Pseudorandom binary sequences: Quality measures and number theoretic constructions

Arne Winterhof (RICAM, Austrian Academy of Sciences, Linz)

Let $\mathcal{S} = (s_n)_{n=0}^{\infty}$, $s_n \in \mathbb{F}_2$, $n = 0, 1, \ldots$, be a binary sequence. For cryptographic applications the unpredictability of such a sequence is crucial. There are several measures of pseudorandomness which can be used to sieve bad sequences including

1. linear complexity,
2. correlation measure of order k,
3. maximum-order complexity,
4. expansion complexity

defined below. These measures are partly not independent and partly complete each other. First we study their relations. Then we analyze these measures for some sequences including the Legendre sequence, the Sidelnikov sequence, the Thue-Morse sequence and the subsequence of the Thue-Morse sequence along squares.

Each talk will focus on one of these four measures of pseudorandomness.

1. Linear complexity

The *Nth linear complexity* $L(\mathcal{S}, N)$ of $\mathcal{S}$ is the smallest positive integer L such that there are constants $c_0, \ldots, c_{L-1} \in \mathbb{F}_2$ with

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

The *linear complexity* $L(\mathcal{S})$ of $\mathcal{S}$ is

$$L(\mathcal{S}) = \sup_{N \in \mathbb{N}} L(\mathcal{S}, N).$$

For a prime $p > 2$ the *Legendre sequence* $\mathcal{L} = (\ell_n)_{n=0}^{\infty}$ is the p-periodic sequence defined by

$$\ell_n = \begin{cases} 1, & n \text{ is a quadratic non-residue modulo } p, \\ 0, & \text{otherwise.} \end{cases}$$

We will determine the values of $L(\mathcal{L})$ first found by R. Turyn. In particular, we have

$$L(\mathcal{L}) \geq (p-1)/2.$$

Dealing with sequences of even period is in general more complicated than with odd period. We explain some difficulties studying the example of the $(q-1)$-periodic *Sidelnikov sequence* $\mathcal{D} = (d_n)_{n=0}^{\infty}$ defined by

$$d_n = \begin{cases} 1, & g^n + 1 \text{ is a quadratic non-residue modulo } p, \\ 0, & \text{otherwise,} \end{cases}$$

where q is the power of an odd prime and g a primitive element of the finite field $\mathbb{F}_q$.

Finally, we study the expected value of the Nth linear complexity of a truly random sequence which is close to $N/2$. This result is obtained by a careful study of the *Berlekamp-Massey algorithm*.

2. Correlation measure

The *Nth correlation measure of order k* of $\mathcal{S}$ introduced by Mauduit and Sárközy is

$$C_k(\mathcal{S}, N) = \max_{M,D} \left| \sum_{n=0}^{M-1} (-1)^{s_{n+d_1}} \cdots (-1)^{s_{n+d_k}} \right|, \quad k \geq 1,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ with integers satisfying $0 \leq d_1 < d_2 < \cdots < d_k$ and $1 \leq M \leq N - d_k$.

First we will prove an upper bound for the correlation measure of order k for the Legendre sequence which is close to the expected value of a truly random sequence.

Then we prove a relation between correlation measures and Nth linear complexity. Roughly speaking, this shows that the correlation measure is a finer quality measure for cryptographic sequences.

Finally, we apply this relation to get a lower bound on the Nth linear complexity of the Legendre sequence which is non-trivial for N at least of order of magnitude $p^{1/2+\varepsilon}$.

3. Maximum order complexity

The *Nth maximum order complexity* $M(\mathcal{S}, N)$ is the smallest positive integer M with

$$s_{n+M} = f(s_{n+M-1}, \ldots, s_n), \quad 0 \leq n \leq N - M - 1,$$

for some mapping $f : \mathbb{F}_2^M \mapsto \mathbb{F}_2$.

Obviously, we have

$$M(\mathcal{S}, N) \leq L(\mathcal{S}, N).$$

We will see that correlation measures and maximum order complexity are also related and use this relation to prove a moderate lower bound on the maximum order complexity of the Legendre sequence.

Then we study the *Thue-Morse sequence* $\mathcal{T} = (t_n)_{n=0}^{\infty}$ defined by

$$t_0 = 0, \quad t_{2n} = t_n, \quad t_{2n+1} = 1 - t_n, \quad n = 0, 1, \ldots$$

It turns out that $M(\mathcal{T}, N)$ is of order of magnitude N. However, this implies that the correlation measure $C_2(\mathcal{T}, N)$ of order 2 is also of order of magnitude N and concerning this measure the Thue-Morse sequence does not behave like a random sequence.

However, taking a certain subsequence of the Thue-Morse sequence may destroy some undesirable structure but still keeps enough structure of the original

sequence which can be used to study the maximum-order complexity. In particular, for the *Thue-Morse sequence along squares* $\mathcal{Q} = (t_{n^2})_{n=0}^{\infty}$ we show that $M(\mathcal{Q}, N)$ is at least of order of magnitude $N^{1/2}$.

4. Expansion complexity

Let

$$G(x) = \sum_{n=0}^{\infty} s_n x^n$$

be the *generating function* of $\mathcal{S}$. Then the *Nth expansion complexity* $E(s_n, N)$ is the least total degree of a nonzero polynomial $h(x, y) \in \mathbb{F}_2[x, y]$ with

$$h(x, G(x)) \equiv 0 \bmod x^N.$$

We can show

$$E(\mathcal{S}, N) \le L(\mathcal{S}, N) + 1.$$

Indeed, the expansion complexity can be much smaller than the linear complexity. For example, the Thue-Morse sequence has linear complexity of order of magnitude N but its generating function $G(x)$ satisfies

$$(1 + x)^3 G(x)^2 + (1 + x)^2 G(x) + x = 0$$

and thus we have

$$E(\mathcal{T}, N) \le 5.$$

On the other hand, the Thue-Morse sequence along the squares satisfies

$$\lim_{N \to \infty} E(\mathcal{Q}, N) \to \infty.$$