<table>
<thead>
<tr>
<th>Title</th>
<th>On the variance of average distance of subsets in the Hamming space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Fu, Fang-Wei; Ling, San; Xing, Chaoping</td>
</tr>
<tr>
<td>Date</td>
<td>2004</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/9840">http://hdl.handle.net/10220/9840</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2004 Elsevier B.V. This is the author created version of a work that has been peer reviewed and accepted for publication by Discrete Applied Mathematics, Elsevier B.V. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<a href="http://dx.doi.org/10.1016/j.dam.2004.08.004">http://dx.doi.org/10.1016/j.dam.2004.08.004</a>].</td>
</tr>
</tbody>
</table>
On the variance of average distance of subsets in the Hamming space

Fang-Wei Fu\textsuperscript{a,1}, San Ling\textsuperscript{b}, Chaoping Xing\textsuperscript{b}

\textsuperscript{a}Temasek Laboratories, National University of Singapore, 5 Sports Drive 2, Singapore 117508, Singapore
\textsuperscript{b}Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543, Singapore

Received 13 September 2002; received in revised form 23 August 2004; accepted 31 August 2004

Abstract

Let $V$ be a finite set with $q$ distinct elements. For a subset $C$ of $V^n$, denote $\text{var}(C)$ the variance of the average Hamming distance of $C$. Let $T(n, M; q)$ and $R(n, M; q)$ denote the minimum and maximum variance of the average Hamming distance of subsets of $V^n$ with cardinality $M$, respectively. In this paper, we study $T(n, M; q)$ and $R(n, M; q)$ for general $q$. Using methods from coding theory, we derive upper and lower bounds on $\text{var}(C)$, which generalize and unify the bounds for the case $q = 2$. These bounds enable us to determine the exact value for $T(n, M; q)$ and $R(n, M; q)$ in several cases.

Keywords: Hamming space; Subsets; Average distance; Variance; Codes; Distance distribution

1. Introduction

Let $V = \{v_1, v_2, \ldots, v_q\}$ be a finite set with $q$ distinct elements, where $q$ is a positive integer. Let $V^n$ be the set of ordered $n$-tuples over $V$. The Hamming distance between two vectors $a$ and $b$ is the number of components where they differ, and is denoted by $d_H(a, b)$. Let $C$ be a subset of $V^n$ with size $|C| = M$. The average Hamming distance of $C$ is defined by

$$\bar{d}(C) = \frac{1}{M^2} \sum_{a \in C} \sum_{b \in C} d_H(a, b).$$

(1.1)

The variance of the average distance of $C$ is defined by

$$\text{var}(C) = \frac{1}{M^2} \sum_{a \in C} \sum_{b \in C} [d_H(a, b) - \bar{d}(C)]^2.$$  \hspace{1cm} (1.2)

It is easy to check that

$$\text{var}(C) = \frac{1}{M^2} \sum_{a \in C} \sum_{b \in C} [d_H(a, b)]^2 - [\bar{d}(C)]^2.$$ \hspace{1cm} (1.3)

1 On leave from the Department of Mathematics, Nankai University, Tianjin 300071, P. R. China.

E-mail addresses: tslufw@nus.edu.sg (F.-W. Fu), matlings@nus.edu.sg (S. Ling), matxcp@nus.edu.sg (C. Xing).
The minimum and maximum average Hamming distance of a subset of $V^n$ with size $M$ are defined by
\[ \beta(n, M; q) = \min \{|d(C)| \mid C \text{ is a subset of } V^n \text{ with size } |C| = M \}, \]
\[ \phi(n, M; q) = \max \{|d(C)| \mid C \text{ is a subset of } V^n \text{ with size } |C| = M \}. \]

The minimum and maximum variance of the average distance of a subset of $V^n$ with size $M$ are defined by
\[ T(n, M; q) = \min \{\text{var}(C) \mid C \text{ is a subset of } V^n \text{ with size } |C| = M \}, \]
\[ R(n, M; q) = \max \{\text{var}(C) \mid C \text{ is a subset of } V^n \text{ with size } |C| = M \}. \]

Ahlswede and Katona [2] first posed the problem of determining $\beta(n, M; q)$ on the extremal combinatorics of Hamming space. There are a number of papers (see [1–4, 8–13, 15, 16]) dealing with this topic thereafter, and some exact values of $\beta(n, M; q)$ are determined. It is still an open problem to determine $\beta(n, M; q)$ for general $n$, $q$, and $1 \leq M \leq q^n$. Ahlswede and Althöfer [1] observed that this problem also occurs in the construction of good codes for write-efficient memories, introduced by Ahlswede and Zhang [3] as a model for storing and updating information on a rewritable medium with cost constraints. Kündgen [12] observed that this problem is equivalent to a covering problem in graph theory. Ahlswede and Katona [2] first mentioned the problem of determining $\phi(n, M; q)$ for $q=2$ and gave a simple solution. Fu and Xing [11] gave a complete solution for the problem of determining $\phi(n, M; q)$ for general $q$. Since the variance is an important digital characteristic for the average distance, Fu and Shen [9] first posed the problem of determining $T(n, M; q)$ for general $q$. Xia [14] and Fu [15] improved the lower and upper bounds of Fu and Shen on $\text{var}(C)$. Furthermore, they determined the exact values of $T(n, 2^{n-1} - 1; 2)$ and $T(n, 2^{n-1} + 1; 2)$.

In this paper, we study $T(n, M; q)$ and $R(n, M; q)$ for general $q$. Using methods from coding theory, we derive upper and lower bounds on $\text{var}(C)$, which generalize and unify the bounds for the case $q=2$. These bounds enable us to determine the exact value for $T(n, M; q)$ and $R(n, M; q)$ in several cases.

Without loss of generality, below we assume that $V = \mathbb{Z}_q = \{0, 1, \ldots, q-1\}$, the abelian group under addition modulo $q$, since we only deal with the Hamming distance in the Hamming space $V^n$. Furthermore, if $q$ is a prime power, we can assume that $V = \mathbb{F}_q$, the finite field of $q$ elements. The Hamming weight $w_H(a)$ of a vector $a$ in $\mathbb{Z}_q^n$ or $\mathbb{F}_q^n$ is the number of nonzero coordinates in $a$. Obviously, for $a, b \in \mathbb{Z}_q^n$ or $\mathbb{F}_q^n$,
\[ d_H(a, b) = w_H(a - b). \]

If $q$ is a prime power, denote
\[ \Gamma = \{(c_1, c_2, \ldots, c_n) \mid c_i \in \mathbb{F}_q \text{ and } c_1 + c_2 = 0\}, \]
\[ \Gamma^+ = \Gamma \cup \{(0, 1, 0, \ldots, 0)\}, \quad \Gamma^- = \Gamma \setminus \{(0, 0, 0, \ldots, 0)\}. \]

If $q$ is a positive integer and $q \geq 2$, denote
\[ A = \mathbb{Z}_q^{n-1} \times \{0\}, \quad A^+ = A \cup \{(0, 0, 0, \ldots, 0, 1)\}, \quad A^- = A \setminus ((0, 0, 0, \ldots, 0, 1)). \]

Our main results in this paper are given as follows.

**Theorem 1.** For $2 \leq q \leq 4$, we have
\[ R(n, q^{n-1}; q) = \text{var}(\Gamma), \] (1.4)
\[ R(n, q^{n-1} + 1; q) = \text{var}(\Gamma^+), \] (1.5)
\[ R(n, q^{n-1} - 1; q) = \text{var}(\Gamma^-). \] (1.6)

**Theorem 2.** If $q \geq 2$, we have
\[ T(n, q^{n-1}; q) = \text{var}(A), \] (1.7)
\[ T(n, q^{n-1} - 1; q) = \text{var}(A^-). \] (1.8)

If $n \geq 3$ or $2 \leq q \leq 4$, we have
\[ T(n, q^{n-1} + 1; q) = \text{var}(A^+). \] (1.9)
The exact values of \( \text{var}(\Gamma), \text{var}(\Gamma^+), \text{var}(\Gamma^-), \text{var}(A), \text{var}(A^+) \) and \( \text{var}(A^-) \) will be computed in Section 3. It seems to be difficult to determine \( T(n, M; q) \) and \( R(n, M; q) \) in general. In particular, it is interesting to know whether Theorem 1 is still true for \( q \) being a prime power and \( q \geq 5 \).

This paper is organized as follows. In Section 2, in order to establish our results, we review some basic properties of distance distributions of codes. In Section 3, we compute \( \text{var}(C) \) for some subsets. In Section 4, we derive an upper bound on \( \text{var}(C) \) for \( 2 \leq q \leq 4 \). Theorem 1 is proved by showing that this upper bound is tight for some cases. In Section 5, we derive a lower bound on \( \text{var}(C) \) for general \( q \). Theorem 2 is proved by showing that this lower bound is tight for some cases.

2. Preliminaries

In this section, we review some basic properties of distance distributions of codes.

For a subset \( C \) of \( V^n \) with size \( |C| = M \), we call \( C \) an \((n, M; q)\) code in coding theory. The distance distribution of \( C \) is defined by

\[
A_i = \frac{1}{M} |\{(a, b) \mid a, b \in C, d_H(a, b) = i\}|, \quad i = 0, 1, \ldots, n.
\]

The dual distance distribution of \( C \) is defined by

\[
B_k = \frac{1}{M} \sum_{j=0}^{n} K_k(j; q) A_j, \quad k = 0, 1, \ldots, n,
\]

where \( K_k(j; q) \) are the \( q \)-ary Krawtchouk numbers defined by

\[
K_k(j; q) = \sum_{i=0}^{k} (-1)^i (q - 1)^{k-i} \binom{j}{i} \binom{n-j}{k-i}.
\]

The distance enumerator of \( C \) is defined as

\[
WC(x) = \sum_{i=0}^{n} A_i x^i
\]

and the dual distance enumerator of \( C \) is defined as

\[
\hat{WC}(x) = \sum_{i=0}^{n} B_i x^i.
\]

The MacWilliams–Delsarte identity (see [14]) gives the relationship between \( WC(x) \) and \( \hat{WC}(x) \):

\[
\hat{WC}(x) = \frac{1}{M} [1 + (q - 1)x]^n WC \left( \frac{1 - x}{1 + (q - 1)x} \right),
\]

\[
WC(x) = \frac{M}{q^n} [1 + (q - 1)x]^n \hat{WC} \left( \frac{1 - x}{1 + (q - 1)x} \right).
\]

It is easy to see that

\[
WC(0) = A_0 = 1, \quad WC(1) = \sum_{i=0}^{n} A_i = M.
\]

By (1.1) and (1.3), the average Hamming distance of \( C \) is given by

\[
\bar{d}(C) = \frac{1}{M} \sum_{i=1}^{n} i A_i
\]

and the variance of \( \bar{d}(C) \) is given by

\[
\text{var}(C) = \frac{1}{M} \sum_{i=1}^{n} i^2 A_i - [\bar{d}(C)]^2.
\]
Delsarte (see [6,7]) showed that

\[ B_k \geq 0, \quad k = 0, 1, \ldots, n. \]  

(2.9)

Let \( x = 0 \) and 1 in (2.4), respectively, we obtain by (2.6) that

Lemma 1.

\[ B_0 = 1, \quad \sum_{i=0}^{n} B_i = \frac{q^n}{M}. \]  

(2.10)

Lemma 2.

\[ \bar{d}(C) = \frac{n(q - 1)}{q} - \frac{B_1}{q}, \]  

(2.11)

\[ \text{var}(C) = \frac{n(q - 1)}{q^2} + \frac{(q - 2)}{q^2} B_1 - \frac{1}{q^2} (B_1)^2 + \frac{2}{q^2} B_2. \]  

(2.12)

Proof. Eq. (2.11) is obtained by differentiating (2.5), putting \( x = 1 \) and combining with (2.7). Eq. (2.12) is obtained by differentiating (2.5) twice, putting \( x = 1 \) and using (2.8) and (2.11). \( \Box \)

Ashikhmin and Simonis [5] showed that

Lemma 3. Let \( C \) be an \( (n, M; q) \) code. The dual distance distribution of \( C \) is given by \( B_0, B_1, \ldots, B_n \). If \( M \equiv x \pmod{q} \), then

\[ B_k \geq \frac{1}{M^2} x(q - x) (q - 1)^{k-1} \binom{n}{k}, \quad k = 1, 2, \ldots, n. \]  

(2.13)

If \( q \) is a prime power and \( V = F_q \), the finite field with \( q \) elements, linear codes over \( F_q \) can be introduced. A \( q \)-ary code \( C \) is called a \( q \)-ary \( [n, k] \) linear code if \( C \) is a \( k \)-dimensional subspace of \( F_q^n \). For two vectors

\[ a = (a_1, a_2, \ldots, a_n) \in F_q^n, \quad b = (b_1, b_2, \ldots, b_n) \in F_q^n, \]

the scalar product of \( a \) and \( b \) is defined as

\[ a \cdot b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n. \]

Let \( C \) be a \( q \)-ary \( [n, k] \) linear code. The set

\[ C^\perp = \{ x \in F_q^n : x \cdot c = 0 \text{ for all } c \in C \} \]

is called the dual code of \( C \). Let \( A_i \) be the number of codewords in \( C \) of Hamming weight \( i \). The sequence of numbers \( A_0, A_1, \ldots, A_n \) is called the weight distribution of \( C \). It is well known in coding theory that for a linear code \( C \), the distance distribution of \( C \) is equal to the weight distribution of \( C \), and the dual distance distribution of \( C \) is equal to the weight distribution of the dual code \( C^\perp \). By using these facts and Lemma 2, sometimes it is more convenient for us to compute \( \bar{d}(C) \) and \( \text{var}(C) \) if \( C \) is a linear code.

3. Computation of \( \text{var}(C) \) for some subsets

In this section, we compute \( \text{var}(C) \) for some subsets. In general, we cannot find a formula to compute \( \text{var}(C \cup \{v\}) \) from \( \text{var}(C) \) or vice versa, but in some cases we can use (1.1), (1.3), and the following proposition to compute \( \text{var}(C \cup \{v\}) \) from \( \text{var}(C) \).

Proposition 1. Let \( C \) be a nonempty subset of \( Z_q^n \) and \( v \in Z_q^n \setminus C \). Then

\[ \sum_{a \in C \cup \{v\}} \sum_{b \in C \cup \{v\}} d_H(a, b) = \sum_{a \in C} \sum_{b \in C} d_H(a, b) + 2 \sum_{a \in C} w_H(a - v). \]  

(3.1)
Hence, by Proposition 1 and (3.7)–(3.10), we have

\[
\sum_{a \in C \cup \{v\}} \sum_{b \in C \cup \{v\}} d_H^2(a, b) = \sum_{a \in C} \sum_{b \in C} d_H^2(a, b) + 2 \sum_{a \in C} w_H^2(a - v). \tag{3.2}
\]

In this section, we will use the following results:

\[
\sum_{b \in Z_q^n} w_H(b) = \sum_{j=0}^{m} \binom{m}{j} (q-1)^j = m(q-1)q^{m-1}, \tag{3.3}
\]

\[
\sum_{b \in Z_q^n} w_H^2(b) = \sum_{j=0}^{m} j^2 \binom{m}{j} (q-1)^j = m(q-1)[1 + m(q-1)]q^{m-2}. \tag{3.4}
\]

Firstly, we obtain the following results by using Proposition 1.

**Proposition 2.** For any \(v \in Z_q^n\), denote \(C = Z_q^n \setminus \{v\}\). Then

\[
\bar{d}(C) = \frac{(q-1)n}{q} - \frac{(q-1)n}{q(q^n-1)^2}, \tag{3.5}
\]

\[
\text{var}(C) = \frac{(q-1)n}{q^2} + \frac{(q-1)n[(q-1)n-1]}{q^2(q^n-1)^2} - \frac{(q-1)^2n^2}{q^2(q^n-1)^4}. \tag{3.6}
\]

In particular, we obtain

\[
T(n, q^n-1; q) = R(n, q^n-1; q) = \frac{(q-1)n}{q^2} + \frac{(q-1)n[(q-1)n-1]}{q^2(q^n-1)^2} - \frac{(q-1)^2n^2}{q^2(q^n-1)^4}.
\]

**Proof.** By (3.3) and (3.4), we have

\[
\sum_{a \in Z_q^n} \sum_{b \in Z_q^n} d_H(a, b) = \sum_{a \in Z_q^n} \sum_{b \in Z_q^n} w_H(b - a) = q^n \sum_{x \in Z_q^n} w_H(x) = n(q-1)q^{2n-1}, \tag{3.7}
\]

\[
\sum_{a \in Z_q^n} \sum_{b \in Z_q^n} d_H^2(a, b) = q^n \sum_{x \in Z_q^n} w_H^2(x) = n(q-1)[1 + n(q-1)]q^{2(n-1)}. \tag{3.8}
\]

Note that \(C = Z_q^n \setminus \{v\}\). By (3.3) and (3.4), we have

\[
\sum_{a \in C} w_H(a - v) = \sum_{b \in Z_q^n \setminus \{0\}} w_H(b) = \sum_{b \in Z_q^n} w_H(b) = n(q-1)q^{n-1}, \tag{3.9}
\]

\[
\sum_{a \in C} w_H^2(a - v) = \sum_{b \in Z_q^n \setminus \{0\}} w_H^2(b) = \sum_{b \in Z_q^n} w_H^2(b) = n(q-1)[1 + n(q-1)]q^{n-2}. \tag{3.10}
\]

Hence, by Proposition 1 and (3.7)–(3.10), we have

\[
\sum_{a \in C} \sum_{b \in C} d_H(a, b) = \frac{n(q-1)}{q - \left(q^n - 1\right)^2 - 1}, \tag{3.11}
\]

\[
\sum_{a \in C} \sum_{b \in C} d_H^2(a, b) = n(q-1)[1 + n(q-1)]q^{n-2}(q^n - 2). \tag{3.12}
\]

Note that \(|C| = q^n - 1\), we obtain (3.5) from (1.1) and (3.11). It is easy to see from (3.5) that \(\bar{d}(C)\) can also be written as

\[
\bar{d}(C) = \frac{(q-1)nq^n(q^n-2)}{q(q^n-1)^2}. \tag{3.13}
\]
Hence, by (1.3), (3.12) and (3.13),

\[ var(C) = \frac{1}{(q^n - 1)^2} \sum_{a \in C} \sum_{b \in C} (d_H(a, b))^2 - \left[ \bar{d}(C) \right]^2 \]

\[ = \frac{n(q - 1)[1 + n(q - 1)]q^n - 2(q^n - 2)}{(q^n - 1)^2} - \frac{(q - 1)^2n^2q^2(q^n - 2)^2}{q^2(q^n - 1)^4} \]

\[ = \frac{(q - 1)n[(q^n - 1)^2 - 1]}{q^2(q^n - 1)^2} \left[ 1 + \frac{(q - 1)n}{(q^n - 1)^2} \right] \]

\[ = \frac{(q - 1)n}{q^2} + \frac{(q - 1)n[(q^n - 1)n - 1]}{q^2(q^n - 1)^2} - \frac{(q - 1)^2n^2}{q^2(q^n - 1)^4}. \]

This completes the proof. □

Secondly, we compute \( var(C) \) if \( C = \Gamma, \Gamma^+, \Gamma^-, A, A^+, A^- \) that are defined in Section 1.

**Proposition 3.** Let \( \Gamma, \Gamma^+, \Gamma^-, A, A^+, A^- \) be the sets defined in Section 1. Then

\[ var(\Gamma) = \frac{(q - 1)(n + 2)}{q^2}, \quad (3.14) \]

\[ var(\Gamma^+) = \frac{(q - 1)(n + 2)}{q^2} - \frac{4q^n + (q - 1)(n + 2) - (q - 1)^2n^2}{q^2(q^n - 1 + 1)^2} - \frac{(q - 1)^2n^2}{q^2(q^n - 1)^4}, \quad (3.15) \]

\[ var(\Gamma^-) = \frac{(q - 1)(n + 2)}{q^2} + \frac{(q - 1)^2n^2 - (q - 1)(n + 2)}{q^2(q^n - 1 - 1)^2} - \frac{(q - 1)^2n^2}{q^2(q^n - 1)^4}, \quad (3.16) \]

\[ var(A) = \frac{(q - 1)(n - 1)}{q^2}, \quad (3.17) \]

\[ var(A^+) = \frac{(q - 1)(n - 1)}{q^2} + \frac{2q^{n+1} + (q - 1)(n - 1)[(q - 1)(n - 1) - 1]}{q^2(q^n - 1 + 1)^2} \]

\[ - \frac{[2q^n - (q - 1)(n - 1)]^2}{q^2(q^n - 1 + 1)^4}, \quad (3.18) \]

\[ var(A^-) = \frac{(q - 1)(n - 1)}{q^2} + \frac{(q - 1)(n - 1)[(q - 1)(n - 1) - 1]}{q^2(q^n - 1 - 1)^2} - \frac{(q - 1)^2(n - 1)^2}{q^2(q^n - 1 - 1)^4}. \quad (3.19) \]

**Proof.** It is easy to see from the definition of \( \Gamma \) that \( \Gamma \) is an \([n, n - 1]\) linear code over \( F_q \) and \( |\Gamma| = q^n - 1 \). The dual code of \( \Gamma \) is given by

\[ \Gamma^\perp = \{(a, a, 0, 0, \ldots, 0) \in F_q^n \mid a \in F_q \}. \]

The dual distance distribution of \( \Gamma \) is equal to the weight distribution of \( \Gamma^\perp \), that is

\[ B_0 = 1, \quad B_1 = 0, \quad B_2 = q - 1, \quad B_i = 0, \quad i \geq 3. \]

Hence, by Lemma 2,

\[ \bar{d}(\Gamma) = \frac{(q - 1)n}{q}, \quad var(\Gamma) = \frac{(q - 1)(n + 2)}{q^2}. \quad (3.20) \]

Furthermore, by (1.1), (1.3) and (3.20), we have

\[ \sum_{a \in \Gamma} \sum_{b \in \Gamma} d_H(a, b) = (q - 1)naq^{2n-3}, \quad (3.21) \]

\[ \sum_{a \in \Gamma} \sum_{b \in \Gamma} d_H^2(a, b) = q^{2(n-2)}[(q - 1)(n + 2) + (q - 1)^2n^2]. \quad (3.22) \]
It follows from the definition of $I'$ and (3.3) that
\[
\sum_{\mathbf{a} \in I'} w_{I'}(\mathbf{a} - (0, 1, 0, \ldots, 0)) \\
= \sum_{(a_1, a_2, a_3, \ldots, a_n) \in I'} [w_{I'}((a_1, a_2 - 1) + w_{I'}((a_3, \ldots, a_n)))] \\
= \sum_{(a_1, a_2, a_3, \ldots, a_n) \in \mathbb{F}_q^n, \ (a_1, a_2, a_3, \ldots, a_n) \in \mathbb{F}_q^n} [1 + w_{I'}((a_3, \ldots, a_n)))] \\
+ \sum_{a_2 \neq 0, \ a_1 + a_2 = 0, \ (a_3, \ldots, a_n) \in \mathbb{F}_q^n} [2 + w_{I'}((a_3, \ldots, a_n)))] \\
= 2 \sum_{\mathbf{b} \in \mathbb{F}_q^{n-2}} [1 + w_{I'}(\mathbf{b})] + (q - 2) \sum_{\mathbf{b} \in \mathbb{F}_q^{n-2}} [2 + w_{I'}(\mathbf{b})] \\
= 2(q - 1)q^{n-2} + q \sum_{\mathbf{b} \in \mathbb{F}_q^{n-2}} w_{I'}(\mathbf{b}) \\
= n(q - 1)q^{n-2}.
\] (3.23)

In the same way, by (3.3) and (3.4), we have
\[
\sum_{\mathbf{a} \in I'} w_{I'}^2(\mathbf{a} - (0, 1, 0, \ldots, 0)) \\
= 2 \sum_{\mathbf{b} \in \mathbb{F}_q^{n-2}} [1 + w_{I'}(\mathbf{b})]^2 + (q - 2) \sum_{\mathbf{b} \in \mathbb{F}_q^{n-2}} [2 + w_{I'}(\mathbf{b})]^2 \\
= [(q - 1)^2n^2 + (q - 1)n - 2]q^{n-3}.
\] (3.24)

Hence, by the definition of $I^+$, (3.1), (3.20) and (3.23),
\[
\bar{d}(I^+) = \frac{q^{(n-1)}}{(q^{n-1} + 1)^2} \bar{d}(I') + \frac{2n(q - 1)q^{n-2}}{(q^{n-1} + 1)^2} = \frac{n(q - 1)q^{n-2}(q^{n-1} + 2)}{(q^{n-1} + 1)^2}.
\] (3.25)

By the definition of $I^+$, (3.2), (3.22) and (3.24),
\[
\sum_{\mathbf{a} \in I^+} \sum_{\mathbf{b} \in I^+} \bar{d}_{I'}^2(\mathbf{a}, \mathbf{b}) = [(q - 1)(n + 2) + (q - 1)^2n^2]q^{2(n-2)} \\
+ 2[(q - 1)^2n^2 + (q - 1)n - 2]q^{n-3}.
\] (3.26)

From (1.3), (3.25) and (3.26), we have
\[
\text{var}(I^+) \\
= \frac{[(q - 1)(n + 2) + (q - 1)^2n^2]q^{2(n-2)} + 2[(q - 1)^2n^2 + (q - 1)n - 2]q^{n-3}}{(q^{n-1} + 1)^2} \\
- \frac{n^2(q - 1)^2q^{2(n-2)}(q^{n-1} + 2)^2}{(q^{n-1} + 1)^4}.
\] (3.27)

Note that
\[
[(q - 1)(n + 2) + (q - 1)^2n^2]q^{2(n-2)} + 2[(q - 1)^2n^2 + (q - 1)n - 2]q^{n-3} \\
= q^{-2}((q - 1)(n + 2) + (q - 1)^2n^2][(q^{n-1} + 1)^2 - 1] - 4q^n),
\] (3.28)
\[
n^2(q - 1)^2q^{2(n-2)}(q^{n-1} + 2)^2 \\
= n^2(q - 1)^2q^{-2}[(q^{n-1} + 1)^4 - 2(q^{n-1} + 1)^2 + 1].
\] (3.29)

Hence, we obtain (3.15) from (3.27)–(3.29).
Now we compute \( \text{var}(I^-) \). Note that \( I^- \) is a linear code over \( F_q \). By the definition of \( I^- \),

\[
\sum_{a \in I^-} \sum_{b \in I^-} d_{H}(a, b) = \sum_{a \in I^-} \sum_{b \in I^-} w_{H}(b - a) \\
= q^{n-1} \sum_{c \in I^-} w_{H}(c) \\
= q^{n-1} \sum_{c \in I^-} w_{H}(c). \tag{3.30}
\]

In the same way, we have

\[
\sum_{a \in I^-} \sum_{b \in I^-} d_{H}^2(a, b) = q^{n-1} \sum_{c \in I^-} w_{H}^2(c). \tag{3.31}
\]

By (3.21), (3.22), (3.30) and (3.31),

\[
\sum_{c \in I^-} w_{H}(c) = (q - 1)nq^{n-2}, \tag{3.32}
\]

\[
\sum_{c \in I^-} w_{H}^2(c) = q^{n-3}[(q - 1)(n + 2) + (q - 1)^2n^2]. \tag{3.33}
\]

By (3.1), (3.2), (3.21), (3.22), (3.32) and (3.33),

\[
\sum_{a \in I^-} \sum_{b \in I^-} d_{H}(a, b) = (q - 1)nq^{n-2}(q^{n-1} - 2), \tag{3.34}
\]

\[
\sum_{a \in I^-} \sum_{b \in I^-} d_{H}^2(a, b) = [(q - 1)(n + 2) + (q - 1)^2n^2]q^{n-3}(q^{n-1} - 2). \tag{3.35}
\]

It follows from (1.1) and (3.34) that

\[
\bar{d}(I^-) = \frac{(q - 1)nq^{n-2}(q^{n-1} - 2)}{(q^{n-1} - 1)^2}. \tag{3.36}
\]

Hence, by (1.3), (3.35) and (3.36),

\[
\text{var}(I^-) = \frac{[(q - 1)(n + 2) + (q - 1)^2n^2]q^{n-3}(q^{n-1} - 2)}{(q^{n-1} - 1)^2} - \frac{(q - 1)^2n^2q^{2(n-2)}(q^{n-1} - 2)^2}{(q^{n-1} - 1)^4}. \tag{3.37}
\]

Note that

\[
q^{n-3}(q^{n-1} - 2) = q^{-2}[(q^{n-1} - 1)^2 - 1], \tag{3.38}
\]

\[
q^{2(n-2)}(q^{n-1} - 2)^2 = q^{-2}[(q^{n-1} - 1)^4 - 2(q^{n-1} - 1)^2 + 1]. \tag{3.39}
\]

Hence, we obtain (3.16) from (3.37)–(3.39).

Now we compute \( \text{var}(A) \). From the definition of \( A \), (3.7) and (3.8), we have

\[
\sum_{a \in A} \sum_{b \in A} d_{H}(a, b) = \sum_{x \in Z_q^{n-1}} \sum_{y \in Z_q^{n-1}} d_{H}(x, y) = (q - 1)(n - 1)q^{2n-3}, \tag{3.40}
\]

\[
\sum_{a \in A} \sum_{b \in A} d_{H}^2(a, b) = \sum_{x \in Z_q^{n-1}} \sum_{y \in Z_q^{n-1}} d_{H}^2(x, y) \\
= (q - 1)(n - 1)[1 + (q - 1)(n - 1)]q^{2n-4}. \tag{3.41}
\]

Hence, by (1.1) and (3.40), we have

\[
\bar{d}(A) = \frac{(q - 1)(n - 1)}{q}. \tag{3.42}
\]
By (1.3), (3.41) and (3.42),
\[ \text{var}(A) = \frac{(q - 1)(n - 1)}{q^2}. \]

Now we compute \( \text{var}(A^+) \). From the definition of \( A \) and (3.3), we have
\[ \sum_{a \in A} w_H(a - (0, \ldots, 0, 1)) = \sum_{x \in \mathbb{Z}_q^{n-1}} [w_H(x) + 1] = q^{n-1} + (q - 1)(n - 1)q^{n-2}. \] (3.43)

In the same way, by (3.3) and (3.4),
\[ \sum_{a \in A} w_H^2(a - (0, \ldots, 0, 1)) = \sum_{x \in \mathbb{Z}_q^{n-1}} [w_H(x) + 1]^2 = q^{n-1} + 2(q - 1)(n - 1)q^{n-2} + (q - 1)(n - 1)[1 + (q - 1)(n - 1)]q^{n-3}. \] (3.44)

Hence, by the definition of \( A^+ \), Proposition 1, (3.40), (3.41), (3.43) and (3.44), we have
\[ \sum_{a \in A^+} \sum_{b \in A^+} d_H(a, b) = (q - 1)(n - 1)q^{2n-3} + 2q^{n-1} + 2(q - 1)(n - 1)q^{n-2} \]
\[ = \frac{(q - 1)(n - 1)q^{n-1}}{q} + 2q^n - (q - 1)(n - 1)q^{n-2}. \] (3.45)

\[ \sum_{a \in A^+} \sum_{b \in A^+} d_H^2(a, b) = (q - 1)(n - 1)[1 + (q - 1)(n - 1)]q^{2n-4} + 2q^{n-1} + 4(q - 1)(n - 1)q^{n-2} + 2(q - 1)(n - 1)[1 + (q - 1)(n - 1)]q^{n-3} \]
\[ = \frac{(q - 1)(n - 1)[1 + (q - 1)(n - 1)]q^{n-1} + 1}{q^2} + \frac{2q^n + 4(q - 1)(n - 1)q^n - (q - 1)(n - 1)[1 + (q - 1)(n - 1)]}{q^2}. \] (3.46)

It follows from (1.1) and (3.45) that
\[ \tilde{d}(A^+) = \frac{(q - 1)(n - 1)}{q} + \frac{2q^n - (q - 1)(n - 1)}{q(n^{n-1} + 1)^2}. \] (3.47)

Hence, by (1.3), (3.46) and (3.47), we have
\[ \text{var}(A^+) = \frac{1}{(n^{n-1} + 1)^2} \sum_{a \in A^+} \sum_{b \in A^+} d_H^2(a, b) - \tilde{d}(A^+))^2 \]
\[ = \frac{(q - 1)(n - 1)}{q^2} + \frac{2q^{n-1} + (q - 1)(n - 1)[(q - 1)(n - 1) - 1]}{q^2(n^{n-1} + 1)^2} \]
\[ - \frac{[2q^n - (q - 1)(n - 1)]^2}{q^2(n^{n-1} + 1)^4}. \]

Now we compute \( \text{var}(A^-) \). From the definitions of \( A \) and \( A^- \) and (3.3), we have
\[ \sum_{a \in A^-} w_H(a) = \sum_{a \in A} w_H(a) = \sum_{x \in \mathbb{Z}_q^{n-1}} w_H(x) = (q - 1)(n - 1)q^{n-2}. \] (3.48)

In the same way, by (3.4),
\[ \sum_{a \in A^-} w_H^2(a) = \sum_{x \in \mathbb{Z}_q^{n-1}} w_H^2(x) = (q - 1)(n - 1)[1 + (q - 1)(n - 1)]q^{n-3}. \] (3.49)
Hence, by the definition of $A^-$, Proposition 1, (3.40), (3.41), (3.48) and (3.49), we have

$$\sum_{a \in A^-} \sum_{b \in A^-} d_\mathcal{H}(a, b) = \frac{(q - 1)(n - 1)}{q} [(q^{n-1} - 1)^2 - 1],$$

and

$$\sum_{a \in A^-} \sum_{b \in A^-} d_\mathcal{H}^2(a, b) = \frac{(q - 1)(n - 1)[1 + (q - 1)(n - 1)]}{q^2} [(q^{n-1} - 1)^2 - 1].$$

It follows from (1.1) and (3.50) that

$$\bar{d}(A^-) = \frac{(q - 1)(n - 1)}{q} - \frac{(q - 1)(n - 1)}{q(q^{n-1} - 1)^2}. \quad (3.52)$$

Hence, by (1.3), (3.51) and (3.52), we have

$$\text{var}(A^-) = \frac{1}{(q^{n-1} - 1)^2} \sum_{a \in A^-} \sum_{b \in A^-} d_\mathcal{H}^2(a, b) - \bar{d}(A^-)^2$$

$$= \frac{(q - 1)(n - 1)}{q^2} + \frac{(q - 1)(n - 1)[(q - 1)(n - 1) - 1]}{q^2(q^{n-1} - 1)^2}$$

$$- \frac{(q - 1)^2(n - 1)^2}{q^2(q^{n-1} - 1)^4}.$$

This completes the proof. \(\square\)

**Remark 1.** There is a typographical error for $T(n, 2^{n-1} + 1; 2)$ in [15]. It follows from Theorem 2 and Proposition 3 that

$$T(n, 2^{n-1} + 1; 2) = \frac{n - 1}{4} + \frac{2^{n+2} + (n - 1)(n - 2)}{4(2^{n-1} + 1)^2} - \frac{[2^{n+1} - n + 1]^2}{4(2^{n-1} + 1)^4}$$

$$= \frac{n - 1}{4} + \frac{n^2 + n + 2}{4(2^{n-1} + 1)^2} + \frac{(2^n - n - 1)(2^{2n} + 2^n + n + 1)}{4(2^{n-1} + 1)^4}.$$

4. **Proof of Theorem 1**

In this section, we first derive a general upper bound on $\text{var}(C)$ for $2 \leq q \leq 4$, then we prove Theorem 1 by observing that this upper bound is tight for $I, I^+$ and $I^-.$

**Theorem 3.** Let $C$ be an $(n, M; q)$ code. If $2 \leq q \leq 4$ and $M \equiv x \pmod{q}$, then

$$\text{var}(C) \leq \frac{(q - 1)n - 2}{q^2} + \frac{2q^{n-2}}{M} - \frac{x^2(q - x)^2}{q^2 M^2}$$

$$- \frac{x(q - x)}{q^2(q - 1)M^2} [2^{q^n - 2} - n(q - 1)(q - 2) - n(n - 1)(q - 1)^2]. \quad (4.1)$$

**Proof.** The dual distance distribution of $C$ is given by $B_0, B_1, \ldots, B_n.$ By Lemmas 1 and 3, we have

$$B_2 = \frac{q^n}{M} - 1 - B_1 - \sum_{k=3}^n B_k$$

$$\leq \frac{q^n}{M} - 1 - B_1 - \frac{x(q - x)}{M^2} \sum_{k=3}^n (q - 1)^{k-1} \binom{n}{k}$$

$$= \frac{q^n}{M} - 1 - B_1 - \frac{x(q - x)}{(q - 1)M^2} \left[ q^n - 1 - (q - 1)n - \frac{n(n - 1)(q - 1)^2}{2} \right]. \quad (4.2)$$
since $B_1 \geq \frac{2(q-2)n}{q^2 M^2}$, by Lemmas 2 and 3 and (4.2), we have for $2 \leq q \leq 4$,

$$\text{var}(C) \leq \frac{(q-1)n-2}{q^2} + \frac{2g^{n-2}}{M} + \frac{(q-4)}{q^2} B_1 - \frac{1}{q^2} (B_1)^2$$

$$- \frac{2(q-2)}{(q-1)q^2 M^2}[2q^n - 2 - (q-1)n - n(n-1)(q-1)]$$

$$\leq \frac{(q-1)n-2}{q^2} + \frac{2g^{n-2}}{M} - \frac{2(q-2)^2 n^2}{q^2 M^4}$$

$$- \frac{2(q-2)}{q^2(q-1)M^2}[2q^n - 2 - n(q-1)(q-2) - n(n-1)(q-1)^2].$$

This completes the proof. □

**Proof of Theorem 1.** Note that

$$|\Gamma| = q^{n-1}, \quad |\Gamma^+| = q^{n-1} + 1, \quad |\Gamma^-| = q^{n-1} - 1.$$    

By Theorem 3 and Proposition 3, we have for $2 \leq q \leq 4$,

$$\text{var}(\Gamma) \leq R(n, q^{n-1}; q) \leq \frac{(q-1)(n+2)}{q^2} = \text{var}(\Gamma),$$

$$\text{var}(\Gamma^+) \leq R(n, q^{n-1} + 1; q)$$

$$\leq \frac{(q-1)(n+2)}{q^2} - \frac{4q^n + (q-1)(n+2) - (q-1)^2 n^2}{q^2(q^{n-1} + 1)^2}$$

$$- \frac{(q-1)^2 n^2}{q^2(q^{n-1} + 1)^4} = \text{var}(\Gamma^+),$$

$$\text{var}(\Gamma^-) \leq R(n, q^{n-1} - 1; q)$$

$$\leq \frac{(q-1)(n+2)}{q^2} + \frac{(q-1)^2 n^2 - (q-1)(n+2)}{q^2(q^{n-1} - 1)^2} - \frac{(q-1)^2 n^2}{q^2(q^{n-1} - 1)^4} = \text{var}(\Gamma^-).$$

Hence, Theorem 1 follows from combining these assertions. □

5. Proof of Theorem 2

In this section, we first derive a general lower bound on $\text{var}(C)$ for all $q$, then we prove Theorem 2 by showing that this lower bound is tight for some cases.

Let $C$ be an $(n, M; q)$ code where $M \equiv x \mod q$. Denote

$$s_q(n, M) = \frac{x(q-x) n}{q M^2},$$

$$t_q(n, M) = \frac{q^{n-1}}{M} - \frac{1}{q} - \frac{x(q-x)[q^n - 1 - (q-1)n]}{q(q-1)M^2}$$

and define

$$F(x) = \frac{(q-2)}{q} x - x^2, \quad x > 0,$$

$$r_q(n, M) = \min\{F(s_q(n, M)), F(t_q(n, M))\}.$$
Theorem 4. Let $C$ be an $(n, M; q)$ code with $M \equiv z \pmod{q}$. Then
\[
\text{var}(C) \geq \frac{(q - 1)n}{q^2} + \frac{z(q - z)(q - 1)n(n - 1)}{q^2 M^2} + r_q(n, M).
\]
Moreover, if $t_q(n, M) \geq \frac{q - 2}{q}$, then $r_q(n, M) = F(t_q(n, M))$.

Proof. The dual distance distribution of $C$ is given by $B_0, B_1, \ldots, B_n$. Denote $x_0 = B_1 q$. By Lemmas 2 and 3, we have
\[
\text{var}(C) = \frac{(q - 1)n}{q^2} + \frac{2}{q^2} B_2 + F(x_0) \geq \frac{(q - 1)n}{q^2} + \frac{z(q - z)(q - 1)n(n - 1)}{q^2 M^2} + F(x_0).
\]
It follows from Lemma 3 that $x_0 \geq s_q(n, M)$. By Lemmas 1 and 3, we have
\[
B_1 = \frac{q^n}{M} - 1 - \sum_{k=2}^{n} B_k \\
\leq \frac{q^n}{M} - 1 - \frac{z(q - z)}{M^2} \sum_{k=2}^{n} (q - 1)^{k-1} \binom{n}{k} \\
= \frac{q^n}{M} - 1 - \frac{z(q - z) (q^n - 1 - (q - 1)n)}{(q - 1) M^2}.
\]
Hence $x_0 \leq t_q(n, M)$. Since for every interval $F(x)$ achieves its minimum at one of its endpoints, we obtain (5.1). Moreover, since $s_q(n, M) \geq 0$, it follows that $F(s_q(n, M)) \geq F(t_q(n, M))$ when $t_q(n, M) \geq \frac{q - 2}{q}$. This completes the proof.

Proof of Theorem 2. Note that
\[
|A| = q^{n-1}, \quad |A^+| = q^{n-1} + 1, \quad |A^-| = q^n - 1.
\]
For $M = q^{n-1}$, we have $x = 0$ and
\[
t_q(n, q^{n-1}) = \frac{q - 1}{q} \geq \frac{q - 2}{q}.
\]
Hence, we have
\[
r_q(n, q^{n-1}) = F\left(\frac{q - 1}{q}\right) = -\frac{q - 1}{q^2}.
\]
Thus, by Theorem 4 and Proposition 3,
\[
\text{var}(A) \geq T(n, q^{n-1}; q) \geq \frac{(q - 1)(n - 1)}{q^2} = \text{var}(A)
\]
which implies (1.7).
For $M = q^{n-1} - 1$, we have $x = q - 1$ and
\[
t_q(n, q^{n-1} - 1) = \frac{q - 1}{q} + \frac{(q - 1)(n - 1)}{q(q^n - 1 - 1)^2}.
\]
It is easy to see that $t_q(n, q^{n-1} - 1) \geq \frac{q - 2}{q}$. Hence, by Theorem 4 and Proposition 3, we have
\[
\text{var}(A^-) \geq T(n, q^{n-1} - 1; q) \geq \frac{(q - 1)(n - 1)}{q^2} + \frac{(q - 1)(n - 1)[(q - 1)(n - 1) - 1]}{q^2 (q^n - 1 - 1)^2} \geq \frac{(q - 1)^2 (n - 1)^2}{q^2 (q^n - 1 - 1)^2} = \text{var}(A^-)
\]
which implies (1.8).
For $M = q^{n-1} + 1$, we have $x = 1$ and

$$t_q(n, q^{n-1} + 1) = \frac{q - 1}{q} - \frac{2q^n - (q - 1)(n - 1)}{q(q^{n-1} + 1)^2}.$$

It is easy to check that $t_q(n, q^{n-1} + 1) \geq \frac{q - 2}{q}$ for $n \geq 3$. If $n = 2$, then $M = q + 1$ and

$$t_q(2, q + 1) = \frac{q - 1}{q} - \frac{2q^2 - q + 1}{q(q + 1)^2},$$

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

It is easy to check that for $n = 2$ and $q = 2, 3$,

$$t_q(2, q + 1) \geq \frac{q - 2}{q}$$

and for $n = 2$ and $q = 4$, $t_4(2, 5) = 0.46$ and $s_4(2, 5) = 0.06$, so that $r_4(2, 5) = F(t_4(2, 5))$. Hence, by Theorem 4 and Proposition 3, we have for $n \geq 3$ or $2 \leq q \leq 4$,

$$\text{var}(A^+) \geq T(n, q^{n-1} + 1; q) \geq \frac{(q - 1)(n - 1)}{q^2} + \frac{2q^{n+1} + (q - 1)(n - 1)[(q - 1)(n - 1) - 1]}{q^2(q^{n-1} + 1)^2} - \frac{[2q^n - (q - 1)(n - 1)]^2}{q^2(q^{n-1} + 1)^4}$$

which implies (1.9). This completes the proof. □

Acknowledgements

The authors would like to thank the anonymous reviewers and the Editor-in-Chief Professor Peter L. Hammer for their valuable suggestions and comments that helped to improve this paper.

This research work is supported in part by the DSTA Research Grant R-394-000-011-422, the National Natural Science Foundation of China under the Grant 60172060, the Trans-Century Training Program Foundation for the Talents by the Education Ministry of China, and the Foundation for University Key Teacher by the Education Ministry of China, and the 100-Talents program of the Chinese Academy of Science.

References