Uniform Distribution Theory **6** (2011), no.1, 127–142

uniform distribution theory

ON THE DISTRIBUTION OF THE SUBSET SUM PSEUDORANDOM NUMBER GENERATOR ON ELLIPTIC CURVES

Simon R. Blackburn — Alina Ostafe — Igor E. Shparlinski

ABSTRACT. Given a prime p, an elliptic curve $\mathcal{E}/\mathbb{F}_p$ over the finite field $\mathbb{F}_p$ of p elements and a binary linear recurrence sequence $(u(n))_{n=1}^{\infty}$ of order r, we study the distribution of the sequence of points

$$\sum_{j=0}^{r-1} u(n+j)P_j, \qquad n = 1, \ldots, N,$$

on average over all possible choices of $\mathbb{F}_p$-rational points $P_1, \ldots, P_r$ on $\mathcal{E}$. For a sufficiently large N we improve and generalise a previous result in this direction due to E. El Mahassni.

Communicated by András Sárközy

1. Introduction

The *knapsack generator* or *subset sum generator* is a pseudorandom number generator introduced by Rueppel and Massey [14] and studied in [12]; see also [10, Section 6.3.2] and [13, Section 3.7.9]. It is defined as follows. For an integer $m \geq 1$ we denote by $\mathbb{Z}_m$ the residue ring modulo m. Let $\big(u(n)\big)_{n=1}^{\infty}$ be a linear recurrence sequence of order r over the field of two elements $\mathbb{F}_2$, see [9, Chapter 8]. Given an r-dimensional vector $\mathbf{z} = (z_0, \ldots, z_{r-1}) \in \mathbb{Z}_m^r$ of weights, we generate a sequence of pseudorandom elements of $\mathbb{Z}_m$ by

$$\sum_{j=0}^{r-1} u(n+j)z_j, \qquad n = 1, 2, \ldots \tag{1}$$

2010 Mathematics Subject Classification: Primary 11K45, 11T71; Secondary 11G05, 11T23, 65C05, 94A60.
Keywords: Pseudorandom numbers, subset sum problem, knapsack, exponential sums.

For cryptographic applications, it is usually recommended to use a linear recurrence sequence of maximal period $\tau = 2^r - 1$ and also the modulus $m = 2^r$. Although the results of [5, 8] suggest that this generator should be used with care, no major attack against it is known. In [2, 6] results on the joint uniform distribution of several consecutive elements of this generator have been obtained (on average over all r-dimensional vectors $\mathbf{z} = (z_0, \ldots, z_{r-1}) \in \mathbb{Z}_m^r$).

El Mahassni [4] has recently considered the *elliptic curve subset sum* generator and obtained some uniformity of distribution results for this generator. More precisely, let p be a prime and let $\mathcal{E}$ be an elliptic curve over the finite field $\mathbb{F}_p$ of p elements. Following [4], given a vector $\mathbf{P} = (P_0, \ldots, P_{r-1}) \in \mathcal{E}(\mathbb{F}_p)^r$ of r points from the group $\mathcal{E}(\mathbb{F}_p)$ of $\mathbb{F}_p$-rational points on $\mathcal{E}$ (see [16] for a background on elliptic curves), we define the sequence:

$$V_{\mathbf{P}}(n) = \sum_{j=0}^{r-1} u(n+j)P_j, \qquad n = 1, 2, \ldots, \tag{2}$$

where the summation symbol refers to the group operation on $\mathcal{E}$; see also [5]. If we fix any function $f : \mathcal{E}(\mathbb{F}_p) \to \mathbb{F}_p$, we can define the output of the elliptic curve subset sum generator to be the sequence $(f(V_{\mathbf{P}}(n)))$. One of the simplest and most natural choices for the function f has been considered in [4], namely $f(P) = x(P)$, the x-coordinate of any affine point $P \in \mathcal{E}(\mathbb{F}_p)$. (We can define $x(O) = 0$ for the point at infinity O.) With this choice for the function f, it is known [4] that for almost all choices of $\mathbf{P} = (P_0, \ldots, P_{r-1}) \in \mathcal{E}(\mathbb{F}_p)^r$, the sequence $x\left(V_{\mathbf{P}}(n)\right)/p$, $n = 1, \ldots, N$, is uniformly distributed modulo 1 for a wide range of N.

In this paper we improve the result of [4] on the distribution of the sequence $x\left(V_{\mathbf{P}}(n)\right)/p$, $n = 1, \ldots, N$, in the case when N is sufficiently large, by adding some combinatorial arguments to the existing techniques. We also establish results on the distribution of the s-dimensional vectors

$$\left(\frac{x\big(V_{\mathbf{P}}(n)\big)}{p}, \ldots, \frac{x\big(V_{\mathbf{P}}(n+s-1)\big)}{p}\right), \qquad n = 1, \ldots, N, \tag{3}$$

for any $s \geq 2$. (Note that we always assume that $\mathbb{F}_p$ is represented by the set $\{0, \ldots, p-1\}$, so the vectors (3) belong to the s-dimensional unit cube.) The methods in [4] do not seem to extend to this case. We note that for small values of N the results of [4] remain the only ones known for the elliptic curve subset sum generator. In particular, full analogues of the results of [2] are still not known.

Throughout the paper, the implied constants in symbols 'O' and '$\ll$' may depend on the integer parameter s. We recall that $U \ll V$ and $U = O(V)$ are both equivalent to the inequality $|U| \le cV$ with some constant $c > 0$.

2. Preliminaries

2.1. Discrepancy and exponential sums

For a real z and an integer $m \ge 1$ we use the notation

$$\mathbf{e}(z) = \exp(2\pi i z) \qquad \text{and} \qquad \mathbf{e}_m(z) = \exp(2\pi i z/m).$$

For a sequence of N points

$$\Gamma = (\gamma_{0,n}, \ldots, \gamma_{s-1,n})_{n=1}^{N} \tag{4}$$

in the s-dimensional unit cube, we denote its *discrepancy* by D_Γ. That is,

$$D_\Gamma = \sup_{B \subseteq [0,1)^s} \left| \frac{\mathcal{T}_\Gamma(B)}{N} - |B| \right|,$$

where $\mathcal{T}_\Gamma(B)$ is the number of points of the sequence Γ in the box

$$B = [\alpha_0, \beta_0) \times \ldots \times [\alpha_{s-1}, \beta_{s-1}) \subseteq [0,1)^s$$

of volume $|B|$ and the supremum is taken over all such boxes.

As we have mentioned, one of our basic tools to study the uniformity of distribution is the Koksma–Szüsz inequality, which we present in a slightly weaker form than that given by Theorem 1.21 of [3].

For an integer vector $\mathbf{a} = (a_0, \ldots, a_{s-1}) \in \mathbb{Z}^s$ we define

$$|\mathbf{a}| = \max_{\nu=0,\ldots,s-1} |a_\nu|, \qquad r(\mathbf{a}) = \prod_{\nu=0}^{s-1} \max\{|a_\nu|, 1\}.$$

Lemma 1. *For any integer $L > 1$ and any sequence Γ of N points (4) for the discrepancy D_Γ we have*

$$D_\Gamma \ll \frac{1}{L} + \frac{1}{N} \sum_{0 < |\mathbf{a}| < L} \frac{1}{r(\mathbf{a})} \left| \sum_{n=1}^{N} \mathbf{e}\left(\sum_{\nu=0}^{s-1} a_\nu \gamma_{\nu,n} \right) \right|,$$

where the sum is taken over all integer vectors

$$\mathbf{a} = (a_0, \ldots, a_{s-1}) \in \mathbb{Z}^s \quad \textit{with} \quad 0 < |\mathbf{a}| < L.$$

For estimation of the corresponding exponential sums with various sequences of pseudorandom numbers, the following special case of the bound of Bombieri [1] is used.

Lemma 2. *For any rational function $f(X,Y) \in \mathbb{F}_p(X,Y)$ of degree d which is not constant on an elliptic $\mathcal{E}$ over $\mathbb{F}_p$, the bound*

$$\sum_{P \in \mathcal{E}(\mathbb{F}_p)}^{*} \mathbf{e}_p\left(f(Q)\right) \ll dp^{1/2}$$

holds, where $\sum^*$ means that the poles of $f(X,Y)$ are excluded from the summation.

We need the orthogonality relation:

$$\sum_{\eta=0}^{m-1} \mathbf{e}_m(\eta\lambda) = \begin{cases} 0 & \text{if } \lambda \not\equiv 0 \pmod m, \\ m & \text{if } \lambda \equiv 0 \pmod m. \end{cases} \tag{5}$$

We also make use of the inequality (which is immediate from [7, Bound (8.6)])

$$\sum_{\eta=0}^{m-1} \left| \sum_{\lambda=1}^{M} \mathbf{e}_m\left(\eta\lambda\right) \right| \ll m \log m, \tag{6}$$

which holds for any integers m and M with $1 \le M \le m$.

2.2. Combinatorial estimates

Let r and s be positive integers such that $s \le r$. Write $\mathbf{e}_1, \ldots, \mathbf{e}_s$ for the standard orthogonal basis vectors of length s and let $\mathbf{0}_s = (0, \ldots, 0)$ be the s-dimensional zero vector. We say that a pair of r-dimensional binary vectors $\mathbf{x} = (x_0, \ldots, x_{r-1})$ and $\mathbf{y} = (y_0, \ldots, y_{r-1})$ is *s-good* if for all $h = 1, \ldots, s$, there exists at least one pair (i,j), $0 \le i, j \le r - s$ such that

$$(x_i, x_{i+1}, \ldots, x_{i+s-1}) = \mathbf{e}_h, \qquad (x_j, x_{j+1}, \ldots, x_{j+s-1}) = \mathbf{0}_s$$

and

$$(y_i, y_{i+1}, \ldots, y_{i+s-1}) = \mathbf{0}_s, \qquad (y_j, y_{j+1}, \ldots, y_{j+s-1}) = \mathbf{e}_h.$$

We say that a pair $(\mathbf{x}, \mathbf{y})$ is *s-bad* if it is not s-good. We wish to obtain a bound on the number $f_s(r)$ of s-bad pairs of vectors of length r.

Lemma 3. *Let s be a fixed positive integer. The number $f_s(r)$ of s-bad pairs of binary vectors of length r is at most*

$$f_s(r) \le 2s4^{s-1}\alpha_s^r,$$

where

$$\alpha_s = \left(4^s - 1\right)^{1/s}.$$

P r o o f. We say that a pair $(\mathbf{x}, \mathbf{y})$ is (s,h)-*bad with respect to* $\mathbf{x}$ if there exists no integer i with $0 \le i \le r-s$ such that

$$(x_i, x_{i+1}, \ldots, x_{i+s-1}) = \mathbf{e}_h \quad \text{and} \quad (y_i, y_{i+1}, \ldots, y_{i+s-1}) = \mathbf{0}.$$

Furthermore, we say that a pair $(\mathbf{x}, \mathbf{y})$ is s-*bad with respect to* $\mathbf{x}$ if and only if it is (s,h)-bad with respect to $\mathbf{x}$ for some h. Note that a pair $(\mathbf{x}, \mathbf{y})$ is s-bad if and only if for some h the pair is (s,h)-bad with respect to either $\mathbf{x}$ or $\mathbf{y}$.

Since there are at most s possibilities for h, and the roles of $\mathbf{x}$ and $\mathbf{y}$ in the definition of s-bad pairs are completely symmetrical, our bound follows if we can prove that for any s and h the number of (s,h)-bad pairs with respect to $\mathbf{x}$ is at most $4^{s-1}\alpha_s^r$.

Let h be fixed. We bound the number of (s,h)-bad pairs $(\mathbf{x}, \mathbf{y})$ with respect to $\mathbf{x}$ as follows. For an integer $m = 0, \ldots, \lfloor r/s \rfloor - 1$, there are at most $2^{2s} - 1$ possibilities for the pair

$$\big((x_{ms}, x_{ms+1}, \ldots, x_{ms+s-1}), (y_{ms}, y_{ms+1}, \ldots, y_{ms+s-1})\big)$$

of subsequences, since this pair of subsequences cannot be equal to $(\mathbf{e}_h, \mathbf{0})$. So there are at most $(2^{2s}-1)^{\lfloor r/s \rfloor} \le (2^{2s}-1)^{r/s} = \alpha_s^r$ possibilities for the pair

$$\big((x_0, x_1, \ldots, x_{\lfloor r/s \rfloor s-1}), (y_0, y_1, \ldots, y_{\lfloor r/s \rfloor s-1})\big).$$

But $r - \lfloor r/s \rfloor s \le s-1$, and so there are at most 4^{s-1} possibilities for the last $r - \lfloor r/s \rfloor s$ positions of $\mathbf{x}$ and $\mathbf{y}$. This establishes our bound. □

In particular, since $\alpha_s < 4$, we see from Lemma 3 that $f_s(r) = o(4^r)$ as $r \to \infty$ with s fixed (and so s-bad pairs are asymptotically rare).

We remark that it is not too difficult to see that $f_s(r)$ is bounded below by $c_s\beta_s^r$ for some positive constants $c_s > 0$ and β_s depending only on s. To see this we may use the Perron–Frobenius Theorem, together with the fact that the number of (s,h)-bad pairs with respect to $\mathbf{x}$ is equal to the number of walks of length $r-s$ in a certain directed graph (namely the tensor product of two copies of a span s binary de Bruijn graph, with a single vertex removed). Indeed, for small values of s computer calculations based on this framework show that $f_s(r) \sim c_s\beta_s^r$ where the value of β_s is given in the following table (to 5 decimal places), with the value of α_s given by our upper bound included for comparison:

s	2	3	4	5	6
α_s	3.87298	3.97906	3.99609	3.99922	3.99984
β_s	3.73205	3.93947	3.98444	3.99615	3.99903

The computer calculations show that the pairs that are (s,h)-bad where

$$h = \lfloor (s-1)/2 \rfloor \quad \text{and} \quad h = \lceil (s-1)/2 \rceil$$

provide the dominant term for $f_s(r)$ for $s \le 6$.

3. Main result

3.1. One dimensional distribution

For $\mathbf{P} = (P_0, \ldots, P_{r-1}) \in \mathcal{E}(\mathbb{F}_p)^r$, we denote by $D_{\mathbf{P}}(N)$ the discrepancy of the points

$$\left(\frac{x\big(V_{\mathbf{P}}(n)\big)}{p}\right), \qquad n = 1, \ldots, N.$$

Theorem 4. *Let the linear recurrence sequence $(u(n))_{n=1}^{\infty}$ be purely periodic with period τ and order $r = O(p^{1/2})$ and let its characteristic polynomial be irreducible over $\mathbb{F}_2$. Then for any $\delta > 0$, and for all except $O(\delta p^r)$ choices for $\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r$, for all $1 \le N \le \tau$, we have*

$$D_{\mathbf{P}}(N) \ll \delta^{-1}\left(N^{-1/2} + 3^{r/2}N^{-1}p^{-1/4} + p^{-1/2}\right)(\log \tau)^2 \log p.$$

Proof. From Lemma 1, used with $L = p$, we derive

$$D_{\mathbf{P}}(N) \ll \frac{1}{p} + \frac{1}{N}\sum_{0<|a|<p}\frac{1}{|a|}\left|\sum_{n=1}^{N}\mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\right|.$$

Let

$$N_\mu = \min\{2^\mu, \tau\}, \quad \mu = 0, 1, \ldots$$

Define k by the inequality $N_{k-1} < N \le N_k$, that is, $k = \lceil \log_2 N \rceil$. Then from (5) we derive

$$\sum_{n=1}^{N}\mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big) = \frac{1}{N_k}\sum_{n=1}^{N_k}\sum_{\lambda=1}^{N}\sum_{\eta=0}^{N_k}\mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\mathbf{e}_{N_k}\big(\eta(n-\lambda)\big).$$

Hence,

$$D_{\mathbf{P}}(N) \ll \frac{1}{p} + \frac{1}{NN_k}\Delta_{\mathbf{P}}(k), \tag{7}$$

where

$$\Delta_{\mathbf{P}}(k) = \sum_{0<|a|<p}\frac{1}{|a|}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}(-\eta\lambda)\right|\left|\sum_{n=1}^{N_k}\mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\mathbf{e}_{N_k}(\eta n)\right|.$$

The celebrated Hasse bound shows that

$$\big(\#\mathcal{E}(\mathbb{F}_p)\big)^r \le (p^{1/2}+1)^{2r} = O(p^r) \tag{8}$$

as we have assumed that $r = O(p^{1/2})$.

Applying the Cauchy inequality, we derive

$$\begin{aligned}
&\left(\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\mathbf{e}_{N_k}(\eta n)\right|\right)^2\\
&\le (p^{1/2}+1)^{2r}\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\mathbf{e}_{N_k}(\eta n)\right|^2\\
&\ll p^r\sum_{n,l=1}^{N_k}\mathbf{e}_{N_k}\big(\eta(n-l)\big)\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\mathbf{e}_p\left(a\Big(x\big(V_{\mathbf{P}}(n)\big)-x\big(V_{\mathbf{P}}(l)\big)\Big)\right).
\end{aligned}$$

For the case $n=l$, we estimate the inner sum trivially as

$$\big(\#\mathcal{E}(\mathbb{F}_p)\big)^r = O(p^r).$$

We split the rest of the sum into two sums: the first over distinct 1-bad pairs of vectors

$$\Big(\big(u(n),\dots,u(n+r-1)\big),\big(u(l)\dots,u(l+r-1)\big)\Big), \tag{9}$$

and the second over 1-good pairs of vectors (9).

Let $\mathcal{B}_r$ be the set of pairs of indices (n,l) such that the pair of vectors (9) is 1-bad, that is the set of vectors for which

$$u(n+i)\ge u(l+i) \quad \text{for all} \quad i=0,\dots,r-1,$$

or

$$u(n+i)\le u(l+i) \quad \text{for all} \quad i=0,\dots,r-1.$$

As the vectors are distinct, there exists an index $i=0,\dots,r-1$ such that we have, for example, $u(n+i)>u(l+i)$, which means that $V_{\mathbf{P}}(l)$ does not depend on the point P_i. The Bombieri bound given by Lemma 2 in the case when $f(X,Y)=X$ shows that for any fixed $c_1\in\mathcal{E}(\mathbb{F}_p)$ and $c_2\in\mathbb{F}_p$

$$\sum_{P\in\mathcal{E}(\mathbb{F}_p)}\mathbf{e}_p\big(ax(c_1+P)+c_2\big)\ll 1+\sum_{P\in\mathcal{E}(\mathbb{F}_p),P\ne -c_1}\mathbf{e}_p\big(ax(c_1+P_i)+c_2\big)=O\big(p^{1/2}\big).$$

So we bound our inner sum by summing over the point $P_i\in\mathcal{E}(\mathbb{F}_p)$ to obtain

$$\begin{aligned}
&\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\mathbf{e}_p\left(a\Big(x\big(V_{\mathbf{P}}(n)\big)-x\big(V_{\mathbf{P}}(l)\big)\Big)\right)\\
&=\sum_{\mathbf{P}_i\in\mathcal{E}(\mathbb{F}_p)^{r-1}}\sum_{P_i\in\mathcal{E}(\mathbb{F}_p)}\mathbf{e}_p\left(a\Big(x\big(F_n(\mathbf{P}_i)+P_i\big)-x\big(F_l(\mathbf{P}_i)\big)\Big)\right)\\
&=O\big(p^{r-1/2}\big),
\end{aligned}$$

where $\mathbf{P}_i$ is the vector obtained from $\mathbf{P}$ by removing the point P_i, and $F_m(P_i) \in \mathcal{E}(\mathbb{F}_p)$ denotes a point on $\mathcal{E}$ that depends only on m and $\mathbf{P}_i$.

It remains to consider the case of $n, l \notin \mathcal{B}_r$. In this case, there exist two indices

$$i, j = 0, \ldots, r-1$$

such that we have, for example,

$$u(n+i) > u(l+i) \quad \text{and} \quad u(n+j) < u(l+j).$$

Thus $V_{\mathbf{P}}(l)$ does not depend on the point P_i and $V_{\mathbf{P}}(n)$ does not depend on the point P_j. Using Lemma 2 again, but this time applied for the sums over the points P_i and P_j and (8), the inner sum becomes

$$\begin{aligned}
&\sum_{\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r} \mathbf{e}_p\Big(a\Big(x\big(V_{\mathbf{P}}(n)\big) - x\big(V_{\mathbf{P}}(l)\big)\Big)\Big) \\
&= \sum_{\mathbf{P}_{i,j} \in \mathcal{E}(\mathbb{F}_p)^{r-2}} \sum_{P_i \in \mathcal{E}(\mathbb{F}_p)} \mathbf{e}_p\Big(a\Big(x\big(G_n(\mathbf{P}_{i,j}) + P_i\big)\Big)\Big) \\
&\quad \times \sum_{P_j \in \mathcal{E}(\mathbb{F}_p)} \mathbf{e}_p\Big(\Big(-ax\big(G_l(\mathbf{P}_{i,j}) + P_j\big)\Big)\Big) \\
&= O(p^{r-1}),
\end{aligned}$$

where $\mathbf{P}_{i,j}$ is the vector obtained from $\mathbf{P}$ by removing the points P_i and P_j, and $G_m(\mathbf{P}_{i,j}) \in \mathcal{E}(\mathbb{F}_p)$ denotes a point on $\mathcal{E}$ that depends only on m and $\mathbf{P}_{i,j}$.

Putting everything together, by Lemma 3, we obtain

$$\begin{aligned}
&\left(\sum_{\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r} \left|\sum_{n=1}^{N_k} \mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\mathbf{e}_{N_k}(\eta n)\right|\right)^2 \\
&\ll p^r \left(N_k p^r + p^{r-1/2} \sum_{\substack{n,l=1 \\ n,l \in \mathcal{B}_r}}^{N_k} 1 + p^{r-1} \sum_{\substack{n,l=1 \\ n,l \notin \mathcal{B}_r}}^{N_k} 1\right) \\
&\ll N_k p^{2r} + 3^r p^{2r-1/2} + N_k^2 p^{2r-1},
\end{aligned}$$

and thus

$$\begin{aligned}
&\sum_{\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r} \left|\sum_{n=1}^{N_k} \mathbf{e}_p\Big(ax\big(V_{\mathbf{P}}(n)\big)\Big)\mathbf{e}_{N_k}(\eta n)\right| \\
&\ll N_k^{1/2} p^r + 3^{r/2} p^{r-1/4} + N_k p^{r-1/2}.
\end{aligned}$$

Using (6), we obtain

$$\begin{aligned}\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r} \Delta_{\mathbf{P}}(k) &\ll \left(N_k^{1/2}p^r + 3^{r/2}p^{r-1/4} + N_k p^{r-1/2}\right)\\ &\quad\times \sum_{0<|a|<p} \frac{1}{|a|}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}\left(-\eta\lambda\right)\right|\\ &\ll \left(N_k^{1/2}p^r + 3^{r/2}p^{r-1/4} + N_k p^{r-1/2}\right) N_k \log N_k \log p\\ &\ll \left(N_k^{3/2}p^r + N_k 3^{r/2}p^{r-1/4} + N_k^2 p^{r-1/2}\right)\log\tau\log p.\end{aligned}$$

Thus, for each $k = 1,\dots,\lceil\log\tau\rceil$, the inequality

$$\Delta_{\mathbf{P}}(k) \ge \delta^{-1}\left(N_k^{3/2} + N_k 3^{r/2}p^{-1/4} + N_k^2 p^{-1/2}\right)(\log\tau)^2\log p \tag{10}$$

holds for at most $O(\delta p^r/\log\tau)$ vectors $\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r$. Therefore, the number of vectors $\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r$ for which (10) holds for at least one $k = 1,\dots,\lceil\log\tau\rceil$ is $O(\delta p^r)$. For all the other points $\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r$, by (7) and taking into account that

$$N_k = 2N_{k-1} \le 2N,$$

we get

$$\begin{aligned}D_{\mathbf{P}}(N) &\ll \delta^{-1}N^{-1}\left(N_k^{1/2} + 3^{r/2}p^{-1/4} + N_k p^{-1/2}\right)(\log\tau)^2\log p\\ &\ll \delta^{-1}\left(N^{-1/2} + 3^{r/2}N^{-1}p^{-1/4} + p^{-1/2}\right)(\log\tau)^2\log p,\end{aligned}$$

which concludes the proof. □

We note that El Mahassni [4] obtained the bound

$$D_{\mathbf{P}}(N) \ll \delta^{-1}\left(N^{-1/2} + p^{-1/4}\right)(\log\tau)^2\log p \tag{11}$$

under the same conditions. Say, in the most interesting case of sequences of maximal period $\tau = 2^r - 1$ and r chosen so that $2^r \ll p \ll 2^r$, Theorem 4 gives a stronger result for

$$\tau \ge N \ge \tau^{0.5\log 3/\log 2} = \tau^{0.79248\dots}.$$

3.2. Multidimensional distribution

For $\mathbf{P} = (P_0, \ldots, P_{r-1}) \in \mathcal{E}(\mathbb{F}_p)^r$, we denote by $D_{\mathbf{P},s}(N)$ the s-dimensional discrepancy of the points (3).

Theorem 5. *Let the linear recurrence sequence $\big(u(n)\big)_{n=1}^{\infty}$ be purely periodic with period τ and order $r = O(p^{1/2})$ and let its characteristic polynomial be irreducible over $\mathbb{F}_2$. Then for any $\delta > 0$, and for all except $O(\delta p^r)$ choices for*

$$\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r, \quad \text{for all } \ 1 \le N \le \tau,$$

we have

$$D_{\mathbf{P},s}(N) \ll \delta^{-1} \left(N^{-1/2} \log p + p^{-1/2} \log p + \alpha_s^{r/2} N^{-1} (\log p)^s\right) (\log \tau)^2,$$

where the implied constant depends only on s and α_s is as in Lemma 3.

P r o o f. Exactly as in the proof of Theorem 4, by Lemma 1 we get

$$D_{\mathbf{P},s}(N) \ \ll \ \frac{1}{p} + \frac{1}{NN_k} \Delta_{\mathbf{P},s}(k), \tag{12}$$

where

$$\Delta_{\mathbf{P},s}(k) \ = \sum_{\substack{0<|\mathbf{a}|<p \\ \mathbf{a}=(a_1,\ldots,a_s)\in\mathbb{Z}^s}} \frac{1}{r(\mathbf{a})} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}(-\eta\lambda) \right| \times \left| \sum_{n=1}^{N_k} \mathbf{e}_p\Big(\sum_{\nu=0}^{s-1} a_\nu x\big(V_{\mathbf{P}}(n+\nu)\big)\Big) \mathbf{e}_{N_k}(\eta n) \right|.$$

We further split the sum $\Delta_{\mathbf{P},s}(k)$ into two parts

$$\Delta_{\mathbf{P},s,1}(k) \quad \text{and} \quad \Delta_{\mathbf{P},s,2}(k),$$

where the summation in $\Delta_{\mathbf{P},s,1}(k)$ is taken over the vectors $\mathbf{a} = (a_1, \ldots, a_s) \in \mathbb{Z}^s$ with only one non-zero component and $\Delta_{\mathbf{P},s,2}(k)$ includes all other terms. Thus

$$\Delta_{\mathbf{P},s}(k) = \Delta_{\mathbf{P},s,1}(k) + \Delta_{\mathbf{P},s,2}(k). \tag{13}$$

As in the proof of Theorem 4 we obtain

$$\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r} \Delta_{\mathbf{P},s,1}(k) \ll \left(N_k^{3/2} p^r + N_k 3^{r/2} p^{r-1/4} + N_k^2 p^{r-1/2}\right) \log \tau \log p. \tag{14}$$

Now, let $\mathcal{A}_s$ be the set of the vectors $\mathbf{a} = (a_1, \ldots, a_s) \in \mathbb{Z}^s$ with $0 < |\mathbf{a}| < p$ and with at least two nonzero components. For $\mathbf{a} \in \mathcal{A}_s$, applying the Cauchy inequality and Hasse bound (8), we derive

$$\begin{aligned}
&\left(\sum_{\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_p \left(\sum_{\nu=0}^{s-1} a_\nu x\big(V_{\mathbf{P}}(n+\nu)\big) \right) \mathbf{e}_{N_k}(\eta n) \right| \right)^2 \\
&\le \big(p^{1/2}+1\big)^{2r} \sum_{\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_p \left(\sum_{\nu=0}^{s-1} a_\nu x\big(V_{\mathbf{P}}(n+\nu)\big) \right) \mathbf{e}_{N_k}(\eta n) \right|^2 \\
&\ll p^r \sum_{n,l=1}^{N_k} \mathbf{e}_{N_k}\big(\eta(n-l)\big) \sum_{\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r} \mathbf{e}_p \left(\sum_{\nu=0}^{s-1} a_\nu \Big(x\big(V_{\mathbf{P}}(n+\nu)\big) - x\big(V_{\mathbf{P}}(l+\nu)\big) \Big) \right).
\end{aligned}$$

Now, as in Theorem 4, we split the sum into two sums, one over pairs (n, l) such that the pair of vectors (9) is s-bad and one over s-good pairs.

Let $\mathcal{B}_{r,s}$ be the set of pairs of indices (n, l) such that the pair of vectors (9) is s-bad. For $(n, l) \in \mathcal{B}_{r,s}$, as in the proof of Theorem 4, we estimate the inner sum over $\mathbf{P}$ trivially as $O(p^r)$.

It remains to consider the case of $(n, l) \notin \mathcal{B}_{r,s}$. Since $\mathbf{a} \in \mathcal{A}_s$ there exist at least two distinct indices $i, j = 0, \ldots, s-1$ such that $a_i, a_j \ne 0$. Since $(n, l) \notin \mathcal{B}_{r,s}$, there exist two indices $i_1, i_2 = 0, \ldots, r-s$ such that

$$\begin{aligned}
\big(u(n+i_1), \ldots, u(n+i_1+s-1)\big) &= \mathbf{e}_i, \\
\big(u(l+i_1), \ldots, u(l+i_1+s-1)\big) &= \mathbf{0}_s,
\end{aligned}$$

and

$$\begin{aligned}
\big(u(n+i_2), \ldots, u(n+i_2+s-1)\big) &= \mathbf{0}_s, \\
\big(u(l+i_2), \ldots, u(l+i_2+s-1)\big) &= \mathbf{e}_i.
\end{aligned}$$

Similarly, there exist two indices $j_1, j_2 = 0, 1, \ldots, r-s$ such that

$$\begin{aligned}
\big(u(n+j_1), \ldots, u(n+j_1+s-1)\big) &= \mathbf{e}_j, \\
\big(u(l+j_1), \ldots, u(l+j_1+s-1)\big) &= \mathbf{0}_s,
\end{aligned}$$

and

$$\begin{aligned}
\big(u(n+j_2), \ldots, u(n+j_2+s-1)\big) &= \mathbf{0}_s, \\
\big(u(l+j_2), \ldots, u(l+j_2+s-1)\big) &= \mathbf{e}_j.
\end{aligned}$$

When $\nu \in \{1, 2, \ldots, s\} \setminus \{i, j\}$, the above equations show that

$$u(n+\nu+i_1) = u(n+\nu+i_2) = u(n+\nu+j_1) = u(n+\nu+j_2) = 0,$$

and so $V_{\mathbf{P}}(n+\nu)$ does not depend on any of $P_{i_1}, P_{i_2}, P_{j_1}, P_{j_2}$. Similarly, $V_{\mathbf{P}}(l+\nu)$ does not depend on any of $P_{i_1}, P_{i_2}, P_{j_1}, P_{j_2}$.

When $\nu = i$ (so $\nu \neq j$), the equations above show that

$$u(n+\nu+i_1) = 1 \quad \text{and} \quad u(n+\nu+i_2) = u(n+\nu+j_1) = u(n+\nu+j_2) = 0,$$

so $V_{\mathbf{P}}(n+\nu) = V_{\mathbf{P}}(n+i)$ depends on P_{i_1}, but does not depend on any of $P_{i_2}, P_{j_1}, P_{j_2}$. Similarly $V_{\mathbf{P}}(l+i)$ depends on P_{i_2}, but none of $P_{i_1}, P_{j_1}, P_{j_2}$; $V_{\mathbf{P}}(n+j)$ depends on P_{j_1}, but none of $P_{i_1}, P_{i_2}, P_{j_2}$; and $V_{\mathbf{P}}(l+j)$ depends on P_{j_2}, but none of $P_{i_1}, P_{i_2}, P_{j_1}$.

Let $\mathbf{P}_{i_1,i_2,j_1,j_2}$ be the vector obtained from $\mathbf{P}$ after discarding the points P_{i_1}, P_{i_2}, P_{j_1} and P_{j_2}. We can apply Lemma 2 to our inner sum as in the one dimensional case, but this time applied for four sums over the points P_{i_1}, P_{i_2}, P_{j_1} and P_{j_2} to obtain

$$\begin{aligned}
\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r} \mathbf{e}_p\left(\sum_{\nu=0}^{s-1} a_\nu\Big(x\big(V_{\mathbf{P}}(n+\nu)\big) - x\big(V_{\mathbf{P}}(l+\nu)\big)\Big)\right)
&= \sum_{\mathbf{P}_{i_1,i_2,j_1,j_2}\in\mathcal{E}(\mathbb{F}_p)^{r-4}} \mathbf{e}_p\big(\Psi_{n,l}(\mathbf{P}_{i_1,i_2,j_1,j_2})\big) \\
&\quad \sum_{P_{i_1}\in\mathcal{E}(\mathbb{F}_p)} \mathbf{e}_p\left(a_i\Big(x\big(G_n(\mathbf{P}_{i_1,i_2,j_1,j_2}) + P_{i_1}\big)\Big)\right) \\
&\quad \sum_{P_{j_1}\in\mathcal{E}(\mathbb{F}_p)} \mathbf{e}_p\left(a_j\Big(x\big(G_n(\mathbf{P}_{i_1,i_2,j_1,j_2}) + P_{j_1}\big)\Big)\right) \\
&\quad \sum_{P_{i_2}\in\mathcal{E}(\mathbb{F}_p)} \mathbf{e}_p\Big(-a_i x\big(G_l(\mathbf{P}_{i_1,i_2,j_1,j_2}) + P_{i_2} br\big)\Big) \\
&\quad \sum_{P_{j_2}\in\mathcal{E}(\mathbb{F}_p)} \mathbf{e}_p\Big(-a_j x\big(G_l(\mathbf{P}_{i_1,i_2,j_1,j_2}) + P_{j_2}\big)\Big) \\
&\quad = O\left(p^{r-4}\left(p^{1/2}\right)^4\right) \\
&\quad = O\left(p^{r-2}\right),
\end{aligned}$$

where

$$\Psi_{n,l}(\mathbf{P}_{i_1,i_2,j_1,j_2}) = \sum_{\substack{\nu=0\\ \nu\neq i,j}}^{s-1} a_\nu\Big(x\big(V_{\mathbf{P}}(n+\nu)\big) - x\big(V_{\mathbf{P}}(l+\nu)\big)\Big)$$

depends only on n, l and $\mathbf{P}_{i_1,i_2,j_1,j_2}$ and $G_m(\mathbf{P}_{i,j}) \in \mathcal{E}(\mathbb{F}_p)$ denotes a point on $\mathcal{E}(\mathbb{F}_p)$ that depends only on m and $\mathbf{P}_{i,j}$. (Note that we are using the fact that a_i and a_j are non-zero at this point.)

Putting everything together, by Lemma 3, we obtain

$$\left(\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_p\left(\sum_{\nu=0}^{s-1}a_\nu x\big(V_{\mathbf{P}}(n+\nu)\big)\right)\mathbf{e}_{N_k}\left(\eta n\right)\right|\right)^2 \ll p^r\left(N_k^2p^{r-2}+\alpha_s^rp^r\right)$$
$$\ll N_k^2p^{2r-2}+\alpha_s^rp^{2r},$$

and thus

$$\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_p\left(\sum_{\nu=0}^{s-1}a_\nu x(V_{\mathbf{P}}(n+\nu))\right)\mathbf{e}_{N_k}\left(\eta n\right)\right| \ll N_kp^{r-1}+\alpha_s^{r/2}p^r.$$

Now using (6), we obtain

$$\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\Delta_{\mathbf{P},s,2}(k) \ll \left(N_kp^{r-1}+\alpha_s^{r/2}p^r\right)\sum_{\substack{0<|\mathbf{a}|<p\\ \mathbf{a}=(a_1,\ldots,a_s)\in\mathbb{Z}^s}}\frac{1}{r(\mathbf{a})}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}\left(-\eta\lambda\right)\right|$$
$$\ll \left(N_kp^{r-1}+\alpha_s^{r/2}p^r\right)N_k\log N_k(\log p)^s$$
$$\ll \left(N_k^2p^{r-1}+\alpha_s^{r/2}N_kp^r\right)\log\tau(\log p)^s.$$

Thus, from (13) and (14), we obtain the inequality

$$\sum_{\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r}\Delta_{\mathbf{P},s,}(k) \ll \left(N_k^{3/2}p^r+N_k3^{r/2}p^{r-1/4}+N_k^2p^{r-1/2}\right)\log\tau\log p$$
$$+\ \left(N_k^2p^{r-1}+\alpha_s^{r/2}N_kp^r\right)\log\tau(\log p)^s$$
$$\ll \left(N_k^{3/2}p^r+N_k^2p^{r-1/2}\right)\log\tau\log p$$
$$+\ \ \alpha_s^{r/2}N_kp^r\log\tau(\log p)^s.$$

Hence, for each $k=1,\ldots,\lceil\log\tau\rceil$, we see that the inequality

$$\Delta_{\mathbf{P},s}(k) \geq \delta^{-1}\left(N_k^{3/2}\log p+N_k^2p^{-1/2}\log p+\alpha_s^{r/2}N_k(\log p)^s\right)(\log\tau)^2 \qquad (15)$$

holds for at most $O(\delta p^r/\log\tau)$ vectors $\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r$. Therefore, the number of vectors $\mathbf{P}\in\mathcal{E}(\mathbb{F}_p)^r$ for which (15) holds for at least one $k=1,\ldots,\lceil\log\tau\rceil$ is $O(\delta p^r)$.

For all the other points $\mathbf{P} \in \mathcal{E}(\mathbb{F}_p)^r$, by (12) and taking into account that

$$N_k = 2N_{k-1} \le 2N,$$

as in the proof of Theorem 4 we get

$$D_{\mathbf{P},s}(N) \ll \delta^{-1} \left(N^{-1/2} \log p + p^{-1/2} \log p + \alpha_s^{r/2} N^{-1} (\log p)^s \right) (\log \tau)^2,$$

which concludes the proof. □

Again, in the most interesting case of sequences of maximal period $\tau = 2^r - 1$ and r chosen so that $2^r \ll p \ll 2^r$, Theorem 5 is nontrivial for

$$\tau \ge N \ge \tau^{\gamma_s + \varepsilon},$$

for any fixed $\varepsilon > 0$ and a sufficiently large p, where

$$\gamma_s = \frac{\log \alpha_s}{2 \log 2} < 1.$$

4. Comments

We remark that the proofs of Theorems 4 and 5 depend only on the fact that the binary vectors

$$(u(n+1), \ldots, u(n+r)), \quad n = 1, \ldots, \tau,$$

are pairwise distinct. Thus the same results hold for many other sequences $(u(n))_{n=1}^{\infty}$, for example, for sequences generated by non-linear recurrence relations. In fact, in this generality, these results are new even in the case of the classical subset sum generator (1) over a residue ring (as the proof in [2] applies only to linear recurrence sequences). Our method also applies to bounds of multiplicative character sums with the sequence (1) on average over the vectors

$$\mathbf{z} = (z_0, \ldots, z_{r-1}) \in \mathbb{Z}_m^r.$$

On the other hand, it is still an open problem to obtain nontrivial results about the multidimensional distribution of the elliptic curve subset sum generator (2) on short segments. Note that the bound (11) is nontrivial starting from the values of N of order $(\log \tau)^4 (\log p)^2$ (which can be further reduced by using the approach of [15]).

ACKNOWLEDGMENTS. The second and the third authors are very grateful to the organisers of the 2nd International Conference on Uniform Distribution Theory, 2010, Strobl (Austria) for the invitation to this meeting, where the idea of this work was born.

During the preparation of this paper, A. O. was supported in part by the Swiss National Science Foundation Grant 133399, I. S. was supported in part by the Australian Research Council Grant DP1092835 and by the National Research Foundation of Singapore Grant CRP2-2007-03.

REFERENCES

[1] BOMBIERI, E.: *On exponential sums in finite fields*, Amer. J. Math. **8**8 (1966), 71–105.

[2] CONFLITTI, A. – SHPARLINSKI, I. E.: *On the multidimensional distribution of the subset sum generator of pseudorandom numbers*, Math. Comp. **7**3 (2004), 1005–1011.

[3] DRMOTA, M. – TICHY, R.: *Sequences, Discrepancies and Applications*, Springer-Verlag, Berlin, 1997.

[4] EL MAHASSNI, E.: *On the distribution of the elliptic subset sum generator of pseudorandom numbers*, Integers **8** (2008), Article #A31.

[5] VON ZUR GATHEN, J. – SHPARLINSKI, I. E.: *Predicting subset sum pseudorandom number generators*, in: *Proc. 11th Workshop on Selected Areas in Cryptography, Waterloo, 2004*, *Lecture Notes in Comput. Sci.*, Springer-Verlag, Berlin, **3357** 2005, pp. 241–251.

[6] VON ZUR GATHEN, J. – SHPARLINSKI, I. E.: , *Subset sum pseudorandom numbers: Fast generation and distribution*, J. Math. Cryptology, **3** (2009), 149–163.

[7] IWANIEC, H. – KOWALSKI, E.: , *Analytic Number Theory*, Amer. Math. Soc., Providence, RI, 2004.

[8] KNELLWOLF, S. – MEIER, W.: *Cryptanalysis of the knapsack generator*, in: *Proc. 18th Workshop on Fast Software Encryption 2011, Lyngby, Denmark, 2011 Lecture Notes in Comput. Sci.*, Springer-Verlag, Berlin (to appear).

[9] LIDL, R. – NIEDERREITER, H.: *Finite Fields*, in: *Encyclopedia of Mathematics and its Applications*, **20**. Cambridge University Press, Cambridge, 1997.

[10] MENEZES, A. J. – VAN OORSCHOT, P. C. – VANSTONE, S. A.: , *Handbook of Applied Cryptography*, CRC Press, Boca Raton, FL, 1996.

[11] NIEDERREITER, H.: *Random Number Generation and Quasi-Monte Carlo Methods*, in: *CBMS-NSF Regional Conference Series in Applied Mathematics* **63**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[12] RUEPPEL, R. A.: *Analysis and Design of Stream Ciphers*, Springer-Verlag, Berlin, 1986.

[13] RUEPPEL, R. A.: *Stream ciphers*, in: *Contemporary Cryptology: The Science of Information Integrity*, IEEE Press, NY, 1992, pp. 65–134.
[14] RUEPPEL, R. A. – MASSEY, J. L.: *Knapsack as a nonlinear function*, in: *IEEE Intern. Symp. of Inform. Theory*, IEEE Press, NY, 1985, p. 46.
[15] SHPARLINSKI, I. E.: , *On the average distribution of pseudorandom numbers generated by nonlinear permutations*, Math. Comp. **80** (2011), 1053–1061.
[16] SILVERMAN, J. H.: *The Arithmetic of Elliptic Curves*, in: *Graduate Text in Mathematics* **106** , Springer-Verlag, Berlin, 2009.

Received February 5, 2011
Accepted April 6, 2011

Simon R. Blackburn
Department of Mathematics
Royal Holloway, University of London
Egham, Surrey, TW20 0EX,
U. K.
E-mail: s.blackburn@rhul.ac.uk

Alina Ostafe
Department of Computing
Macquarie University
Sydney, NSW 2109,
AUSTRALIA
E-mail: alina.ostafe@mq.edu.au

Igor E. Shparlinski
Department of Computing
Macquarie University
Sydney, NW 2109
AUSTRALIA
E-mail: igor.shparlinski@mq.edu.au