Linear Complexity of Periodically Repeated Random Sequences

Zong-Duo Dai *
Dept. of Math., RHBNC, University of London,
Egham Hill, Egham, Surrey, TW20 0EX, U.K.

Jun-Hui Yang
Computing Center, Academia Sinica,
P.O.Box 2719. 100080, Beijing, China

Abstract

On the linear complexity $\Lambda(\tilde{z})$ of a periodically repeated random bit sequence $\tilde{z}$, R. Rueppel proved that, for two extreme cases of the period T, the expected linear complexity $E[\Lambda(\tilde{z})]$ is almost equal to T, and suggested that $E[\Lambda(\tilde{z})]$ would be close to T in general [6, pp. 33-52] [7, 8]. In this note we obtain bounds of $E[\Lambda(\tilde{z})]$, as well as bounds of the variance $Var[\Lambda(\tilde{z})]$, both for the general case of T, and we estimate the probability distribution of $\Lambda(\tilde{z})$. Our results on $E[\Lambda(\tilde{z})]$ quantify the closeness of $E[\Lambda(\tilde{z})]$ and T, in particular, formally confirm R. Rueppel's suggestion.

Keywords: Linear Complexity, Random Sequences.

1 Introduction

The linear complexity [8, p. 32] (or linear equivalence [1, p.199]) of a sequence is the length of the shortest linear shift register (LFSR) by which the given

*On leave from Graduate School, Academia Sinica, 100039-08, Beijing, China, with this work supported by SERC grant GR/F 72727.

sequence could be generated. Since there exists an efficient algorithm for finding the shortest LFSR which generates a given sequence (the Berlekamp-Massey LFSR synthesis algorithm [5]), the linear complexity is particularly important as a measure of the unpredictability of sequences. The statistical properties of the linear complexity of a periodically repeated random bit string are of considerable practical interest [6, pp. 33-52] [7, 8], since deterministically generated key streams in cipher systems must be ultimately periodic.

Given T, let $z^T = z_0, z_1, \ldots, z_{T-1}$ be a binary sequence where z_i $(0 \leq i \leq T-1)$ is selected according to a fair coin tossing experiment, and let $\tilde{z}$ be the semi-infinite sequence by periodically repeating the random bit string z^T. Let $\mathcal{Z}$ be the sample space consisting of all the possible semi-infinite periodically repeated random sequences $\tilde{z}$. The elements in $\mathcal{Z}$ are equiprobable. Since $|\mathcal{Z}| = 2^T$, where $|\mathcal{Z}|$ denote the size of $\mathcal{Z}$, so the probability of the occurrence of each $\tilde{z}$ is equal to $1/|\mathcal{Z}| = 2^{-T}$. Let $\Lambda(\tilde{z})$ denote the linear complexity of $\tilde{z}$, then $\Lambda(\tilde{z})$ is a random variable on the sample space $\mathcal{Z}$. Let $E[\Lambda(\tilde{z})]$ be the expected linear complexity of $\tilde{z}$, and $Var[\Lambda(\tilde{z})]$ the variance of the linear complexity $\Lambda(\tilde{z})$.

R. Rueppel computed $E[\Lambda(\tilde{z})]$ in two extreme cases: when $T = 2^n - 1$ (any prime n) and when $T = 2^m$ (any m) [6, pp. 33-52] [7, 8]. In both cases he proved that $E[\Lambda(\tilde{z})]$ is almost equal to T, or more precisely, $E[\Lambda(\tilde{z})] \geq\simeq e^{-1/n}(2^n - 3/2)$ when $T = 2^n - 1$, and

$$E[\Lambda(\tilde{z})] = 2^m - 1 + 2^{-2^m} \tag{1}$$

when $T = 2^m$, and suggested that in the general case $E[\Lambda(\tilde{z})]$ would be close to T.

D. Gollmann [2] proved that, when $T = p^n$, $p > 2$ prime, and p^2 is not a factor of $2^{p-1} - 1$,

$$E[\Lambda(\tilde{z})] = p^n - \frac{1}{2} - (p-1)\sum_{i=0}^{n-1} p^i 2^{-n_p p^i}, \tag{2}$$

where n_p is the degree of the irreducible polynomials with period p over $GF(2)$.

In this note we consider $E[\Lambda(\tilde{z})]$, as well as $Var[\Lambda(\tilde{z})]$, both for the general case. We obtain expressions for $E[\Lambda(\tilde{z})]$ and for $Var[\Lambda(\tilde{z})]$, and we bound $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ in terms of the arithmetic function $d(T)$, and then we bound $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ in terms of analytic functions, or more precisely, we show that for any $\varepsilon > 0$, (i) $E[\Lambda(\tilde{z})] > T - T^{\varepsilon}$ and $Var[\Lambda(\tilde{z})] < T^{\varepsilon}$, provided T is large enough, (ii) $E[\Lambda(\tilde{z})] > T - T^{(1+\varepsilon)\log 2/\log\log T}$ and $Var[\Lambda(\tilde{z})] < T^{(1+\varepsilon)\log 2/\log\log T}$, provided T is large enough, and (iii) $E[\Lambda(\tilde{z})] > T - (\log T)^{(1+\varepsilon)\log 2}$ and $Var[\Lambda(\tilde{z})] < \log_2(1+T)(\log T)^{(1+\varepsilon)\log 2}$ for almost all T (see Remark 1 in section 4). We also estimate the probability distribution of $\Lambda(\tilde{z})$, for any $\varepsilon > \delta > 0$ we get that $Prob.(\Lambda(\tilde{z}) > T - T^{\varepsilon}) > 1 - T^{-2\varepsilon+\delta}$ for large enough T. Our results on $E[\Lambda(\tilde{z})]$ quantify the closeness of $E[\Lambda(\tilde{z})]$ and T, and in particular formally confirm R. Rueppel's suggestion.

In this paper the base of the logarithms is e, *i.e.*, $\log = \log_e$, unless indicated otherwise.

2 Expressions for $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$

We identify the sequence $\tilde{z}$ with its generating function $\tilde{z}(x)$, defined over the binary field $GF(2)$, as $\tilde{z}(x) = \sum_{j=0}^{\infty} z_j x^j$. It is known that $\tilde{z}(x)$ is equal to a rational fraction $\tilde{z}(x) = z^*(x)/(1-x^T) = P(\tilde{z},x)/C(\tilde{z},x)$, where $z^*(x) = \sum_{j=0}^{T-1} z_j x^j$, $P(\tilde{z},x)$ and $C(\tilde{z},x)$ are coprime to each other. It is also known that $C(\tilde{z},x)$ is the minimal polynomial [1, p.201][8, p. 26] of $\tilde{z}$, and $\Lambda(\tilde{z}) = degC(\tilde{z},x)$, where $degC(\tilde{z},x)$ is the degree of $C(\tilde{z},x)$.

The range of $C(\tilde{z},x)$ depends on the factorization of $1-x^T$. If $T = 2^m T_1$, $\gcd(2,T_1) = 1$, it is known [4, pp. 64-65] that $1 - x^T = \prod_{d|T_1} \prod_{j=1}^{\phi(d)/n_d} C_{d,j}^{2^m}(x)$, where for any given d, $C_{d,j}(x)$ ($0 \le j \le \phi(d)/n_d$) are all the distinct monic irreducible polynomials with period d over $GF(2)$, and of the same degree n_d, where n_d is the order of 2 modulo d, (*i.e.*, the least positive integer such that $2^{n_d} = 1 \,(\bmod\, d)$), $\phi(d)$ is the Euler's function, (*i.e.*, the number of the integers $i, 1 \le i \le d$, coprime to d). As a factor of $1 - x^T$, $C(\tilde{z},x)$ must be of the form $C(\tilde{z},x) = \prod_{d|T_1} \prod_{j=1}^{\phi(d)/n_d} C_{d,j}^{e_{d,j}(\tilde{z})}(x)$, $0 \le e_{d,j}(\tilde{z}) \le 2^m$. The exponent $e_{d,j}(\tilde{z})$ is a random variable defined on $\mathcal{Z}$ with range $[0,\ 2^m]$. Now we have $\Lambda(\tilde{z}) = \sum_{d|T_1} \sum_{j=1}^{\phi(d)/n_d} n_d e_{d,j}(\tilde{z})$.

Lemma 1 .
1. The random variable $e_{d,j}(\tilde{z})$ *has the following probability density function*

$$Prob.(e_{d,j}(\tilde{z}) = e) = \begin{cases} 2^{-n_d 2^m} & e = 0, \\ 2^{-n_d 2^m}(2^{n_d e} - 2^{n_d(e-1)}) & e > 0. \end{cases}$$

2. All the random variables $e_{d,j}(\tilde{z})$, $d \mid T_1$, $1 \leq j \leq \phi(d)/n_d$, *are mutually independent.*

Observe that the probability density function of $e_{d,j}(\tilde{z})$ is not dependent on the parameter j, we denote by E_d the expected value of $e_{d,j}(\tilde{z})$, and by V_d the variance of $e_{d,j}(\tilde{z})$.

Lemma 2

$$E_d = 2^m - \frac{2^{n_d 2^m} - 1}{2^{n_d 2^m}(2^{n_d} - 1)} ,$$

and

$$V_d = \frac{2^{n_d(2^{m+1}+1)} - (2^{m+1}+1)(2^{n_d(2^m+1)} - 2^{n_d 2^m}) - 1}{2^{n_d 2^{m+1}}(2^{n_d} - 1)^2} .$$

Theorem 1 **(Expressions)** *Let* $T = 2^m T_1$, $\gcd(2, T_1) = 1$. *Then*

$$E[\Lambda(\tilde{z})] = T - \sum_{d|T_1} \frac{\phi(d)(2^{n_d 2^m} - 1)}{2^{n_d 2^m}(2^{n_d} - 1)} ,$$

and

$$Var[\Lambda(\tilde{z})] = \sum_{d|T_1} \frac{\phi(d) n_d [2^{n_d(2^{m+1}+1)} - (2^{m+1}+1)(2^{n_d(2^m+1)} - 2^{n_d 2^m}) - 1]}{2^{n_d 2^{m+1}}(2^{n_d} - 1)^2} .$$

Theorem 1 gives a way to calculate $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ based on the factorization of T case by case. In the special case when $T = 2^m$ this is straightforward. Both of the summations in Theorem 1 contain only one term with $d = 1$, from which one obtains (1), as well as

$$Var[\Lambda(\tilde{z})] = \frac{2^{2^{m+1}+1} - (2^{m+1}+1)(2^{(2^m+1)} - 2^{2^m}) - 1}{2^{2^{m+1}}} < 2 .$$

For another exampe, when $T = p^n$, $p > 2$ prime, and p^2 is not a factor of $2^{p-1} - 1$, from $E[\Lambda(\tilde{z})]$'s expression, in which the summation contains $n + 1$ terms with $d = p^i, 0 \leq i \leq n$, and $n_{p^i} = n_p p^{i-1}, 1 \leq i \leq n$, one obtains (2). But the real significance of Theorem 1 is that from it one may bound $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ in terms of the arithmetic function $d(n)$, which is defined to be the number of all possible positive factors of n, *i.e.*, $d(n) = \sum_{d|n} 1$.

3 Bounds for $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ by $d(n)$

Theorem 2 *Let* $T = 2^m T_1$, $\gcd(2, T_1) = 1$. *Then*

$$E[\Lambda(\tilde{z})] > T - d(T_1) \geq T - d(T),$$

and

$$Var[\Lambda(\tilde{z})] < d(T_1)(1 + \log_2(1 + T_1)) \leq d(T)(1 + \log_2(1 + T)).$$

With Theorem 2 and the factorization of T, the evaluation for both of $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ becomes easier. In fact, if $T_1 = \prod_{i=1}^{s} p_i^{e_i}$, where $p_i, 1 \leq i \leq s$, are distinct prime factors, then $d(T_1) = \prod_{i=1}^{s}(1+e_i)$ [2, p. 238]. Hence $E[\Lambda(\tilde{z})] > T - \prod_{i=1}^{s}(1+e_i)$ and $Var[\Lambda(\tilde{z})] < (1+\log_2(1+T_1)) \prod_{i=1}^{s}(1+e_i)$. What is more interesting is that from Theorem 2 we shall get analytic bounds for $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ based on the orders of $d(n)$.

4 Bounds for $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$ by Analytic Functions

Lemma 3 [2, pp. 259-261, p. 361] *If* $\varepsilon > 0$, *then we have*

1. $d(n) < n^{\varepsilon}$ *for all* $n > n_{\varepsilon}$, *where* n_{ε} *depends on* ε.
2. $d(n) < n^{(1+\epsilon)\log 2/\log\log n}$ *for all* $n > n_{\varepsilon}$, *where* n_{ε} *depends on* ε.
3. $d(n) < (\log n)^{(1+\epsilon)\log 2}$ *for almost all numbers* n.

Remark 1 A property P of positive integers n is said to be true for **almost all numbers** if $\lim_{x\to\infty} N(x)/x = 1$, where $N(x)$ is the number of positive integers less than x which satisfy P.

Remark 2 Lemma 3 provides three kinds of bounds for $d(n)$. The bounds given in item *1* and item *2* hold for large enough n. The bound given in item*2*, a kind of power of n with the exponent tending slowly to zero when n goes to infinity, is tighter than the bound given in item *1*, but the latter looks much simpler. The bound given in item *3* is the tightest one, but it holds only for almost all n.

From Theorem 2 and Lemma 3 we may obtain immediately three kinds of bounds for $E[\Lambda(\tilde{z})]$.

Theorem 3 (Bounds for $E[\Lambda(\tilde{z})]$) *If* $\varepsilon > 0$, *then we have*

1. $E[\Lambda(\tilde{z})] > T - T^{\varepsilon}$ *for all* $T > T_{\varepsilon}$, *where* T_{ε} *depends on* ε.

2. $E[\Lambda(\tilde{z})] > T - T^{(1+\varepsilon)\log 2/\log\log T}$ *for all* $T > T_{\varepsilon}$, *where* T_{ε} *depends on* ε.

3. $E[\Lambda(\tilde{z})] > T - (\log T)^{(1+\varepsilon)\log 2}$ *for almost-all* T.

Remark 3 The bounds on $E[\Lambda(\tilde{z})]$ shown in Theorem 3 quantify the closeness of $E[\Lambda(\tilde{z})]$ and T, and in particular, the expected linear complexity $E[\Lambda(\tilde{z})]$ and the period T are of the same asymptotical order, *i.e.*, $\lim_{T\to\infty} E[\Lambda(\tilde{z})]/T = 1$, hence formally confirm R. Rueppel's suggestion.

Theorem 4 (Bounds for $Var[\Lambda(\tilde{z})]$) *If* $\varepsilon > 0$, *then we have*

1. $Var[\Lambda(\tilde{z})] < T^{\varepsilon}$, *for all* $T > T_{\varepsilon}$, *where* T_{ε} *depends on* ε.

2. $Var[\Lambda(\tilde{z})] < T^{(1+\varepsilon)\log 2/\log\log T}$, *for all* $T > T_{\varepsilon}$, *where* T_{ε} *depends on* ε.

3. $Var[\Lambda(\tilde{z})] < (\log T)^{(1+\varepsilon)\log 2}\log_2(1+T)$, *for almost-all* T.

5 Probability Distribution of $\Lambda(\tilde{z})$

Based on the knowlege on $E[\Lambda(\tilde{z})]$ and $Var[\Lambda(\tilde{z})]$, we prove that the linear complexity $\Lambda(\tilde{z})$ distributes very close to the length T with a probability almost equal to 1, provided T is large enough, as shown in the following theorem.

Theorem 5 *If $\varepsilon > \delta > 0$, then for large enough T we have*

$$Prob.(\Lambda(\tilde{z}) > T - T^{\varepsilon}) > 1 - T^{-2\varepsilon+\delta}$$

6 From GF(2) to GF(q)

With the same arguments the results above can be generalized to the semi-infinite periodically repeated random sequences over any given finite field $GF(q)$, $q = p^m$, p prime.

Given T, let $z^T = z_0, z_1, \ldots, z_{T-1}$ be a random sequence of length T over $GF(q)$, and $\tilde{z}$ the semi-infinite sequence by periodically repeating z^T. Let $\mathcal{Z}$ be the sample space consisting of all the possible semi-infinite periodically repeated random sequences $\tilde{z}$, then $|\mathcal{Z}| = q^T$. We assume the elements in $\mathcal{Z}$ are equiprobable, *i.e.*, the probability of the occurrence of each $\tilde{z}$ is equal to q^{-T}. Now let n_d denote the order of q modulo d, then Theorem 1 extends to

Theorem 6 *Let $T = p^m T_1$, $\gcd(p, T_1) = 1$. Then*

$$E[\Lambda(\tilde{z})] = T - \sum_{d|T_1} \frac{\phi(d)(q^{n_d p^m} - 1)}{q^{n_d p^m}(q^{n_d} - 1)},$$

and

$$Var[\Lambda(\tilde{z})] = \sum_{d|T_1} \frac{\phi(d) n_d [q^{n_d(2p^m+1)} - (2p^m + 1)(q^{n_d(p^m+1)} - q^{n_d p^m}) - 1]}{q^{2n_d p^m}(q^{n_d} - 1)^2}.$$

And Theorem 2 extends to

Theorem 7 *Let $T = p^m T_1$, $\gcd(p, T_1) = 1$. Then*

$$E[\Lambda(\tilde{z})] > T - d(T_1) \geq T - d(T),$$

and

$$Var[\Lambda(\tilde{z})] < d(T_1)(1 + \log_q(1 + T_1)) \leq d(T)(1 + \log_q(1 + T)).$$

Hence all the other theorems over $GF(2)$ above can be extended to over $GF(q)$.

Acknowledgement

The authors are very grateful to Fred Piper for his invitation of visiting RHBNC, University of London, and for his valuable comments and suggestions about this work. The authors would like to thank the hospitality of the Department of Mathematics, RHBNC, University of London, where some of this work was undertaken. Thanks are also due to Mike Burmester for his help in improving the English of this paper.

References

[1] H. Beker and F. Piper, "Ciper Systems", Northwood Books, London, 1982.

[2] D. Gollman, "Linear Complexity of Sequences with Period p^n", Eurocrypt'86, A Workshop on the Theory and Application of Chryptographic Techniques, May 20-22, 1986, in Linköping, Sweden, pp. 3.2-3.3.

[3] G. H. Hardy and E. M. Wright, "An Introduction To The Theory of Numbers", Oxford, The Clarendon Press, 1938.

[4] R. Lidl and H. Niederreiter, "Finite Fields", *Encyclopaedia of Mathematics and its Applications* **20**, Reading, Mass, Addison-Wesley, 1983.

[5] J. L. Massey, "Shift-Register Synthesis and BCH decoding", IEEE Trans. on Info. Theory, Vol. IT-15, pp. 122-127, Jan. 1969.

[6] R. A. Rueppel, "New Approaches to Stream Ciphers", Ph.D.dissertation, Inst. of Telecommunications, Swiss Federal Inst. of Technol., Zurich, Dec. 1984.

[7] R. A. Rueppel, "Linear Complexity and Random Sequences", Presented at Eurocrypt'85

[8] R. A. Rueppel, "Analysis and Design of Stream Ciphers", Springer, Berlin-Heidelberg-New York-London-Paris-Tokyo: Springer-Verlag. 1986.