Discrete Applied Mathematics 214 (2016) 153–168

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

The cross-correlation measure of families of finite binary sequences: Limiting distributions and minimal values

László Mérai

Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstr. 69, 4040 Linz, Austria

ARTICLE INFO

Article history:
Received 24 August 2015
Received in revised form 21 June 2016
Accepted 24 June 2016
Available online 21 July 2016

Keywords:
Pseudorandom sequences
Binary sequence
Correlation measure
Cross-correlation measure

ABSTRACT

Gyarmati, Mauduit and Sárközy introduced the *cross-correlation measure* $\Phi_k(G)$ *of order* k to measure the level of pseudorandom properties of families of finite binary sequences. In an earlier paper we estimated the cross-correlation measure of a random family of binary sequences. In this paper, we sharpen these earlier results by showing that for random families, the cross-correlation measure converges strongly, and so has limiting distribution. We also give sharp bounds to the minimum values of the cross-correlation measure, which settles a problem of Gyarmati, Mauduit and Sárközy nearly completely.

1. Introduction

Recently, in a series of papers the pseudorandomness of *finite binary sequences* $E_N = (e_1, \dots, e_N) \in \{-1, 1\}^N$ has been studied. In particular, measures of pseudorandomness have been defined and investigated; see [3,6,9,11] and the references therein.

For example, Mauduit and Sárközy [11] introduced the *correlation measure* $C_k(E_N)$ *of order* k of the binary sequence E_N. Namely, for a k-tuple $D = (d_1, \dots, d_k)$ with non-negative integers $0 \le d_1 < \cdots < d_k < N$ and $M \in \mathbb{N}$ with $M + d_k \le N$ write

$$V_k(E_N, M, D) = \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_k}.$$

Then $C_k(E_N)$ is defined as

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_k} \right|.$$

This measure has been widely studied, see, for example [1–3,6,8,12,17]. In particular, Alon, Kohayakawa, Mauduit, Moreira and Rödl [3] obtained the typical order of magnitude of $C_k(E_N)$. They proved that, if E_N is chosen uniformly from $\{-1, +1\}^N$, then for all $0 < \varepsilon < 1/16$ the probability that

$$\frac{2}{5}\sqrt{N \log \binom{N}{k}} < C_k(E_N) < \frac{7}{4}\sqrt{N \log \binom{N}{k}}$$

E-mail address: merai@cs.elte.hu.

http://dx.doi.org/10.1016/j.dam.2016.06.024
0166-218X/

holds for every integer $2 \le k \le N/4$ is at least $1-\varepsilon$ if N is large enough. (Here, and in what follows, we write log for the natural logarithm, and $\log_a$ for the logarithm to base a.)

They also showed in [3], that the correlation measure $C_k(E_N)$ is concentrated around its mean $\mathbb{E}[C_k]$. Namely, for all $\varepsilon > 0$ and integer function $k = k(N)$ with $2 \le k \le \log N - \log\log N$ the probability that

$$1-\varepsilon < \frac{C_k(E_N)}{\mathbb{E}[C_k]} < 1+\varepsilon$$

holds is at least $1-\varepsilon$ if N is large enough.

Recently, K.-U. Schmidt studied the limiting distribution of $C_k(E_N)$ [17]. He showed that if $e_1, e_2, \ldots \in \{-1,+1\}$ are chosen independently and uniformly, then for fixed k

$$\frac{C_k(E_N)}{\sqrt{2N\log\binom{N}{k-1}}} \to 1 \quad \text{almost surely},$$

as $N \to \infty$, where $E_N = (e_1, \ldots, e_N)$.

Let us now turn to the minimal value of $C_k(E_N)$. Clearly,

$$\min\{C_k(E_N) : E_N \in \{-1,+1\}\} = 1 \quad \text{for odd } k,$$

where the minimum is reached by the alternating sequence $(1,-1,1,-1,\ldots)$. However, for even order, Alon, Kohayakawa, Mauduit, Moreira and Rödl [2] showed that

$$\min\{C_{2k}(E_N) : E_N \in \{-1,+1\}\} > \sqrt{\frac{1}{2}\left\lfloor\frac{N}{2k+1}\right\rfloor}, \tag{1}$$

see also [17].

In order to study the pseudorandomness of *families* of finite binary sequences instead of single sequences, Gyarmati, Mauduit and Sárközy [10] introduced the notion of the *cross-correlation measure* (see also the survey paper [15]).

Definition 1. For positive integers N and S, consider a map

$$G_{N,S} : \{1,2,\ldots,S\} \to \{-1,+1\}^N,$$

and write $G_{N,S}(s) = (e_1(s), \ldots, e_N(s)) \in \{-1,1\}^N$ $(1 \le s \le S)$.

The *cross-correlation measure* $\Phi_k\left(G_{N,S}\right)$ *of order* k of $G_{N,S}$ is defined as

$$\Phi_k\left(G_{N,S}\right) = \max\left|\sum_{n=1}^{M} e_{n+d_1}(s_1)\cdots e_{n+d_k}(s_k)\right|,$$

where the maximum is taken over all integers $M, d_1, \ldots, d_k$ and $1 \le s_1, \ldots, s_k \le S$ such that $0 \le d_1 \le d_2 \le \cdots \le d_k < M + d_k \le N$ and $d_i \ne d_j$ if $s_i = s_j$.

We remark that in [10] only injective maps $G_{N,S}$ were considered, and the cross-correlation measure is defined for the families $\mathcal{F} = \{G_{N,S}(s) : s = 1,2,\ldots,S\}$ of size S.

The typical order of magnitude of $\Phi_k\left(G_{N,S}\right)$ was established in [14] for large range of k and for random maps $G_{N,S}$, i.e. when all $e_n(s) \in \{-1,+1\}$ $(1 \le n \le N, 1 \le s \le S)$ are chosen independently and uniformly.

Theorem 1. *For a given $\varepsilon > 0$, there exists N_0, such that if $N > N_0$ and $1 \le \log_2 S < N/12$, then we have with probability at least $1-\varepsilon$, that*

$$\frac{2}{5}\sqrt{N\left(\log\binom{N}{k} + k\log S\right)} < \Phi_k\left(G_{N,S}\right) < \frac{5}{2}\sqrt{N\left(\log\binom{N}{k} + k\log S\right)}$$

for every integer k with $2 \le k \le N/(6\log_2 S)$.

Our first result tells that analogously to the correlation measure of binary sequences, the cross-correlation measure of families $\Phi_k\left(G_{N,S}\right)$ is concentrated around its mean $\mathbb{E}\left[\Phi_k\left(G_{N,S}\right)\right]$ if k is small enough.

Theorem 2. *For any fixed constant $\varepsilon > 0$ and any integer function $k = k(N)$ with $2 \le k \le (\log N + \log S)/\log\log N$, there is a constant $N_0 \ge 12\log_2 S$ for which the following holds. If $N \ge N_0$, then the probability that*

$$1-\varepsilon < \frac{\Phi_k(G_{N,S})}{\mathbb{E}\left[\Phi_k(G_{N,S})\right]} < 1+\varepsilon$$

holds is at least $1-\varepsilon$.