Measures of pseudorandomness: Arithmetic autocorrelation and correlation measure

Richard Hofer, László Mérai, Arne Winterhof

Dedicated to Robert F. Tichy on the occasion of his 60th birthday.

Abstract We prove a relation between two measures of pseudorandomness, the arithmetic autocorrelation and the correlation measure of order k. Roughly speaking, we show that any binary sequence with small correlation measure of order k up to a sufficiently large k cannot have a large arithmetic correlation. We apply our result to several classes of sequences including Legendre sequences defined with polynomials.

1 Introduction

Pseudorandom numbers are generated by deterministic algorithms and are not random at all. However, in contrast to truly random numbers they guarantee certain randomness properties. Their desirable features depend on the application area. For example, unpredictable sequences are needed for cryptography and uncorrelated sequences for wireless communication or radar. Some corresponding quality measures are linear complexity and expansion complexity for unpredictability and autocorrelation or more general correlation measure of order k.

Finding relations between different measures of pseudorandomness is an important goal. For example, the linear complexity provides essentially the same quality measure as certain lattice tests coming from the area of Monte Carlo methods, see [4, 20]. The correlation measure of order k is a rather general measure of pseudorandomness introduced by Mauduit and Sárközy [16]. A relation between linear complexity and the correlation measure of order k is given in [2]. Hence, we may

Richard Hofer, László Mérai, Arne Winterhof
Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenberger Str. 69, 4040 Linz, Austria
e-mail: {richard.hofer,laszlo.merai,arne.winterhof}@oeaw.ac.at

roughly say that correlation measure is a stronger measure than linear complexity. Expansion complexity introduced in [3] is another measure which is essentially the same as linear complexity in the periodic case but finer in the aperiodic case [18] (see also [14, 19]). There are many other related measures of pseudorandomness for sequences, see [11, 21, 22], and analyzing their hierarchy is very important.

In this paper we analyze the relation between another figure of merit coming from coding theory, the *arithmetic autocorrelation*, and the *correlation measures* of higher orders. Roughly speaking, we show that any binary sequence with small correlation measure of order k up to a sufficiently large k cannot have a large arithmetic autocorrelation. Correlation measure of order k and arithmetic autocorrelation are defined in the following paragraphs.

For a (purely) T-periodic binary sequence (a_n),

$$C_k(a_n) = \max_{0<d_1<\ldots<d_{k-1}<T} \left| \sum_{n=0}^{T-1} (-1)^{a_n+a_{n+d_1}+\cdots+a_{n+d_{k-1}}} \right|.$$

denotes the (periodic) *correlation measure of order* $k \geq 1$ of (a_n). $C_2(a_n)$ is also called autocorrelation of (a_n).

In this article we study a different notion of autocorrelation, the *arithmetic* autocorrelation introduced by Mandelbaum [15]. Sequences with small arithmetic autocorrelation can be used to define good error-correcting codes over the integers (instead of finite fields). Also, see the recent monograph by Goresky and Klapper [9] for more background and results on arithmetic correlations.

For an eventually T-periodic binary sequence (s_n) with preperiod T_0, that is $s_{n+T} = s_n$ for all $n \geq T_0$, the *imbalance* $Z(s_n)$ is defined by

$$Z(s_n) = N_0 - N_1,$$

where

$$N_i = |\{T_0 \leq n \leq T_0 + T - 1 : s_n = i\}|, \qquad i = 0, 1.$$

(Note that for a purely periodic sequence (a_n) we have $C_1(a_n) = |Z(a_n)|$.) The *arithmetic autocorrelation function* $A(t)$ of a (purely) T-periodic binary sequence (a_n) is defined as follows. For $t \in \{1, 2, \ldots, T-1\}$ let (a_{n+t}) be the shift of (a_n) by lag t. Put

$$x_t = \sum_{n=0}^{T-1} a_{n+t} 2^n \quad \text{and} \quad \alpha_t = \sum_{n=0}^{\infty} a_{n+t} 2^n, \qquad 0 \leq t < T.$$

Since α_t converges in the 2-norm of $\mathbb{Q}$ it holds (see [13])

$$\alpha_t = -\frac{x_t}{2^T - 1}, \qquad 0 \leq t < T.$$

We write

$$\alpha_0 - \alpha_t = \sum_{n=0}^{\infty} s_{n,t} 2^n \tag{1}$$

with unique $s_{n,t} \in \{0,1\}$. Note that $(s_{n,t})$ is (purely) periodic with period T if $x_0 \geq x_t$ and eventually periodic with period T from T on if $x_0 < x_t$ (see [13] for more details). In both cases we define

$$A(t) = Z(s_{n,t}), \qquad 1 \leq t \leq T-1.$$

In Section 2 we estimate the arithmetic autocorrelation of a binary sequence of period T in terms of correlation measures. In Section 3 we apply this result to several classes of sequences including Legendre sequences defined with polynomials.

2 A bound on the arithmetic autocorrelation

Theorem 1. *Put*

$$\Gamma_s = \max_{1 \leq l \leq s} C_l(a_n).$$

Then the arithmetic autocorrelation function of a T-periodic binary sequence (a_n) satisfies

$$A(t) \ll \min\left\{T^{1/2}\Gamma_{\lfloor \log T \rfloor}^{1/2}, 2^r \Gamma_{\lfloor \log T \rfloor} \log T\right\}$$

where $r = \min\{t, T-t\}$ for $1 \leq t \leq T-1$.

Proof. By the symmetry $A(t) = A(T-t)$ of the arithmetic autocorrelation (see [13, Proposition 1]) we may assume $1 \leq t \leq \lfloor T/2 \rfloor$. In the following we derive a lower bound on the number N_1 of ones in a period of the T-periodic sequence $(s_{n,t})$ defined by (1).

For a binary vector $(i_0, i_1, \ldots, i_{k-1}) \in \{0,1\}^k$ and lags $0 = d_1 < \ldots < d_{k-1} < T$, put

$$N = |\{0 \leq n \leq T-1 : a_n = i_0, a_{n+d_1} = i_1, \ldots, a_{n+d_{k-1}} = i_{k-1}\}|.$$

Similarly as in [17, Theorem 2] it can be shown that

$$\left|N - \frac{T}{2^k}\right| \leq \frac{1}{2^k} \sum_{l=1}^{k} \binom{k}{l} C_l(a_n). \tag{2}$$

Now take $c \in \{0,1\}$. For $k = 1, \ldots, m-1$ and $n = T, \ldots, 2T-1$ assume

$$\begin{aligned} &(a_{n-k}, a_{n-k+t}) = (c, 1-c), \\ &a_{n-k+j} = a_{n-k+j+t}, \quad j = 1, \ldots, k-1, \\ &(a_n, a_{n+t}) \in \{0,1\}^2. \end{aligned} \tag{3}$$

First we assume $m+1 \leq t \leq \lfloor T/2 \rfloor$. From (2) we know that (for fixed c) the number of patterns

$$\begin{pmatrix} a_{n-k} & a_{n-k+1} & \dots & a_{n-1} & a_n \\ a_{n-k+t} & a_{n-k+t+1} & \dots & a_{n-1+t} & a_{n+t} \end{pmatrix} \tag{4}$$

satisfying the assumptions (3) in

$$\begin{matrix} a_{T-k} & a_{T-k+1} & \dots & a_{T-1} & a_T & \dots & a_{2T-2} & a_{2T-1} \\ a_{t+T-k} & a_{t+T-k+1} & \dots & a_{t+T-1} & a_{t+T} & \dots & a_{t+2T-2} & a_{t+2T-1} \end{matrix} \tag{5}$$

is at least $T/2^{2k+2} - \Gamma_{2k+2}$. We have to distinguish between two cases.

If $c = 1$, then $(a_{n-k}, a_{n-k+t}) = (1,0)$. The subtraction of 0 from 1 gives no carry, no matter if there was a carry in the previous step. Hence

$$s_{n,t} = \begin{cases} 1 & \text{if } a_n \neq a_{n+t}, \\ 0 & \text{if } a_n = a_{n+t}. \end{cases}$$

Since there are 2^k possible choices for the pattern (4) we count at least $T/2^{k+2} - 2^k\Gamma_{2k+2}$ different $T \le n < 2T$ with $s_{n,t} = 1$.

If $c = 0$, then $(a_{n-k}, a_{n-k+t}) = (0,1)$. The subtraction of 1 from 0 gives a carry, no matter if there was a carry in the previous step. Hence

$$s_{n,t} = \begin{cases} 1 & \text{if } a_n = a_{n+t}, \\ 0 & \text{if } a_n \neq a_{n+t}. \end{cases}$$

Just as before there are 2^k possible choices for the pattern (4) and so we get at least $T/2^{k+2} - 2^k\Gamma_{2k+2}$ additional n with $s_{n,t} = 1$.

Thus in total we have at least $T/2^{k+1} - 2^{k+1}\Gamma_{2k+2}$ different $T \le n < 2T$ with $a_{n-k} \neq a_{n-k+t}$, $(a_{n-k+j}, a_{n-k+j+t}) \in \{(0,0),(1,1)\}$ for $j = 1, \dots, k-1$ and $s_{n,t} = 1$.

Summing up all the contributions we get

$$\begin{aligned} N_1 &\ge \frac{1}{2}\left(\sum_{k=1}^{m-1} 2^{-k}\right) T - 2\left(\sum_{k=1}^{m-1} 2^k \Gamma_{2k+2}\right) \ge \frac{1}{2}\left(\sum_{k=1}^{m-1} 2^{-k}\right) T - 2\Gamma_{2m}\left(\sum_{k=1}^{m-1} 2^k\right) \\ &\ge \frac{T}{2} - 2^{-m}T - 2^{m+1}\Gamma_{2m} \end{aligned}$$

where we used $\Gamma_s = \Gamma_{2k+2} \le \Gamma_{2m}$ since $k \le m-1$. Analogously N_0 can be bounded below by

$$N_0 \ge \frac{T}{2} - 2^{-m}T - 2^{m+1}\Gamma_{2m}$$

and therefore

$$|A(t)| = |N_0 - N_1| \le 2^{-m+1}T + 2^{m+2}\Gamma_{2m}$$

since

$$\begin{aligned} N_0 - N_1 &= T - 2N_1 \le 2^{-m+1}T + 2^{m+2}\Gamma_{2m}, \\ N_0 - N_1 &= 2N_0 - T \ge -(2^{-m+1}T + 2^{m+2}\Gamma_{2m}). \end{aligned}$$

Now we assume $1 \le t \le m$, that means some indices in (4) coincide and so we have to deal with shorter patterns. From (2) we know that (for fixed c) the number of patterns (4) satisfying the assumptions (3) in (5) is at least

$$\begin{aligned} &T/2^{2k+2} - \Gamma_{2k+2}, \qquad && k \le t-1, \\ &T/2^{k+t+1} - \Gamma_{k+t+1}, \qquad && k \ge t. \end{aligned}$$

Similarly as before if $c = 1$, then

$$s_{n,t} = \begin{cases} 1 & \text{if } a_n \ne a_{n+t}, \\ 0 & \text{if } a_n = a_{n+t}, \end{cases}$$

and if $c = 0$, then

$$s_{n,t} = \begin{cases} 1 & \text{if } a_n = a_{n+t}, \\ 0 & \text{if } a_n \ne a_{n+t}. \end{cases}$$

For each case we have 2^k possible choices for the pattern (4) if $k \le t-1$ and 2^{t-1} possible choices if $k \ge t$ and thus in total we count at least

$$\begin{aligned} &T/2^{k+1} - 2^{k+1}\Gamma_{2k+2}, \qquad && k \le t-1, \\ &T/2^{k+1} - 2^{t}\Gamma_{k+t+1}, \qquad && k \ge t, \end{aligned}$$

different $T \le n < 2T$ with $a_{n-k} \ne a_{n-k+t}$, $(a_{n-k+j}, a_{n-k+j+t}) \in \{(0,0),(1,1)\}$ for $j = 1, \ldots, k-1$ and $s_{n,t} = 1$.

Put $m' = 2m - t$. Summing up all the contributions we get

$$\begin{aligned} N_1 &\ge \frac{1}{2}\left(\sum_{k=1}^{m'-1} 2^{-k}\right) T - 2\left(\sum_{k=1}^{t-1} 2^k \Gamma_{2k+2} + 2^{t-1}\sum_{k=t}^{m'-1} \Gamma_{k+t+1}\right) \\ &\ge \frac{1}{2}\left(\sum_{k=1}^{m'-1} 2^{-k}\right) T - 2\Gamma_{2m}(2^t - 2) - 2^t \Gamma_{m+t}(m' - t) \\ &\ge \frac{T}{2} - 2^{-m'}T - 2^{t+1}\Gamma_{2m} - 2^t\Gamma_{2m}(m' - t) \\ &\ge \frac{T}{2} - 2^{-2m+t}T - 2^{t+1}(m - t + 1)\Gamma_{2m} \end{aligned}$$

where we used $\Gamma_{2k+2} \le \Gamma_{2m}$ or $\Gamma_{k+t+1} \le \Gamma_{m+t} \le \Gamma_{2m}$, respectively. Analogously N_0 can be bounded below by

$$N_0 \ge \frac{T}{2} - 2^{-2m+t}T - 2^{t+1}(m - t + 1)\Gamma_{2m}$$

and therefore

$$|A(t)| = |N_0 - N_1| \le 2^{-2m+t+1}T + 2^{t+2}(m - t + 1)\Gamma_{2m}.$$

Choosing

$$m = \left\lfloor \frac{1}{2} \log \frac{T}{\Gamma_{\lfloor \log T \rfloor}} \right\rfloor$$

we obtain the result. (Note that we may assume $\Gamma_{\lfloor \log T \rfloor} = o(T)$ and thus $m \geq 1$ since otherwise the result is trivial.) □

3 Applications

Now we apply our result to several classes of sequences.

For a prime $p > 2$ and a squarefree polynomial $f \in \mathbb{F}_p[x]$ with positive degree d let the p-periodic sequences (ℓ_n) be defined by

$$\ell_n = \begin{cases} 1 & \text{if } \left(\frac{f(n)}{p}\right) = 1, \\ 0 & \text{otherwise,} \end{cases} \qquad n \geq 0, \tag{6}$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol modulo p. For $f(n) = n$ these sequences are the Legendre sequences.

Corollary 1. *If $d < 0.5 \log p / \log\log p$ or 2 is a primitive root modulo p, then the arithmetic autocorrelation function of the p-periodic sequences (ℓ_n) defined by (6) satisfies*

$$A(t) \ll \min\left\{ d^{1/2} p^{3/4}, 2^r d p^{1/2} \log p \right\}$$

where $r = \min\{t, p-t\}$ for $1 \leq t \leq p-1$.

This result immediately follows from Theorem 1 and from the following proposition.

Proposition 1. *If f has no multiple zeros in the algebraic closure of $\mathbb{F}_p$ and*

(i) $k < p$ and 2 is a primitive root modulo p, or
(ii) $(4k)^d < p$,

then the (periodic) correlation measure of order k satisfies $C_k(\ell_n) \ll k d p^{1/2}$.

Proof. Similar to the proof of Theorem 1 in [10] (since we consider the periodic correlation measure of order k instead of the aperiodic one we lose the $\log p$ term). □

Let q be the power of an odd prime, α a primitive element of $\mathbb{F}_q$, and let η denote the quadratic character of $\mathbb{F}_q$. We denote by (a_n) the $(q-1)$-periodic *Sidelnikov-Lempel-Cohn-Eastman sequence* defined by

$$a_n = \begin{cases} 1 & \text{if } \eta(\alpha^n + 1) = 1, \\ 0 & \text{otherwise,} \end{cases} \qquad n \geq 0.$$

Similar to the proof in [2, Lemma 1] it follows that the (periodic) correlation measure of order k satisfies $C_k \ll kq^{1/2}$. From Theorem 1 we get

$$A(t) \ll \min\left\{q^{3/4}, 2^r q^{1/2}\log q\right\},$$

where $r = \min\{t, q-1-t\}$ for $1 \le t \le q-2$.

Let p be an odd prime, $\lambda \in \mathbb{F}_p^*$ of multiplicative order T and $f \in \mathbb{F}_p[x]$ a polynomial of positive degree d not of the form $cx^\alpha(g(x))^2$ with $c \in \mathbb{F}_p$, $\alpha \in \mathbb{N}$ and $g(x) \in \mathbb{F}_p[x]$. Define the T-periodic sequence (b_n) by

$$b_n = \begin{cases} 1 & \text{if } \left(\frac{f(\lambda^n)}{p}\right) = 1, \\ 0 & \text{otherwise,} \end{cases} \quad n \ge 0.$$

Similar to the proof in [12, Theorem 2] it follows that if T is a prime and either $(4k)^d \le T$ or 2 is a primitive root modulo T, then the (periodic) correlation measure of order k satisfies $C_k \ll kdp^{1/2}$. From Theorem 1 we get that if $d \le \log p/\log\log p$ or 2 is a primitive root modulo T, then

$$A(t) \ll \min\left\{d^{1/2}p^{1/4}T^{1/2}, 2^r dp^{1/2}\log T\right\}$$

where $r = \min\{t, T-t\}$ for $1 \le t \le T-1$.

Let $q = 2^k$ such that $q-1$ is prime, $f \in \mathbb{F}_q[x]$ be a polynomial of positive odd degree $d \ge \log q$ such that the coefficients of its terms are zero if and only if the term has an even exponent. Let $\mathrm{Tr} : \mathbb{F}_{2^k} \to \mathbb{F}_2$ denote the absolute trace function of $\mathbb{F}_{2^k}$ and α be a primitive element of $\mathbb{F}_{2^k}$. Define the $q-1$-periodic sequence (d_n) by

$$d_n = \mathrm{Tr}(f(\alpha^n)), \quad n \ge 0.$$

Similar to the proof in [5, Theorem 4] it follows that if $k \le d+1$ the (periodic) correlation measure of order k satisfies $C_k \ll dq^{1/2}$. In the same way as Corollary 1 we get

$$A(t) \ll \min\left\{d^{1/2}q^{3/4}, 2^r dq^{1/2}\log q\right\}$$

where $r = \min\{t, q-1-t\}$ for $1 \le t \le q-2$.

4 Final remarks

For fixed $1 \le t < T$, Goresky and Klapper [7, 8] proved that the expected arithmetic autocorrelation, averaged over all binary sequences of period T, is

$$\frac{T}{2^{T-\gcd(t,T)}}.$$

Sequences with *ideal arithmetic autocorrelation* equal to zero for all nontrivial shifts t are known, see [6]. However, the maximum absolute value of the (classical) autocorrelation of these so-called *ℓ-sequences* equals the period since the second half of a period is the bit-wise complement of the first half [6, Proposition 1]. Hence, these sequences are far away from looking random. In contrast to these sequences, the sequences studied in Section 3 still guarantee a rather small arithmetic autocorrelation with respect to their periods if the period is sufficiently large.

Actually, the correlation measure of order k was defined for finite sequences. Analogs of our results for finite sequences can be easily obtained with the obvious definition of arithmetic autocorrelation. Moreover, for a truly random sequence of length T, Alon et al. [1] showed that the correlation measure of order k is of order of magnitude $k^{1/2}T^{1/2}\log^{1/2}T$ and thus its arithmetic autocorrelation is at most of order of magnitude $T^{3/4}\log^{1/4}T$.

The correlation measure of order k was also introduced and analyzed for non-binary sequences, see [17], as well as the arithmetic autocorrelation for non-binary sequences, see [7]. A similar relation as Theorem 1 can be easily obtained in a similar way, that is, reduce first the problem of estimating the arithmetic autocorrelation to the pattern distribution of the sequence which can be estimated by the correlation measure of order k.

Acknowledgements The authors are partially supported by the Austrian Science Fund FWF Project 5511-N26 which is part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications".

References

1. Alon, N., Kohayakawa, Y., Mauduit, C., Moreira, C. G., Rödl, V.: Measures of pseudorandomness for finite sequences: typical values. Proc. Lond. Math. Soc. (3) **95**, 778–812 (2007)
2. Brandstätter, N., Winterhof, A.: Linear complexity profile of binary sequences with small correlation measure. Period. Math. Hungar. **52**, 1–8 (2006)
3. Diem, C.: On the use of expansion series for stream ciphers. LMS, J. Comput. Math. **15**, 326–340 (2012)
4. Dorfer, G., Winterhof A.: Lattice structure and linear complexity profile of nonlinear pseudorandom number generators. Appl. Algebra Engrg. Comm. Comput. **13**, 499–508 (2003)
5. Folláth, J.: Construction of pseudorandom binary sequences using additive characters over $GF(2^k)$. Period. Math. Hungar. **57**, 73–81 (2008)
6. Goresky, M., Klapper, A.: Arithmetic crosscorrelations of feedback with carry shift register sequences. IEEE Trans. Inform. Theory **43**, 1342–1345 (1997)
7. Goresky, M., Klapper, A.: Some results on the arithmetic correlation of sequences (extended abstract). Sequences and their applications–SETA 2008, 71–80, Lecture Notes in Comput. Sci. **5203**, Springer, Berlin (2008)
8. Goresky, M., Klapper, A.: Statistical properties of the arithmetic correlation of sequences. Internat. J. Found. Comput. Sci. **22**, 1297–1315 (2011)
9. Goresky, M., Klapper, A.: Algebraic shift register sequences. Cambridge University Press, Cambridge (2012)
10. Goubin, L., Mauduit, C., Sárközy A.: Construction of large families of pseudorandom binary sequences. J. Number Theory **106**, 56–69 (2004)

11. Gyarmati, K.: Measures of pseudorandomness. In: Charpin, P., Pott, A., Winterhof, A. (eds) Finite fields and their applications, volume 11 of Radon Ser. Comput. Appl. Math., pp. 43–64. de Gruyter, Berlin (2013)
12. Gyarmati, K., Pethő, A., Sárközy, A.: On linear recursion and pseudorandomness. Acta Arith. **118**, 359–374 (2005)
13. Hofer, R., Winterhof, A.: On the arithmetic autocorrelation of the Legendre sequence. Adv. Math. Commun. (to appear)
14. Hofer, R., Winterhof, A.: Linear complexity and expansion complexity of some number theoretic sequences. Lecture Notes in Comput. Sci. (to appear)
15. Mandelbaum, D.: Arithmetic codes with large distance. IEEE Trans. Inform. Theory **13**, 237–242 (1967)
16. Mauduit, C., Sárközy, A.: On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the Legendre symbol. Acta Arith. **82**, 365–377 (1997)
17. Mauduit, C., Sárközy, A.: On finite pseudorandom sequences of k symbols. Indag. Math. **13**, 89–101 (2002)
18. Mérai, L., Niederreiter, H., Winterhof, A.: Expansion complexity and linear complexity of sequences over finite fields. Cryptogr. Commun. (2016) doi: 10.1007/s12095-016-0189-2
19. Mérai, L., Winterhof, A.: On the Nth linear complexity of p-automatic sequences over $\mathbb{F}_p$. (Preprint 2016)
20. Niederreiter, H., Winterhof, A.: Lattice structure and linear complexity of nonlinear pseudorandom numbers. Appl. Algebra Engrg. Comm. Comput. **13**, 319–326 (2002)
21. Rivat, J., Sárközy, A.: On pseudorandom sequences and their application. In: General theory of information transfer and combinatorics, Lecture Notes in Comput. Sci. **4123** pp. 343–361, Springer, Berlin (2006)
22. Topuzoğlu, A., Winterhof, A.: Pseudorandom sequences. In Topics in geometry, coding theory and cryptography, Algebr. Appl. **6**, pp. 135–166, Springer, Dordrecht (2007)