Periodica Mathematica Hungarica Vol. 54 (2), 2007, pp. 163–173
DOI: 10.1007/s-10998-007-2163-9

ON PSEUDORANDOMNESS IN FAMILIES OF SEQUENCES DERIVED FROM THE LEGENDRE SYMBOL

ANDRÁS SÁRKÖZY[1] and C.L. STEWART[2]

[Communicated by Attila Pethő]

[1]Department of Algebra and Number Theory, Eötvös University
H-1117 Budapest, Pázmány Péter sétány 1/c, Hungary
E-mail: sarkozy@cs.elte.hu

[2]Department of Pure Mathematics, University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
E-mail: cstewart@uwaterloo.ca

(*Received February 17, 2006; Accepted October 3, 2006*)

Abstract

We construct a family of finite binary sequences which has a remarkable uniformity with respect to specification of several terms and which also has the property that every sequence in the family has small measures of normality, well distribution in arithmetical progressions and multiple correlations. We also construct a pseudorandom bit generator whose output consists of members of the family.

1. Introduction

Let p be an odd prime and consider the sequence of Legendre symbols $E(p) = ((\frac{1}{p}), (\frac{2}{p}), \dots, (\frac{p-1}{p}))$. One half of the terms of $E(p)$ are 1 and the other half are -1 and, apart from a central symmetry, the distribution of 1's appears to be chaotic when $E(p)$ is calculated for various small primes p. One might ask about the apparently random behaviour of the sequences $E(p)$ and a first step would be to check whether or not given patterns $(\varepsilon_1, \dots, \varepsilon_k)$ with ε_i from $\{-1, 1\}$ occur with the expected frequency in $E(p)$. In 1906 Jacobstahl showed this to be the case when k is 2 or 3. In Davenport's first paper [11] he treated the cases $k = 4$ or 5 and two years later, in 1933 [12], he extended the work to cover all positive integers k less than 10. Let $E_p(\varepsilon_1, \dots, \varepsilon_k)$ denote the number of occurences of $(\varepsilon_1, \dots, \varepsilon_k)$ as

Mathematics subject classification number: 11Y16, 11L40, 94A55, 94A60.

Key words and phrases: pseudorandom sequence, character sums, normality measure, correlations.

0031-5303/2007/$20.00

Akadémiai Kiadó, Budapest
Springer, Dordrecht

consecutive terms of $E(p)$. Further progress was made by Gelfond and Linnik in [13] in 1965, Bach [4] in 1987 and Peralta [25] in 1992 by means of the Weil bounds [29], [30] for exponential sums. In particular, it follows from their work that

$$E_p(\varepsilon_1, \ldots, \varepsilon_k) = \frac{p}{2^k} + O(kp^{1/2}). \tag{1}$$

What other tests of randomness might one apply to the sequences $E(p)$? In 1997, following on earlier work of Knuth [20] on pseudorandomness of finite sequences, Mauduit and Sárközy [23] introduced several measures of randomness for finite sequences. In particular, they introduced measures of normality, well distribution in arithmetical progressions and multiple correlations. Let N be a positive integer and let $E_N = (e_1, \ldots, e_N)$ be a sequence of terms from $\{-1, 1\}$. Let k be a positive integer and let $X = (\varepsilon_1, \ldots, \varepsilon_k)$ be a sequence of terms from $\{-1, 1\}$. Let M be a positive integer and put

$$T(E_N, M, X) = |\{n : 0 \le n < M, (e_{n+1}, \ldots, e_{n+k}) = X\}|,$$

where for any set Y, we denote its cardinality by $|Y|$. The normality measure of order $k, N_k(E_N)$, is defined by

$$N_k(E_N) = \max_{X \varepsilon \{-1,1\}^k} \max_{0<M\le N+1-k} \left|T(E_N, M, X) - M/2^k\right|.$$

Note that (1) gives us information on $N_k(E(p))$.

The well distribution measure of $E_N, W(E_N)$, is defined by

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

where the maximum is taken over all positive integers a, b, t such that $1 \le a < a + (t-1)b \le N$. Further, the correlation measure of order k of $E_N, C_k(E_N)$, is defined by

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

where the maximum is taken over all k-tuples of distinct non-negative integers $D = (d_1, \ldots, d_k)$ and positive integers M for which $M + d_k \le N$.

For each $\varepsilon > 0$ the probability exceeds $1 - \varepsilon$ that a randomly chosen sequence E_N from $\{-1, 1\}^N$ will have $W(E_N)$ between two positive multiples, depending on ε, of $N^{1/2}$ and will have $C_k(E_N)$ between two positive multiples, depending on ε, of $(kN \log N)^{1/2}$ for N sufficiently large, see [9] and [21]. Further, Mauduit and Sárközy [23] proved that the normality measure could be bounded from above by the correlation measures. They proved that for all N, E_N and $k < N$,

$$N_k(E_N) \le \max_{1\le t\le k} |C_t(E_N)|. \tag{2}$$

As a consequence they focussed on the measures W and C_k with the objective of proving upper bounds for them for various sequences E_N which approached in strength those for a typical random sequence and this work initiated much further study, see for example [8], [15] and [27]. Furthermore, they proved that there exist positive numbers c_1 and c_2 such that

$$W(E(p)) < c_1 p^{1/2} \log p \qquad \text{and} \qquad C_k(E(p)) < c_2 k p^{1/2} \log p.$$

In [14], Goubin, Mauduit and Sárközy considered a generalization of the sequence $E(p)$. Let f be a polynomial in $\mathbb{F}_p[X]$ where we identify $\mathbb{F}_p$ with $\mathbb{Z}/p\mathbb{Z}$. Put $E_p(f) = (e_1, \dots, e_p)$ where

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

They proved that if f has degree d (> 0) and f has no multiple zero in an algebraic closure of $\mathbb{F}_p$ then

$$W(E_p(f)) < 10 d p^{1/2} \log p. \tag{3}$$

Further they proved that provided k is 2, $(4d)^k < p$ or 2 is a primitive root modulo p,

$$C_k(E_p(f)) < 10 k d p^{1/2} \log p. \tag{4}$$

In [1], Ahlswede, Khachatrian, Mauduit and Sárközy used an argument based on Lagrange interpolation to show that the family of sequences $E_p(f)$ where f is of degree at most d and has no multiple zero in an algebraic closure of $\mathbb{F}_p$ is quite complex. In particular they showed that it was possible to specify any d of the p terms and find a sequence from the family which satisfies the specification. Our objective in this paper is to introduce families of sequences $E_p(f)$ with a remarkable uniformity with respect to specifications of up to d terms when d is small relative to p. In particular, we shall show that within our families each possible specification is roughly equally likely.

Let us now introduce our families of sequences. Let p be a prime and let ℓ be an integer with $1 \le \ell \le p$. For each subset A of $\{1, \dots, p\}$ we define the polynomial $f_A(x)$ by

$$f_A(x) = \prod_{a \in A} (x - a). \tag{5}$$

Further we define the sequence $E_{p,\ell}(f_A)$ by

$$E_{p,\ell}(f_A) = (e_1, \dots, e_\ell),$$

where

$$e_i = \begin{cases} \left(\frac{f_A(i)}{p}\right) & \text{when } i \notin A, \\ (-1)^j & \text{when } i \in A \text{ and } |\{1, \dots, i\} \cap A| = j. \end{cases} \tag{6}$$

For each integer d with $1 \le d \le p$ and positive integer m with $\ell + m \le p$ we put

$$\mathcal{F}(p, \ell, m, d) = \{E_{p,\ell}(f_A) : |A| = d, A \subset \{1, \dots, m\}\}. \tag{7}$$

Note that since $\ell + m \le p$, the sequences $E_{p,\ell}(f_A)$ are subsequences of consecutive terms of the sequences studied by Goubin, Mauduit and Sárközy, apart from the small modification made at the values of n where $f(n)$ is congruent to 0 modulo p. It can be checked that this change does not affect the estimates (3) and (4). Further the upper bounds for the well distribution and correlation measures continue to hold when one passes to subsequences of consecutive terms. Therefore when $\ell + m \le p$ and $|A| = d$,

$$W(E_{p,\ell}(f_A)) < 10dp^{1/2} \log p, \tag{8}$$

and, provided that k is 2, $(4d)^k < p$ or 2 is a primitive root modulo p,

$$C_k(E_{p,\ell}(f_A)) < 10kdp^{1/2} \log p. \tag{9}$$

DEFINITION. A specification S of size t from $\{1, \dots, \ell\}$ is defined to be a sequence $(\varepsilon_1, \dots, \varepsilon_t)$ of t terms from $\{1, -1\}$ together with an index set $(i_1, \dots, i_t)$ where $1 \le i_1 < i_2 < \cdots < i_t \le \ell$.

For brevity we put $\mathcal{F} = \mathcal{F}(p, \ell, m, d)$. Let $\mathcal{F}(S)$ denote the set of sequences $(e_1, \dots, e_\ell)$ in $\mathcal{F}$ for which $e_{i_j} = \varepsilon_j$ for $j = 1, \dots, t$. Thus $\mathcal{F}(S)$ consists of the elements of $\mathcal{F}$ which satisfy the specification given by S.

THEOREM 1. *Let p be a prime and let ℓ, m and d be positive integers with $\ell + m \le p$,*

$$\min(\ell, m) > 20dp^{1/2} \log p, \tag{10}$$

and $d < p^{1/2}$. Let $\mathcal{F} = \mathcal{F}(p, \ell, m, d)$ be defined as in (7) and let t be a positive integer. Then, for any specification S of size t from $\{1, \dots, \ell\}$,

$$\left| |\mathcal{F}(S)| - \frac{|\mathcal{F}|}{2^t} \right| \le \frac{12tp^{1/2} \log p}{m} |\mathcal{F}|.$$

The uniform nature of our result has implications for the cryptographic security of a pseudorandom number generator which we shall introduce and discuss in §3.

2. Proof of Main Theorem

The key tool we require for the proof of our main theorem is a consequence of the Weil bounds [29], [30]. We shall use it in two different ways. The first application of the Weil bounds allows us to conclude that different subsets A of $\{1,\dots,m\}$ of cardinality d give rise to different sequences $E(A) = E_{p,\ell}(f_A)$ provided that the length ℓ of the sequences satisfies (10). The second application allows us to control an average taken over all members of our family $\mathcal{F}$. Indeed the Weil bounds are also used to obtain the estimates (8) and (9) for the well distribution and correlation measures and so they are fundamental in this context.

LEMMA 1. *Let p be a prime number and let χ be a non-principal Dirichlet character modulo p of order h. Let f be polynomial in $\mathbb{F}_p[x]$ of degree d (≥ 1) with the multiple of at least one of the zeros of f, in the algebraic closure of $\mathbb{F}_p$, coprime with h. Let X and Y be real numbers with $0 < Y \leq p$. Then*

$$\Big|\sum_{X<n\leq X+Y} \chi(f(n))\Big| < 9dp^{1/2}\log p.$$

PROOF. This is a consequence of Weil's Theorem; see, for example, Lemma 2 of [27] and Corollary 1 of [23].

□

PROOF OF THEOREM 1. Our first step is to show that if ℓ satisfies (10), then $E(A_1)$ and $E(A_2)$ are distinct whenever A_1 is different from A_2. Accordingly we shall assume that $E(A_1) = E(A_2)$ and show that this leads to a contradiction. Let $(e_1,\dots,e_\ell)$ be $E(A_1)$. Further, let χ be the character defined by the Legendre symbol modulo p so that

$$\chi(n) = \begin{cases} (\frac{n}{p}) & \text{if } (n,p) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Put $f_0(x) = f_{A_1}(x)f_{A_2}(x)$. Since A_1 is different from A_2, f_0 has at least one root of odd multiplicity. Thus, by Lemma 1,

$$\Big|\sum_{i=1}^{\ell} \chi(f_0(i))\Big| < 9(|A_1| + |A_2|)p^{1/2}\log p. \tag{11}$$

But, by the definition of e_i, recall (6),

$$\ell = \sum_{i=1}^{\ell} e_i^2 = w + \sum_{i=1}^{\ell} \chi(f_0(i)),$$

where w is an integer with $0 \le w \le |A_1 \cup A_2|$. Since $|A_1| = |A_2| = d$ we have

$$\ell < 18dp^{1/2}\log p + 2d < 20dp^{1/2}\log p,$$

which contradicts (10). Therefore distinct subsets of $\{1,\dots,m\}$ give rise to distinct sequences.

Let S be a specification of t terms $(\varepsilon_1,\dots,\varepsilon_t)$ from $\{1,-1\}$ together with an index set $(i_1,\dots,i_t)$ where $1 \le i_1 < i_2 < \cdots < i_t \le \ell$. Then

$$|\mathcal{F}(S)| = \frac{\varepsilon_1\dots\varepsilon_t}{2^t}\sum_{E(A)\in\mathcal{F}}\prod_{j=1}^{t}(e_{i_j}+\varepsilon_j)$$

where $E(A) = (e_1,\dots,e_\ell)$. Thus

$$|\mathcal{F}(S)| = \frac{|\mathcal{F}|}{2^t} + \frac{1}{2^t}\sum_{r=1}^{t}\sum_{1\le j_1<\cdots<j_r\le t}\varepsilon_{j_1}\dots\varepsilon_{j_r}\sum_{E(A)\in\mathcal{F}}\prod_{\substack{1\le s\le t\\ s\notin\{j_1,\dots,j_r\}}}e_{i_s}$$

so

$$\Big||\mathcal{F}(S)| - \frac{|\mathcal{F}|}{2^t}\Big| \le \frac{1}{2^t}\sum_{u=1}^{t}\sum_{1\le v_1<\cdots<v_u\le t}\Big|\sum_{E(A)\in\mathcal{F}}\prod_{z=1}^{u}e_{i_{v_z}}\Big|. \tag{12}$$

Thus it suffices to estimate $\Big|\sum_{E(A)\in\mathcal{F}}\prod_{z=1}^{u}e_{i_{v_z}}\Big|$. Let $S_d(m) = \{A\subset\{1,\dots,m\} : |A| = d\}$. Since different subsets A give rise to different sequences $E(A)$ we see that

$$\begin{aligned}\Big|\sum_{A\in S_d(m)}\prod_{z=1}^{u}\chi(f_A(i_{v_z})) &- \sum_{E(A)\in\mathcal{F}}\prod_{z=1}^{u}e_{i_{v_z}}\Big| \\ &\le \sum_{\substack{A\in S_d(m)\\ p\mid\prod_{z=1}^{u}f_A(i_{v_z})}}\Big|\prod_{z=1}^{u}e_{i_{v_z}}\Big| = \big|\{A\in S_d(m) : p\mid f_A(i_{v_1})\cdots f_A(i_{v_u})\}\big|. \\ &= \Big|\Big\{A\in S_d(m) : p\mid\prod_{a\in A}\prod_{z=1}^{u}(i_{v_z}-a)\Big\}\Big|.\end{aligned} \tag{13}$$

Each A in $S_d(m)$ is a subset of $\{1,\dots,m\}$ and hence of $\{1,\dots,p\}$ and thus $p \mid \prod_{a\in A}\prod_{z=1}^{u}(i_{v_z}-a)$ if and only if there exists y with $1\le y\le u$ for which i_{v_y} is in A. In particular $\Big|\Big\{A\in S_d(m) : p\mid\prod_{a\in A}\prod_{z=1}^{u}(i_{v_z}-a)\Big\}\Big|$ is the number of subsets A of $\{1,\dots,m\}$ with $|A| = d$ which contain at least one member of $\{i_{v_1},\dots,i_{v_u}\}$. This number is at most $u\binom{m-1}{d-1}$. Since $u\le t$ we see, from (13), that

$$\Big|\sum_{A\in S_d(m)}\prod_{z=1}^{u}\chi(f_A(i_{v_z})) - \sum_{E(A)\in\mathcal{F}}\prod_{z=1}^{u}e_{i_{v_z}}\Big| \le \frac{td}{m}\binom{m}{d}. \tag{14}$$

For any integer c from $\{1,\dots,m\}$ let $S_{d-1}(m,c)$ denote the set of all $(d-1)$-element subsets of $\{1,\dots,m\}$ which do not contain c. We have

$$\begin{aligned} d\sum_{A\in S_d(m)}\prod_{z=1}^{u}\chi(f_A(i_{v_z})) &= \sum_{c=1}^{m}\sum_{A_1\in S_{d-1}(m,c)}\chi\Big(\prod_{z=1}^{u}(i_{v_z}-c)f_{A_1}(i_{v_z})\Big)\\ &= \sum_{A\in S_{d-1}(m)}\sum_{\substack{c=1\\ c\notin A}}^{m}\chi\Big(\prod_{z=1}^{u}(i_{v_z}-c)f_A(i_{v_z})\Big)\\ &= \sum_{A\in S_{d-1}(m)}\chi\Big(\prod_{z=1}^{u}f_A(i_{v_z})\Big)\sum_{\substack{c=1\\ c\notin A}}^{m}\chi\Big(\prod_{z=1}^{u}(i_{v_z}-c)\Big). \end{aligned}$$

Put $g(x)=\prod_{z=1}^{u}(i_{v_z}-x)$. Then

$$\Big|d\sum_{A\in S_d(m)}\prod_{z=1}^{u}\chi(f_A(i_{v_z}))\Big| \le \sum_{A\in S_{d-1}(m)}\Big|\sum_{\substack{c=1\\ c\notin A}}^{m}\chi(g(c))\Big|. \tag{15}$$

Plainly,

$$\Big|\sum_{\substack{c=1\\ c\notin A}}^{m}\chi(g(c))\Big| \le \Big|\sum_{c=1}^{m}\chi(g(c))\Big| + d - 1. \tag{16}$$

Since g is a polynomial of degree u with u distinct linear factors in $\mathbb{F}_p[x]$ we deduce from Lemma 1 that

$$\Big|\sum_{c=1}^{m}\chi(g(c))\Big| < 9up^{1/2}\log p \le 9tp^{1/2}\log p. \tag{17}$$

It follows from (15), (16) and (17) that

$$\Big|\sum_{A\in S_d(m)}\prod_{z=1}^{u}\chi(f_A(i_{v_z}))\Big| \le \frac{1}{d}\binom{m}{d-1}10tp^{1/2}\log p.$$

Therefore, from (14),

$$\Big|\sum_{E(A)\in\mathcal{F}}\prod_{z=1}^{u}e_{i_{v_z}}\Big| \le \Big(\frac{td}{m}+\frac{10tp^{1/2}\log p}{(m-d+1)}\Big)\binom{m}{d},$$

and so by (12),

$$\Big||\mathcal{F}(S)|-\frac{|\mathcal{F}|}{2^t}\Big| \le \Big(\frac{td}{m}+\frac{10tp^{1/2}\log p}{(m-d+1)}\Big)\binom{m}{d}.$$

Since $|\mathcal{F}|=\binom{m}{d}$, $d<p^{1/2}$ and $m>20d$ our result follows.

□

3. A pseudorandom bit generator and its cryptographic significance

A pseudorandom bit generator is a deterministic algorithm which, given a random binary sequence of length k, outputs a binary sequence of length ℓ, larger than k, which appears to be random. The input is known as the seed and the output is a pseudorandom sequence, [24]. The importance of such generators comes from the fact that it is often difficult to obtain a truly random seed and it is desirable to be able to stretch random seeds to much longer, apparently random, sequences. There are a number of different pseudorandom number generators which have been proposed which appeal to techniques from number theory, see [22]. They are deemed to be pseudorandom with respect to certain statistical tests. Those which pass all polynomial-size statistical tests are viewed as cryptographically secure. No such pseudorandom bit generators have been determined. However, it can be proved that if problems such as factoring or the discrete logarithm problem are not solvable in polynomial time then cryptographically secure pseudorandom bit generators exist and they have been developed in, for example, [2], [6] and [7]. Our objective in this section is to describe a pseudorandom bit generator whose output we can prove always possesses small well distribution measure, correlation measure and normality measure. In addition, the uniform nature of our main theorem allows us to deduce certain cryptographic aspects of our generator unconditionally.

We now describe the generator. Let p be a prime and let ℓ, m and d be positive integers with $d \le \ell$ and $\ell + m \le p$. Choose a subset A of $\{1, \dots, m\}$ of cardinality d in a random way. A constitutes the random seed. Form the polynomial f_A as in (5). The output is the sequence $E(A) = E_{p,\ell}(f_A)$ as in (6).

Hoffstein and Lieman [18] have proposed such a generator with f_A replaced by f where f is any polynomial in $\mathbb{F}_p[x]$ which is squarefree and neither even nor odd. Further, their generator is of the sort introduced by Anshel and Goldfeld [3] involving sequences of coefficients of certain zeta functions from the Selberg class [28]. Earlier Damgård [10] had proposed subsequences of the sequence of Legendre symbols modulo a prime as a basis for a pseudorandom generator. However, in the above constructions the claim that the output sequences are pseudorandom is supported only by empirical evidence or unproved conjectures.

Let θ be a real number with $\theta < 1/2$. A good choice for d, ℓ and m is to take $\ell = m = \left[\frac{p}{4}\right]$ and

$$d = [p^\theta].$$

For any subset A of $\{1, \dots, \ell\}$ of cardinality d we put $E(A) = E_{p,\ell}(f_A)$ as in (6). Then, since $f_A(x)$ factors into distinct linear factors over $\mathbb{F}_p$, by (8),

$$W(E(A)) < c_3 \ell^{1/2+\theta} \log \ell; \tag{18}$$

here $c_3, c_4, \ldots$ denote positive constants. Further, let k be a positive integer. If p is a prime for which 2 is a primitive root, k is 2 or $(4d)^k$ is less than p then, by (9),

$$C_k(E(A)) < c_4 k \ell^{1/2+\theta} \log \ell. \tag{19}$$

Now, provided that p is sufficiently large that (10) holds, the family $\mathcal{F} = \{E(A) : A$ a subset of $\{1, \ldots, \ell\}$ of cardinality $d\}$ has cardinality $\binom{\ell}{d}$. Furthermore, if any t of the ℓ terms of our output are specified or determined then provided that

$$12t2^{t+4} < p^{1/2}/\log p,$$

by our main theorem there are at least $\binom{\ell}{d}/2^{t+1}$ members of $\mathcal{F}$ with the same specification. Thus the family is still large after specification of a small constant times $\log p$ terms, and so determination of $E(A)$ from the information given by the specified terms is difficult under any attack provided that p and thus $\binom{\ell}{d}$ is sufficiently large.

The general problem of reconstructing a hidden monic polynomial f of degree $d \geq 1$ over a finite field $\mathbb{F}_p$ of p elements given a black box which, for any $x \in \mathbb{F}_p$, evaluates the quadratic character of $f(x)$ has been considered by Russell and Shparlinski [26]. They prove that for any $\varepsilon > 0$ and d with $1 \leq d < p^{1/2}/(\log p)^2$ and an oracle which evaluates the quadratic character of f at any x in $\mathbb{F}_p$ one can find any polynomial f of degree d over $\mathbb{F}_p$ in $O(d^2 p^{d+\varepsilon})$ binary operations. This is the sharpest result known and no further improvement is apparent if one restricts one's attention to polynomials which factor into linear polynomials over $\mathbb{F}_p$. Thus, with the present state of knowledge, there is no way to efficiently reconstruct A from the output sequence $E(A)$.

Recall that if p is a prime for which 2 is a primitive root then estimate (19) holds for the correlation measures. In 1927, Artin conjectured that a positive proportion of all primes have 2 as a primitive root and this was established by Hooley, in 1967 [17], under the assumption of the Generalized Riemann Hypothesis. While no unconditional proof of Artin's conjecture is known, it is easy to test whether 2 is a primitive root for a given prime, see [5], and as a consequence it is easy to find primes of an appropriate size for which 2 is a primitive root. Further, the reason that the subset A is drawn from the first $[\frac{p}{4}]$ integers and that the length ℓ is taken to be $[\frac{p}{4}]$ is to avoid the region of symmetry about $\frac{p}{2}$ in the sequence of Legendre symbols $((\frac{1}{p}), (\frac{2}{p}), \ldots, (\frac{p-1}{p}))$. The study of measures of randomness with respect to symmetry has been undertaken by Gyarmati [16] in this context.

Suppose that we wish to generate a sequence of length $\ell = [\frac{p}{4}]$. Let θ be less than $1/2$ and put $d = [p^\theta]$. (While our estimates for the well distribution measure (18), the correlation measures (19) and, by (2), the normality measures become weaker as θ approaches $1/2$, the cryptographic security presumably increases when θ increases.) In order to determine A, and hence our sequence $E(A)$, we choose d

distinct integers at random from $\{1, \dots, \ell\}$. This can be done with a small multiple of $d \log \ell$ random binary digits with high probability. Thus our seed usually has a small multiple of $d \log \ell$ random digits and the length ℓ of our output $E(A)$ is slightly less than the $(\frac{1}{\theta})$-th power of the length of the seed. Notice also that the size of the family $\mathcal{F}$ of sequences $E(A)$ is $\binom{\ell}{d}$ and therefore is exponential in terms of the length of the seed.

Acknowledgments

The research of András Sárközy was partially supported by the Hungarian National Foundation for Scientific Research, Grant No. T043623. The research of C.L. Stewart was supported in part by Grant A3528 from the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chairs Program.

References

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

[2] W. Alexi, B. Chor, O. Goldreich and O.P. Schnarr, RSA and Rabin functions: Certain parts are as hard as the whole, *SIAM J. Comp.*, **17** (1988), 194–209.

[3] M. Anshel and D. Goldfeld, Zeta functions, one-way functions, and pseudorandom number generators, *Duke Math. J.*, **88** (1997), 371–390.

[4] E. Bach, Realistic analysis of some randomized algorithms, 19*th ACM Symp. on Theory of Comp.*, 1987.

[5] E. Bach and J. Shallit, *Algorithmic Number Theory, Volume I: Efficient Algorithms*, MIT Press, Cambridge, Mass., 1996.

[6] M. Blum and S. Micali, How to generate cryptographically strong sequences of pseudorandom bits, *SIAM J. Comp.*, **13** (1984), 850–864.

[7] L. Blum, M. Blum and M. Shub, A simple unpredictable pseudorandom number generator, *SIAM J. Comp.*, **15** (1986), 364–383.

[8] J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat and A. Sárközy, On finite pseudorandom binary sequences III: The Liouville function. I, *Acta Arith.*, **87** (1999), 367–390.

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

[10] I. Damgård, On the randomness of Legendre and Jacobi sequences, *Lect. Notes in Comp. Sci.*, 403, Springer-Verlag, Berlin, 1990, 163–172.

[11] H. Davenport, On the distribution of quadratic residues (mod p), *J. London Math. Soc.*, **6** (1931), 49–54.

[12] H. Davenport, On the distributions of quadratic residues (mod p), *J. London Math. Soc.*, **8** (1933), 46–52.

[13] A. O. Gelfond and Yu. V. Linnik, *Elementary methods in analytic number theory*, Rand McNally, Chicago, Illinois, 1965.

[14] L. Goubin, C. Mauduit and A. Sárközy, Construction of large families of pseudorandom binary sequences, *J. Number Theory*, **106** (2004), 56–69.
[15] K. Gyarmati, On a family of pseudorandom binary sequences, *Period. Math. Hungar.*, **49** (2004), 45–63.
[16] K. Gyarmati, On a pseudorandom property of binary sequences, *Ramanujan J.*, **8** (2004), 289–302.
[17] C. Hooley, On Artin's conjecture, *J. Reine Angew. Math.*, **225** (1967), 209–220.
[18] J. Hoffstein and D. Lieman, The distribution of the quadratic symbol in function fields and a faster mathematical stream cipher, *Progress in Computer Science and Applied Logic*, 20, Birkhäuser Verlag, Basel, Switzerland, 2001, 59–68.
[19] E. Jacobstahl, Anwendungen einer Formel aus der Theorie der quadratischen Reste, *Dissertation*, Berlin, 1906, 26–32.
[20] D. E. Knuth, *The Art of Computer Programming*, Vol. 2, 2nd ed., Addison-Wesley, Reading, Mass., 1981.
[21] Y. Kohayakawa, C. Mauduit, C. Moreira and V. Rödl, Measures of pseudorandomness for finite sequences: minimum and typical values, *Proceedings of WORDS'03*, TUCS Gen. Publ., 27, Turku Cent. Comput. Sci., Turku, 2003, 159–169.
[22] J. Lagarias, Pseudorandom number generators in cryptography and number theory, *Cryptology and Computational Number Theory* (C. Pomerance, ed.), Proceedings of Symposia in Applied Mathematics, 42, Amer. Math. Soc., 1990, 115–143.
[23] C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol, *Acta Arith.*, **82** (1997), 365–377.
[24] A. Menezes, P. van Oorshot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, 1997.
[25] R. Peralta, On the distribution of quadratic residues and nonresidues modulo a prime number, *Math. Comp.*, **58** (1992), 433–440.
[26] A. Russell and I. Shparlinski, Classical and quantum function reconstruction via character evaluation, *J. Compl.*, **20** (2004), 404–422.
[27] A. Sárközy, A finite pseudorandom binary sequence, *Studia Sci. Math. Hungar.*, **38** (2001), 377–384.
[28] A. Selberg, Old and new conjectures and results about a class of Dirichlet series, *Collected Papers*, Vol. 2, Springer-Verlag, Berlin, 1991, 47–63.
[29] A. Weil, Sur les courbes algébriques et les variétés qui s'en déduisent, *Act. Sci. Ind.*, **1041** (1948), Hermann, Paris.
[30] A. Weil, On some exponential sums, *Proc. Nat. Acad. Sci. U.S.A.*, **34** (1948), 204–207.