STATISTICAL INDEPENDENCE OF A NEW CLASS OF INVERSIVE CONGRUENTIAL PSEUDORANDOM NUMBERS

JÜRGEN EICHENAUER-HERRMANN

Abstract. Linear congruential pseudorandom numbers show several undesirable regularities which can render them useless for certain stochastic simulations. This was the motivation for important recent developments in nonlinear congruential methods for generating uniform pseudorandom numbers. It is particularly promising to achieve nonlinearity by employing the operation of multiplicative inversion with respect to a prime modulus. In the present paper a new class of such inversive congruential generators is introduced and analyzed. It is shown that they have excellent statistical independence properties and model true random numbers very closely. The methods of proof rely heavily on Weil-Stepanov bounds for rational exponential sums.

1. Introduction

The outcome of a stochastic simulation strongly depends on the quality of the pseudorandom numbers. General background material on pseudorandom number generation can be found in the book of Knuth [23] and the survey article of Niederreiter [28]. The classical standard method of generating uniform pseudorandom numbers in the interval [0, 1) is the linear congruential method. Theoretical results on the structural and statistical properties of the generated sequences indicate that a reasonable behavior can be obtained if a judicious choice of parameters is made which depends on the dimension of the simulation problem (cf. [27–29]). Hence, a considerable computational effort has to be expended to guarantee acceptable properties, at least for a very modest range of dimensions (cf. [1, 19, 20]). However, linear congruential sequences show an unfavorable coarse lattice structure which stems from the simple nature of the underlying linear recursion and cannot be overcome even by the most judicious choice of parameters (cf. [25, 26, 37]).

This state of affairs provided the motivation for recent work on nonlinear congruential methods in order to overcome the deficiencies of the linear congruential method (cf. [2–9, 11–18, 30–33]). A review of the development of this area is given in [10] and in Niederreiter's excellent survey articles [34, 35]. The key idea behind these methods is the use of different nonlinear recursions in modular arithmetic instead of the simple linear recursion. Understandably, all these techniques are somewhat slower than the linear congruential method.
However, nowadays the computer time taken for pseudorandom number generation in a typical stochastic simulation can almost always be neglected (cf. [21, 36]).

The most promising results have been obtained for prime moduli. For a (large) prime $p$ put $\mathbb{Z}_p = \{0, 1, \ldots, p - 1\}$ and $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$. The following general class of nonlinear congruential generators was introduced in [2]: A \textit{nonlinear congruential sequence} $(y_n)_{n \geq 0}$ of elements of $\mathbb{Z}_p$ is generated by

$$y_{n+1} = f(y_n), \quad n \geq 0,$$

where $f: \mathbb{Z}_p \rightarrow \mathbb{Z}_p$ is a function such that $(y_n)_{n \geq 0}$ is purely periodic with maximal period length $p$. Niederreiter [30] pointed out that there exists a uniquely determined permutation polynomial $g$ of degree $s$ with $1 \leq s \leq p - 2$ over the finite field $\mathbb{Z}_p$ such that

$$y_n = g(n), \quad n \geq 0.$$

In Niederreiter [31] it is shown that any nonlinear congruential generator has excellent statistical independence properties for all dimensions $d \leq s$ provided the degree $s$ of $g$ is small relative to $p^{1/2}$.

In contrast, the present paper deals with certain polynomials $g$ of maximal degree $s = p - 2$ and establishes favorable statistical independence properties for all dimensions $d < p$. In the following the abbreviation $\overline{z} \equiv z^{p-2} \pmod{p}$, $\overline{z} \in \mathbb{Z}_p$, is used for integers $z$. Note that $\overline{z}$ is the multiplicative inverse of $z$ modulo $p$ if $z \not\equiv 0 \pmod{p}$. A method for its efficient calculation is based on the Euclidean algorithm with the integers $z$ and $p$ (cf. [10]). The standard inversive congruential method, which was introduced in [4], generates a sequence $(y_n)_{n \geq 0}$ of elements of $\mathbb{Z}_p$ by the recursion $y_{n+1} \equiv a\overline{y}_n + b \pmod{p}$ for $n \geq 0$. In the present paper the following new class of inversive congruential generators is considered which has even better structural and statistical independence properties than the standard type. For integers $a$, $b \in \mathbb{Z}_p$ with $a \neq 0$ an \textit{inversive congruential sequence} $(y_n)_{n \geq 0}$ of elements of $\mathbb{Z}_p$ is defined by

$$y_n = an + b, \quad n \geq 0.$$

A sequence $(x_n)_{n \geq 0}$ of \textit{inversive congruential pseudorandom numbers} in the interval $[0, 1)$ is obtained by the normalization $x_n = y_n/p$ for $n \geq 0$. Obviously, any inversive congruential sequence is purely periodic with maximal period length $p$, i.e., $\{y_0, y_1, \ldots, y_{p-1}\} = \mathbb{Z}_p$, which guarantees that the one-dimensional distribution of the corresponding pseudorandom numbers is as good as possible. Hence, any inversive congruential generator passes the \textit{uniformity test} for equidistribution in $[0, 1)$.

Statistical independence properties of pseudorandom numbers are at least as important for stochastic simulations as uniformity properties. A reliable theoretical test for statistical independence is the \textit{serial test}, which employs the discrepancy of tuples of pseudorandom numbers. For a given dimension $k \geq 2$ and for $N$ arbitrary points $t_0, t_1, \ldots, t_{N-1} \in [0, 1)^k$ the \textit{discrepancy} is defined by

$$D_N(t_0, t_1, \ldots, t_{N-1}) = \sup_J |F_N(J) - V(J)|,$$

where the supremum is extended over all subintervals $J$ of $[0, 1)^k$, $F_N(J)$ is $N^{-1}$ times the number of terms among $t_0, t_1, \ldots, t_{N-1}$ falling into $J$, and $V(J)$
denotes the volume of \( J \). In the present paper, for a sequence of inversive congruential pseudorandom numbers \((x_n)_{n \geq 0}\), the points
\[
x_n = (x_{n+n_1}, x_{n+n_2}, \ldots, x_{n+n_k}) \in [0, 1)^k, \quad 0 \leq n < p,
\]
are considered, where \( n_1, n_2, \ldots, n_k \) are arbitrary integers with \( 0 = n_1 < n_2 < \cdots < n_k < p \), and the abbreviation
\[
D_p^{(k)} = D_p(x_0, x_1, \ldots, x_{p-1})
\]
is used for their discrepancy. An inversive congruential generator passes the \( k \)-dimensional serial test if \( D_p^{(k)} \) is reasonably small. Since an exact calculation of the discrepancy \( D_p^{(k)} \) is impossible, one is interested in bounds for \( D_p^{(k)} \).

In the present paper upper and lower bounds for \( D_p^{(k)} \) are established, which are essentially best possible. In §2 the main results are stated precisely and the behavior of inversive congruential generators under the serial test is discussed. Section 3 contains several auxiliary results. The proof of the main results is given in §4. The methods of proof rely heavily on Weil-Stepanov bounds for rational exponential sums. Extensive background material on exponential sums can be found in [24].

2. Main results

**Theorem 1.** Let \( 2 \leq k < p \). Then the discrepancy \( D_p^{(k)} \) for any inversive congruential generator satisfies
\[
D_p^{(k)} < 2p^{-1/2} \left( (k-1) \left( \frac{2}{\pi} \log p + \frac{7}{5} \right)^k + 1 \right) + kp^{-1}.
\]

**Theorem 2.** Let \( 0 < t < 1 \). Then there exist more than \( A_p(t)(p-1) \) values of \( a \in \mathbb{Z}_p^* \) such that the discrepancy \( D_p^{(k)} \) for any corresponding inversive congruential generator satisfies
\[
D_p^{(k)} \geq \frac{t}{2(\pi + 2)} p^{-1/2}
\]
for all dimensions \( k \geq 2 \), where
\[
A_p(t) = \frac{(1 - t^2)p}{(4 - t^2)p + 12p^{1/2} + 9}.
\]

Theorem 1 shows that \( D_p^{(k)} = O(p^{-1/2} (\log p)^k) \) for any inversive congruential generator, where the implied constant is absolute. It should be observed that this bound is independent not only of the specific choice of the parameters \( a, b \) in the inversive congruential method, but also of the parameters \( n_2, \ldots, n_k \). This is a remarkable contrast to the linear congruential method, where the behavior under the serial test strongly depends on these quantities (and on the dimension \( k \)).

Theorem 2 implies that a positive proportion of the inversive congruential generators has a discrepancy \( D_p^{(k)} \) which is at least of the order of magnitude \( p^{-1/2} \) for all dimensions \( k \geq 2 \). Therefore the upper bound in Theorem 1 is in general best possible up to the logarithmic factor.

Theorems 1 and 2 show that in the inversive congruential method the discrepancy \( D_p^{(k)} \) has on the average an order of magnitude between \( p^{-1/2} \) and
It is in this range of magnitudes where one also finds the discrepancy of \( p \) independent and uniformly distributed random points from \([0, 1)^k\), which should be roughly \( p^{-1/2}(\log \log p)^{1/2} \) according to the law of the iterated logarithm for discrepancies (cf. [22]). In this sense, inversive congruential pseudorandom numbers model true random numbers very closely.

3. Auxiliary results

First, some further notation is necessary. For integers \( k \geq 1 \) and \( q \geq 2 \) let \( C_k(q) \) be the set of all nonzero lattice points \((h_1, \ldots, h_k) \in \mathbb{Z}^k\) with \(-q/2 < h_j \leq q/2\) for \(1 \leq j \leq k\). Put

\[
r(h, q) = \begin{cases} 
1, & \text{for } h = 0, \\
q \sin \pi h/q, & \text{for } h \in C_1(q),
\end{cases}
\]

and define

\[
r(h, q) = \prod_{j=1}^{k} r(h_j, q)
\]

for \( h = (h_1, \ldots, h_k) \in C_k(q) \). For \( t \in \mathbb{R} \) the abbreviation \( e(t) = e^{2\pi i t} \) is used, for integers \( z \) we put \( \chi(z) = e(z/p) \), and \( u \cdot v \) stands for the standard inner product of \( u, v \in \mathbb{R}^k \).

Below, four known results are stated. The first two lemmas follow from Lemmas 2.2 and 2.3 in [27], the third lemma is a special case of a classical result of Weil [39] (cf. [32, 38]), and the last lemma is a special version of Lemma 1 in [33].

**Lemma 1.** Let \( N \geq 1 \) and \( q \geq 2 \) be integers. Suppose that \( y_0, y_1, \ldots, y_{N-1} \in \mathbb{Z}^k \). Then the discrepancy of the points \( t_n = q^{-x}y_n \in [0, 1)^k \) for \( 0 \leq n < N \) satisfies

\[
D_N(t_0, t_1, \ldots, t_{N-1}) \leq \frac{k}{q} + \frac{1}{N} \sum_{h \in C_k(q)} \frac{1}{r(h, q)} \left| \sum_{n=0}^{N-1} e(h \cdot t_n) \right|.
\]

**Lemma 2.** Let \( q \geq 2 \) be an integer. Then

\[
\sum_{h \in C_k(q)} \frac{1}{r(h, q)} < \frac{2}{\pi} \log q + \frac{2}{5}.
\]

**Lemma 3.** Let \( P, Q \) be polynomials over the finite field \( \mathbb{Z}_p \) with \( 1 \leq \deg(Q) < \deg(P) < p \). Let \( r \) denote the number of distinct poles of \( PQ \) in the algebraic closure of \( \mathbb{Z}_p \) (including the point at infinity) and let \( m_1, \ldots, m_r \) be the multiplicities of the poles. Then

\[
\left| \sum_{z \in \mathbb{Z}_p} \chi(P(z)Q(z)) \right| \leq \left( r - 2 + \sum_{i=1}^{r} m_i \right) p^{1/2}.
\]

**Lemma 4.** The discrepancy of \( N \) arbitrary points \( t_0, t_1, \ldots, t_{N-1} \in [0, 1)^k \) satisfies
for any lattice point \( h = (h_1, h_2, 0, \ldots, 0) \in \mathbb{Z}^k \) with \( h_1 h_2 \neq 0 \).

Lemmas 1 and 4 show that the exponential sums \( S(h) = \sum_{n \in \mathbb{Z}_p} e(h \cdot x_n) \) for \( h \in C_k(p) \) are the crucial quantities for the analysis of the discrepancy \( D_p^{(k)} \). Put \( J(h) = \{1 \leq j \leq k | h_j \neq 0\} \) for \( h = (h_1, \ldots, h_k) \in C_k(p) \). The following technical result is used for the proof of Theorems 1 and 2.

**Lemma 5.** Let \( h \in C_k(p) \). Then \( |S(h)| \leq m(2p^{1/2} + 1) - (2p^{1/2} - 1) \), where \( m \) denotes the number of nonzero coordinates of \( h \).

**Proof.** The definition of an inversive congruential sequence implies that

\[
S(h) = \sum_{n \in \mathbb{Z}_p} \chi \left( \sum_{j \in J(h)} h_j (a(n + n_j) + b) \right)
\]

for \( h = (h_1, \ldots, h_k) \in C_k(p) \). If \( m = 1 \), then

\[
S(h) = \sum_{z \in \mathbb{Z}_p} \chi(z) = 0,
\]

which proves the desired inequality. From now on, \( m \geq 2 \) is assumed. Put

\[
N(h) = \{n \in \mathbb{Z}_p | a(n + n_j) + b \equiv 0 \pmod{p} \text{ for some } j \in J(h)\}.
\]

Then one obtains

\[
|S(h)| \leq m + \sum_{n \in \mathbb{Z}_p \atop \notin N(h)} \chi \left( \sum_{j \in J(h)} h_j (a(n + n_j) + b) \right)
\]

\[
= m + \sum_{n \in \mathbb{Z}_p \atop \notin N(h)} \chi \left( \sum_{j \in J(h)} h_j \prod_{l \in J(h) \atop l \neq j} (a(n + n_l) + b) \right)
\]

\[
\cdot \left( \prod_{j \in J(h)} (a(n + n_j) + b) \right).
\]

Now, let \( j_0 \in J(h) \) be fixed and put

\[
Z(h) = \{a(n_{j_0} - n_j) \in \mathbb{Z}_p | j \in J(h)\}.
\]
Then the transformation \( z = a(n + n_{jo}) + b \) yields

\[
|S(h)| \leq m + \left| \sum_{z \in \mathbb{Z}_p} \chi \left( \left( \sum_{j \in J(h)} h_j \prod_{l \in J(h), l \neq j} (\hat{z} + a(n_l - n_{jo})) \right) \right) \cdot \left( \prod_{j \in J(h)} (\hat{z} + a(n_j - n_{jo})) \right) \right|
\]

\[
= m + \left| \sum_{z \in \mathbb{Z}_p} \chi \left( z \left( \sum_{j \in J(h)} h_j \prod_{l \in J(h), l \neq j} (a(n_l - n_{jo})z + 1) \right) \right) \cdot \left( \prod_{j \in J(h), j \neq j_0} (a(n_j - n_{jo})z + 1) \right) \right|
\]

\[
\leq m + 1 + \left| \sum_{z \in \mathbb{Z}_p} \chi(P(z)Q(z)) \right|
\]

with the polynomials

\[
P(z) = z \left( \sum_{j \in J(h)} h_j \prod_{l \in J(h), l \neq j} (a(n_l - n_{jo})z + 1) \right)
\]

and

\[
Q(z) = \prod_{j \in J(h), j \neq j_0} (a(n_j - n_{jo})z + 1)
\]

for \( z \in \mathbb{Z}_p \). Since \( \deg(P) = m \) and \( \deg(Q) = m - 1 \geq 1 \), Lemma 3 can be applied with \( r = \deg(Q) + 1 = m \) and \( m_1 = \cdots = m_r = 1 \), which implies that \(|S(h)| \leq m + 1 + (2m - 2)p^{1/2} \), i.e., the desired result. \( \Box \)

Finally, some further prerequisites are necessary in order to prove Theorem 2. First, a short calculation and the transformation \( z = n + \bar{a}b \) (mod \( p \)) show that \( S(h) = K_{K_2}(\bar{a}) \) for \( h = (1, -1, 0, \ldots, 0) \in \mathbb{Z}^k \), where

\[
K_\gamma(c) = \sum_{z \in \mathbb{Z}_p} \chi(c(\bar{z} - (z + \gamma)))
\]

for \( c \in \mathbb{Z}_p^* \) and \( \gamma \in \mathbb{Z}_p^* \). Hence, Lemma 5 implies that \(|K_\gamma(c)| \leq 2p^{1/2} + 3\) for \( c \in \mathbb{Z}_p^* \) and \( \gamma \in \mathbb{Z}_p^* \).

Lemma 6. Let \( \gamma \in \mathbb{Z}_p^* \). Then

\[
\sum_{c \in \mathbb{Z}_p^*} |K_\gamma(c)|^2 \geq p(p - 1).
\]
Proof. Easy calculations show that

\[
\sum_{c \in \mathbb{Z}_p^2} |K_\gamma(c)|^2 = \sum_{y \in \mathbb{Z}_p} \sum_{z \in \mathbb{Z}_p} \chi(c(\bar{y} - (y + \gamma) - \bar{z} + (z + \gamma)))
\]

\[
= \sum_{y \in \mathbb{Z}_p} \sum_{z \in \mathbb{Z}_p} \chi(c(\bar{y} - (y + \gamma) - \bar{z} + (z + \gamma))) - p^2
\]

\[
= p\left\{(y, z) \in \mathbb{Z}_p^2 \mid \bar{y} - (y + \gamma) \equiv \bar{z} - (z + \gamma) \pmod{p}\right\} - p^2
\]

\[
\geq p\left\{(y, z) \in \mathbb{Z}_p^2 \mid y = z \text{ or } y \equiv -(z + \gamma) \pmod{p}\right\} - p^2
\]

\[
\geq p(2p - 1) - p^2 = p(p - 1). \quad \Box
\]

4. Proof of the main results

Proof of Theorem 1. First, Lemma 1 is applied with \(N = q = p\) and \(t_n = x_n\) for \(0 \leq n < p\). This yields

\[
D^p_k \leq \frac{k}{p} + \frac{1}{p} \sum_{h \in C_k(p)} \frac{1}{r(h, p)} |S(h)|
\]

\[
= \frac{k}{p} + \frac{1}{p} \sum_{m=1}^{k} \sum_{J \subseteq \{1, \ldots, k\}} \sum_{h \in C_k(p)} \frac{1}{r(h, p)} |S(h)|.
\]

Now, Lemma 5 can be used in order to obtain

\[
D^p_k \leq \frac{k}{p} + \frac{1}{p} \sum_{m=1}^{k} (m(2p^{1/2} + 1) - (2p^{1/2} - 1)) \left(\begin{array}{c} k \\ m \end{array}\right) \left(\sum_{h \in C_k(p)} \frac{1}{r(h, p)}\right)^m.
\]

Therefore, Lemma 2 implies that

\[
D^p_k < \frac{k}{p} + \frac{1}{p} \sum_{m=1}^{k} (m(2p^{1/2} + 1) - (2p^{1/2} - 1)) \left(\begin{array}{c} k \\ m \end{array}\right) \left(\frac{2}{\pi} \log p + \frac{2}{5}\right)^m
\]

\[
= \frac{k}{p} + \frac{1}{p} \left(2p^{1/2} + 1\right)k \left(\frac{2}{\pi} \log p + \frac{2}{5}\right) \left(\frac{2}{\pi} \log p + \frac{7}{5}\right)^{k-1}
\]

\[
- (2p^{1/2} - 1) \left(\left(\frac{2}{\pi} \log p + \frac{7}{5}\right)^k - 1\right)
\]

\[
= 2p^{-1/2} \left(k - 1\right) \left(\frac{2}{\pi} \log p + \frac{7}{5}\right)^k + 1 + (k - 1)p^{-1}
\]

\[
- p^{-1} \left(\frac{2}{\pi} \log p + \frac{7}{5}\right)^{k-1} \left(2kp^{1/2} - k \left(\frac{2}{\pi} \log p + \frac{2}{5}\right) - \left(\frac{2}{\pi} \log p + \frac{7}{5}\right)\right)
\]

\[
< 2p^{-1/2} \left(k - 1\right) \left(\frac{2}{\pi} \log p + \frac{7}{5}\right)^k + 1 + kp^{-1}
\]

\[
- 2kp^{-1} \left(\frac{2}{\pi} \log p + \frac{7}{5}\right)^{k-1} \left(p^{1/2} - \frac{2}{\pi} \log p - \frac{9}{10}\right),
\]

which yields the desired result, since \(p^{1/2} - \frac{2}{\pi} \log p - \frac{9}{10} > 0\). \(\Box\)
Proof of Theorem 2. First, Lemma 4 is applied with $N = p$, $t_n = x_n$ for $0 \leq n < p$, and $h = (1, -1, 0, \ldots, 0) \in \mathbb{Z}^k$. This yields
\[ D_p^{(k)} \geq \frac{1}{2(\pi + 2)^p} |K_{n_2}(\overline{a})|. \]

Now, it is proved by contradiction that for every fixed $\gamma \in \mathbb{Z}_p^*$ and $0 < t \leq 1$ there exist more than $A_p(t)(p - 1)$ values of $c \in \mathbb{Z}_p^*$ such that $|K_\gamma(c)| \geq tp^{1/2}$, which completes the proof. Suppose that $|K_\gamma(c)| \geq tp^{1/2}$ for at most $A_p(t)(p - 1)$ values of $c \in \mathbb{Z}_p^*$. Then $|K_\gamma(c)| < tp^{1/2}$ for at least $(1 - A_p(t))(p - 1)$ values of $c \in \mathbb{Z}_p^*$. Since $|K_\gamma(c)| \leq 2p^{1/2} + 3$ for all $c \in \mathbb{Z}_p^*$, it follows that
\[ \sum_{c \in \mathbb{Z}_p^*} |K_\gamma(c)|^2 < (1 - A_p(t))(p - 1)t^2p + A_p(t)(p - 1)(2p^{1/2} + 3)^2 = p(p - 1), \]
which is a contradiction to Lemma 6. \(\Box\)

Acknowledgment
The author would like to thank the referee for valuable suggestions.

Bibliography

19. G. S. Fishman, Multiplicative congruential random number generators with modulus $2^\beta$: An exhaustive analysis for $\beta = 32$ and a partial analysis for $\beta = 48$, Math. Comp. 54 (1990), 331–344.
32. _, The serial test for congruential pseudorandom numbers generated by inversions, Math. Comp. 52 (1989), 135–144.

FACHBEREICH MATHEMATIK, TECHNISCHE HOCHSCHULE DARMSTADT, SCHLOSSGARTENSTRASSE 7, D–6100 DARMSTADT, GERMANY