Available online at www.sciencedirect.com
ScienceDirect
Discrete Mathematics 308 (2008) 6203–6209
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

A technique to study the correlation measures of binary sequences

Venkat Anantharam*

EECS Department, University of California, Berkeley, CA 94720, USA

Received 21 November 2005; received in revised form 19 June 2007; accepted 24 November 2007
Available online 27 December 2007

Abstract

Let $E^N = (e_1, e_2, \ldots, e_N)$ be a binary sequence with $e_i \in \{+1, -1\}$. For $2 \le k \le N$, the correlation measure of order k of the sequence is defined by Mauduit and Sárközy as

$$C_k(E^N) = \max_{M, d_1, \ldots, d_k} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \ldots e_{n+d_k} \right|$$

where the maximum is taken over all $M \ge 1$ and $0 \le d_1 < d_2 < \ldots < d_k$ such that $M + d_k \le N$. These measures have been extensively studied over the last decade. Several inequalities for these measures (that hold for all E^N for all large enough N) have been proved, and others conjectured. Further, these measures have been estimated for various special sequences E^N.

Fix $M \ge 1$ and $L \ge 1$. For $1 \le a \le L$, let $E^M[a] = (e_1[a], \ldots, e_M[a])$ be a binary sequence with $e_i[a] \in \{+1, -1\}$. For $2 \le k \le L$ we define the correlation measure of order k of the family of sequences $E^M[1 : L] = \{E^M[1], \ldots, E^M[L]\}$ as

$$C_k(E^M[1 : L]) = \max_{1 \le a_1 < a_2 < \cdots < a_k \le L} \left| \sum_{i=1}^{M} e_i[a_1] e_i[a_2] \ldots e_i[a_k] \right|.$$

We use these new correlation measures as a vehicle to study the correlation measures introduced by Mauduit and Sárközy.

Alon, Kohayakawa, Mauduit, Moreira, and Rödl recently proved that for each $k \ge 1$ there is an absolute constant $c_{2k} > 0$ such that $C_{2k}(E^N) \ge c_{2k}\sqrt{N}$ for all E^N for all large enough N. thus answering a question of Cassaigne, Mauduit, and Sárközy (in stronger form than an earlier result of Kohayakawa, Mauduit, Moreira, and Rödl). We prove a lower bound on the even correlation measures $C_{2k}(E^M[1 : L])$ when $L > k(2k-1)M$ and use it to provide an alternate proof of this result. The constant c_{2k} in our proof is better than that of Alon, Kohayakawa, Mauduit, Moreira, and Rödl for $k = 1$, but poorer for all $k \ge 2$.

We study $C_3(E^N)$ via $C_3(E^M[1 : L])$. This allows us to strengthen a recent result of Gyarmati which relates $C_3(E^N)$ and $C_2(E^N)$. We prove that given any $\kappa > 0$ there is an associated $c > 0$ (depending only on κ) such that, for all sufficiently large N, if $C_2(E^N) \le \kappa N^{2/3}$ we have $C_3(E^N) \ge c\sqrt{N}$. This also answers a question of Gyarmati.

Finally, the study of $C_3(E^M[1 : L])$ allows us to verify a conjecture of Mauduit. We prove that there is an absolute constant $c > 0$ such that $C_2(E^N)C_3(E^N) \ge cN$ for all E^N for all large enough N.

Keywords: Binary sequences; Correlation measures; Pseudorandomness

* Tel.: +1 510 643 8435(O); fax: +1 510 642 2845.
E-mail address: ananth@eecs.berkeley.edu.

0012-365X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2007.11.043

1. Discussion

Let $E^N = (e_1, e_2, \ldots, e_N)$ be a binary sequence with $e_i \in \{+1, -1\}$. For $2 \leq k \leq N$, the correlation measure of order k of the sequence is defined by Mauduit and Sárközy [7] as

$$C_k(E^N) = \max_{M, d_1, \ldots, d_k} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \ldots e_{n+d_k} \right|$$

where the maximum is taken over all $M \geq 1$ and $0 \leq d_1 < d_2 < \ldots < d_k$ such that $M + d_k \leq N$. In this paper we develop a simple technique for proving some lower bounds on and relations between these correlation measures.

Fix $1 \leq M \leq N$ and $1 \leq L \leq N - M + 1$. For $1 \leq a \leq L$ define the binary sequence $E^M[a] = (e_1[a], \ldots, e_M[a])$ by setting

$$e_i[a] = e_{a+i-1}. \tag{1}$$

For $2 \leq k \leq L$ and $1 \leq a_1 < a_2 < \ldots < a_k \leq L$ observe that

$$e_i[a_1] e_i[a_2] \ldots e_i[a_k] = e_{a_1+i-1} e_{a_2+i-1} \ldots e_{a_k+i-1}.$$

Thus

$$C_k(E^N) \geq \left| \sum_{i=1}^{M} e_i[a_1] e_i[a_2] \ldots e_i[a_k] \right|.$$

This motivates us to make the following definition. For arbitrary $M \geq 1$ and $L \geq 1$, for $1 \leq a \leq L$, let $E^M[a] = (e_1[a], \ldots, e_M[a])$ be an arbitrary binary sequence with $e_i[a] \in \{+1, -1\}$. Then, for $2 \leq k \leq L$ we define the correlation measure of order k of the family of sequences $E^M[1 : L] = \{E^M[1], \ldots, E^M[L]\}$ as

$$C_k(E^M[1 : L]) = \max_{1 \leq a_1 < a_2 < \ldots < a_k \leq L} \left| \sum_{i=1}^{M} e_i[a_1] e_i[a_2] \ldots e_i[a_k] \right|.$$

We thus have

$$C_k(E^N) \geq C_k(E^M[1 : L]) \tag{2}$$

for any $1 \leq M \leq N$ and $k \leq L \leq N - M + 1$, when $E^M[1 : L]$ is constructed from E^N as in Eq. (1). Hence, finding estimates on and relations between the correlation measures of the type $C_k(E^M[1 : L])$ for arbitrary $E^M[1 : L]$ will yield corresponding results for the correlation measures $C_k(E^N)$ of Mauduit and Sárközy.

It should be mentioned that other notions of pseudorandomness for families of binary sequences have been introduced by Ahlswede, Khachatrian, Mauduit, and Sárközy [1].

In the following, we will occasionally use the notation $f(E^N) \gg g(N)$ where $f(E^N)$ is a nonnegative function of correlation measures of E^N and $g(N)$ is a nonnegative function. This should be understood to mean that there is an absolute constant $c > 0$ such that $f(E^N) \geq cg(N)$ for all E^N for all sufficiently large N.

2. Proof of $C_{2k}(E^N) \gg \sqrt{N}$

We will first illustrate the power of the viewpoint provided by the newly defined correlation measures by giving an elementary proof that for each $k \geq 1$ there is an absolute constant $c_{2k} > 0$ with $C_{2k}(E^N) \geq c_{2k}\sqrt{N}$ for all E^N for all large enough N. As mentioned in the abstract, this result is not new. It was conjectured by Cassaigne, Mauduit, and Sárkozy [3] (see Problem 2 on pg. 107 and the discussion on pp. 109–110 of [3]) that for some absolute constants $d > 0$ and $c > 0$ we have $C_2(E^N) \geq cN^d$ for all E^N for all large enough N. Recently, Alon, Kohayakawa, Mauduit, Moreira, and Rödl [2] proved that $C_{2k}(E^N) \geq \sqrt{\frac{1}{2}\lfloor \frac{N}{2k+1} \rfloor}$ for all E^N for $1 \leq k \leq \lfloor \frac{N}{2} \rfloor$, thus answering the question of [3] in stronger form than an earlier result of Kohayakawa, Mauduit, Moreira, and Rödl [5].

We get a better constant for $k = 1$ than that of [2], but poorer constants for all $k \geq 2$. In general we have not concerned ourselves with optimizing constants. It should be noted that our proof is elementary and the result is broader in that it also yields, for all $L > k(2k-1)M$, lower bounds on $C_{2k}(E^M[1 : L])$ that apply for all $E^M[1 : L]$.