Randomness and Complexity of Sequences over Finite Fields

Harald Niederreiter, FAMS

RICAM Linz and University of Salzburg (Austria)

— Introduction

— A hierarchy of complexities

— Complexity and random sequences

— ∞-distribution

— Complexity and ∞-distribution

— Expansion complexity

Introduction

Sequences of random (or pseudorandom) elements are needed for various applications (cryptography, simulation methods, probabilistic algorithms,...). To assess randomness, we can use complexity-theoretic properties and statistical properties of sequences.

We focus on sequences over a finite field $\mathbb{F}_q$ with q an arbitrary prime power.

A hierarchy of complexities

We first consider finite sequences. Let S_N be a finite sequence over $\mathbb{F}_q$ of length N.

Def. 1. The linear complexity $L(S_N)$ of S_N is the least order of a linear recurrence relation over $\mathbb{F}_q$ that generates S_N (with $L(S_N) = 0$ if S_N is the zero sequence).

Equivalently, $L(S_N)$ is the length of the shortest linear feedback shift register (FSR) that generates S_N.

Def. 2. The quadratic complexity $Q(S_N)$ of S_N is the length of the shortest FSR with feedback function of degree ≤ 2 that generates S_N.

We always have $0 \leq Q(S_N) \leq L(S_N) \leq N$. Similarly, we can define the cubic complexity, quartic complexity,.... At the end of this chain of complexities is the following one.

Def. 3 (Jansen, 1989). The maximum order complexity $M(S_N)$ of S_N is the length of the shortest (arbitrary) FSR that generates S_N.

$M(S_N)$ is a lower bound for all polynomial complexities.

$M(S_N)$ can be computed by Blumer's algorithm in graph theory, by reformulating the problem of computing $M(S_N)$ as a problem for directed acyclic graphs.

Recall that $L(S_N)$ can be efficiently computed by the Berlekamp-Massey algorithm.

In a different vein, we have the Kolmogorov complexity which is important in theoretical computer science.

Def. 4 (Kolmogorov, 1965). The Kolmogorov complexity $K(S_N)$ of S_N is the length of the shortest Turing machine program (in an alphabet of size q, say) that generates S_N.

We always have $K(S_N) \leq N + c$ with an absolute constant c (just list the terms of S_N).
Actually, most results hold for the more refined self-delimiting Kolmogorov complexity. These two Kolmogorov complexities differ at most by an additive term of logarithmic order.

For an infinite sequence S over $\mathbb{F}_q$ we define complexity profiles. Let S_N denote the finite sequence consisting of the first N terms of S. Write $L_N(S) = L(S_N)$, etc.

Def. 5. The sequence $L_1(S), L_2(S), \ldots$ is called the linear complexity profile of S.

Similarly for the other complexities. Each complexity profile is a nondecreasing sequence of nonnegative integers.

Complexity and random sequences

Let $\mathbb{F}_q^\infty$ denote the sequence space over $\mathbb{F}_q$ and let μ_q be the natural probability measure on $\mathbb{F}_q^\infty$, i.e., the complete product measure of the uniform probability measure on $\mathbb{F}_q$ (the latter measure assigns measure q^{-1} to each element of $\mathbb{F}_q$).

We say that a property of sequences S over $\mathbb{F}_q$ is satisfied μ_q-almost everywhere if the property holds for all $S \in \mathcal{R} \subseteq \mathbb{F}_q^\infty$ with $\mu_q(\mathcal{R}) = 1$. Such properties can be viewed as typical properties of random sequences.

The behavior of the linear complexity profile of random sequences is well understood.

Theorem 1 (H.N., 1988). We have μ_q-almost everywhere

$$\lim_{N \to \infty} \frac{L_N(S)}{N} = \frac{1}{2}. \tag{1}$$

More precisely, we have μ_q-almost everywhere

$$\limsup_{N \to \infty} \frac{|L_N(S) - N/2|}{\log_q N} = \frac{1}{2}.$$

Problem 1. Determine the behavior of the quadratic (cubic,...) complexity profile of random sequences.

Theorem 2 (Jansen, 1989). For the expected value of the maximum order complexity we have

$$\lim_{N\to\infty} \frac{E(M_N(S))}{\log_q N} = 2.$$

Problem 2. Determine whether $\lim_{N\to\infty} M_N(S)/\log_q N$ exists for random sequences, i.e., μ_q-almost everywhere.

Theorem 3 (Martin-Löf, 1966). For every $\varepsilon > 0$ we have μ_q-almost everywhere

$$K_N(S) \geq N - \log_q N - (1+\varepsilon)\log_q \log N$$

for all sufficiently large N. Moreover, for every sequence S there are infinitely many N for which

$$K_N(S) \leq N - \log_q N.$$

∞-distribution

This concept was popularized by the book of Knuth, The Art of Computer Programming, Vol. 2, at least in the binary case.

Let S be the sequence $s_1, s_2, \ldots$ over $\mathbb{F}_q$. For integers $k \geq 1$ and $n \geq 1$, write

$$\mathbf{s}_n^{(k)} = (s_n, s_{n+1}, \ldots, s_{n+k-1}) \in \mathbb{F}_q^k.$$

For an integer $N \geq 1$ and a given block $\mathbf{b} = (b_0, b_1, \ldots, b_{k-1}) \in \mathbb{F}_q^k$, let

$$A(\mathbf{b}, S; N) = \#\{1 \leq n \leq N : \mathbf{s}_n^{(k)} = \mathbf{b}\}.$$

Def. 6. For an integer $k \geq 1$, the sequence S over $\mathbb{F}_q$ is *k-distributed* in $\mathbb{F}_q$ if

$$\lim_{N \to \infty} \frac{A(\mathbf{b}, S; N)}{N} = \frac{1}{q^k} \qquad \text{for all } \mathbf{b} \in \mathbb{F}_q^k.$$

The sequence S over $\mathbb{F}_q$ is *∞-distributed* (or *completely uniformly distributed*) in $\mathbb{F}_q$ if it is k-distributed in $\mathbb{F}_q$ for all $k \geq 1$.

Theorem 4. Sequences over $\mathbb{F}_q$ are ∞-distributed in $\mathbb{F}_q$ μ_q-almost everywhere.

Complexity and ∞-distribution

∞-distribution is clearly a statistical randomness property. Does this property imply, or is it implied by, complexity-theoretic randomness properties?

Theorem 5 (Martin-Löf, 1971). If $K_N(S) \geq N - c$ for some constant c and infinitely many N, then S is ∞-distributed.

Thus, Kolmogorov complexity is a sufficiently strong concept to yield ∞-distribution.

How about linear complexity? Does the limit relation (1) imply ∞-distribution? Answer: **no**.

Theorem 6 (H.N., 2012). We can construct a sequence S over $\mathbb{F}_q$ which satisfies (1), but is not 1-distributed in $\mathbb{F}_q$.

Proof. Consider the sequence S whose generating function (in the variable x^{-1}) is the power series $\sigma \in \mathbb{F}_q[[x^{-1}]]$ with continued fraction expansion

$$\sigma = 1/(x + 1/(x + \cdots)),$$

where all partial quotients are equal to x.

Problem 3. If $M_N(S) \approx 2\log_q N$, check whether this implies that S is ∞-distributed.

For Theorem 6 there is also a result in the opposite direction.

Theorem 7 (H.N., 2012). We can construct a sequence S over $\mathbb{F}_q$ which is ∞-distributed, but for which $L_N(S) = O((\log N)^2)$.

Problem 4. If S is ∞-distributed, then it is trivial that

$$\lim_{N\to\infty} L_N(S) = \infty.$$

Determine the minimal growth rate of $L_N(S)$ as $N \longrightarrow \infty$. Presumably it is smaller than $(\log N)^2$ in Theorem 7.

Expansion complexity

Let S be the sequence $s_1, s_2, \ldots$ over $\mathbb{F}_q$. Let its generating function be

$$\gamma(x) = \sum_{i=1}^{\infty} s_i x^{i-1} \in \mathbb{F}_q[[x]].$$

Def. 7 (Diem, preprint 2011). The expansion complexity $E_N(S)$ is the least degree of a nonzero polynomial $h(x, y) \in \mathbb{F}_q[x, y]$ with

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

Remark. If $s_i = 0$ for $1 \leq i \leq N$, define $E_N(S) = 0$. Then $0 \leq E_N(S) \leq N$ for any sequence S.

We can also introduce the expansion complexity profile $E_1(S), E_2(S), \ldots$ of S.

Problem 5. Determine the relationship between the maximum order complexity $M_N(S)$ and the expansion complexity $E_N(S)$.

Problem 6. Determine the behavior of the expansion complexity profile of random sequences. A partial result is that μ_q-almost everywhere $E_N(S)$ grows at least at the rate $N^{1/2}$.