THE PROBABILISTIC THEORY OF LINEAR COMPLEXITY

Harald Niederreiter

Mathematical Institute, Austrian Academy of Sciences
Dr.-Ignaz-Seipel-Platz 2
A-1010 Vienna, Austria

1. INTRODUCTION

Linear complexity is a widely accepted measure for unpredictability and randomness of keystream sequences in the context of stream ciphers (see Rueppel [10], [11, Ch. 4]). In this paper we develop a detailed probabilistic theory of linear complexity and linear complexity profiles for sequences of elements of a finite field. The basic tools are the connection between linear complexity and continued fractions for formal Laurent series established in Niederreiter [8] as well as techniques from probability theory and the theory of dynamical systems.

In practice, keystream sequences are sequences of bits, and we identify bits with elements of the binary field F_2. However, the methods of this paper work for arbitrary finite fields. We denote by F_q the finite field with q elements, where q is an arbitrary prime power. A sequence $s_1, s_2, \ldots$ of elements of F_q is called a kth-order (linear feedback) <u>shift register sequence</u> if there exist constant coefficients $a_k, \ldots, a_0 \in F_q$ with $a_k \neq 0$ such that

$$a_k s_{i+k} + \cdots + a_1 s_{i+1} + a_0 s_i = 0 \quad \text{for} \quad i = 1,2,\ldots . \tag{1}$$

The zero sequence $0,0,\ldots$ is viewed as a shift register sequence of order 0. A kth-order shift register sequence is uniquely determined by the recursion (1) and by the initial values $s_1, s_2, \ldots, s_k$.

<u>Definition 1.</u> Let S be an arbitrary sequence $s_1, s_2, \ldots$ of elements of F_q and let n be a positive integer. Then the <u>linear complexity</u> $L_n(S)$ is defined as the least k such that $s_1, s_2, \ldots, s_n$ form the first n terms of a kth-order shift register sequence.

<u>Definition 2.</u> With the notation of Definition 1, the sequence $L_1(S), L_2(S), \ldots$ is called the <u>linear complexity profile</u> of S.

It is clear that $0 \leq L_n(S) \leq n$ and $L_n(S) \leq L_{n+1}(S)$ for all n and S. Therefore the linear complexity profile is a nondecreasing sequence of nonnegative integers. Rueppel [10], [11, Ch. 4] proposed the linear complexity profile as a test

C.G. Guenther (Ed.): Advances in Cryptology - EUROCRYPT '88, LNCS 330, pp. 191-209, 1988.

for randomness and set up the following stochastic model. Let n be fixed and consider $L_n(S)$ for random sequences of bits. Since $L_n(S)$ just depends on the first n terms of S, it suffices to consider the linear complexity for all choices of $s_1, s_2, \ldots, s_n$ from F_2. Then the linear complexity can be viewed as a random variable on F_2^n, where each string $s_1, s_2, \ldots, s_n$ is equiprobable. It turns out that the expected value of this random variable is $\frac{n}{2} + c_n$ with $0 \leq c_n \leq \frac{5}{18}$ and its variance is roughly $\frac{86}{81}$. This suggests that $L_n(S)$ should be close to $\frac{n}{2}$ for a random sequence of bits.

To arrive at a statistically meaningful use of the linear complexity profile, the following question has to be answered: for a randomly chosen and then fixed sequence S, what is the behavior of $L_n(S)$ as n varies? We settle this question for sequences S of elements of F_q and also discuss related questions. The necessary background and basic results on continued fractions and dynamical systems are established in Sections 2 and 3. These results yield, first of all, the probabilistic limit theorems for continued fractions in Section 4. Exploiting the connection between continued fractions and linear complexity, we deduce the probabilistic limit theorems for linear complexity in Section 5. These limit theorems describe the asymptotic behavior of $L_n(S)$ as $n \to \infty$ and the deviations from the asymptotic behavior for random S. In Section 6 we study frequency distributions associated with the linear complexity for random S. The detailed information on the behavior of $L_n(S)$ for random S is used in Section 7 to set up new types of randomness tests for keystream sequences.

2. CONTINUED FRACTIONS

We use the approach in Niederreiter [8] which is based on identifying a sequence S of elements $s_1, s_2, \ldots$ of F_q with its generating function $S = \sum_{i=1}^{\infty} s_i x^{-i}$. As in [8] we view S as an element of the field $G = F_q((x^{-1}))$ of formal Laurent series in x^{-1} over F_q. For $S \in G$ let $\mathrm{Pol}(S)$ be its <u>polynomial part</u> and $\mathrm{Fr}(S) = S - \mathrm{Pol}(S)$ its <u>fractional part</u>. Thus $\mathrm{Fr}(S)$ is the part of S containing the negative powers of x. We introduce the <u>valuation</u> v on G which extends the degree function on the polynomial ring $F_q[x]$ as follows. For $S \in G$, $S \neq 0$, we put

$$v(S) = -r \quad \text{if} \quad S = \sum_{i=r}^{\infty} s_i x^{-i} \quad \text{and} \quad s_r \neq 0.$$

For $S = 0$ we put $v(S) = -\infty$. We have the following properties for $S_1, S_2 \in G$:

$$v(S_1 S_2) = v(S_1) + v(S_2),$$
$$v(S_1 + S_2) \leq \max(v(S_1), v(S_2)),$$
$$v(S_1 + S_2) = \max(v(S_1), v(S_2)) \quad \text{if} \quad v(S_1) \neq v(S_2).$$

For $p_1,p_2 \in F_q[x], p_2 \neq 0$, we have $v(p_1/p_2) = \deg(p_1) - \deg(p_2)$.

Let H be the set of all generating functions, thus $H = \{S \in G: v(S) < 0\}$. Every $S \in H$ has a unique continued fraction expansion of the form

$$S = 0 + 1/(A_1(S) + 1/(A_2(S) + \ldots)) = :[A_1(S),A_2(S),\ldots],$$

where $A_j(S) \in F_q[x]$ and $\deg(A_j(S)) \geq 1$ for $j \geq 1$. This expansion is finite for rational S and infinite for irrational S. The polynomials $A_j(S)$ are obtained recursively by the following algorithm:

$$A_0(S) = 0, \qquad B_0(S) = S$$

$$A_{j+1}(S) = \mathrm{Pol}(B_j(S)^{-1}), \qquad B_{j+1}(S) = \mathrm{Fr}(B_j(S)^{-1}) \quad \text{for } j \geq 0,$$

which can be continued as long as $B_j(S) \neq 0$. If the continued fraction expansion is broken off after the term $A_j(S)$, we get the rational convergent $P_j(S)/Q_j(S)$. The polynomials $P_j(S)$ and $Q_j(S)$ can be calculated recursively by

$$P_{-1}(S) = 1,\ P_0(S) = 0,\ P_j(S) = A_j(S)P_{j-1}(S) + P_{j-2}(S) \quad \text{for } j \geq 1,$$

$$Q_{-1}(S) = 0,\ Q_0(S) = 1,\ Q_j(S) = A_j(S)Q_{j-1}(S) + Q_{j-2}(S) \quad \text{for } j \geq 1.$$

We have then

$$\deg(Q_j(S)) = \sum_{m=1}^{j} \deg(A_m(S)) \quad \text{for } j \geq 1. \tag{2}$$

For rational S we interpret $\deg(A_j(S)) = \deg(Q_j(S)) = \infty$ whenever $A_j(S)$ and $Q_j(S)$ do not exist. From [8] we note the formula

$$v(Q_j(S)S - P_j(S)) = -v(Q_{j+1}(S)) \quad \text{for } j \geq 0. \tag{3}$$

For $S \in H$ we write $L_n(S)$ for the linear complexity of the sequence which corresponds to the generating function S. The following is a special case of a result in [8].

<u>Lemma 1.</u> For any $n \geq 1$ and $S \in H$ we have $L_n(S) = \deg(Q_j(S))$, where $j \geq 0$ is uniquely determined by the condition

$$\deg(Q_{j-1}(S)) + \deg(Q_j(S)) \leq n < \deg(Q_j(S)) + \deg(Q_{j+1}(S)).$$

With the metric $d(S_1,S_2) = 2^{v(S_1 - S_2)}$ for $S_1,S_2 \in H$, the set H is a compact ultrametric space. Since H is also an additive subgroup of G and addition is a continuous operation in this metric topology, it follows that H is a compact abelian group. Let $\mathcal{B}$ be the σ-algebra of Borel sets in H. Then there exists a unique Haar measure h on H, i.e. a translation-invariant probability measure defined on $\mathcal{B}$. If $D(S_0;r): = \{S \in H: v(S - S_0) < -r\}$, $S_0 \in H$, $r = 0,1,\ldots,$ is a disk, then the translation invariance of h implies that

$$h(D(S_0;r)) = q^{-r}. \tag{4}$$

We write P for the set of polynomials over F_q of positive degree.

<u>Lemma 2.</u> For $A_1,\ldots,A_k \in P$ let $R(A_1,\ldots,A_k) = \{S \in H: A_j(S) = A_j \text{ for } 1 \leq j \leq k\}$.

Then

$$h(R(A_1,\dots,A_k)) = q^{-2(\deg(A_1) + \dots + \deg(A_k))}.$$

Proof. For any $S \in R(A_1,\dots,A_k)$ we have the same value of $P_k(S) = P_k$ and $Q_k(S) = Q_k$, thus

$$v(S - \frac{P_k}{Q_k}) = -2v(Q_k) - v(A_{k+1}(S)) < -2v(Q_k)$$

by (3). Conversely, if $v(S - P_k/Q_k) < -2v(Q_k)$, then $v(Q_k S - P_k) < -v(Q_k)$, and by [8, Lemma 3] we get $Q_k = CQ_n(S)$ and $P_k = CP_n(S)$ for some $n \geqq 1$ and $C \in F_q[x]$. Then

$$[A_1,\dots,A_k] = \frac{P_k}{Q_k} = \frac{P_n(S)}{Q_n(S)} = [A_1(S),\dots,A_n(S)],$$

so from the uniqueness of the continued fraction expansion we obtain $n = k$ and $A_j(S) = A_j$ for $1 \leqq j \leqq k$. Thus we have shown $R(A_1,\dots,A_k) = D(P_k/Q_k;2v(Q_k))$, and the desired result follows from (2) and (4). $\square$

3. DYNAMICAL SYSTEMS

We recall that a dynamical system is a probability space together with a measure-preserving transformation acting on it. We consider now the transformation T on $(H,\mathcal{B},h)$ defined by $T(S) = Fr(S^{-1})$ for $S \neq 0$ and $T(0) = 0$.

Lemma 3. T is measure preserving with respect to h.

Proof. We have to prove $h(T^{-1}(B)) = h(B)$ for all $B \in \mathcal{B}$, where $T^{-1}(B)$ is the inverse image of B under T. By [1, Theorem 1.1] it suffices to show this for every disk $D = D(S_0;r)$. For $X \neq 0$ we have $X \in T^{-1}(D)$ if and only if $v(X^{-1} - S_0 - p) < -r$ for some $p \in P$. The latter condition can only be satisfied if $v(X^{-1}) = v(S_0 + p)$, and from this we see that for fixed $p \in P$ we have $v(X^{-1} - S_0 - p) < -r$ if and only if $X \in D((S_0 + p)^{-1};r + 2v(p))$. If $D(W_1^{-1};r + 2v(p_1)) \cap D(W_2^{-1};r + 2v(p_2)) \neq \emptyset$ with $W_1 = S_0 + p_1$, $W_2 = S_0 + p_2$, and $p_1 \neq p_2$ in P, then $v(W_1) = v(p_1), v(W_2) = v(p_2)$, and

$$v(W_1^{-1} - W_2^{-1}) < -r - 2\min(v(W_1),v(W_2)).$$

On the other hand,

$$v(W_1^{-1} - W_2^{-1}) = v(W_2 - W_1) - v(W_1) - v(W_2) \geqq -2\min(v(W_1),v(W_2)),$$

where the last inequality is seen by distinguishing the cases $v(W_1) \neq v(W_2)$ and $v(W_1) = v(W_2)$. This contradiction shows that the disks $D((S_0 + p)^{-1};r + 2v(p))$ are pairwise disjoint as p ranges over P. Since such a disk has h-measure $q^{-r-2v(p)}$ by (4) and since for fixed $d \geqq 1$ there are exactly $(q - 1)q^d$ polynomials $p \in P$

with $v(p) = d$, we obtain

$$h(T^{-1}(D)) = \sum_{p \in P} q^{-r-2v(p)} = (q-1)q^{-r} \sum_{d=1}^{\infty} q^{-d} = q^{-r} = h(D). \square$$

Lemma 3 shows that $(H, \mathcal{B}, h, T)$ is a dynamical system. A second dynamical system is obtained as follows. Let μ be the probability measure defined on the power set $\mathcal{P}$ of P and determined by $\mu(p) = q^{-2\deg(p)}$ for $p \in P$. We consider the cartesian product $P^{\infty} = \prod_{n=1}^{\infty} P_n$ with $P_n = P$ for all n and the corresponding product probability space $(P^{\infty}, \mathcal{P}^{\infty}, \mu^{\infty})$. On this space the transformation T_1 is defined by

$$T_1(p_1, p_2, \ldots) = (p_2, p_3, \ldots) \quad \text{for} \quad (p_1, p_2, \ldots) \in P^{\infty}.$$

Then $(P^{\infty}, \mathcal{P}^{\infty}, \mu^{\infty}, T_1)$ is a dynamical system, called the one-sided (or unilateral) <u>Bernoulli shift</u> on P^{∞}. See Krengel [3, Sec. 1.4] for general information on Bernoulli shifts. We use the following concept of isomorphism for dynamical systems from Billingsley [1, p. 53].

<u>Definition 3.</u> The dynamical systems $(\Omega, \mathcal{F}, m, \tau)$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{m}, \tilde{\tau})$ are said to be <u>isomorphic</u> if there exist sets Ω_0 in $\mathcal{F}$ and $\tilde{\Omega}_0$ in $\tilde{\mathcal{F}}$ of measure 1 and a bijection ϕ of Ω_0 onto $\tilde{\Omega}_0$ with the following properties:

(i) If $A \subseteq \Omega_0$ and $\tilde{A} = \phi(A)$, then $A \in \mathcal{F}$ if and only if $\tilde{A} \in \tilde{\mathcal{F}}$, in which case $m(A) = \tilde{m}(\tilde{A})$;

(ii) $\tau(\Omega_0) \subseteq \Omega_0$ and $\tilde{\tau}(\tilde{\Omega}_0) \subseteq \tilde{\Omega}_0$;

(iii) $\phi(\tau(\omega)) = \tilde{\tau}(\phi(\omega))$ for all $\omega \in \Omega_0$.

<u>Theorem 1.</u> The dynamical system $(H, \mathcal{B}, h, T)$ is isomorphic to the one-sided Bernoulli shift on P^{∞}.

<u>Proof.</u> We use Definition 3 with $(\Omega, \mathcal{F}, m, \tau) = (P^{\infty}, \mathcal{P}^{\infty}, \mu^{\infty}, T_1)$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{m}, \tilde{\tau}) = (H, \mathcal{B}, h, T)$. We take $\Omega_0 = P^{\infty}$ and $\tilde{\Omega}_0 = I$, the set of irrationals in H. Since there are just countably many rationals in H, we have $h(I) = 1$. The mapping ϕ from P^{∞} onto I is defined by

$$\phi(p_1, p_2, \ldots) = [p_1, p_2, \ldots] \in I \quad \text{for} \quad (p_1, p_2, \ldots) \in P^{\infty}.$$

It follows from the uniqueness of the continued fraction expansion that ϕ is a bijection.

To prove (i) in Definition 3, we first show that if $A \in \mathcal{P}^{\infty}$, then $\tilde{A} \in \mathcal{B}$ and $\mu^{\infty}(A) = h(\tilde{A})$. It suffices to prove this for cylinder sets $A = \{(p_1, p_2, \ldots) \in P^{\infty} : p_j = A_j$ for $1 \leq j \leq k\}$, where $k \geq 1$ and $A_1, \ldots, A_k \in P$ are fixed. But then $\tilde{A} = R(A_1, \ldots, A_k) \cap I$, and since we have shown in the proof of Lemma 2 that $R(A_1, \ldots, A_k)$ is a disk, we get $\tilde{A} \in \mathcal{B}$. Furthermore by Lemma 2,

$$\mu^{\infty}(A) = \prod_{j=1}^{k} \mu(A_j) = \prod_{j=1}^{k} q^{-2\deg(A_j)} = h(R(A_1,\dots,A_k)) = h(\tilde{A}).$$

Now we have to show that if $\tilde{A} \subseteq I$ and $\tilde{A} \in \mathcal{B}$, then $A = \phi^{-1}(\tilde{A}) \in \mathcal{P}^{\infty}$. It suffices to prove this for sets $\tilde{A}$ that are intersections of I with a disk. We first consider the special case where

$$\tilde{A} = \{S \in I\colon v(S - S_0) \leqq - v(Q_k(S_0)) - v(Q_{k+1}(S_0))\}$$

with $k \geqq 0$ and $S_0 \in I$. If $S \in \tilde{A}$, then

$$v(S - \frac{P_k(S_0)}{Q_k(S_0)}) \leqq \max(v(S - S_0), v(S_0 - \frac{P_k(S_0)}{Q_k(S_0)})) < - 2v(Q_k(S_0))$$

by (3), and so S has the continued fraction expansion

$$S = [A_1(S_0),\dots,A_k(S_0),A_{k+1}(S),\dots]$$

by an argument in the proof of Lemma 2. Now

$$- v(Q_k(S_0)) - v(Q_{k+1}(S_0)) \geqq v(S - \frac{P_k(S_0)}{Q_k(S_0)}) = v(S - \frac{P_k(S)}{Q_k(S)}) = - v(Q_k(S)) - v(Q_{k+1}(S))$$

and $Q_k(S) = Q_k(S_0)$ imply $v(A_{k+1}(S)) \geqq v(A_{k+1}(S_0)) =: n$. Conversely, if S has a continued fraction expansion as above with $v(A_{k+1}(S)) \geqq n$, then it is seen immediately that $S \in \tilde{A}$. Thus

$$\tilde{A} = \Big(\bigcup_{\substack{A_{k+1} \in P \\ v(A_{k+1}) \geqq n}} R(A_1(S_0),\dots,A_k(S_0),A_{k+1}) \Big) \cap I,$$

hence $\phi^{-1}(\tilde{A}) = \{(p_1,p_2,\dots) \in P^{\infty}\colon p_j = A_j(S_0)$ for $1 \leqq j \leqq k$ and $v(p_{k+1}) \geqq n\}$ is a countable union of cylinder sets and so in $\mathcal{P}^{\infty}$. Now we consider the general case where $\tilde{A} = D \cap I$ with a disk $D = \{S \in H\colon v(S - S_0) \leqq - r\}$, $S_0 \in H$, $r \geqq 0$. Since any element of D can serve as the center of D (H is ultrametric!), we can assume that S_0 is irrational. For every $U \in \tilde{A}$ and every integer $k \geqq 0$ with $v(Q_k(U)) + v(Q_{k+1}(U)) \geqq r$ we define

$$D_k(U) = \{S \in H\colon v(S - U) \leqq - v(Q_k(U)) - v(Q_{k+1}(U))\}.$$

Every disk $D_k(U)$ is contained in D. We claim that the family of all $D_k(U)$ covers D. For this it suffices to show that every rational $S \in D$ lies in some $D_k(U)$. Let $S = [A_1(S),A_2(S),\dots,A_t(S)]$ and $S \in D$ (if $S = 0$, put $t = 0$ and $Q_0(S) = 1$ in the following). If $v(Q_t(S)) \geqq r/2$, put

$$U = [A_1(S),A_2(S),\dots,A_t(S),x,x,\dots].$$

Then

$$v(S - U) = v(\frac{P_t(U)}{Q_t(U)} - U) = - v(Q_t(U)) - v(Q_{t+1}(U))$$

and $v(Q_t(U)) + v(Q_{t+1}(U)) > 2v(Q_t(S)) \geqq r$, thus $S \in D_t(U)$ and $U \in \tilde{A}$. If $v(Q_t(S)) < r/2$, put

$$U = [A_1(S),A_2(S),\dots,A_t(S),A_{t+1}(S_0),x,x,\dots].$$

We have

$$v(S - S_0) = v(\frac{P_t(S)}{Q_t(S)} - S_0) \leq - r < - 2v(Q_t(S)),$$

and so $A_j(S_0) = A_j(S)$ for $1 \leq j \leq t$ by an argument in the proof of Lemma 2. It follows that

$$v(S - U) = v(\frac{P_t(U)}{Q_t(U)} - U) = - v(Q_t(U)) - v(Q_{t+1}(U)) =$$

$$= - v(Q_t(S_0)) - v(Q_{t+1}(S_0)) = v(\frac{P_t(S_0)}{Q_t(S_0)} - S_0) = v(S - S_0) \leq - r,$$

hence $S \in D_t(U)$ and $U \in \tilde{A}$. Thus we have shown that the closed (and also open) disks $D_k(U)$ form an open cover of the compact set D, and so finitely many of the sets $D_k(U)$, say $E_1, \ldots, E_b$, already cover D. Therefore

$$\tilde{A} = D \cap I = (\bigcup_{i=1}^{b} E_i) \cap I = \bigcup_{i=1}^{b} (E_i \cap I).$$

Each $E_i \cap I$ is of the special form considered earlier, thus $\phi^{-1}(\tilde{A}) = \bigcup_{i=1}^{b} \phi^{-1}(E_i \cap I) \in \mathcal{P}^{\infty}$ as a finite union of elements of $\mathcal{P}^{\infty}$. Property (ii) in Definition 3 is trivially satisfied and (iii) follows from an easy calculation using the algorithm for the $A_j(S)$ and $B_j(S)$ in Section 2. □

4. LIMIT THEOREMS FOR CONTINUED FRACTIONS

It follows from Theorem 1 that $(H, \mathcal{B}, h, T)$ inherits all dynamical properties of the one-sided Bernoulli shift on P^{∞} (compare with [1, Ch. 2]). In particular, since every one-sided Bernoulli shift is ergodic (see [3, Sec. 1.4], [4, p. 183]), we obtain that T is ergodic with respect to h, i.e. $T^{-1}(B) = B$ for some $B \in \mathcal{B}$ implies that $h(B) = 0$ or 1. The individual ergodic theorem, in the form given in [4, p. 183], yields the following result. Here and in the following we say that a stated property holds h-almost everywhere (h-a.e.) if the property holds for a set of $S \in H$ of h-measure 1.

<u>Theorem 2.</u> For any h-integrable function f on H we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(S)) = \int_H f dh \qquad \text{h-a.e.}$$

We note that since T^j denotes the jth iterate of T (with T^0 the identity mapping), we have $T^j(S) = B_j(S)$ for all $j \geq 0$ and $S \in I$. Rational S can be ignored since they form a set of h-measure 0.

<u>Theorem 3.</u> For any function g on P with $\sum_{p\in P} |g(p)| q^{-2\deg(p)} < \infty$ we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} g(A_j(S)) = \sum_{p\in P} g(p)q^{-2\deg(p)} \qquad \text{h-a.e.}$$

<u>Proof.</u> We apply Theorem 2 with $f(S) = g(\mathrm{Pol}(S^{-1}))$ for $S \neq 0, f(0) = 0$. For $S \in I$ we have then $f(T^j(S)) = f(B_j(S)) = g(A_{j+1}(S))$ for all $j \geq 0$. In particular $f(S) = g(A_1(S))$, hence

$$\int_H f\,dh = \sum_{p\in P} g(p)h(R(p)) = \sum_{p\in P} g(p)q^{-2\deg(p)}$$

by Lemma 2. The condition on g guarantees that f is h-integrable on H. □

<u>Corollary 1.</u> $\lim_{n\to\infty} \frac{1}{n}\deg(Q_n(S)) = \frac{q}{q-1}$ h-a.e.

<u>Proof.</u> This follows from Theorem 3 with $g(p) = \deg(p)$ for $p \in P$. We also use (2) and the identity $\sum_{d=1}^{\infty} dz^d = z(1-z)^{-2}$ with $z = q^{-1}$. □

<u>Corollary 2.</u> We have h-a.e.

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \leq j \leq n:\ A_{j+i-1}(S) = A_i \ \text{ for } \ 1 \leq i \leq k\} = q^{-2(\deg(A_1) + \dots + \deg(A_k))}$$

for all $k \geq 1$ and all $A_1,\dots,A_k \in P$.

<u>Proof.</u> We apply Theorem 2 with f being the characteristic function of the set $R(A_1,\dots,A_k)$ and use Lemma 2. Since there are just countably many choices for $A_1,\dots,A_k$, the result follows. □

For $k = 1$ Corollary 2 gives the distribution of the partial quotients $A_j(S)$ in the continued fraction expansion of a random generating function S.

<u>Lemma 4.</u> Let g be an arbitrary real-valued function on P. If $X_j(S) = g(A_j(S))$ for $j \geq 1$, then $X_1, X_2, \dots$ is a sequence of independent and identically distributed random variables on $(H, \mathcal{B}, h)$.

<u>Proof.</u> Strictly speaking, X_j is only defined on I, but we may define X_j arbitrarily on the set of h-measure 0 formed by the rationals. For $S \in I$ and any $j \geq 1$ we have

$$X_j(S) = g(\mathrm{Pol}(B_{j-1}(S)^{-1})) = g(A_1(B_{j-1}(S))) = X_1(B_{j-1}(S)) = X_1(T^{j-1}(S)),$$

hence Lemma 3 implies that the X_j are identically distributed. To prove that

$X_1,\dots,X_k$ are independent, it suffices to show that the events $A_1(S) = A_1,\dots,A_k(S) = A_k$ are independent for any $A_1,\dots,A_k \in P$, and this follows from Lemma 2. □

Theorem 4 (Law of the Iterated Logarithm for Continued Fractions). Let g be a non-constant real-valued function on P with $\sum_{p\in P} g(p)^2 q^{-2\deg(p)} < \infty$. Put

$$E = \sum_{p\in P} g(p) q^{-2\deg(p)}, \quad \sigma = \Big(\sum_{p\in P} g(p)^2 q^{-2\deg(p)} - E^2\Big)^{1/2}.$$

Then h-a.e.

$$\overline{\lim_{n\to\infty}} \frac{1}{\sigma(2n\log\log n)^{1/2}} \Big(\sum_{j=1}^{n} g(A_j(S)) - nE\Big) = 1,$$

$$\underline{\lim_{n\to\infty}} \frac{1}{\sigma(2n\log\log n)^{1/2}} \Big(\sum_{j=1}^{n} g(A_j(S)) - nE\Big) = -1.$$

Proof. Let the random variables X_j be as in Lemma 4. Then E is the expected value and σ the standard deviation of X_j, and the conditions on g guarantee that the second moment of X_j exists and $\sigma > 0$. The result follows then from the Hartman-Wintner law of the iterated logarithm in the form given in Bingham [2]. □

Corollary 3. We have h-a.e.

$$\overline{\lim_{n\to\infty}} \frac{q-1}{(2qn\log\log n)^{1/2}} \Big(\deg(Q_n(S)) - \frac{qn}{q-1}\Big) = 1,$$

$$\underline{\lim_{n\to\infty}} \frac{q-1}{(2qn\log\log n)^{1/2}} \Big(\deg(Q_n(S)) - \frac{qn}{q-1}\Big) = -1.$$

Proof. We apply Theorem 4 with $g(p) = \deg(p)$ for $p \in P$. Then $E = q/(q-1)$ by the identity in the proof of Corollary 1. The identity $\sum_{d=1}^{\infty} d^2 z^d = (z^2 + z)(1-z)^{-3}$ with $z = q^{-1}$ yields

$$\sigma^2 = \frac{q^2+q}{(q-1)^2} - \frac{q^2}{(q-1)^2} = \frac{q}{(q-1)^2}.$$

Together with (2) the result follows. □

Theorem 5 (Central Limit Theorem for Continued Fractions). Let g, E, σ be as in Theorem 4. Then for any $a < b$ (where we can have $a = -\infty$ or $b = \infty$),

$$\lim_{n\to\infty} h\Big(\Big\{S \in H:\ a\sigma\sqrt{n} \leq \sum_{j=1}^{n} g(A_j(S)) - nE \leq b\sigma\sqrt{n}\Big\}\Big) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

Proof. We proceed as in the proof of Theorem 4 and use the central limit theorem for independent and identically distributed random variables (see [9, pp. 22-23]). □

Theorem 6. Let f be a nonnegative function on the positive integers. If $\sum_{j=1}^{\infty} q^{-f(j)} < \infty$, then h-a.e. we have $\deg(A_j(S)) \leq f(j)$ for all sufficiently large j. If $\sum_{j=1}^{\infty} q^{-f(j)} = \infty$, then h-a.e. we have $\deg(A_j(S)) > f(j)$ for infinitely many j.

Proof. The events $\deg(A_j(S)) > f(j)$ for $j = 1,2,\ldots$ are independent by Lemma 4. If $k(j)$ is the least integer $> f(j)$, then these events are identical with the events $\deg(A_j(S)) \geq k(j)$. For each j we have

$$h(\{S \in H: \deg(A_j(S)) \geq k(j)\}) = \sum_{\substack{p \in P \\ \deg(p) \geq k(j)}} q^{-2\deg(p)} = q^{1-k(j)}$$

by Lemma 2. Since $\sum_{j=1}^{\infty} q^{1-k(j)}$ converges (resp. diverges) if and only if $\sum_{j=1}^{\infty} q^{-f(j)}$ converges (resp. diverges), the theorem follows from the Borel zero-one law (see [6, p. 228]). □

Corollary 4. $\overline{\lim_{j\to\infty}} \frac{\deg(A_j(S))}{\log j} = \frac{1}{\log q}$ h-a.e.

5. LIMIT THEOREMS FOR LINEAR COMPLEXITY

Because of the connection between continued fractions and linear complexity expressed in Lemma 1, the results in Section 4 have implications for the linear complexity $L_n(S)$.

Theorem 7. $\lim_{n\to\infty} \frac{L_n(S)}{n} = \frac{1}{2}$ h-a.e.

Proof. If n and j are related as in Lemma 1, then from this result we get

$$|L_n(S) - \tfrac{n}{2}| \leq \tfrac{1}{2} \max(\deg(A_j(S)), \deg(A_{j+1}(S))). \tag{5}$$

Since $n \geq \deg(Q_j(S))$, it follows that

$$\left|\frac{L_n(S)}{n} - \frac{1}{2}\right| \leq \frac{1}{2} \max\left(1 - \frac{\deg(Q_{j-1}(S))}{\deg(Q_j(S))}, \frac{\deg(Q_{j+1}(S))}{\deg(Q_j(S))} - 1\right).$$

Corollary 1 yields

$$\lim_{j\to\infty} \frac{\deg(Q_{j+1}(S))}{\deg(Q_j(S))} = \lim_{j\to\infty} \frac{\deg(Q_{j+1}(S))/(j+1)}{\deg(Q_j(S))/j} \cdot \frac{j+1}{j} = 1 \qquad \text{h-a.e.},$$

hence the desired result follows. □

The deviation of $L_n(S)$ from its asymptotic expected value $\frac{n}{2}$ is described more precisely by the following results.

Theorem 8. Let f be a nonnegative nondecreasing function on the positive integers with $\sum_{n=1}^{\infty} q^{-f(n)} < \infty$. Then h-a.e.

$$|L_n(S) - \tfrac{n}{2}| \leq \tfrac{1}{2} f(n) \quad \text{for all sufficiently large } n.$$

Proof. Theorem 6 shows that h-a.e. we have $\deg(A_j(S)) \leq f(j)$ for all sufficiently large j. For such an S we deduce from (5) that

$$|L_n(S) - \tfrac{n}{2}| \leq \tfrac{1}{2} f(j+1) \quad \text{for all sufficiently large } n.$$

Now $n \geq \deg(Q_{j-1}(S)) + \deg(Q_j(S)) \geq 2j - 1 \geq j + 1$ for all $j \geq 2$, and so $f(j+1) \leq f(n)$. □

Theorem 9. Let f be a nonnegative nondecreasing function on the positive integers with $\sum_{n=1}^{\infty} q^{-f(n)} = \infty$. Then h-a.e.

$$L_n(S) > \tfrac{n}{2} + \tfrac{1}{2} f(n) \quad \text{for infinitely many } n,$$

$$L_n(S) < \tfrac{n}{2} - \tfrac{1}{2} f(n) \quad \text{for infinitely many } n.$$

Proof. From the conditions on f we get $\sum_{n=1}^{\infty} q^{-f(5n)} = \infty$. Thus Theorem 6 implies that h-a.e. we have $\deg(A_j(S)) > f(5j)$ for infinitely many j. For such S and j we take $n = \deg(Q_{j-1}(S)) + \deg(Q_j(S))$, then

$$L_n(S) - \tfrac{n}{2} = \tfrac{1}{2} \deg(A_j(S)) > \tfrac{1}{2} f(5j)$$

by Lemma 1. By Corollary 1 we can assume that S satisfies $\lim_{j\to\infty} \deg(Q_j(S))/j = q/(q-1)$. Then

$$\tfrac{1}{j} \deg(Q_j(S)) < \tfrac{5}{2} \quad \text{for all sufficiently large } j.$$

Thus for infinitely many j we have $n = \deg(Q_{j-1}(S)) + \deg(Q_j(S)) < 2\deg(Q_j(S)) < 5j$, hence

$$L_n(S) - \tfrac{n}{2} > \tfrac{1}{2} f(5j) \geq \tfrac{1}{2} f(n)$$

for infinitely many n. The second part is shown similarly, using that h-a.e. we have $\deg(A_{j+1}(S)) > f(5j + 5) + 1$ for infinitely many j and taking $n = \deg(Q_j(S)) + \deg(Q_{j+1}(S)) - 1$. □

Theorem 10 (Law of the Logarithm for Linear Complexity). We have h-a.e.

$$\overline{\lim_{n\to\infty}} \frac{L_n(S) - (n/2)}{\log n} = \frac{1}{2 \log q},$$

$$\underline{\lim_{n\to\infty}} \frac{L_n(S) - (n/2)}{\log n} = -\frac{1}{2 \log q}.$$

Proof. We use Theorem 8 with $f(n) = (1 + \varepsilon)(\log n)/\log q$ for arbitrary $\varepsilon > 0$ and Theorem 9 with $f(n) = (\log n)/\log q$. □

6. FREQUENCY DISTRIBUTIONS FOR LINEAR COMPLEXITY

For any integers c and N with $N \geq 1$ let $Z(N;c;S)$ be the number of n, $1 \leq n \leq N$, with $L_n(S) = (n + c)/2$. We note that the cases $c = 0$ and $c = 1$ correspond to perfect linear complexity (compare with [8], [10], [11]).

Theorem 11. We have h-a.e.

$$\lim_{N\to\infty} \frac{Z(N;c;S)}{N} = \frac{q-1}{2q^{|c-(1/2)|+(1/2)}} \quad \text{for all integers } c.$$

Proof. From Corollary 1 we get

$$\lim_{j\to\infty} \frac{j}{\deg(Q_{j-1}(S)) + \deg(Q_j(S))} = \frac{q-1}{2q} \quad \text{h-a.e.}$$

Let $j(N,S)$ be the largest index j with $\deg(Q_{j-1}(S)) + \deg(Q_j(S)) \leq N$. Then with $j' = j(N,S)$ we have

$$\deg(Q_{j'-1}(S)) + \deg(Q_{j'}(S)) \leq N < \deg(Q_{j'}(S)) + \deg(Q_{j'+1}(S)),$$

and so

$$\lim_{N\to\infty} \frac{j(N,S)}{N} = \frac{q-1}{2q} \quad \text{h-a.e.} \tag{6}$$

Now let $c \geq 1$. Whenever $\deg(Q_{j-1}(S)) + \deg(Q_j(S)) \leq n < \deg(Q_j(S)) + \deg(Q_{j+1}(S))$, then Lemma 1 shows that $L_n(S) = (n + c)/2$ if and only if $n = 2\deg(Q_j(S)) - c$ with $j \geq 1$. This value of n lies in the indicated range if and only if $\deg(Q_{j-1}(S)) + \deg(Q_j(S)) \leq 2\deg(Q_j(S)) - c$, which is equivalent to $\deg(A_j(S)) \geq c$. Therefore

$$Z(N;c;S) = B(j(N,S);c;S) - \varepsilon(N;c;S),$$

where $B(r;c;S)$ denotes the number of j, $1 \leq j \leq r$, with $\deg(A_j(S)) \geq c$ and where $\varepsilon(N;c;S) = 0$ or 1. Let g be the function on P defined by $g(p) = 1$ if $\deg(p) \geq c$ and $g(p) = 0$ otherwise. Then Theorem 3 yields

$$\lim_{r\to\infty} \frac{B(r;c;S)}{r} = \lim_{r\to\infty} \frac{1}{r} \sum_{j=1}^{r} g(A_j(S)) = (q-1)\sum_{d=c}^{\infty} q^{-d} = q^{1-c} \qquad \text{h-a.e.}$$

It follows from (6) that h-a.e.

$$\lim_{N\to\infty} \frac{Z(N;c;S)}{N} = \lim_{N\to\infty} \frac{B(j(N,S);c;S)}{j(N,S)} \cdot \frac{j(N,S)}{N} = \frac{q-1}{2q^c}.$$

For $c \leq 0$ the result is shown similarly. □

For $c = 0$ and $c = 1$ we define $Y_n^{(c)}(S), n = 1,2,\ldots,$ by $Y_n^{(c)}(S) = 1$ if $L_{2n-c}(S) = n$ and $Y_n^{(c)}(S) = 0$ if $L_{2n-c}(S) \neq n$.

<u>Lemma 5.</u> If $c = 0$ or $c = 1$, then $Y_1^{(c)}, Y_2^{(c)}, \ldots$ is a sequence of independent and identically distributed random variables on $(H, \mathcal{B}, h)$.

<u>Proof.</u> It follows from Lemma 1 that $L_{2n-c}(S) = n$ if and only if $\deg(Q_j(S)) = n$ for some $j \geq 1$. Since the last condition is independent of c, we have $Y_n^{(0)} = Y_n^{(1)}$, and we write Y_n for $Y_n^{(c)}$. We have

$$h(\{S \in H: Y_n(S) = 1\}) = \sum_{j=1}^{n} h(\{S \in H: \deg(Q_j(S)) = n\}).$$

For fixed j, $1 \leq j \leq n$, we obtain from (2) and Lemma 2:

$$h(\{S \in H: \deg(Q_j(S)) = n\}) = \sum_{\substack{d_1,\ldots,d_j \geq 1 \\ d_1+\ldots+d_j = n}} h(\{S \in H: \deg(A_m(S)) = d_m \text{ for } 1 \leq m \leq j\})$$

$$= \sum_{\substack{d_1,\ldots,d_j \geq 1 \\ d_1+\ldots+d_j = n}} (q-1)q^{d_1} \ldots (q-1)q^{d_j} q^{-2(d_1 + \ldots + d_j)}$$

$$= (q-1)^j q^{-n} \sum_{\substack{d_1,\ldots,d_j \geq 1 \\ d_1+\ldots+d_j = n}} 1 = (q-1)^j q^{-n} \binom{n-1}{j-1}.$$

Thus

$$h(\{S \in H: Y_n(S) = 1\}) = (q-1)q^{-n} \sum_{j=0}^{n-1} \binom{n-1}{j}(q-1)^j = \frac{q-1}{q}, \tag{7}$$

which shows in particular that the Y_n are identically distributed. To prove that $Y_1, \ldots, Y_k$ are independent, we choose $\varepsilon_1, \ldots, \varepsilon_k \in \{0,1\}$ arbitrarily and let $1 \leq r_1 < r_2 < \ldots < r_t \leq k$ be exactly those indices for which $\varepsilon_{r_i} = 1$. By the remark at the beginning of the proof we have $Y_1(S) = \varepsilon_1, \ldots, Y_k(S) = \varepsilon_k$ if and only if $r_1, \ldots, r_t$ appear as values of $\deg(Q_j(S))$ for some $j \geq 1$ and the other elements of $\{1,2,\ldots,k\}$ do not. This condition is equivalent to $\deg(Q_1(S)) = r_1, \ldots, \deg(Q_t(S)) = r_t, \deg(Q_{t+1}(S)) > k$, which is in turn equivalent to $\deg(A_1(S)) = r_1, \deg(A_2(S)) = r_2 - r_1, \ldots, \deg(A_t(S)) = r_t - r_{t-1}, \deg(A_{t+1}(S)) > k - r_t$, where we put $r_0 = 0$ if $t = 0$. Therefore Lemma 2 yields

$$h(\{S \in H: Y_1(S) = \varepsilon_1, \ldots, Y_k(S) = \varepsilon_k\}) =$$

$$= (q-1)q^{r_1}(q-1)q^{r_2 - r_1} \ldots (q-1)q^{r_t - r_{t-1}} q^{-2r_t} \sum_{m=k-r_t+1}^{\infty} (q-1)q^{-m}$$

$$= (q-1)^{t+1} q^{-r_t} \sum_{m=k-r_t+1}^{\infty} q^{-m} = (q-1)^t q^{-k}.$$

On the other hand, it follows from (7) that

$$\prod_{n=1}^{k} h(\{S \in H: Y_n(S) = \varepsilon_n\}) = \left(\frac{q-1}{q}\right)^t \left(\frac{1}{q}\right)^{k-t} = (q-1)^t q^{-k},$$

and so $Y_1, \ldots, Y_k$ are independent. □

<u>Theorem 12</u> (Law of the Iterated Logarithm for Perfect Linear Complexity, First Version). For $c = 0$ and $c = 1$ we have h-a.e.

$$\overline{\lim_{N \to \infty}} \frac{1}{(N \log \log N)^{1/2}} \left(Z(N;c;S) - \frac{(q-1)N}{2q}\right) = \frac{(q-1)^{1/2}}{q},$$

$$\underline{\lim_{N \to \infty}} \frac{1}{(N \log \log N)^{1/2}} \left(Z(N;c;S) - \frac{(q-1)N}{2q}\right) = -\frac{(q-1)^{1/2}}{q}.$$

<u>Proof.</u> By (7) the expected value of Y_n is $(q-1)/q$ and the variance of Y_n is

$$\sigma^2 = \int_H Y_n^2 \, dh - \left(\frac{q-1}{q}\right)^2 = \frac{q-1}{q} - \left(\frac{q-1}{q}\right)^2 = \frac{q-1}{q^2}.$$

It follows from Lemma 5 and the Hartman-Wintner law of the iterated logarithm that

$$\lim_{n \to \infty} \begin{matrix} \sup \\ \inf \end{matrix} \frac{1}{\sigma(2n \log \log n)^{1/2}} \left(\sum_{i=1}^{n} Y_i(S) - \frac{(q-1)n}{q}\right) = \begin{matrix} 1 \\ -1 \end{matrix} \qquad \text{h-a.e.} \tag{8}$$

Putting $n = \lfloor (N+c)/2 \rfloor$, where $\lfloor t \rfloor$ denotes the greatest integer $\leq t$, and using

$$Z(N;c;S) = \sum_{i=1}^{\lfloor (N+c)/2 \rfloor} Y_i(S) \tag{9}$$

for $c = 0$ and $c = 1$, we obtain the theorem. □

<u>Theorem 13</u> (Law of the Iterated Logarithm for Perfect Linear Complexity, Second Version). If $W(N;S)$ is the number of n, $1 \leq n \leq N$, with $L_n(S) = \frac{n}{2}$ or $\frac{n+1}{2}$, then

h-a.e.

$$\overline{\lim_{N\to\infty}} \frac{1}{(N \log\log N)^{1/2}} \left(W(N;S) - \frac{(q-1)N}{q}\right) = \frac{2(q-1)^{1/2}}{q},$$

$$\underline{\lim_{N\to\infty}} \frac{1}{(N \log\log N)^{1/2}} \left(W(N;S) - \frac{(q-1)N}{q}\right) = -\frac{2(q-1)^{1/2}}{q}.$$

Proof. We put $n = \lfloor N/2 \rfloor$ in (8) and use

$$W(N;S) = Z(N;0;S) + Z(N;1;S) = 2 \sum_{i=1}^{\lfloor N/2 \rfloor} Y_i(S) + \theta(N;S) \tag{10}$$

with $\theta(N;S) = 0$ or 1, as follows from (9). □

Theorem 14 (Central Limit Theorem for Perfect Linear Complexity, First Version). For $c = 0$ and $c = 1$ we have for any $a < b$ (where we can have $a = -\infty$ or $b = \infty$),

$$\lim_{N\to\infty} h\left(\left\{S \in H: \frac{a}{q}\sqrt{\frac{(q-1)N}{2}} \le Z(N;c;S) - \frac{(q-1)N}{2q} \le \frac{b}{q}\sqrt{\frac{(q-1)N}{2}}\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

Proof. The expected value and the variance of Y_n have been calculated in the proof of Theorem 12. From Lemma 5 and the central limit theorem we obtain

$$\lim_{n\to\infty} h\left(\left\{S \in H: a\sigma\sqrt{n} \le \sum_{i=1}^{n} Y_i(S) - \frac{(q-1)n}{q} \le b\sigma\sqrt{n}\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt. \tag{11}$$

Applying this with $n = \lfloor (N+c)/2 \rfloor$ and using (9) we get

$$\lim_{N\to\infty} h(B_N(a,b,c)) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt,$$

where

$$B_N(a,b,c) = \left\{S \in H: a\sigma\sqrt{\left\lfloor \frac{N+c}{2} \right\rfloor} \le Z(N;c;S) - \frac{q-1}{q}\left\lfloor \frac{N+c}{2} \right\rfloor \le b\sigma\sqrt{\left\lfloor \frac{N+c}{2} \right\rfloor}\right\}.$$

Put

$$A_N(a,b,c) = \left\{S \in H: a\sigma\sqrt{\frac{N}{2}} \le Z(N;c;S) - \frac{(q-1)N}{2q} \le b\sigma\sqrt{\frac{N}{2}}\right\}.$$

For given $\varepsilon > 0$ we have $A_N(a,b,c) \subseteq B_N(a-\varepsilon, b+\varepsilon, c)$ for all sufficiently large N, hence

$$\overline{\lim_{N\to\infty}} h(A_N(a,b,c)) \le \overline{\lim_{N\to\infty}} h(B_N(a-\varepsilon, b+\varepsilon, c)) = \frac{1}{\sqrt{2\pi}} \int_{a-\varepsilon}^{b+\varepsilon} e^{-t^2/2}\, dt.$$

With $\varepsilon \to 0+$ we obtain

$$\overline{\lim_{N\to\infty}} h(A_N(a,b,c)) \le \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

Using $B_N(a+\varepsilon, b-\varepsilon, c) \subseteq A_N(a,b,c)$ for all sufficiently large N, we get similarly

$$\underline{\lim}_{N\to\infty} h(A_N(a,b,c)) \geq \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt,$$

and the desired result follows. □

Theorem 15 (Central Limit Theorem for Perfect Linear Complexity, Second Version). If $W(N;S)$ is as in Theorem 13, then we have for any $a < b$ (where we can have $a = -\infty$ or $b = \infty$),

$$\lim_{N\to\infty} h(\{S \in H: \tfrac{a}{q}\sqrt{2(q-1)N} \leq W(N;S) - \frac{(q-1)N}{q} \leq \tfrac{b}{q}\sqrt{2(q-1)N}\}) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

Proof. We apply (11) with $n = \lfloor N/2 \rfloor$, use (10), and proceed as in the proof of Theorem 14. □

Theorem 16. We have h-a.e.

$$\overline{\lim}_{N\to\infty} \frac{1}{(N \log\log N)^{1/2}} \left(Z(N;c;S) - \frac{(q-1)N}{2q^{|c-(1/2)|+(1/2)}}\right) \leq \frac{(q^{|c-(1/2)|+(1/2)}-q)^{1/2}+1}{q^{|c-(1/2)|+(1/2)}} (q-1)^{1/2},$$

$$\underline{\lim}_{N\to\infty} \frac{1}{(N \log\log N)^{1/2}} \left(Z(N;c;S) - \frac{(q-1)N}{2q^{|c-(1/2)|+(1/2)}}\right) \geq - \frac{(q^{|c-(1/2)|+(1/2)}-q)^{1/2}+1}{q^{|c-(1/2)|+(1/2)}} (q-1)^{1/2}$$

for all integers c.

Proof. For $c = 0$ and $c = 1$ this follows from Theorem 12. Now let $c \geq 2$. From the proof of Theorem 11 we obtain

$$Z(N;c;S) \leq B(j(N,S);c;S) \tag{12}$$

with $B(r;c;S) = \sum_{j=1}^{r} g(A_j(S))$, where g is the function on P defined by $g(p) = 1$ if $\deg(p) \geq c$ and $g(p) = 0$ otherwise. By Theorem 4 we have

$$\overline{\lim}_{r\to\infty} \frac{1}{\sigma(2r \log\log r)^{1/2}} (B(r;c;S) - rq^{1-c}) = 1 \quad \text{h-a.e.},$$

where

$$\sigma^2 = \sum_{p\in P} g(p)^2 q^{-2\deg(p)} - q^{2-2c} = q^{1-c} - q^{2-2c} = q^{1-2c}(q^c - q).$$

For an $S \in H$ with the property above and for a given $0 < \varepsilon < 1$ we therefore get

$$B(j(N,S);c;S) - j(N,S)q^{1-c} \leq (1+\varepsilon)\sigma(2j(N,S)\log\log j(N,S))^{1/2} \tag{13}$$

for all sufficiently large N. By Corollary 3 we can assume that the $S \in H$ under consideration satisfies

$$\deg(Q_n(S)) \geq \frac{qn}{q-1} - \frac{1+\varepsilon}{q-1}(2qn \log\log n)^{1/2}$$

for all sufficiently large n. By the definition of $j(N,S)$ in the proof of Theorem 11 we have

$$N \geq \deg(Q_{j(N,S)-1}) + \deg(Q_{j(N,S)})$$

$$\geq \frac{2qj(N,S)}{q-1} - \frac{2+3\varepsilon}{q-1}(2qj(N,S)\log\log j(N,S))^{1/2}$$

for all sufficiently large N. Put

$$F(j) = \frac{2qj}{q-1} - \frac{2+3\varepsilon}{q-1}(2qj\log\log j)^{1/2}.$$

Then $F(j)$ is an increasing function of j for sufficiently large j and it is easily checked that

$$F\left(\frac{(q-1)N}{2q} + \frac{1+2\varepsilon}{q}((q-1)N\log\log N)^{1/2}\right) > N$$

for all sufficiently large N. It follows that

$$j(N,S) \leq \frac{(q-1)N}{2q} + \frac{1+2\varepsilon}{q}((q-1)N\log\log N)^{1/2} \tag{14}$$

for all sufficiently large N. In particular, we have $j(N,S) \leq (1+\varepsilon)^2(q-1)N/(2q)$ for all sufficiently large N. Now (12), (13), and (14) yield

$$Z(N;c;S) - \frac{(q-1)N}{2q^c} \leq$$

$$\leq B(j(N,S);c;S) - j(N,S)q^{1-c} + \left(j(N,S) - \frac{(q-1)N}{2q}\right)q^{1-c}$$

$$\leq (1+\varepsilon)^2\sigma\left(\frac{q-1}{q}N\log\log N\right)^{1/2} + \frac{1+2\varepsilon}{q^c}((q-1)N\log\log N)^{1/2}$$

$$\leq (1+3\varepsilon)\frac{(q^c-q)^{1/2}+1}{q^c}(q-1)^{1/2}(N\log\log N)^{1/2}$$

for all sufficiently large N, and so the first part of the theorem is shown for $c \geq 2$. The remaining cases are proved similarly. □

7. CONTINUED FRACTION TESTS

From Lemma 1 we see that a linear complexity profile always has the following form:

$$0,\ldots,0,d_1,\ldots,d_1,d_1+d_2,\ldots,d_1+d_2,\ldots, \tag{15}$$

with 0 repeated $d_1 - 1$ times and $\sum_{i=1}^{j} d_i$ repeated $d_j + d_{j+1}$ times for all $j \geq 1$, where $d_1, d_2, \ldots$ are positive integers given by $d_j = \deg(A_j(S))$. Therefore, prescribing a linear complexity profile is equivalent to prescribing $d_1, d_2, \ldots$. If an arbitrary sequence $d_1, d_2, \ldots$ of positive integers is given, then the following algorithm in Niederreiter [8] generates a sequence $s_1, s_2, \ldots$ of elements of F_q whose linear complexity profile is as in (15). We put $q_j = \sum_{i=1}^{j} d_i$ for $j \geq 1$. We recall that the polynomial $a_k x^k + \ldots + a_1 x + a_0$ associated with the linear recursion

(1) is called the characteristic polynomial of the linear recursion.

Algorithm

Initialization: $Q_0 = 1$ (considered as a polynomial over F_q).

Step 1: Choose a polynomial A_1 over F_q with $\deg(A_1) = d_1$ and let $Q_1 = A_1$. Calculate the terms s_i with $1 \leq i \leq q_1 + q_2 - 1$ by the linear recursion with characteristic polynomial Q_1 and initial values $s_i = 0$ for $1 \leq i \leq q_1 - 1$, $s_i = c^{-1}$ for $i = q_1$, where c is the leading coefficient of Q_1.

Step j (for $j \geq 2$): Suppose the polynomials $Q_1,\ldots,Q_{j-1}$ and the terms s_i with $1 \leq i \leq q_{j-1} + q_j - 1$ have already been calculated. Choose a polynomial A_j over F_q with $\deg(A_j) = d_j$ and let $Q_j = A_j Q_{j-1} + Q_{j-2}$. Calculate the terms s_i with $q_{j-1} + q_j \leq i \leq q_j + q_{j+1} - 1$ from the previously calculated terms by the linear recursion with characteristic polynomial Q_j.

If this procedure is continued indefinitely, it yields a nonperiodic sequence with the prescribed linear complexity profile. If the procedure is broken off after finitely many steps, then a minor modification in the last step is needed (see [8]).

Let S be an arbitrary sequence of elements of F_q and let $A_j(S), j = 1,2,\ldots,$ as usual be the polynomials appearing in the continued fraction expansion of the generating function S. If we put $d_j(S) = \deg(A_j(S))$, then each d_j can be viewed as a random variable on the probability space $(H,\mathcal{B},h)$ and the values of d_j are positive integers. By Lemma 4 the random variables $d_1,d_2,\ldots$ are independent and identically distributed. For every positive integer m, the probability that $d_j = m$ is equal to $(q - 1)q^{-m}$ by Lemma 2. Thus, in a statistical sense we can say that <u>the linear complexity profile of a random sequence of elements of F_q has the form (15), where $d_1,d_2,\ldots$ are independent and identically distributed with the probability distribution $\mathrm{Prob}(d_j = m) = (q - 1)q^{-m}$ for all positive integers m.</u> We note that each d_j has expected value $q/(q - 1)$ and variance $q/(q - 1)^2$, as shown in the proof of Corollary 3. In particular, in (15) we can expect an average step height of $q/(q - 1)$ and an average step length of $2q/(q - 1)$. For $q = 2$ this agrees with a result of Rueppel [11, p. 45] that was proved by a different method.

This description of the linear complexity profile of a random sequence of elements of F_q can serve as the basis for new types of randomness tests. For a concretely given sequence S, we can calculate $d_j = d_j(S)$ by the Berlekamp-Massey algorithm (see [5, Ch. 6], [7]). The sequence $d_1,d_2,\ldots$ is then subjected to conventional statistical tests for randomness, the null hypothesis being that $d_1,d_2,\ldots$ are independent and identically distributed with the probability distribution given above. More generally, we can calculate the $A_j(S)$ by the continued fraction algorithm or the Berlekamp-Massey algorithm, take an arbitrary real-valued function g on P, and use the independent and identically distributed random variables X_j in

Lemma 4 as the basis for a randomness test. These types of randomness tests may be called continued fraction tests.

Other types of randomness tests may be based on the independent and identically distributed random variables $Y_n = Y_n^{(c)}$ in Lemma 5 for which the probability distribution is given by $\mathrm{Prob}(Y_n = 0) = 1/q, \mathrm{Prob}(Y_n = 1) = (q - 1)/q$ according to (7).

REFERENCES

[1] P. Billingsley: Ergodic Theory and Information, Wiley, New York, 1965.

[2] N. H. Bingham: Variants on the law of the iterated logarithm, Bull. London Math. Soc. 18, 433-467 (1986).

[3] U. Krengel: Ergodic Theorems, de Gruyter, Berlin, 1985.

[4] L. Kuipers and H. Niederreiter: Uniform Distribution of Sequences, Wiley, New York, 1974.

[5] R. Lidl and H. Niederreiter: Introduction to Finite Fields and Their Applications, Cambridge Univ. Press, Cambridge, 1986.

[6] M. Loève: Probability Theory, 3rd ed., Van Nostrand, New York, 1963.

[7] J. L. Massey: Shift-register synthesis and BCH decoding, IEEE Trans. Information Theory 15, 122-127 (1969).

[8] H. Niederreiter: Sequences with almost perfect linear complexity profile, Advances in Cryptology - EUROCRYPT '87 (D. Chaum and W. L. Price, eds.), Lecture Notes in Computer Science, Vol. 304, pp. 37-51, Springer, Berlin, 1988.

[9] M. Rosenblatt: Random Processes, 2nd ed., Springer, New York, 1974.

[10] R. A. Rueppel: Linear complexity and random sequences, Advances in Cryptology - EUROCRYPT '85 (F. Pichler, ed.), Lecture Notes in Computer Science, Vol. 219, pp. 167-188, Springer, Berlin, 1986.

[11] R. A. Rueppel: Analysis and Design of Stream Ciphers, Springer, Berlin, 1986.

The author gratefully acknowledges support for this research project by the Austrian Ministry for Science and Research.