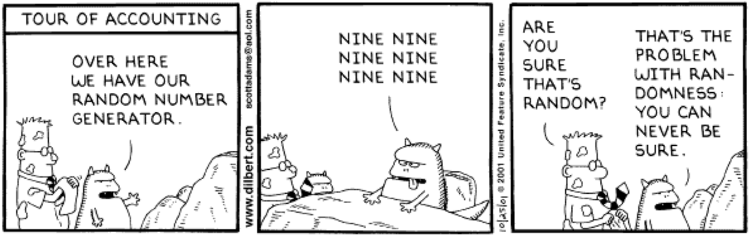
Pseudorandom Sequences I: Linear Complexity and Related Measures

Arne Winterhof

Austrian Academy of Sciences
Johann Radon Institute for Computational and Applied Mathematics
Linz

Carleton University 2010

Why pseudorandom and not 'truly' random sequences?



Pseudorandom Sequences

Sequences which are generated by a deterministic algorithm and 'look random' are called pseudorandom.

Desirable 'randomness properties' depend on the application!

cryptography: unpredictability
numerical integration (quasi-Monte Carlo): uniform distribution
radar: distinction from reflected signal
gambling: a good lawyer

Linear Complexity

The linear complexity $L(s_n)$ of a periodic sequence (s_n) over a field $\mathbb{F}$ is the smallest positive integer L such that there are constants $c_0, \ldots, c_{L-1} \in \mathbb{F}$ with

$$s_{n+L} = c_{L-1}s_{n+L-1} + \ldots + c_0 s_n, \ \ n \geq 0.$$

For a positive integer N the Nth linear complexity $L(s_n, N)$ of a sequence (s_n) over $\mathbb{F}$ is the smallest positive integer L such that there are constants $c_0, \ldots, c_{L-1} \in \mathbb{F}$ satisfying

$$s_{n+L} = c_{L-1}s_{n+L-1} + \ldots + c_0 s_n,$$

$0 \leq n \leq N - L - 1.$

Cryptographic Background

A low Nth linear complexity has turned out to be undesirable for cryptographical applications as stream ciphers.

Example (Stream Cipher)

We consider a message $m_0, m_1, \dots$ represented as a sequence over $\mathbb{F}$. In a *stream cipher* each message symbol m_j is enciphered with an element x_j of another sequence $x_0, x_1, \dots$ over $\mathbb{F}$, the *key stream*, by

$$c_j = m_j + x_j.$$

The cipher text $c_0, c_1, \dots$ can be deciphered by subtracting the key stream

$$m_j = c_j - x_j.$$

Relation to Quasi-Monte Carlo Methods

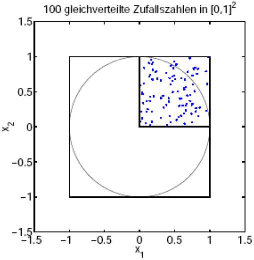
Sequences with low linear complexity are shown to be unsuitable for some applications using quasi-Monte Carlo methods as well. The following example describes a typical quasi-Monte Carlo application.

Example (Quasi-Monte-Carlo Calculation of π)

1. Choose N pairs of a sequence (x_n) in $[0, 1)$

$$(x_n, x_{n+1}) \in [0, 1)^2, \quad n = 0, \ldots, N-1.$$

2. Count the number K of pairs (x_n, x_{n+1}) in the unit circle.
3. Approximate π by $\frac{4K}{N}$.



k	π
10	3.2000000
100	2.9600000
1000	3.2040000
10000	3.1196000
100000	3.1390000
1000000	3.1460440
10000000	3.1416592

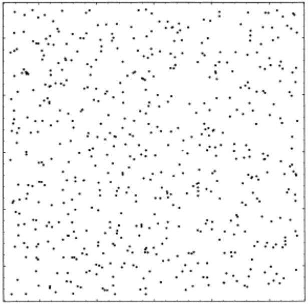
Marsaglia's Lattice test, 1972

(η_n) T-periodic sequence over $\mathbb{F}_p$
For $s \geq 1$ we say that (η_n) passes the s-dimensional lattice test if the vectors $\{\mathbf{u}_n - \mathbf{u}_0 : 1 \leq n < T\}$ span $\mathbb{F}_p^s$, where

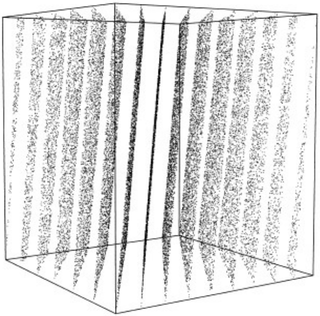
$$\mathbf{u}_n = (\eta_n, \eta_{n+1}, \dots, \eta_{n+s-1}), \quad 0 \leq n < T.$$

$$S(\eta_n) = \max\left\{s : \langle \mathbf{u}_n - \mathbf{u}_0,\ 1 \leq n < T\rangle = \mathbb{F}_p^s\right\}$$

$s = 2$:



$s = 3$:



Niederreiter/W., 2002: $L(\eta_n) = S(\eta_n)$ or $= S(\eta_n) + 1$

Dorfer/W., 2003: Lattice test for parts of the period, $S(\eta_n, N)$
We have either

$$S(\eta_n, N) = \min(L(\eta_n, N), N + 1 - L(\eta_n, N))$$

or

$$S(\eta_n, N) = \min(L(\eta_n, N), N + 1 - L(\eta_n, N)) - 1.$$

Relation to Information and Coding Theory

The Kolmogorov complexity of a sequence over $\mathbb{F}$ is the length of a shortest Turing machine that generates it.

Beth/Dai, 1989: $\mathbb{F} = \mathbb{F}_2$: Linear complexity and Kolmogorov complexity are the same for almost all binary sequences.

In general there is no algorithm for calculating the Kolmogorov complexity.
In contrast we have the Berlekamp-Massey Algorithm for calculating the linear complexity.
This algorithm stems from coding theory.

A Consequence of the Berlekamp-Massey Algorithm

Theorem

If $L(s_n, N) > N/2$, *then we have*

$$L(s_n, N+1) = L(s_n, N).$$

If $L(s_n, N) \leq N/2$, *then we have either*

$$L(s_n, N+1) = L(s_n, N)$$

or

$$L(s_n, N+1) = N + 1 - L(s_n, N).$$

The Expected Value

$\mathbb{F} = \mathbb{F}_q$

Theorem

The expected value for $L(s_n, N)$ *is*

$$\begin{cases} \frac{N}{2} + \frac{q}{(q+1)^2} - q^{-N}\frac{N(q+1)+q}{(q+1)^2} & \text{for even } N, \\ \frac{N}{2} + \frac{q^2+1}{2(q+1)^2} - q^{-N}\frac{N(q+1)+q}{(q+1)^2} & \text{for odd } N. \end{cases}$$

Lower Bounds

In case of a p-periodic sequence (ξ_n) over $\mathbb{F}_p$, where p is a prime, linear complexity is related to the degree of the polynomial $g(X) \in \mathbb{F}_p[X]$ representing the sequence (ξ_n), i.e., $g(X)$ is the unique polynomial which satisfies $\deg g \leq p-1$ and

$$\xi_n = g(n), \quad 0 \leq n \leq p-1.$$

These sequences are called *explicit nonlinear congruential generators* and we have

$$L(\xi_n) = \deg g + 1.$$

High linear complexity but low Nth linear complexity

Example:

$$\xi_n = 1 - (n+1)^{p-1}, \quad 0 \le n \le p-1$$

$$(\xi_0, \xi_1, \ldots, \xi_{p-2}, \xi_{p-1}) = (0, 0, \ldots, 0, 1)$$

$$L(\xi_n) = p$$

$$L(\xi_n, N) = 0, \quad 1 \le N \le p-1$$

highly predictable

The *explicit inversive congruential generator* (z_n) is produced by the relation

$$z_n = (an+b)^{p-2}, \quad n = 0, \ldots, p-1, \quad z_{n+p} = z_n, \; n \geq 0,$$

with $a, b \in \mathbb{F}_p$, $a \neq 0$, and $p \geq 5$. We have

$$L(z_n, N) \geq \begin{cases} (N-1)/3, & 1 \leq N \leq (3p-7)/2, \\ N-p+2, & (3p-5)/2 \leq N \leq 2p-3, \\ p-1, & N \geq 2p-2. \end{cases}$$

$c_L = -1$, $N \le p$

$$\sum_{l=0}^{L} c_l z_{n+l} = 0, \quad 0 \le n \le N - L - 1$$

$a(n+l) + b \ne 0$, $0 \le l \le L$:

$$\sum_{l=0}^{L} c_l (a(n+l)+b)^{-1} = 0$$

$$F(X) = \sum_{l=0}^{L} c_l \prod_{\substack{j=0 \\ j \ne L}}^{L} (a(X+j)+b)$$

has at least $N - L - (L+1)$ zeros and degree at most L.

$$F(-a^{-1}b - L) = c_L \prod_{j=0}^{L-1} (a(j-L)) \ne 0$$

$N - 2L - 1 \le L$ and thus $L \ge (N-1)/3$

$z_n = (an + b)^{p-2}$ is still highly predictable since inversion is cheap and $a = z_{n+1}^{-1} - z_n^{-1}$ for all but two n.

Open problem: Define a modified linear complexity with inversions and analyze it.

Let $p > 2$ be a prime. The *Legendre-sequence* (l_n) is defined by

$$l_n = \begin{cases} 1, & \left(\frac{n}{p}\right) = -1, \\ 0, & \text{otherwise}, \end{cases} \qquad n \geq 0,$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre-symbol.

Theorem

The linear complexity of the Legendre sequence is

$$L(l_n) = \begin{cases} (p-1)/2, & p \equiv 1 \bmod 8, \\ p, & p \equiv 3 \bmod 8, \\ p-1, & p \equiv 5 \bmod 8, \\ (p+1)/2, & p \equiv 7 \bmod 8. \end{cases}$$

Theorem

The Nth linear complexity of the Legendre sequence satisfies

$$L(l_n, N) > \frac{\min\{N, p\}}{1 + p^{1/2}(1 + \log p)} - 1, \quad N \geq 1.$$

Weil:

Let $f(X) \in \mathbb{F}_p[X]$ a (monic) polynomial which is not a square and $a \in \mathbb{F}_p^*$ then we have

$$\left| \sum_{x \in \mathbb{F}_p} \left(\frac{af(x)}{p} \right) \right| \leq (\deg(f) - 1)p^{1/2}.$$

$$\sum_{k=0}^{L} c_k l_{n+k} = 0 \in \mathbb{F}_2, \quad 0 \leq n \leq N - L - 1, \quad (c_L = 1)$$

$$(-1)^{l_n} = \left(\frac{n}{p} \right), \quad n \neq 0,$$

$$(-1)^{\sum_{k=0}^{L} c_k l_{n+k}} = \left(\frac{\prod_{l=0}^{L} (n+l)^{c_l}}{p} \right) = 1$$

for at least $\min\{N, p\} - (L+1)$ different n

Summing over n:

$$\min\{N, p\} - (L+1) \le \sum_{n=0}^{N-1} \left(\frac{\prod_{l=0}^{L} (n+l)^{c_l}}{p} \right)$$

$$< (L+1)p^{1/2}(1 + \log p)$$

Relation to Wireless Communication

The correlation measure of order k of a binary sequence (s_n) is introduced as

$$C_k(s_n) = \max_{M,D} \left| \sum_{n=1}^{M} (-1)^{s_{n+d_1}} \cdots (-1)^{s_{n+d_k}} \right|, \quad k \geq 1,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ with non-negative integers $d_1 < d_2 < \cdots < d_k$ and M such that $M - 1 + d_k \leq T - 1$.

$$L(s_n, N) \geq N - \max_{1 \leq k \leq L(s_n,N)+1} C_k(s_n), \quad 2 \leq N \leq t - 1.$$

Examples.

a) $b_n = 0$, $0 \leq n \leq t-2$, $b_{t-1} = 1$
$L(b_n) = t$, one change $L(b'_n) = 0$

b) $c_{n+4} = c_n$, $n \geq 0$, with $c_0 = c_1 = c_2 = 1$, $c_3 = 0$
over $\mathbb{F}_2$: $L(c_n) = 4$
over $\mathbb{F}_3$: $L(c_n) = 3$ since $c_{n+3} = 2c_{n+2} + 2c_{n+1} + 2c_n$, $n \geq 0$

Desirable:

1. high linear complexity even if we change a few elements
2. high linear complexity over different fields

Let (s_n) be a sequence over $\mathbb{F}$, with period t. The *k-error linear complexity* $L_k(s_n)$ of (s_n) is defined as

$$L_k(s_n) = \min_{(y_n)} L(y_n),$$

where the minimum is taken over all t-periodic sequences (y_n) over $\mathbb{F}$, for which the Hamming distance of the vectors $(s_0, s_1, \ldots, s_{t-1})$ and $(y_0, y_1, \ldots, y_{t-1})$ is at most k.

Theorem

Let $L_k(l_n)$ denote the k-error linear complexity over $\mathbb{F}_p$ of the Legendre sequence (l_n). Then,

$$L_k(l_n) = \begin{cases} p, & k = 0, \\ (p+1)/2, & 1 \le k \le (p-3)/2, \\ 0, & k \ge (p-1)/2. \end{cases}$$

$$l_n = 2^{-1}(n^{p-1} - n^{(p-1)/2}) \in \mathbb{F}_p, \quad n \geq 0$$

$$l_n = 2^{-1}(1 - n^{(p-1)/2}) =: h(n), \quad n \neq 0$$

$l_n = f(n)$ implies $L(l_n) = \deg(f) + 1$
Let (y_n) be obtained from (l_n) by at most k changes.
Case I: $(y_n) = (l_n)$: $L(y_n) = p$
Case II: $(y_n) = h(n)$: $L(y_n) = (p+1)/2$
Case III: $y_n = g(n)$, $g(X) \neq h(X)$

$$\deg(g - h) \geq p - k - 1 \geq (p+1)/2 \quad \text{if } k \leq (p-3)/2$$

Shparlinski/W.,2006: linear complexity over $\mathbb{F}_k$, k prime:

$$L(l_n) \geq \frac{1}{2\log k}\min\left\{\frac{p}{p^{1/2}\log p+2}-1, 2k-1\right\}.$$

Open Problem

Find more sequences with high (Nth, k-error) linear complexity.

For example, study recursive sequences.

Linear Pseudorandom Number Generators

$\mathbb{F}_q$ finite field of q elements, $a, b, x_0 \in \mathbb{F}_q$, $a \neq 0$

$$\boxed{x_{n+1} = ax_n + b, \quad n \geq 0}$$

$q = p$ prime, $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$: $y_n = x_n/p \in [0, 1)$, $n \geq 0$

Nice features:
– long period can be easily obtained
– uniform distribution in dimension 1
flaws:
– predictable ($L(x_n) \leq 2$)
– coarse structure

Nonlinear Pseudorandom Numbers

$f \in \mathbb{F}_q[X]$, $2 \leq \deg(f) \leq q-1$, $x_0 \in \mathbb{F}_q$

$$\boxed{x_{n+1} = f(x_n), \quad n \geq 0}$$

(purely) periodic with period $t \leq q$
$q = p$ prime: $y_n = x_n/p \in [0,1)$

Lower Bound on the Linear Complexity Profile

Gutierrez/Shparlinski/W., 2003:
The linear complexity profile of a nonlinear sequence (x_n) defined by

$$x_{n+1} = f(x_n), \quad n = 0, 1, \ldots,$$

with a polynomial $f \in \mathbb{F}_q[X]$ of degree $d \geq 2$, purely periodic with period t, satisfies

$$L(x_n, N) \geq \min \left\{ \lceil \log_d(N - \lfloor \log_d N \rfloor) \rceil, \lceil \log_d t \rceil \right\}.$$

Proof.
$F_0(X) := X, \quad F_i(X) := F_{i-1}(f(X)),\ i \geq 1$
$\deg(F_i) = d^i$, $x_{n+j} = F_j(x_n)$
$x_{n+L} = a_{L-1}x_{n+L-1} + \ldots + a_0 x_n$,
$0 \leq n \leq N - L - 1$

$$F(X) := -F_L(X) + a_{L-1}F_{L-1}(X) + \ldots + a_0 F_0(X)$$

has degree d^L and at least $\min\{N - L, t\}$ zeros, namely, x_n with $0 \leq n \leq \min\{N - L - 1, t - 1\}$.

$$d^L \geq \min\{N - L, t\}$$

□

Inversive Generators

$a, b, y_0 \in \mathbb{F}_q$, $a \neq 0$

$$y_{n+1} = ay_n^{q-2} + b = \begin{cases} ay_n^{-1} + b, & y_n \neq 0, \\ b, & y_n = 0. \end{cases}$$

Gutierrez/Shparlinski/W., 2003:

$$L(y_n, N) \geq \min\left\{ \left\lceil \frac{N-1}{3} \right\rceil, \left\lceil \frac{t-1}{2} \right\rceil \right\}$$

Reason for better result:
$f(X) = \frac{bX+a}{X}$, $F_j(X) = \frac{a_jX+b_j}{c_jX+d}$

Dickson and Power Generator

The Dickson polynomial $D_e(X, a) \in \mathbb{F}_q[X]$ is defined by the following recurrence relation

$$D_e(X, a) = XD_{e-1}(X, a) - aD_{e-2}(X, a), \quad e = 2, 3, \ldots,$$

with initial values

$$D_0(X, a) = 2, \quad D_1(X, a) = X,$$

where $a \in \mathbb{F}_q$. Obviously, the degree of D_e is e. Moreover, if $a \in \{0, 1\}$ then we have $D_e(D_f(X, a), a) = D_{ef}(X, a)$.

$a = 0$:
$D_e(X, 0) = X^e$, $e \geq 2$

$$p_{n+1} = p_n^e, \quad n \geq 0$$

power generator
Griffin/Shparlinski, 2000: ($q = p$ prime)

$$L(p_n, N) \geq \min\left\{\frac{N^2}{4(p-1)}, \frac{t^2}{p-1}\right\}, \quad N \geq 1.$$

Reason for better result: $F_k(X) = X^{e^k \bmod p-1}$

$a = 1$:

$$D_e(x + x^{-1}, 1) = x^e + x^{-e}, \quad x \in \mathbb{F}_{q^2}$$

$$u_{n+1} = D_e(u_n, 1), \quad n \geq 0,$$

with some initial value u_0 and $e \geq 2$.
Dickson generator

Aly/W., 2006:

$$L(u_n, N) \geq \frac{\min\{N^2, 4t^2\}}{16(p+1)} - (p+1)^{1/2}$$

Redéi generator

Suppose that

$$r(X) = X^2 - \alpha X - \beta \in \mathbb{F}_p[X]$$

is an irreducible quadratic polynomial with the two different roots ξ and $\zeta = \xi^p$ in $\mathbb{F}_{p^2}$. Then any polynomial $b(X) \in \mathbb{F}_{p^2}[X]$ can uniquely be written in the form $b(X) = g(X) + h(X)\xi$ with $g(X), h(X) \in \mathbb{F}_p[X]$. We consider the elements

$$(X + \xi)^e = g_e(X) + h_e(X)\xi.$$

e is the degree of the polynomial $g_e(X)$, and $h_e(X)$ has degree at most $e - 1$. The Rédei function $f_e(X)$ of degree e is then given by

$$f_e(X) = \frac{g_e(X)}{h_e(X)}.$$

$$u_{n+1} = f_e(u_n), \quad n \geq 0,$$

with a Rédei permutation $f_e(X)$ and some initial element $u_0 \in \mathbb{F}_p$.

Meidl/W., 2007:

$$L(u_n, N) \geq \frac{\min\{N^2, 4t^2\}}{20(p+1)^{3/2}}, \quad N \geq 2.$$

p-Weight Degree

n nonnegative integer
p-weight of n:

$$\sigma_p\left(\sum_{i=0}^{l} n_i p^i\right) = \sum_{i=0}^{l} n_i, \quad 0 \le n_i < p.$$

$0 \le e_1 < e_2 < \cdots < e_l$ integers, $q = p^r$, $f(X) = \sum_{i=1}^{l} \gamma_i X^{e_i} \in \mathbb{F}_q[X]$ nonzero polynomial over $\mathbb{F}_q$ with $\gamma_i \neq 0$, $i = 1, \ldots, l$
p-weight degree of f:

$$\mathrm{w}_p(f) = \max\{\sigma_p(e_i) : 1 \le i \le l\}.$$

$\mathrm{w}_p(f) \le \deg(f)$

If $g(X) \in \mathbb{F}_q[X]$ and $\{\beta_1, \ldots, \beta_r\}$ is a fixed ordered $\mathbb{F}_p$-basis of $\mathbb{F}_q$, we define

$$G(X_1, \ldots, X_r) = \mathrm{Tr}(g(X_1\beta_1 + \ldots + X_r\beta_r)),$$

where $\mathrm{Tr}(X) = X + X^p + \ldots + X^{p^{r-1}}$ is the absolute trace function of $\mathbb{F}_q$. Then the transformed polynomial $G_R(X_1, \ldots, X_r)$ of $g(X)$ is the unique polynomial with all local degrees smaller than p such that

$$G_R(X_1, \ldots, X_r) \equiv G(X_1, \ldots, X_r) \mod (X_1^p - X_1, \ldots, X_r^p - X_r).$$

The interest of this construction relies on the fact that, under certain assumptions, the total degree of $G_R(X_1, \ldots, X_r)$ coincides with the p-weight degree of $g(X)$.

$$f(X) = \alpha X^d + \tilde{f}(X) \in \mathbb{F}_q[X] \text{ with } \alpha \neq 0, \quad \mathrm{w}_p(\tilde{f}) < \sigma_p(d). \qquad (1)$$

If the sequence (x_n) given by $x_{n+1} = f(x_n)$, $n \geq 0$, with a polynomial $f(X) \in \mathbb{F}_q[X]$ of the form (1) satisfying

$$\gcd\left(d, \frac{q-1}{p-1}\right) \leq \sigma_p(d)^{r/2},$$

with p-weight degree $w = \sigma_p(d) > 1$, is purely periodic with period t, then for $N \geq 1$,

$$L(x_n, N) \geq \frac{\min\left\{\log(N/p^{r-1} - \log(N/p^{r-1})/\log w), \log(t/p^{r-1})\right\}}{\log w}.$$

(Ibeas/W., 2010)

Polynomial Systems

Let $\{F_1, \ldots, F_r\}$ be a system of $r \geq 2$ polynomials $F_i \in \mathbb{F}_q[X_i, \ldots, X_m],\ i = 1, \ldots, r$, defined in the following way:

$$\begin{aligned}
F_1(X_1, \ldots, X_r) &= X_1 G_1(X_2, \ldots, X_r) + H_1(X_2, \ldots, X_r),\\
F_2(X_1, \ldots, X_r) &= X_2 G_2(X_3, \ldots, X_r) + H_2(X_3, \ldots, X_r),\\
&\ldots\\
F_{r-1}(X_1, \ldots, X_r) &= X_{r-1} G_{r-1}(X_r) + H_{r-1}(X_r),\\
F_r(X_1, \ldots, X_r) &= g_r X_r + h_r.
\end{aligned}$$

Using the following vector notation

$$\vec{F} = (F_1(X_1, \ldots, X_r), \ldots, F_r(X_1, \ldots, X_r)),$$

we define the following vector sequence

$$\vec{w}_{n+1} = \vec{F}(\vec{w}_n), \quad n = 0, 1, \ldots.$$

Identifying the r dimensional vectors over $\mathbb{F}_q$ with elements of $\mathbb{F}_{q^r}$ we get

$$L(\vec{w}_n, N) \gg \frac{N^{1/(r-1)}}{q}, \qquad 1 \le N \le t.$$

(Ostafe, Shparlinski, W., 2010)

Open Problem

Find more good nonlinear generators.

Thank you for your attention.