i_2}\dots x_{i_j}.$$

Denote the coefficients of $p_{j,M}$ by $c_{i,j,M}$:

$$p_{j,M}(x) = c_{j,j,M}x^j + c_{j-1,j,M}x^{j-1} + \cdots + c_{0,j,M}.$$

Then $c_{i,j,M} = 0$ if $i \not\equiv j \pmod 2$, and $(-1)^{(j-i)/2}c_{i,j,M} \geq 0$ if $i \equiv j \pmod 2$. If j is even we also have:

$$c_{0,j,M} = (-1)^{j/2}\binom{M/2}{j/2}.$$

Proof of Lemma 2. We will prove this lemma by induction on j. $p_{1,M}(x) = x$ trivially. Since $x_i^2 = 1$, $p_{2,M}(x) = \frac{1}{2}x^2 - \frac{M}{2}$ because

$$\begin{aligned}\frac{1}{2}(x_1 + \cdots + x_M)^2 - \frac{M}{2} &= \frac{1}{2}\left((x_1 + \cdots + x_M)^2 - x_1^2 - \cdots - x_M^2\right)\\ &= \sum_{1\leq i<j\leq M} x_i x_j.\end{aligned}$$

Thus

$$\begin{aligned}&c_{0,1,M} = 0,\ c_{1,1,M} = 1\\ &c_{0,2,M} = -M/2,\ c_{1,2,M} = 0,\ c_{2,2,M} = 1/2.\end{aligned} \tag{5.6}$$

Suppose that the polynomials $p_{1,M}, p_{2,M}, \ldots, p_{j-1,M}$ exist. From this we will prove that $p_{j,M}$ also exists.

Using again $x_i^2 = 1$ we get:

$$\begin{aligned}\sum_{1\leq i_1<i_2<\cdots<i_j\leq M} x_{i_1}x_{i_2}\ldots x_{i_j} &= \frac{1}{j}\sum_{1\leq i_1<i_2<\cdots<i_{j-1}\leq M} x_{i_1}x_{i_2}\ldots x_{i_{j-1}}(x_1 + \cdots + x_M)\\ &\quad - \frac{M-(j-2)}{j}\sum_{1\leq i_1<i_2<\cdots<i_{j-2}\leq M} x_{i_1}x_{i_2}\ldots x_{i_{j-2}}.\end{aligned}$$

Thus for $j \geq 3$ we have

$$p_{j,M}(x) = \frac{1}{j}\,xp_{j-1,M}(x) - \frac{M-(j-2)}{j}\,p_{j-2,M}(x).$$

From this we obtain that the following holds for the coefficients $c_{i,j,M}$:

$$c_{i,j,M} = \frac{1}{j}\,c_{i-1,j-1,M} - \frac{M-(j-2)}{j}\,c_{i,j-2,M}. \tag{5.7}$$

By induction on j, Lemma 2 follows immediately from this recursion. I leave the details to the reader.

By Lemma 2

$$S_1 - S_2 = 0$$

is equivalent with

$$\sum_{1\le d_1<\cdots<d_{2\ell-1}\le N-(2k+1)} p_{2k+1,N-d_{2\ell-1}}\left(\sum_{n=1}^{N-d_{2\ell-1}} e_n e_{n+d_1}\dots e_{n+d_{2\ell-1}}\right)$$
$$-\sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} p_{2\ell,N-d_{2k}}\left(\sum_{n=1}^{N-d_{2k}} e_n e_{n+d_1}\dots e_{n+d_{2k}}\right)=0.$$

So:

$$\sum_{1\le d_1<\cdots<d_{2\ell-1}\le N-(2k+1)} p_{2k+1,N-d_{2\ell-1}}\left(\sum_{n=1}^{N-d_{2\ell-1}} e_n e_{n+d_1}\dots e_{n+d_{2\ell-1}}\right)$$
$$-\sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell}\left(p_{2\ell,N-d_{2k}}\left(\sum_{n=1}^{N-d_{2k}} e_n e_{n+d_1}\dots e_{n+d_{2k}}\right)-c_{0,2\ell,N-d_{2k}}\right)$$
$$=\sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} c_{0,2\ell,N-d_{2k}}.$$

Using the triangle inequality we get:

$$\sum_{1\le d_1<\cdots<d_{2\ell-1}\le N-(2k+1)}\left|p_{2k+1,N-d_{2\ell-1}}\left(\sum_{n=1}^{N-d_{2\ell-1}} e_n e_{n+d_1}\dots e_{n+d_{2\ell-1}}\right)\right|$$
$$+\sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell}\left|p_{2\ell,N-d_{2k}}\left(\sum_{n=1}^{N-d_{2k}} e_n e_{n+d_1}\dots e_{n+d_{2k}}\right)-c_{0,2\ell,N-d_{2k}}\right|$$
$$\ge\left|\sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} c_{0,2\ell,N-d_{2k}}\right|. \tag{5.8}$$

We will give estimates for both side of (5.8). In order to estimate the right hand side of (5.8), we need upper bounds for the coefficients of the polynomials $p_{j,M}$.

Definition 1 *Let*

$$d_{0,1}=0,\ d_{1,1}=1$$
$$d_{0,2}=1/2,\ d_{1,2}=0,\ d_{2,2}=1/2.$$

If $i < 0$ or $j < i$ let $d_{i,j} = 0$.
For $i > 2$ let

$$d_{i,j} = \frac{1}{j}\left(d_{i-1,j-1} + d_{i,j-2}\right). \tag{5.9}$$

Lemma 3 *If $j \leq M$ then*

$$|c_{i,j,M}| \leq d_{i,j} M^{(j-i)/2}.$$

Proof of Lemma 3. We will prove the lemma by induction on j. For $j = 1, 2$ by (5.6) the assertion is trivial. If the lemma holds for $j \leq k-1$ then it also holds for $j = k$ because of triangle-inequality and (5.7):

$$\begin{aligned}|c_{i,k,M}| &\leq \frac{1}{k}|c_{i-1,k-1,M}| + \frac{M-(k-2)}{k}|c_{i,k-2,M}| \leq \frac{1}{k}|c_{i-1,k-1,M}| + \frac{M}{k}|c_{i,k-2,M}| \\ &\leq \frac{1}{k} d_{i-1,k-1} M^{(k-i)/2} + \frac{M}{k} d_{i,k-2} M^{(k-i-2)/2} = M^{(k-i)/2} d_{i,k}.\end{aligned}$$

Thus Lemma 3 is proved.

Next we give an upper bound for the polynomial $p_{j,M}$.

Lemma 4 *Let $w_j \stackrel{def}{=} d_{0,j} + d_{1,j} + \cdots + d_{j,j}$, $j \leq M$*
(i) If $|x| \leq y$, $v > 0$, $y > \sqrt{\frac{N}{3(v+1)}}$ and $M \leq N$ then

$$|p_{j,M}(x)| \leq (3(v+1))^{j/2} w_j |y|^j.$$

(ii) If j is even $|x| \leq \sqrt{N}$ and $M \leq N$ then

$$|p_{j,M}(x) - c_{0,j,M}| \leq w_j N^{(j-2)/2} x^2.$$

Proof of Lemma 4. (i) By Lemma 3

$$|c_{i,j,M}| \leq d_{i,j} M^{(j-i)/2} \leq d_{i,j} N^{(j-i)/2}. \tag{5.10}$$

Using this and $|x| \leq y$ we obtain:

$$\begin{aligned}p_{j,M}(x) &\leq d_{j,j} y^j + d_{j-1,j} N^{1/2} y^{j-1} + d_{j-2,j} N y^{j-2} + \cdots + d_{0,j} N^{j/2} \\ &= y^j \left(d_{j,j} + d_{j-1,j} \frac{N^{1/2}}{y} + \cdots + d_{0,j} \left(\frac{N^{1/2}}{y}\right)^j \right).\end{aligned}$$

By $y > \sqrt{\frac{N}{3(v+1)}}$ we have

$$p_{j,M}(x) \le y^j \left(d_{j,j} + d_{j-1,j}(3(v+1))^{1/2} + \cdots + d_{0,j}(3(v+1))^{j/2}\right)$$
$$\le (3(v+1))^{j/2}(d_{j,j} + d_{j-1,j} + \cdots + d_{0,j})y^j = (3(v+1))^{j/2} w_j y^j.$$

which proves (i).

(ii) Since j is even, by Lemma 2 we have $c_{1,j,M} = 0$. Using again (5.10) we get

$$|p_{j,M}(x) - c_{0,j,M}| \le d_{j,j}x^j + d_{j-1,j}N^{1/2}x^{j-1} + \cdots + d_{2,j}N^{(j-2)/2}x^2$$
$$= x^2\left(d_{j,j}x^{j-2} + d_{j-1,j}N^{1/2}x^{j-3} + \cdots + d_{2,j}N^{(j-2)/2}\right)$$

Because of $x \le N^{1/2}$ we have

$$|p_{j,M}(x) - c_{0,j,M}| \le w_j N^{(j-2)/2}x^2.$$

This completes the proof of Lemma 4.

Using Lemma 4 we are able to estimate the right hand-side of (5.8). Indeed, by the definition of the correlation measure and Theorem 2 (which was proved in [3]) we have

$$\left|\sum_{n=1}^{N-d_{2\ell-1}} e_n e_{n+d_1} \dots e_{n+d_{2\ell-1}}\right| \le C_{2\ell}(E_N),\ C_{2\ell}(E_N) > \sqrt{\frac{N}{3(2\ell+1)}}.$$

Thus by Lemma 4 (i) we have

$$\left|p_{2k+1,N-d_{2\ell-1}}\left(\sum_{n=1}^{N-d_{2\ell-1}} e_n e_{n+d_1} \dots e_{n+d_{2\ell-1}}\right)\right| \le (3(2\ell+1))^{(2k+1)/2} w_{2k+1} C_{2\ell}^{2k+1}(E_N). \tag{5.11}$$

On the other hand by (5.4) we have

$$\left|\sum_{n=1}^{N-d_{2k}} e_n e_{n+d_1} \dots e_{n+d_{2k}}\right| \le C_{2k+1}(E_N) \le \sqrt{N}.$$

Using Lemma 4 (ii) we get

$$\left|p_{2\ell,N-d_{2k}}\left(\sum_{n=1}^{N-d_{2k}} e_n e_{n+d_1} \dots e_{n+d_{2k}}\right) - c_{0,2\ell,N-d_{2k}}\right| \le w_{2\ell} N^{\ell-1} C_{2k+1}^2(E_N). \tag{5.12}$$

We also have

$$\sum_{1\le d_1<\cdots<d_{2\ell-1}\le N-(2k+1)} 1 = \binom{N-(2k+1)}{2\ell-1} \le \frac{N^{2\ell-1}}{(2\ell-1)!},$$
$$\sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} 1 = \binom{N-2\ell}{2k} \le \frac{N^{2k}}{(2k)!}, \tag{5.13}$$

By (5.8), (5.11), (5.12) and (5.13) we have

$$(3(2\ell+1))^{(2k+1)/2} w_{2k+1} \frac{N^{2\ell-1}}{(2\ell-1)!} C_{2\ell}^{2k+1} + w_{2\ell} \frac{N^{2k+\ell-1}}{(2k)!} C_{2k+1}^2(E_N)$$
$$\ge \left| \sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} c_{0,2\ell,N-d_{2k}} \right|. \tag{5.14}$$

The following lemma gives an upper bound for w_j.

Lemma 5

$$w_j \le \frac{1}{[j/2]!}.$$

Proof of Lemma 5. The lemma is true for $j = 1, 2$. We will prove that if it is true for $j \le k-1$ then it is also true for $j = k$. By the recursion (5.9) we get

$$w_k = \frac{1}{k}(w_{k-1} + w_{k-2})$$

Thus by the inductive hypothesis we have

$$w_k \le \frac{1}{k}\left(\frac{1}{[(k-1)/2]!} + \frac{1}{[(k-2)/2]!}\right) \le \frac{1}{[k/2]!}$$

which completes the proof of Lemma 5.

Using Lemma 5, from (5.14) we get:

$$(3(2\ell+1))^{(2k+1)/2} \frac{N^{2\ell-1}}{k!(2\ell-1)!} C_{2\ell}^{2k+1} + \frac{N^{2k+\ell-1}}{\ell!(2k)!} C_{2k+1}^2(E_N)$$
$$\ge \left| \sum_{1\le d_1<\cdots<d_{2k}\le N-2\ell} c_{0,2\ell,N-d_{2k}} \right| \stackrel{\text{def}}{=} L. \tag{5.15}$$

In order to prove Theorem 1 we need a lower bound for the right hand-side of (5.15). By Lemma 2 we have

$$L = \left| \sum_{1 \le d_1 < \cdots < d_{2k} \le N-2\ell} c_{0,2\ell,N-d_{2k}} \right| = \sum_{1 \le d_1 < \cdots < d_{2k} \le N-2\ell} \binom{(N-d_{2k})/2}{\ell}$$
$$= \sum_{d_{2k}=2k}^{N-2\ell} \left(\sum_{1 \le d_1 < \cdots < d_{2k-1} \le d_{2k}-1} 1 \right) \binom{(N-d_{2k})/2}{\ell}$$
$$= \sum_{d_{2k}=2k}^{N-2\ell} \binom{d_{2k}-1}{2k-1} \binom{(N-d_{2k})/2}{\ell}. \tag{5.16}$$

We will use the following lemma

Lemma 6 *If $a > 2\ell^2$ then*

$$\binom{a}{\ell} \ge \frac{a^\ell}{e\ell!}$$

and

$$\binom{a/2}{\ell} \ge \frac{1}{e2^\ell} \binom{a}{\ell}.$$

Proof of Lemma 6. By $a \ge \ell^2 - 1$ and $1 + x \le e^x$ we get:

$$\binom{a}{\ell} \ge \frac{(a+1-\ell)^\ell}{\ell!} = \frac{a^\ell}{\ell! \left(1 + \frac{\ell-1}{a-(\ell-1)}\right)^\ell} \ge \frac{a^\ell}{\ell! \left(1 + \frac{\ell-1}{(\ell^2-1)-(\ell-1)}\right)^\ell}$$
$$= \frac{a^\ell}{\ell! \left(1 + \frac{1}{\ell}\right)^\ell} \ge \frac{a^\ell}{e\ell!}.$$

On the other hand

$$\binom{a/2}{\ell} / \binom{a}{\ell} = \frac{a(a-2)\dots(a-2(\ell-1))}{2^\ell a(a-1)\dots(a-(\ell-1))}. \tag{5.17}$$

By $a \ge 2\ell^2 \ge \ell^2 + \ell - 2$ for $1 \le i \le \ell - 1$ we have

$$\frac{a-2i}{a-i} = 2 - \frac{a}{a-i} \ge 2 - \frac{a}{a-(\ell-1)} = 1 - \frac{\ell-1}{a-(\ell-1)} \ge 1 - \frac{\ell-1}{\ell^2-1} = \frac{1}{\left(1+\frac{1}{\ell}\right)}. \tag{5.18}$$

By (5.17) and (5.18) we have

$$\binom{a/2}{\ell} / \binom{a}{\ell} \ge \frac{1}{2^\ell \left(1 + \frac{1}{\ell}\right)^\ell} \ge \frac{1}{e2^\ell}$$

which completes the proof of Lemma 6.

Let

$$H \stackrel{\text{def}}{=} \frac{1}{e2^{\ell}} \sum_{d_{2k}=N-2\ell^2+1}^{N-\ell} \binom{d_{2k}-1}{2k-1}\binom{N-d_{2k}}{\ell}. \tag{5.19}$$

By Lemma 6 from (5.16) we obtain

$$\begin{aligned} L &\geq \sum_{d_{2k}=2k}^{N-2\ell^2} \binom{d_{2k}-1}{2k-1}\binom{(N-d_{2k})/2}{\ell} \geq \frac{1}{e2^{\ell}} \sum_{d_{2k}=2k}^{N-2\ell^2} \binom{d_{2k}-1}{2k-1}\binom{N-d_{2k}}{\ell} \\ &\geq \frac{1}{e2^{\ell}} \sum_{d_{2k}=2k}^{N-\ell} \binom{d_{2k}-1}{2k-1}\binom{N-d_{2k}}{\ell} - H. \end{aligned} \tag{5.20}$$

Consider how many ways we can choose from the integers $1, 2, \ldots, N$ exactly $2k+\ell$ pieces. This is trivially $\binom{N}{2k+\ell}$. On the other hand if we fixed the value of the $2k$-th largest integer from these $2k+\ell$ pieces, let it be d_{2k}, then the number of the possibilities is $\binom{d_{2k}-1}{2k-1}\binom{N-d_{2k}}{\ell}$. Therefore

$$\binom{N}{2k+\ell} = \sum_{d_{2k}=2k}^{N-\ell} \binom{d_{2k}-1}{2k-1}\binom{N-d_{2k}}{\ell}. \tag{5.21}$$

By Lemma 6 we have

$$\binom{N}{2k+\ell} \geq \frac{N^{2k+\ell}}{e(2k+\ell)!}. \tag{5.22}$$

By (5.20), (5.21) and (5.22) we have

$$L \geq \frac{N^{2k+\ell}}{e^2 2^{\ell}(2k+\ell)!} - H. \tag{5.23}$$

Lemma 7

$$H = \frac{1}{e2^{\ell}} \sum_{d_{2k}=N-2\ell^2+1}^{N-\ell} \binom{d_{2k}-1}{2k-1}\binom{N-d_{2k}}{\ell} \leq \frac{N^{2k+\frac{2}{3}\ell}}{e2^{\ell}(2k+\ell)!}.$$

Proof of Lemma 7. By the Stirling-formula if $d_{2k} \geq N - 2\ell^2 + 1$ we have:

$$\binom{N-d_{2k}}{\ell} \leq \binom{2\ell^2}{\ell} < \frac{(2\ell^2)^{\ell}}{\ell!} \leq \frac{(2\ell^2)^{\ell}}{\left(\frac{\ell}{e}\right)^{\ell}} \leq (2e\ell)^{\ell}. \tag{5.24}$$

On the other hand

$$\binom{d_{2k}-1}{2k-1} \le \frac{N^{2k-1}}{(2k-1)!} = \frac{1}{(2k+\ell)!}\frac{(2k+\ell)!}{(2k-1)!}N^{2k-1} \le \frac{(2k+\ell)^{\ell+1}}{(2k+\ell)!}N^{2k-1}. \tag{5.25}$$

By $\ell \le k$ and $67k^3 \le N$:

$$\begin{aligned} H &\le \frac{1}{e2^\ell}\left(2\ell^2(2e\ell)^\ell\frac{(2k+\ell)^{\ell+1}}{(2k+\ell)!}N^{2k-1}\right) \le \frac{1}{e2^\ell(2k+\ell)!}\left(\sqrt{6e}k\right)^{2\ell+3}N^{2k-1} \\ &\le \frac{1}{e2^\ell(2k+\ell)!}N^{2k+\frac{2}{3}\ell} \end{aligned}$$

which proves Lemma 7.

By Lemma 7, (5.23) and $N > 67$ we have

$$L \ge \frac{N^{2k+\ell}}{e^2 2^\ell(2k+\ell)!}\left(1-\frac{e}{N^{\ell/3}}\right) \ge \frac{N^{2k+\ell}}{23\cdot 2^\ell(2k+\ell)!}. \tag{5.26}$$

From (5.15) and (5.26) and $2^\ell \le 2^k \le (\sqrt{2})^{2k+1}$ we have

$$\begin{aligned} &(3\sqrt{2}(2\ell+1))^{(2k+1)/2}\frac{(2k+\ell)!}{k!(2\ell-1)!}C_{2\ell}^{2k+1} + 2^\ell\frac{(2k+\ell)!}{\ell!(2k)!}N^{2k-\ell}C_{2k+1}^2(E_N) \\ &\ge \frac{N^{2k-\ell+1}}{23}. \end{aligned}$$

Here,

$$\begin{aligned} \frac{(2k+\ell)!}{k!(2\ell-1)!} &\le \frac{(2k+\ell)^{2k-\ell+1}}{k!} \le \frac{(3k)^{2k}}{\left(\frac{k}{e}\right)^k} \le (9ek)^{(2k+1)/2} \\ \frac{(2k+\ell)!}{\ell!(2k)!} &\le \frac{(2k+\ell)^\ell}{\ell!} \le \frac{(3k)^\ell}{\left(\frac{\ell}{e}\right)^\ell} \le \left(8.16\frac{k}{\ell}\right)^\ell. \end{aligned}$$

Thus

$$\left(11\sqrt{k(2\ell+1)}\,C_{2\ell}\right)^{2k+1} + \left(17\,\frac{2k+1}{2\ell}\right)^\ell N^{2k-\ell}C_{2k+1}^2 \ge \frac{1}{23}N^{2k-\ell+1}, \tag{5.27}$$

which was to be proved.

I would like to thank Professors Julien Cassaigne and András Sárközy for the valuable discussions.

Chapter 6

An inequality between the measures of pseudorandomness

6.1 Introduction

In this paper I will improve on a generalization of an inequality of Mauduit and Sárközy [22]. They introduced the following measures of pseudorandomness in [21]:

For a binary sequence

$$E_N = \{e_1, \ldots, e_N\} \in \{-1, +1\}^N,$$

write

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

and, for $D = (d_1, \ldots, d_k)$ with non-negative integers $0 \leq d_1 < \cdots < d_k$,

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

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

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

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

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

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

In [22] Mauduit and Sárközy proved that for all sequences $E_N \in \{-1, +1\}^N$ we have $W(E_N) \le \sqrt{NC_2(E_N)}$. Later in [11] this inequality was generalized by me to correlation measure of any even order: If $3\ell^2 \le N$ and $E_N \in \{-1, +1\}^N$ then $W(E_N) \le 3\ell N^{1-1/(2\ell)} (C_{2\ell}(E_N))^{1/(2\ell)}$. In the present paper I will improve on the factor 3ℓ showing that this inequality even holds with an absolute constant factor:

Theorem 1 *If* $\varepsilon > 0$, $N \ge 18\ell/\varepsilon^2$, *then for all* $E_N \in \{-1, +1\}^N$ *we have*

$$W(E_N) \le (\sqrt{2} + \varepsilon) N^{1-1/(2\ell)} C_{2\ell}(E_N)^{1/(2\ell)}.$$

Mauduit and Sárközy [22] also proved that their inequality is sharp by using probabilistic arguments. In [11] I presented an explicit construction for which the generalized inequality is sharp apart from a $\sqrt{\ell}$ factor. This construction was based on the notion of index (discrete logarithm): Denote ind n the index of n modulo p, defined as the unique integer with

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

and $1 \le \text{ind } n \le p-1$, where g is a fixed primitive root modulo p. Let $\text{ind}^* n$ be the modulo m residue of ind n:

$$\text{ind}^* n \equiv \text{ind } n \pmod{m} \tag{6.1}$$

with $1 \le \text{ind}^* n \le m$.

Construction 1 *Let* $m \mid p-1$ *and* $\text{ind}^* n$ *be the function defined by* (6.1). *Then let the sequence* $E_{p-1} = \{e_1, \dots, e_{p-1}\}$ *be*

$$e_n = \begin{cases} +1 & \text{if } 1 \le \text{ind}^* f(n) \le \frac{m}{2}, \\ -1 & \text{if } \frac{m}{2} < \text{ind}^* f(n) \le m \text{ or } p \mid f(n), \end{cases} \tag{6.2}$$

where $f(x) \in \mathbb{F}_p[x]$ *is a polynomial with the degree* k.

In Theorem 1 and 3 in [11] I gave estimates for the well-distribution measure and correlation measures of this sequence E_{p-1} if some, not too restrictive conditions hold on the polynomial $f(x)$. Then

$$W(E_{p-1}) \gg \frac{1}{\sqrt{\ell k^{\ell+1}}} p^{1-1/(2\ell)} \left(C_{2\ell}(E_{p-1})\right)^{1/(2\ell)} \tag{6.3}$$

follows from these theorems, where the implied constant factor is absolute.

This inspired me to consider the simplest polynomial $f(x) = x$ in Construction 1, hoping that inequality (6.3) holds with a factor larger than $\frac{1}{\sqrt{\ell}}$. Indeed we will study the following sequence:

Construction 2 *Let $m \mid p-1$ and $\mathrm{ind}^* n$ be the function defined by* (6.1). *Then let the sequence $E_{p-1} = \{e_1, \dots, e_{p-1}\}$ be*

$$e_n = \begin{cases} +1 & \text{if } 1 \le \mathrm{ind}^* n \le \frac{m}{2}, \\ -1 & \text{if } \frac{m}{2} < \mathrm{ind}^* n \le m. \end{cases} \tag{6.4}$$

For this sequence we have:

Theorem 2 *If m is even then the sequence in Construction 2 satisfies*

$$W(E_{p-1}) \le 36 p^{1/2} \log p \log(m+1)$$

while for odd m we have

$$W(E_{p-1}) = \frac{p-1}{m} + O(p^{1/2} \log p \log(m+1)).$$

Indeed, this is Theorem 1 in [11] in the special case when k, the degree of the polynomial is 1.

In case of the correlation measure we will give slightly better upper bound than in Theorem 3 (in the special case $k = 1$) in [11]:

Theorem 3 *If m is even then the sequence in Construction 2 satisfies:*

$$C_\ell(E_{p-1}) \le 9\ell 4^\ell p^{1/2} \log p \, (\log m)^\ell,$$

while for odd m we have

$$C_\ell(E_{p-1}) = \frac{p}{m^\ell} + O(5^\ell p^{1/2} \log p (\log m)^\ell).$$

It follows from Theorems 2 and 3:

Corollary 1 *For every $\varepsilon > 0$ there exist positive constants $p_0(\varepsilon)$ and $c_0(\varepsilon)$ such that if $p > p_0(\varepsilon)$ and m is an odd divisor of $p-1$ with*

$$m < c_0(\varepsilon)\frac{p^{1/(2\ell)}}{\ell\,(\log p)^{1+1/\ell}} \tag{6.5}$$

(so $\frac{p}{m^\ell} \gg 5^\ell p^{1/2}\log p\,(\log m)^\ell$), then

$$W(E_{p-1}) \geq (1-\varepsilon)p^{1-1/(2\ell)}\left(C_{2\ell}(E_{p-1})\right)^{1/(2\ell)}. \tag{6.6}$$

I remark that to make sure that condition (6.5) holds, first we fix an odd integer m, and after this we look for a prime number p with $m \mid p-1$ and (6.5). This is possible by Dirichlet's theorem on primes in arithmetic progressions.

So, indeed Theorem 1 is best possible apart from a constant factor. The interesting feature of this proof is that it is explicit, we give a sequence for which (6.6) holds. In the most cases there is only an existence proof for the sharpness of an inequality between pseudorandom measures.

6.2 Proofs of Theorem 1 and 3

Proof of Theorem 1

It follows from the definition of $W(E_N)$ that there exist $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ with $1 \leq a+b < a+tb \leq N$ such that

$$W(E_N) = \Big|\sum_{\substack{a+b\leq i\leq a+tb\\ i\equiv a+b \pmod b}} e_i\Big|. \tag{6.7}$$

For $0 \leq h < b$ let

$$D_h \stackrel{\text{def}}{=} \Big(\sum_{\substack{a+b\leq i\leq a+tb\\ i\equiv h \pmod b}} e_i\Big)^{2\ell} - 2\ell! \sum_{\substack{a+b\leq i_1<\cdots<i_{2\ell}\leq a+tb\\ h\equiv i_1\equiv\cdots\equiv i_{2\ell} \pmod b}} e_{i_1}\ldots e_{i_{2\ell}}. \tag{6.8}$$

Using the multinomial theorem we get that D_h is a sum of products of the form $c \cdot e_{j_1} \dots e_{j_r}$ where $c \geq 0$. Thus D_h takes his maximum when all e_i's are $+1$ (or all e_i's are -1). So:

$$D_h \leq \Big(\sum_{\substack{a+b \leq i \leq a+tb \\ i \equiv h \pmod b}} 1 \Big)^{2\ell} - 2\ell! \sum_{\substack{a+b \leq i_1 < \dots < i_{2\ell} \leq a+tb \\ h \equiv i_1 \equiv \dots \equiv i_{2\ell} \pmod b}} 1$$
$$\leq t^{2\ell} - (t-1)(t-2)\dots(t-2\ell) \leq t^{2\ell} - (t-2\ell)^{2\ell} \leq 4\ell^2 t^{2\ell-1}.$$

By this, (6.7) and (6.8) we have

$$\begin{aligned}
(W(E_N))^{2\ell} &\leq \sum_{h=0}^{b-1} \Big(\sum_{\substack{a+b \leq i \leq a+tb \\ i \equiv h \pmod b}} e_i \Big)^{2\ell} \\
&= \sum_{h=0}^{b-1} \Big(D_h + 2\ell! \sum_{\substack{a+b \leq i_1 < \dots < i_{2\ell} \leq a+tb \\ h \equiv i_1 \equiv \dots \equiv i_{2\ell} \pmod b}} e_{i_1} \dots e_{i_{2\ell}} \Big) \\
&\leq \sum_{h=0}^{b-1} \Big(4\ell^2 t^{2\ell-1} + 2\ell! \sum_{\substack{a+b \leq i_1 < \dots < i_{2\ell} \leq a+tb \\ h \equiv i_1 \equiv \dots \equiv i_{2\ell} \pmod b}} e_{i_1} \dots e_{i_{2\ell}} \Big) \\
&= 4b\ell^2 t^{2\ell-1} + 2\ell! \sum_{\substack{a+b \leq i_1 < \dots < i_{2\ell} \leq a+tb \\ i_1 \equiv \dots \equiv i_{2\ell} \pmod b}} e_{i_1} \dots e_{i_{2\ell}}.
\end{aligned}$$

From this replacing i_2 by $i_1 + d_1$, i_3 by $i_1 + d_2$ and so on, finally $i_{2\ell}$ by $i_1 + d_{2\ell-1}$ we obtain

$$(W(E_N))^{2\ell} \leq 4b\ell^2 t^{2\ell-1} + 2\ell! \sum_{\substack{1 \leq d_1 < \dots < d_{2\ell-1} \leq (t-1)b \\ d_1 \equiv \dots \equiv d_{2\ell-1} \equiv 0 \pmod b}} \sum_{i_1 = a+b}^{a+tb-d_{2\ell-1}} e_{i_1} e_{i_1+d_1} \dots e_{i_1+d_{2\ell-1}}. \tag{6.9}$$

By the definition of the correlation measure we have

$$\Big| \sum_{i_1 = a+b}^{a+tb-d_{2\ell-1}} e_{i_1} e_{i_1+d_1} \dots e_{i_1+d_{2\ell-1}} \Big| \leq C_{2\ell} E_N. \tag{6.10}$$

By $tb \leq a+tb \leq N$ we have $4b\ell^2 t^{2\ell-1} = 4\ell^2 (tb) t^{2\ell-2} \leq 4\ell^2 N^{2\ell-1}$, and so from

(6.9) and (6.10) we obtain

$$(W(E_N))^{2\ell} \le 4\ell^2 N^{2\ell-1} + 2\ell! \frac{N^{2\ell-1}}{(2\ell-1)!} C_{2\ell}(E_N)$$
$$= 2\ell \left(1 + \frac{2\ell}{C_{2\ell}(E_N)}\right) N^{2\ell-1} C_{2\ell}(E_N).$$

From this by the binomial theorem we get:

$$W(E_N) \le (2\ell)^{1/(2\ell)} \left(1 + \frac{1}{C_{2\ell}(E_N)}\right) N^{1-1/(2\ell)} \left(C_{2\ell}(E_N)\right)^{1/(2\ell)}.$$

Kohayakawa, Mauduit, Moreira and V. Rödl [3] proved that $C_{2\ell}(E_N) > \sqrt{\frac{N}{3(2\ell+1)}}$ holds for all $E_N \in \{-1,+1\}^N$ by this and since $(2\ell)^{1/(2\ell)} \le \sqrt{2}$ we get:

$$W(E_N) \le \sqrt{2} \left(1 + \sqrt{\frac{3(2\ell+1)}{N}}\right) N^{1-1/(2\ell)} \left(C_{2\ell}(E_N)\right)^{1/(2\ell)}.$$

If $N \ge 18\ell/\varepsilon^2 \ge 6(2\ell+1)/\varepsilon^2$ then this completes the proof of the theorem.

Proof of Theorem 3

The proof of the theorem is very similar to the proof of Theorem 1 in [10]. By the formula

$$\frac{1}{m} \sum_{\chi:\chi^m=1} \overline{\chi}^j(a)\chi(b) = \begin{cases} 1 & \text{if } m \mid \text{ind } a - \text{ind } b, \\ 0 & \text{if } m \nmid \text{ind } a - \text{ind } b, \end{cases}$$

we obtain

$$e_n = 2 \sum_{\substack{1 \le i \le m/2 \\ i \equiv \text{ind } n \pmod m}} 1 - 1 = \frac{2}{m} \sum_{1 \le i \le m/2} \sum_{\chi:\chi^m=1} \overline{\chi}(n)\chi(g^i) - 1.$$

Thus

$$e_n = \frac{2}{m} \left(\sum_{1 \le i \le m/2} \sum_{\chi \ne \chi_0 : \chi^m = 1} \overline{\chi}(n)\chi(g^j) + \frac{(-1)^m - 1}{4} \right). \tag{6.11}$$

To prove Theorem 3, consider any $\mathcal{D} = \{d_1, d_2, \ldots, d_\ell\}$ with non-negative integers $d_1 < d_2 < \cdots < d_\ell$ and positive integer M with $M + d_\ell \le p - 1$.

Then arguing as in [26, p. 382] with m in place of $p-1$ from (6.11) we obtain:

$$V(E_N,M,D)=\frac{2^\ell}{m^\ell}\sum_{n=1}^{M}\prod_{j=1}^{\ell}\left(\sum_{1\le i\le m/2}\sum_{\substack{\chi_j\neq\chi_0,\\ \chi_j^m=1}}\overline{\chi_j}(n+d_j)\chi_j(g^i)+\frac{(-1)^m+1}{4}\right)$$

$$=\frac{2^\ell}{m^\ell}\left(\sum_{k=0}^{\ell}\sum_{1\le j_1<\cdots<j_k\le\ell}\left(\frac{(-1)^m+1}{4}\right)^{\ell-k}\sum_{\substack{\chi_{j_1}\neq\chi_0,\\ \chi_{j_1}^m=1}}\cdots\sum_{\substack{\chi_{j_k}\neq\chi_0,\\ \chi_{j_k}^m=1}}\right.$$

$$\left.\sum_{n=1}^{M}\overline{\chi}_{j_1}(n+d_{j_1})\ldots\overline{\chi}_{j_k}(n+d_{j_k})\prod_{t=1}^{k}\left(\sum_{1\le\ell_t\le m/2}\chi_{j_i}(g^{\ell_t})\right)\right). \tag{6.12}$$

Let $S_0=M$, $V_0=\left(\frac{1}{2}\right)^\ell$ and for $1\le k\le\ell$ let

$$S_k=\max_{\substack{\chi_1\neq\chi_0,\ldots,\chi_k\neq\chi_0\\ 1\le j_1<\cdots<j_k\le\ell}}\left|\sum_{n=1}^{M}\overline{\chi}_1(n+d_{j_1})\ldots\overline{\chi}_k(n+d_{j_k})\right| \tag{6.13}$$

and

$$V_k=\sum_{1\le j_1<\cdots<j_k\le\ell}\left(\frac{1}{2}\right)^{\ell-k}\sum_{\substack{\chi_{j_1}\neq\chi_0,\\ \chi_{j_1}^m=1}}\cdots\sum_{\substack{\chi_{j_k}\neq\chi_0,\\ \chi_{j_k}^m=1}}\prod_{t=1}^{k}\left|\sum_{1\le\ell_t\le m/2}\chi_{j_i}(g^{\ell_t})\right|. \tag{6.14}$$

Then by the triangle-inequality, the value of $\frac{(-1)^m+1}{4}$ and (6.12) we obtain that if m is even then

$$|V(E_N,M,D)|\le\frac{2^\ell}{m^\ell}S_\ell V_\ell \tag{6.15}$$

and

$$V(E_N,M,D)=\frac{2^\ell}{m^\ell}S_0V_0+O\left(\frac{2^\ell}{m^\ell}\sum_{k=1}^{\ell}S_kV_k\right) \tag{6.16}$$

Next we give an upper bound for S_k. In order to do this we will use the following lemma:

Lemma 1 *Suppose that p is a prime, χ is a non-principal character modulo p of order z, $f \in \mathbb{F}_p[x]$ has s distinct roots in $\overline{F}_p$, and it is not a constant multiple of a z-th power of a polynomial over $\mathbb{F}_p$. Let y be a real number with $0 < y \leq p$. Then for any $x \in \mathbb{R}$:*

$$\left| \sum_{x < n \leq x+y} \chi(f(n)) \right| < 9sp^{1/2} \log p.$$

Poof of Lemma 1

This is a trivial consequence of Lemma 1 in [2]. Indeed, there this result is deduced from Weil's theorem, see [28].

Now let χ be a modulo p character of order m; for simplicity we will choose χ as the character uniquely defined by $\chi(g) = e\left(\frac{1}{m}\right)$.

Returning to the estimate of S_k, let $\overline{\chi_u} = \chi^{\delta_u}$ for $u = 1, 2, \ldots, \ell$, whence by $\chi_1 \neq \chi_0, \ldots, \chi_\ell \neq \chi_0$, we may take

$$1 \leq \delta_u < m.$$

Thus in (6.13) we have

$$\left| \sum_{n=1}^{M} \overline{\chi}_1(n + d_{j_1}) \ldots \overline{\chi}_k(n + d_{j_k}) \right| = \left| \sum_{n=1}^{M} \chi^{\delta_1}(n + d_{j_1}) \ldots \chi^{\delta_\ell}(n + d_{j_k}) \right|$$
$$= \left| \sum_{n=1}^{M} \chi\left((n + d_{j_1})^{\delta_1} \ldots (n + d_{j_k})^{\delta_k} \right) \right|.$$

Since $(n + d_{j_1})^{\delta_1} \ldots (n + d_{j_k})^{\delta_k}$ is not a perfect m-th power, this sum can be estimated by Lemma 1, whence

$$S_k \leq 9kp^{1/2} \log p. \tag{6.17}$$

By (6.14) we have

$$V_k = \sum_{1 \leq j_1 \cdots \leq j_k \leq \ell} \left(\frac{1}{2}\right)^{\ell - k} \left(\sum_{\substack{\chi \neq \chi_0, \\ \chi^m = 1}} \left| \sum_{j=1}^{[m/2]} \chi^j(g) \right| \right)^k$$

Lemma 2

$$\sum_{\substack{\chi\neq\chi_0,\\ \chi^m=1}} \left| \sum_{j=1}^{[m/2]} \chi^j(g) \right| \leq \sum_{\substack{\chi\neq\chi_0,\\ \chi^m=1}} \frac{2}{|1-\chi(g)|} < 2m\log(m+1).$$

Proof of Lemma 2 This is Lemma 3 in [10] with m in place of d and $m/2$ in place of $(p-1)/2$, and it can be proved in the same way.

Using Lemma 2 we obtain

$$V_k \leq \sum_{1\leq j_1\cdots\leq j_k\leq \ell} \left(\frac{1}{2}\right)^{\ell-k} \left(2m\left(\log(m+1)\right)^k\right) = \frac{4^k}{2^\ell}\binom{\ell}{k} m^k \left(\log(m+1)\right)^k \tag{6.18}$$

By (6.15), (6.16), (6.17) and (6.18) we obtain that if m is even then

$$|V(E_N, M, D)| \leq 9\ell 4^\ell p^{1/2}\log p\left(\log(m+1)\right)^\ell,$$

and if m is odd then

$$\begin{aligned} V(E_N, M, D) &= \frac{M}{m^\ell} + O\left(\frac{9p^{1/2}\log p}{m^\ell}\sum_{k=1}^{\ell} k\binom{\ell}{k} 4^k m^k \left(\log(m+1)\right)^k\right) \\ &= \frac{M}{m^\ell} + O\left(\frac{9\ell p^{1/2}\log p}{m^\ell}\left(4m\log(m+1)\right)^\ell\right) \\ &= \frac{M}{m^\ell} + O\left(5^\ell p^{1/2}\log p\left(\log(m+1)\right)^\ell\right), \end{aligned}$$

which completes the proof of the theorem.

Bibliography

[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] R. Ahlswede, C. Mauduit, A. Sárközy, *Large families of pseudorandom sequences of k symbols and their complexity, Part I, Part II.*, Proceedings on General Theory of Information Transfer and Combinatorics, to appear.

[3] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, *Measures of pseudorandomness for finite sequences: minimum values*, Combinatorics, Probability and Computation, to appear.

[4] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, *Measures of pseudorandomness for finite sequences: typical values*, Combinatorics, Probability and Computation, preprint.

[5] E. R. Berlekamp, *Algebraic Coding Theory*, McGraw Hill, New-York, 1968.

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

[7] S. W. Golomb, Shift Register Sequences, Holden-Day, San Francisco, 1967. Reprinted by Aegean Park Press, 1982.

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

[9] K. Gyarmati, *An inequality between the measures of pseudorandomness*, Annales Univ. Sci. Budapest. Eötvös, to appear.

[10] K. Gyarmati, *On a family of pseudorandom binary sequences*, Periodica Math. Hungar., 49 (2), 2004 pp. 1-19.

[11] K. Gyarmati, *On a fast version of a pseudorandom generator*, General Theory of Information Transfer and Combinatorics, Conference Proceedings, to appear.

[12] K. Gyarmati, *On a pseudorandom property of binary sequences*, The Ramanujan Journal, Vol. 8, No. 3 (2004), 289-302.

[13] K. Gyarmati, *On the correlation of binary sequences*, Studia Sci. Math. Hungar. to appear.

[14] D. R. Heath-Brown, *Zero-Free Regions for Dirichlet L-Functions and the Least Prime in an Arithmetic Progression*, Proc. London Math. Soc. 64, (1992), 265-338.

[15] D. E. Knuth, *The Art of Computer Programming*, Vol. 2, 2nd ed., Addision-Wesley, Reading Mass., 1981.

[16] N. Koblitz, *A course in number theory and cryptography*, Graduate Texts in Mathematics 114, Springer-Verlag, New-York, 1994.

[17] J. L. Massey, *Shift-register synthesis and BCH decoding*, IEEE Transactions on Information Theory, 15 (1969), 122-127.

[18] J. Matoušek, J. Spencer, *Discrepancy in arithmetic progression*, J. Amer.b Math Soc. 9 (1996), 196-204.

[19] C. Mauduit, *Construction of pseudorandom finite sequences*, unpublished lecture notes to the conference, Information Theory and Some Friendly Neighbours- ein Wunschkonzert, Bielefeld, 2003.

[20] C. Mauduit, J. Rivat, A. Sárközy, *Construction of pseudorandom binary sequences using additive characters*, Monatshefte Math., to appear.

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

[22] C. Mauduit, A. Sárközy, *On the measures of pseudorandomness of binary sequences*, Discrete Math. 271 (2003), 195-207.

[23] A. J. Menezes, P. C. Oorschot, S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996.

[24] S. Pohlig, M. Hellman, *An Improved Algorithm for Computing Logarithms over GF(p) and its Cryptographic Significance*, IEEE Trans. Info. Th. 24 (1978), 106-110.

[25] K. F. Roth, *Remark concerning integer sequences*, Acta Arith. 9 (1964), 257-260.

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

[27] W. M. Schmidt, *Equation over Finite Fields, An Elementary Approach*, Lecture Notes in Math. 536, Springer, 1976.

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

Mademoiselle **GYARMATI Katalin**

DOCTORAT de l'UNIVERSITE HENRI POINCARE, NANCY 1

en MATHEMATIQUES

VU, APPROUVÉ ET PERMIS D'IMPRIMER N° 1037

Nancy, le 7 février 2005

Le Président de l'Université

UNIVERSITÉ HENRI POINCARÉ, NANCY 1 — LE PRÉSIDENT

J.P. FINANCE

Université Henri Poincaré, Nancy I
24-30 rue Lionnois - B.P. 3069 - 54013 NANCY Cédex
Tél. : 03 83 68 20 00 - Fax : 03 83 68 21 00