ON THE MULTIDIMENSIONAL DISTRIBUTION
OF THE SUBSET SUM GENERATOR
OF PSEUDORANDOM NUMBERS

ALESSANDRO CONFLITTI AND IGOR E. SHPARLINSKI

ABSTRACT. We show that for a random choice of the parameters, the subset
sum pseudorandom number generator produces a sequence of uniformly and
independently distributed pseudorandom numbers. The result can be useful for
both cryptographic and quasi-Monte Carlo applications and relies on bounds
of exponential sums.

1. INTRODUCTION

For an integer $m \geq 1$ we denote by $\mathbb{Z}_m$ the residue ring modulo m. For integers
s and $m \geq 1$ we denote by $\lfloor s \rfloor_m$ the remainder of s on division by m.

Let $(u(x))$ be a linear recurrence sequence of order r over the field of two elements
$\mathbb{F}_2$; see [2, Chapter 8].

For an integer $m \geq 1$ one can consider the following *subset sum* generator of
pseudorandom numbers. Given an r-dimensional vector $\mathbf{z} = (z_1, \ldots, z_r) \in \mathbb{Z}_m^r$ of
weights, one can consider the sequence

$$V_{\mathbf{z}}(n) = \left\lfloor \sum_{j=1}^{r} u(n+j-1)z_j \right\rfloor_m, \qquad n = 1, 2, \ldots,$$

of elements of $\mathbb{Z}_m$.

This generator, which is also known as the *knapsack generator*, has been intro-
duced in [7] and studied in [5]; see also [3, Section 6.3.2] and [6, Section 3.7.9].

For cryptographic applications, it is usually recommended to use a linear re-
currence sequence of maximal period $\tau = 2^r - 1$ and also the modulus $m = 2^r$;
however, here we consider more general settings.

Here we study some statistical properties of the subset sum generator and show
that for any fixed dimension ν with $1 \leq \nu \leq r$, for almost all choices of weights
$\mathbf{z} = (z_1, \ldots, z_r) \in \mathbb{Z}_m^r$, the vectors $(V_{\mathbf{z}}(n), \ldots, V_{\mathbf{z}}(n + \nu - 1))$ are uniformly and
independently distributed. We formulate this result as a more general statement
about the deviation of the distribution of these vectors from the perfectly uniform
ν-dimensional distribution. In fact we use the classical number-theoretic notion

of the multidimensional *discrepancy* to give a quantitative form of this property. It can also be reformulated in terms of the ε-bias of the most significant bits of the elements of the generating sequences, which is more common in cryptographic literature.

We remark that in the special case of weights $z_j = 2^{r-j}$, $j = 1, \ldots, r$, and the modulus $m = 2^r$ this generator is well known in the theory of quasi-Monte Carlo methods. An exhaustive survey of known results about the distribution of this and more general generators can be found in [4, Chapter 9]. Although, for this (deterministic) choice of weights some uniformity of distribution results are known, they are weaker than our results which, however, apply only to randomized choice of weights. It is also clear that this choice of weights corresponds to very easy instances of the knapsack problem and thus is probably not suitable for cryptographic applications.

Our method is based on some simple bounds on exponential sums and the famous *Koksma–Szüsz inequality* (see Lemma 2.2 below) which relates the deviation from uniformity of distribution, that is, the discrepancy, and the corresponding exponential sums.

Throughout the paper, the implied constants in symbols "O" may depend on the integer parameter $\nu \geq 1$.

2. Preparations

Here we present several necessary technical tools.

We say that a linear recurrence sequence $u(x)$ of elements of $\mathbb{F}_2$ is of order r with characteristic polynomial

$$f(T) = T^r + c_{r-1}T^{r-1} + \ldots + c_1 T + c_0 \in \mathbb{F}_2[T]$$

if

$$u(x+r) + c_{r-1}u(x+r-1) + \ldots + c_1 u(x+1) + c_0 u(x) = 0, \qquad x = 1, 2, \ldots,$$

and it does not satisfy any shorter linear relation; see [2, Chapter 8].

It is easy to see that the set of all sequences with the same characteristic polynomial f form a linear space $\mathcal{L}(f)$ over $\mathbb{F}_2$.

We also need the following property of sequences from $\mathcal{L}(f)$ with irreducible f which is essentially [2, Theorem 8.28].

Lemma 2.1. *If $f \in \mathbb{F}_2[T]$ is irreducible over $\mathbb{F}_2$, then all nonzero sequences from $\mathcal{L}(f)$ are purely periodic with the same period.*

For a real z and an integer q we use the notation

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

We need the identity (see Exercise 11.a in Chapter 3 of [8])

$$(2.1) \qquad \sum_{\eta=0}^{q-1} \mathbf{e}_q(\eta\lambda) = \begin{cases} 0, & \text{if } \lambda \not\equiv 0 \pmod{q}, \\ q, & \text{if } \lambda \equiv 0 \pmod{q}. \end{cases}$$

We also make use of the inequality

$$(2.2) \qquad \sum_{\eta=0}^{q-1} \left| \sum_{\lambda=1}^{M} \mathbf{e}_q(\eta u) \right| = O(q \log q),$$

which holds for any integers q and M, $1 \le M \le q$; see [8, Chapter III, Exercise 11c].

For a sequence of N points

$$\Gamma = (\gamma_{0,x}, \dots, \gamma_{\nu-1,x})_{x=1}^{N} \tag{2.3}$$

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

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

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

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

and the supremum is taken over all such boxes.

As we have mentioned, one of our basic tools for studying 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 [1].

For an integer vector $\mathbf{a} = (a_1, \dots, a_\nu) \in \mathbb{Z}^\nu$ we define

$$|\mathbf{a}| = \max_{j=1,\dots,\nu} |a_j|, \qquad r(\mathbf{a}) = \prod_{j=1}^{\nu} \max\{|a_j|, 1\}. \tag{2.4}$$

Lemma 2.2. *For any integer $L > 1$ and any sequence Γ of N points* (2.3) *the bound*

$$D_\Gamma = O\left(\frac{1}{L} + \frac{1}{N} \sum_{0 < |\mathbf{a}| < L} \frac{1}{r(\mathbf{a})} \left| \sum_{x=1}^{N} \mathbf{e}\left(\sum_{j=0}^{\nu-1} a_j \gamma_{j,x} \right) \right| \right)$$

on the discrepancy D_Γ holds, where $|\mathbf{a}|$, $r(\mathbf{a})$ are defined by (2.4) *and the sum is taken over all integer vectors*

$$\mathbf{a} = (a_0, \dots, a_{\nu-1}) \in \mathbb{Z}^\nu$$

with $0 < |\mathbf{a}| < L$.

3. Main result

We denote by $D_{\mathbf{z}}^\nu(N)$ the discrepancy of the points

$$\left(\frac{V_{\mathbf{z}}(n)}{m}, \dots, \frac{V_{\mathbf{z}}(n+\nu-1)}{m} \right), \qquad n = 1, \dots, N.$$

Theorem 3.1. *Let the linear recurrence sequence $(u(x))$ be purely periodic with period τ and order r and let its characteristic polynomial be irreducible over $\mathbb{F}_2$. Then for any $\delta > 0$, and any $\nu \le r$ for all $\mathbf{z} \in \mathbb{Z}_m^r$ except at most $O(\delta m^r)$ of them, for all $1 \le N \le \tau$ the bound*

$$D_{\mathbf{z}}^\nu(N) = O\left(\delta^{-1} N^{-1/2} \log^\nu m \log^2 \tau \right)$$

holds.

Proof. From Lemma 2.2, used with $L = \lfloor m/\nu \rfloor$, we derive

$$D_{\mathbf{z}}^\nu(N) = O\left(\frac{1}{m} + \frac{1}{N} \sum_{0 < |\mathbf{a}| < m/\nu} \frac{1}{r(\mathbf{a})} \left| \sum_{n=1}^{N} \mathbf{e}_m\left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \right| \right).$$

Let $N_\mu = 2^\mu$, $\mu = 0, 1, \ldots$. Define k by the inequality $N_{k-1} < N \le N_k$, that is, $k = \lceil \log_2 N \rceil$. Then from (2.1) we derive

$$\sum_{n=1}^{N} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right)$$

$$= \frac{1}{N_k} \sum_{n=1}^{N_k} \sum_{\lambda=1}^{N} \sum_{\eta=0}^{N_k} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \mathbf{e}_{N_k} \left(\eta(n-\lambda) \right).$$

Hence,

$$(3.1) \qquad D_{\mathbf{z}}^{\nu}(N) = O \left(\frac{1}{m} + \frac{1}{NN_k} \Delta_{\mathbf{z}}^{\nu}(k) \right)$$

where

$$\Delta_{\mathbf{z}}^{\nu}(k) = \sum_{0 < |\mathbf{a}| < m/\nu} \frac{1}{r(\mathbf{a})} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k} \left(-\eta\lambda \right) \right|$$

$$\times \left| \sum_{n=1}^{N_k} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \mathbf{e}_{N_k} \left(\eta n \right) \right|.$$

Applying the Cauchy inequality, we derive

$$\left(\sum_{\mathbf{z} \in \mathbb{Z}_m^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \mathbf{e}_{N_k} \left(\eta n \right) \right| \right)^2$$

$$\le m^r \sum_{\mathbf{z} \in \mathbb{Z}_m^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \mathbf{e}_{N_k} \left(\eta n \right) \right|^2$$

$$= m^r \sum_{n,l=1}^{N_k} \mathbf{e}_{N_k} \left(\eta(n-l) \right) \sum_{\mathbf{z} \in \mathbb{Z}_m^r} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j \left(V_{\mathbf{z}}(n+j) - V_{\mathbf{z}}(l+j) \right) \right).$$

By definition of $V_{\mathbf{z}}(n)$ we have

$$\sum_{\mathbf{z} \in \mathbb{Z}_m^r} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j \left(V_{\mathbf{z}}(n+j) - V_{\mathbf{z}}(l+j) \right) \right)$$

$$= \prod_{h=1}^{r} \sum_{z_h \in \mathbb{Z}_m} \mathbf{e}_m \left(z_h \sum_{j=0}^{\nu-1} a_j \left(u(n+j+h-2) - u(l+j+h-2) \right) \right).$$

The product is equal to m^r if for every $h = 1, \ldots, r$

$$(3.2) \qquad \sum_{j=0}^{\nu-1} a_j \left(u(n+j+h-2) - u(l+j+h-2) \right) \equiv 0 \pmod{m};$$

otherwise it vanishes.

Therefore

$$(3.3) \qquad \sum_{\mathbf{z} \in \mathbb{Z}_m^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \mathbf{e}_{N_k}(\eta n) \right| \leq m^r T_k^{1/2}$$

where T_k is the number of pairs (n, l), $1 \leq n, l \leq N_k$, for which (3.2) holds for every $h = 1, \ldots, r$.

Because $u(x) \in \{0, 1\}$ for all integers $x \geq 1$ and $0 \leq |a_j| < m/\nu$, the congruence (3.2) becomes an equation

$$\sum_{j=0}^{\nu-1} a_j \left(u(n+j+h-2) - u(l+j+h-2) \right) = 0, \qquad h = 1, \ldots, r.$$

Let us write $a_j = 2^\alpha b_j$ where 2^α is the largest power of 2 which divides every a_j, $j = 0, \ldots, \nu - 1$. In particular, at least one b_j is odd. Then the previous equation becomes

$$(3.4) \qquad \sum_{j=0}^{\nu-1} b_j \left(u(n+j+h-2) - u(l+j+h-2) \right) = 0, \qquad h = 1, \ldots, r.$$

Considering the equation (3.4) in $\mathbb{F}_2$, we derive

$$w(n+h) \equiv w(l+h) \pmod{2}, \qquad h = 1, \ldots, r,$$

where

$$w(x) = \sum_{j=0}^{\nu-1} b_j u(x+j-2)$$

is a non-zero sequence over $\mathbb{F}_2$ because at least one b_j, $j = 0, \ldots, \nu - 1$, is odd and $\nu \leq r$. Taking into account that $w(x)$ is a linear recurrence sequence of order r (with the same characteristic polynomial as $u(x)$), we obtain

$$(3.5) \qquad w(n+x) \equiv w(l+x) \pmod{2}, \qquad x = 1, 2, \ldots.$$

Because the characteristic polynomial of u is irreducible, by Lemma 2.1 the linear recurrence sequence $w(x)$ has the same period τ. Therefore (3.5) implies that $n \equiv l \pmod{\tau}$ which yields the inequality $T_k \leq N_k(\lfloor N_k/\tau \rfloor + 1) \leq 2N_k$, because $N_k = 2N_{k-1} < 2N \leq 2\tau$.

Thus by (3.3) we have

$$\sum_{\mathbf{z} \in \mathbb{Z}_m^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_m \left(\sum_{j=0}^{\nu-1} a_j V_{\mathbf{z}}(n+j) \right) \mathbf{e}_{N_k}(\eta n) \right| \leq 2^{1/2} m^r N_k^{1/2}.$$

Hence recalling (2.2) we obtain

$$
\begin{aligned}
\sum_{\mathbf{z}\in\mathbb{Z}_m^r}\Delta_{\mathbf{z}}^{\nu}(k) &= \sum_{0<|\mathbf{a}|<m/\nu}\frac{1}{r(\mathbf{a})}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}\left(-\eta\lambda\right)\right| \\
&\qquad\times\sum_{\mathbf{z}\in\mathbb{Z}_m^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_m\left(\sum_{j=0}^{\nu-1}a_jV_{\mathbf{z}}(n+j)\right)\mathbf{e}_{N_k}\left(\eta n\right)\right| \\
&= 2^{1/2}m^rN_k^{1/2}\sum_{0<|\mathbf{a}|<m/\nu}\frac{1}{r(\mathbf{a})}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}\left(-\eta\lambda\right)\right| \\
&= O\left(m^rN_k^{3/2}k\sum_{0<|\mathbf{a}|<m/\nu}\frac{1}{r(\mathbf{a})}\right) \\
&= O\left(m^rN_k^{3/2}k\log^{\nu}m\right) \\
&= O\left(m^rN_k^{3/2}\log^{\nu}m\log\tau\right),
\end{aligned}
$$

because $k = O(\log\tau)$.

This implies that for any k the number of vectors $\mathbf{z}\in\mathbb{Z}_m^r$ with

$$
\Delta_{\mathbf{z}}^{\nu}(k) \geq \delta^{-1}N_k^{3/2}\log^{\nu}m\log^2\tau
$$

is at most $O(\delta m^r\log^{-1}\tau)$. Therefore, we have that the number of vectors $\mathbf{z}\in\mathbb{Z}_m^r$ with

$$
\Delta_{\mathbf{z}}^{\nu}(k) \geq \delta^{-1}N_k^{3/2}\log^{\nu}m\log^2\tau
$$

for at least one $k = 1,\ldots,\lceil\log\tau\rceil$ is at most $O(\delta m^r)$. For other $\mathbf{z}\in\mathbb{Z}_m^r$, from (3.1), we obtain

$$
D_{\mathbf{z}}^{\nu}(N) = O\left(\frac{1}{m} + \frac{1}{NN_k}\Delta_{\mathbf{z}}^{\nu}(k)\right) = O\left(\delta^{-1}N^{-1}N_k^{1/2}\log^{\nu}m\log^2\tau\right).
$$

Taking into account the inequality $N^{-1}N_k^{1/2} \leq 2N^{-1/2}$, we obtain the desired result. □

We remark that the result of Theorem 3.1 can be extended to more general classes of characteristic polynomials. However, as we have mentioned, the case of the most practical interest is $\tau = 2^r - 1$ which implies that the characteristic polynomial is primitive, and thus irreducible, over $\mathbb{F}_2$.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Drmota and R. Tichy, *Sequences, discrepancies and applications*, Springer-Verlag, Berlin, 1997. MR **98j:**11057
[2] R. Lidl and H. Niederreiter, *Finite fields*, Cambridge University Press, Cambridge, 1997. MR **97i:**11115
[3] A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of applied cryptography*, CRC Press, Boca Raton, FL, 1996. MR **99g:**94015

[4] H. Niederreiter, *Random number generation and Quasi–Monte Carlo methods*, SIAM Press, 1992. MR **93h:**65008

[5] R. A. Rueppel, *Analysis and design of stream ciphers*, Springer-Verlag, Berlin, 1986. MR **88h:**94002

[6] R. A. Rueppel, 'Stream ciphers', *Contemporary cryptology: The science of information integrity*, IEEE Press, NY, 1992, 65–134.

[7] R. A. Rueppel and J. L. Massey, 'Knapsack as a nonlinear function', *IEEE Intern. Symp. of Inform. Theory*, IEEE Press, NY, 1985, 46.

[8] I. M. Vinogradov, *Elements of number theory*, Dover Publ., New York, 1954. MR **15:**933e

Dipartimento di Matematica, Università degli Studi di Roma "Tor Vergata", Via della Ricerca Scientifica, I-00133 Roma, Italy

E-mail address: `conflitt@mat.uniroma2.it`

Department of Computing, Macquarie University, Sydney, New South Wales 2109, Australia

E-mail address: `igor@ics.mq.edu.au`