

On the Distance Distribution of Codes

Gil Kalai and Nathan Linial

Abstract—The distance distribution of a binary code C is the sequence $(B_i)_{i=0}^n$ defined as follows: Let $B_i(w)$ be the number of codewords at distance i from the codeword w , and let B_i be the average of $B_i(w)$ over all w in C . In this correspondence we study the distance distribution for codes of length n and minimal distance δn , with $\delta > 0$ fixed and $n \rightarrow \infty$. Our main aim is to relate the size of the code with the distribution of distances near the minimal distance.

Index Terms—Binary codes, distance distribution, linear programming bounds.

I. INTRODUCTION

Let V_n be the set of all 0–1 vectors of length n . The Hamming distance, $d(u, v)$ of two vectors v, u in V_n is the number of coordinates in which they differ. A binary code C of length n is a subset of V_n , and elements in C are also called *codewords*. The minimum distance of C is the least Hamming distance between two distinct codewords. One of the main open problems in coding theory is to determine the largest cardinality, $A(n, d)$ of a binary code of length n and minimal distance d . For more information on coding theory see [12]–[14].

Our main concern is with the case where d is proportional to n . When n tends to infinity and d/n tends to $\delta < 1/2$, then $A(n, d)$ is exponential in n . The determination of the basis for this exponential function is a difficult question of fundamental importance for coding theory.

We need some notation now: The rate $R(C)$ of a code C is

$$R(C) = \frac{\log(|C|)}{n}.$$

(Here and elsewhere in the paper $\log x$ stands for $\log_2 x$.) Let

$$R(n, d) = \log A(n, d) \cdot n^{-1}$$

be the maximum rate of a code of length n and minimal distance d . Next, for every real number $0 \leq \delta \leq 1$ let

$$R(\delta) = \limsup_{n \rightarrow \infty} R(n, d_n)$$

where $d_n = \delta n(1 + o(1))$. (Here and elsewhere in the correspondence all $o(1)$ terms are taken for $n \rightarrow \infty$.) As usual, the entropy function is

$$H(x) = -x \log x - (1 - x) \log (1 - x).$$

The best known lower bound for $R(\delta)$ goes back to Gilbert [7]

$$R(\delta) \geq 1 - H(\delta). \tag{1}$$

Gilbert’s proof of this bound is simply to “grow” a code, by always adding new codewords subject only to the constraint that no distances smaller than δn occur. Despite its extreme simplicity,

Manuscript received August 29, 1993; revised Nov. 18, 1994. 1991 Mathematics Subject Classification. Primary 94B65, Secondary 05B40, 52C17. This work was supported in part by the Binational Israel–US Science Foundation and by Israeli Academy of Sciences and Humanities.

G. Kalai is with the Institute of Mathematics, Hebrew University, Jerusalem 91904, Israel.

N. Linial is with the Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel.

IEEE Log Number 9413357.

this argument has never been improved, and some researchers believe that no asymptotic improvement is possible. Thus one of the principal problems of coding theory is as follows.

Problem 1: Is it true that $R(\delta) = 1 - H(\delta)$?

The best known upper bound on $R(\delta)$ for binary codes was achieved by McEliece, Rodemich, Rumsey, and Welch (henceforth MRRW), [11] using Delsarte’s linear programming method (see below). They showed that

$$R(\delta) \leq \mu(\delta) = H(1/2 - \sqrt{\delta(1 - \delta)}). \tag{2}$$

Using more general inequalities by Delsarte for constant-weight codes the same authors proved an even stronger upper bound for $R(\delta)$, which applies for $\delta < 0.273$, see [11].

The number of codewords at distance i from a codeword w is denoted $B_i(w)$ and $B_i(= B_i(C))$ is the average of $B_i(w)$ over all w in C . For $0 \leq s \leq 1$ we write

$$b_s(= b_s(C)) = n^{-1} \log (B_{[sn]}(C)).$$

We also define

$$R_s(\delta) = \limsup_{n \rightarrow \infty} b_{z_n}(C)$$

where $z_n = (s + o(1)) \cdot n$ and the supremum is taken over all codes of length n and minimal distance $d_n = (\delta + o(1)) \cdot n$.

It is of considerable interest to study the possible distance distributions of codes, and our paper is a contribution to this area. Let us remark that if in the proof of the Gilbert bound one selects the next codeword at random, while maintaining a minimal distance of $\geq \delta n$, then the resulting code C achieves the Gilbert bound and almost surely satisfies $b_s(C) = H(s) - H(\delta) + o(1)$, for every $s > \delta$, see, e.g., [6]. The same distance distribution is obtained (almost surely) if one selects at random

$$2^n / \left(3 \sum_{i=1}^{[\delta n]-1} \binom{n}{i} \right)$$

vectors in V_n and deletes all pairs of vectors of distance $< [\delta n]$. No known family of codes meets the Gilbert bound and has asymptotically a different distance distribution. So a natural extension of Problem 1 is as follows:

Problem 2: Is it true that for every s, δ , $0 < \delta < s \leq 1/2$, $R_s(\delta) = H(s) - H(\delta)$?

Our main result is that if C is a code of length n and

$$b_s(C) \leq H(s) - H(\delta) + o(1)$$

for every s in a certain neighborhood $[\delta, u(\delta))$ of δ , then $R(C) \leq 1 - H(\delta) + o(1)$.

In other words, a family of codes, whose cardinalities exponentially exceed the Gilbert lower bound must have “many” pairs of codewords whose distance is close to the minimum. Specifically, the distance distribution of such codes must exceed those of the “random” Gilbert code in a certain neighborhood of δn .

The neighborhood of the minimal distance is given by the following function. Define

$$u_1(\delta) = 2\delta - 2\delta^2 \tag{3}$$

$$u_2(\delta) = \mu^{-1}(1 - H(\delta)). \tag{4}$$

$$u(\delta) = \min(u_1(\delta), u_2(\delta)). \tag{5}$$

The functions u_1, u_2, u are tabulated in Table I. $u_1(x)$ is smaller than $u_2(x)$ For $x < 0.082 \dots$.

TABLE I

δ	$u_1(\delta)$	$u_2(\delta)$
0.02	0.03920	0.04988
0.04	0.07680	0.08666
0.06	0.11280	0.11868
0.08	0.14720	0.14764
0.10	0.18000	0.17437
0.12	0.21120	0.19930
0.14	0.24080	0.22276
0.16	0.26880	0.24495
0.18	0.29520	0.26603
0.20	0.32000	0.28611
0.22	0.34320	0.30530
0.24	0.36480	0.32365
0.26	0.38480	0.34124
0.28	0.40320	0.35810
0.30	0.42000	0.37429
0.32	0.43520	0.38982
0.34	0.44880	0.40472
0.36	0.46080	0.41901
0.38	0.47120	0.43269
0.40	0.48000	0.44576
0.42	0.48720	0.45820
0.44	0.49280	0.46999
0.46	0.49680	0.48106

Theorem 1.1: Let C be a binary code of length n and minimal distance δn , where $1/2 > \delta > 0$. Then

$$R(C) \leq 1 - H(\delta) + \sup\{b_s(C) - (H(s) - H(\delta)) : \delta \leq s < u(\delta)\} + o(1).$$

Corollary 1.2:

$$R(\delta) - (1 - H(\delta)) \leq \sup\{R_s(\delta) - (H(s) - H(\delta)) : \delta \leq s < u(\delta)\}.$$

In particular, we have

Corollary 1.3: If $R_s(\delta) \leq (H(s) - H(\delta))$ for every s , $\delta \leq s < u(\delta)$, then $R(\delta) = 1 - H(\delta)$. Namely, Gilbert's bound is tight for that δ .

Thus in order to prove that $R(0.01) = 1 - H(0.01)$ it would suffice to prove that $R_s(0.01) \leq H(s) - H(0.01)$ for $0.01 \leq s < 0.0198$. The equality $R(0.3) = 1 - H(0.3)$ would follow from $R_s(0.3) \leq H(s) - H(0.3)$ for $0.3 \leq s < 0.375$.

Note that Theorem 1.1 sharpens the MRRW bound (2) which says:

- If $b_s(C) = 0$ for every $s < \delta$, then $R(C) \leq \mu(\delta) + o(1)$.

Theorem 1.1 yields the same conclusion from weaker assumptions: Consider a code C of length n and minimal distance βn . Define $\delta = u_2(\beta)$ whence $1 - H(\beta) = \mu(\delta)$. Apply Theorem 1.1 with β instead of δ . The maximum is taken over the interval

$$[\beta, u(\beta)] \subseteq [\beta, u_2(\beta)] = [\beta, \delta].$$

If $b_s(C) - (H(s) - H(\delta)) \leq 0$ throughout this interval, the conclusion is

$$R(C) \leq 1 - H(\beta) + o(1) = \mu(\delta) + o(1).$$

So indeed the conclusion of MRRW is obtained from weaker assumptions:

Theorem 1.4: Let C be a code of length n and let $\delta = u_2(\beta)$ be real. If $b_s(C) = 0$ for $s < \beta$ and $b_s(C) \leq H(s) - H(\beta) + o(1)$ for $\beta \leq s < \delta$ then $R(C) \leq \mu(\delta) + o(1)$.

The proof of Theorem 1.1 consists of two separate arguments, involving the functions $u_1(\delta)$ and $u_2(\delta)$, respectively. The proof for $u_1(\delta)$ is based on a simple double counting argument, and is given in Section II. The proof for $u_2(\delta)$ is based on a variant of the linear programming method as applied in the proof of (2) and is given in Section III. Both proofs give, in fact, a slightly stronger statement, namely, that $R(C) \leq 1 - H(\delta) + o(1)$ already follows if

$$b_s(C) \leq H(s) - H(\delta) + w_s(s) + o(1)$$

for every s in the interval $[\delta, u(\delta))$, where $w_s(s)$ is a certain nonnegative decreasing function of s . The actual function w_s as obtained in the two proofs is given in Sections II and IV, respectively. (The asymptotic analysis of Krawtchouk polynomials in Section IV may be of independent interest.)

Both arguments described here apply to other types of codes and give analogous results for constant-weight codes, for codes over larger alphabets, and for spherical codes.

What remains a mystery is the behavior of the distance distribution of codes near the minimum. We conjecture, for example, that $R_s(\delta) = 0$ for every δ . Several open problems on the behavior of binary and spherical codes near the minimal distance are discussed in the final Section V. We also suggest a possible way to get upper bounds on the individual B'_i 's via a certain hypercontractive inequality of Beckner.

II. AN AVERAGING ARGUMENT

Proposition 2.1: For every binary code C of length n and every s , $0.5 > s > 0$

$$R(C) - (1 - H(s)) \leq \max\{b_t(C) - (H(t) - H(s)) - w_s(t) : s \leq t \leq 2s\} + o(1). \quad (6)$$

where

$$w_s(t) = H(s) - \left[sH\left(\frac{t}{2s}\right) + (1-s)H\left(\frac{t}{2(1-s)}\right) \right]$$

is always nonnegative.

Proof: Let $S_a(z)$ be the Hamming sphere with radius a centered at z . By Cauchy-Schwartz:

$$\begin{aligned} |C| \binom{n}{a} &= \sum_{z \in V_n} |S_a(z) \cap C| \leq \sqrt{2^n \sum_z |S_a(z) \cap C|^2} \\ &= \sqrt{2^n \sum_{x_1, x_2 \in C} |S_a(x_1) \cap S_a(x_2)|}. \end{aligned}$$

If $d(x_1, x_2) = b$, then

$$|S_a(x_1) \cap S_a(x_2)| = \binom{b}{b/2} \binom{n-b}{a-b/2}.$$

(In particular, b is even, or else the set is empty.) There are $|C| \cdot B_b(C)$ pairs of codewords (x_1, x_2) at distance b , and the inequality simplifies to

$$|C| \binom{n}{a}^2 \leq 2^n \sum_b \binom{b}{b/2} \binom{n-b}{a-b/2} \cdot B_b(C).$$

Whence

$$|C| \binom{n}{a}^2 \leq 2^n \cdot n \cdot \max_b \binom{b}{b/2} \binom{n-b}{a-b/2} \cdot B_b(C).$$

It is easily verified that

$$\binom{b}{b/2} \binom{n-b}{a-b/2} = \binom{n}{a} \binom{a}{b/2} \binom{n-a}{b/2} / \binom{n}{b}$$

so

$$|C| \binom{n}{a} 2^{-n} \leq n \cdot \max_b B_b(C) \frac{\binom{n}{a} \binom{a}{b/2} \binom{n-a}{b/2}}{\binom{n}{b} \binom{n}{a}}.$$

Taking logarithms and dividing by n , the proposition follows. To see that $w_s \geq 0$, observe that

$$\sum_j \binom{a}{j} \binom{n-a}{j} = \binom{n}{a}$$

so

$$\binom{a}{b/2} \binom{n-a}{b/2} \leq \binom{n}{a}.$$

Again the conclusion follows by taking logarithms and dividing by n . Equality $w_s(t) = 0$ holds for $t = 2s(1-s)$ and only there.

We now strengthen Proposition 2.1, in that we replace the interval $s \leq t \leq 2s$ on which the maximum is taken, by a subinterval $s \leq t \leq u_1(s)$. Recall Markov's inequality: if X is a nonnegative random variable, then $\Pr(X \leq c \cdot E(X)) \geq 1 - \frac{1}{c}$, for every $c > 1$. Our probability space consists of all triples $\{x, y, z\}$ with $z \in V_n$, $x, y \in C$, and $d(x, z) = d(y, z) = a$. The random variable X equals $d(x, y)$ on this triple. In particular, as we saw

$$\Pr(X = b) = \frac{\binom{b}{b/2} \binom{n-b}{a-b/2} \cdot B_b(C)}{\sum_j \binom{j}{j/2} \binom{n-j}{a-j/2} \cdot B_j(C)}.$$

We claim that $E(X) \leq 2a - 2a^2/n$. In fact, this inequality holds even conditional on any fixed $z \in V_n$. Having fixed z , all relevant codewords form (a translation by z of) a code of constant weight a and we need the following easy fact (see, e.g., [10]):

Proposition 2.2: Let Γ be a code of length n and constant weight a . Then the average distance of two codewords in Γ is at most $2a - 2a^2/n$.

Proof: Let p_i be the fraction of codewords w in Γ with $w_i = 1$, then $\sum p_i = a$. Therefore, two randomly chosen codewords in Γ differ in their i th coordinate with probability $2p_i(1-p_i)$. It follows that the expected Hamming distance between two randomly chosen codewords is

$$2 \sum p_i(1-p_i) \leq 2a - 2a^2/n$$

(since $\sum p_i = a$). \square

Now X takes only integer values, and it is easy to observe that there is no integer between $2a - 2a^2/n$ and $(1 + \frac{1}{n^2})(2a - 2a^2/n)$, so

$$\Pr\left(X \leq \left(1 + \frac{1}{n^2}\right)(2a - 2a^2/n)\right) = \Pr(X \leq 2a - 2a^2/n).$$

Apply Markov's inequality with $c = 1 + \frac{1}{n^2}$ to conclude

$$\sum_j \binom{j}{j/2} \binom{n-j}{a-j/2} \cdot B_j(C) \leq (n^2 + 1) \sum_{j \leq 2a - 2a^2/n} \binom{j}{j/2} \binom{n-j}{a-j/2} \cdot B_j(C).$$

Substituting $s = \delta$ in Proposition 2.1 in its strong form we get

Theorem 2.3: For C a binary code of length n , and minimal distance $(\delta + o(1))n$

$$R(C) \leq 1 - H(\delta) + \max\{b_s - (H(s) - H(\delta) + w_s(s)) : \delta \leq s \leq u_1(\delta)\} + o(1).$$

Remark: In order to replace the maximum over the interval $[\delta, u_1(\delta)]$ by the supremum over $[\delta, u_1(\delta))$, apply Theorem 2.3 for a sequence $\delta_m \nearrow \delta$.

III. A VARIANT OF THE LINEAR PROGRAMMING BOUND

In 1972 Delsarte [3], [4] found (as part of a much more general theory of association schemes) a system of linear inequalities satisfied by the distance distribution of every binary code. For linear codes, Delsarte's inequalities reduce to identities which go back to MacWilliams.

Delsarte's *linear programming method* calls for deriving an upper bound on the size of the code, by maximizing the sum of the B_i 's (which is the size of the code) subject to his system of inequalities.

The Krawtchouk polynomials $K_k^{(n)}$ are defined as follows:

$$K_k^{(n)}(x) = \sum_j (-1)^j \binom{x}{j} \binom{n-x}{k-j}. \quad (7)$$

Whenever the value of n is clear from the context, we omit it and write $K_k(x)$ for $K_k^{(n)}(x)$.

The MacWilliams-Delsartes system of inequalities for binary codes of length n and minimal distance d are

$$B_0 = 1 \quad (8)$$

$$B_i = 0, \quad \text{for } i = 1, 2, \dots, d-1$$

$$B_i \geq 0, \quad \text{for } i = d, d+1, \dots, n$$

$$\sum_{i=0}^n B_i K_k^n(i) \geq 0, \quad \text{for } k = 0, 1, \dots, n.$$

Delsarte's *linear programming method* is to derive an upper bound on the size of the code, by maximizing the sum of the B_i 's (which is the size of the code) subject to this system of inequalities.

It is convenient to work with the dual linear program which has the following simple form.

Theorem 3.1: For every polynomial

$$\beta(x) = 1 + \sum \beta_k K_k(x)$$

with $\beta_k \geq 0$ for $1 \leq k \leq n$, such that $\beta(j) \leq 0$ for $j = d, d+1, \dots, n$

$$\sum_{i=0}^n B_i \leq \beta(0). \quad (9)$$

We will discuss now the effect of adding to the Delsarte's inequalities, upper bounds for the individual B_i 's and derive our main theorem for $u_2(\delta)$.

Proposition 3.1: For codes C of length n and minimal distance d and for every polynomial

$$\beta(x) = \beta_0 + \sum \beta_k K_k(x)$$

with $\beta_0 > 0$ and $\beta_k \geq 0$ for $1 \leq k \leq n$, such that $\beta(j) \leq 0$ for $j = m, m+1, \dots, n$

$$\sum_{i=0}^n B_i \leq (\beta_0)^{-1} \cdot \left[\beta(0) + \sum_{i=d}^{m-1} B_i \beta(i) \right]. \quad (10)$$

Proof: The coefficients β_k are a feasible solution to the dual of the linear program: $\max \sum B_i$ under Delsarte's inequalities. This is an instance of the fact that any dual feasible solution provides an upper bound to the optimum of the primal LP. \square

Now define

$$k(a, b) = \limsup \left\{ \frac{1}{n} \log |K_j^n(x)| : j = (a + o(1))n \right. \\ \left. \text{and } x = (b + o(1))n \right\}. \quad (11)$$

(Note that $k(a, 0) = H(a)$.)

Let $x_1^{(m)}$ denote the first zero of the Krawtchouk polynomial $K_m^{(n)}(x)$. It is known [11] that if $m = (s + o(1))n$ for some $0 < s < 1/2$, then $x_1^{(m)} = (\alpha(s) + o(1))n$ where $\alpha(s) = \frac{1}{2} - \sqrt{s(1-s)}$.

Proposition 3.2: For binary codes of length n , minimal distance δn and for every s ,

$$R(C) \leq \max \{ H(\alpha(s)), \max \{ b_x(C) + 2k(\alpha(s), x) - H(\alpha(s)) \\ : \delta \leq x \leq s \} \} + o(1). \quad (12)$$

Proof: Apply the previous proposition with a choice of $\beta(x)$ much like that of MRRW, namely

$$\beta(x) = (a - x)^{-1} (K_t(a)K_{t+1}(x) - K_{t+1}(a)K_t(x))^2. \quad (13)$$

However, here t and a are selected as follows: t is the largest integer for which $x_1^{(t)} < sn$, and a is the (unique) point in the interval $(x_1^{(t+1)}, x_1^{(t)})$ for which $K_t(a) = -K_{t+1}(a)$. As observed in [11], $\beta(x)$ is a nonnegative combination of Krawtchouk polynomials.

Now, apply the previous Proposition for $m = sn$. With this choice $t = (\alpha(s) + o(1))n$. As before, it suffices to consider the largest term on the right-hand side of (10), which we proceed to do. As shown in [11] (see also [14, p. 67])

$$\beta_0 = -\frac{2}{t+1} \binom{n}{t} K_t(a)K_{t+1}(a).$$

Therefore

$$\beta(0) \cdot \beta_0^{-1} = \frac{(n+1)^2}{2a(t+1)} \binom{n}{t}$$

and

$$n^{-1} \log(\beta(0) \cdot \beta_0^{-1}) = H(t) + o(1) = H(\alpha(s)) + o(1).$$

Denote $i = x \cdot n$ and calculate the i th term in the sum

$$n^{-1} \log(B_i \beta(i) \cdot \beta_0^{-1}) = b_x(C) + 2k(\alpha(s), x) - H(t) + o(1).$$

By the previous proposition

$$\sum_{i=1}^n B_i \leq n \cdot \max \{ (\beta_0)^{-1} \cdot \beta(0), \max \{ (\beta_0^{-1}) \cdot \beta(i) \cdot B_i : \delta n \leq i \leq sn \} \}.$$

Taking logarithms of both sides and dividing by n we get the statement of the proposition. \square

Theorem 3.2: For C a binary code of length n , and minimal distance $(\delta + o(1))n$,

$$R(C) \leq 1 - H(\delta) + \max \{ b_x(C) - (H(s) - H(\delta) - \bar{w}_s(s)) \\ : \delta \leq s \leq u_2(\delta) \} + o(1) \quad (14)$$

where

$$\bar{w}_s(x) = 2 - H(x) - H(\delta) - 2k(\alpha(u_2(\delta)), x) \quad (15)$$

is a nonnegative function of s in the interval $[\delta, u_2(\delta)]$.

Proof: Apply the previous proposition with $s = u_2(\delta)$. With this choice, $H(\alpha(s)) = 1 - H(\delta)$. We get that

$$R(C) - (1 - H(\delta)) \leq \max \{ 0, b_x(C) - (H(x) - H(\delta)) \\ + (H(x) - H(\delta)) - (1 - H(\delta)) + 2k(\alpha(u_2(\delta)), x) \\ - H(\alpha(u_2(\delta))) : \delta \leq x \leq u_2(\delta) \}.$$

We get (14) with

$$\bar{w}_s(x) = -[(H(x) - H(\delta)) - (1 - H(\delta)) \\ + 2k(\alpha(u_2(\delta)), x) - H(\alpha(u_2(\delta)))] \\ = 1 - H(x) + H(\alpha(u_2(\delta))) - 2k(\alpha(u_2(\delta)), x)$$

which simplifies to (15). To show that \bar{w} is nonnegative we need the following

Proposition 3.3: For every $0 \leq a, b \leq 1$,

$$1 + H(a) - H(b) - 2k(a, b) \geq 0.$$

Proof: This follows from the following orthogonality relation of Krawtchouk polynomials (see, e.g., [14]):

$$\sum (K_k^{(n)}(j))^2 \binom{n}{j} = 2^n \cdot \binom{n}{k}.$$

IV. ASYMPTOTICS OF KRAWTCHOUK POLYNOMIALS

In this section we derive an explicit expression for $k(a, b)$, hence also for $\bar{w}_s(s)$.

In what follows we assume that both j and x grow linearly with n . To get the asymptotic behavior of Krawtchouk polynomials, we recall the following identity ([11, eq. A.14]):

$$(n-x)K_j(x+1) - (n-2j)K_j(x) + xK_j(x-1) = 0. \quad (16)$$

Recall also [11] that all j zeros of K_j are in the interval

$$\left[\frac{n}{2} - (1 + o(1))\sqrt{j(n-j)}, \frac{n}{2} + (1 + o(1))\sqrt{j(n-j)} \right]$$

and that the leading coefficient in K_j is $\frac{(-2)^j}{j!}$. Therefore, we may write

$$K_j(x) = \frac{2^j}{j!} \prod (x_i - x)$$

where x_i are the roots of K_j . We want to compare the terms

$$\frac{K_j(x+1)}{K_j(x)}$$

and

$$\frac{K_j(x)}{K_j(x-1)}.$$

The above expression for K_j yields

$$\frac{K_j(x+1)K_j(x-1)}{K_j^2(x)} = \prod \frac{(x_i - x - 1)(x_i - x + 1)}{(x_i - x)^2} \\ = 1 + O\left(\frac{1}{n}\right). \quad (17)$$

This is because we get $O(n)$ terms, each of which is $1 + O(\frac{1}{n^2})$.
Therefore, if we let

$$z := \frac{K_j(x+1)}{K_j(x)}$$

whence

$$\frac{K_j(x)}{K_j(x-1)} = \left(1 + O\left(\frac{1}{n}\right)\right)z$$

we may rewrite the basic identity (16) as a quadratic equation

$$(n-x)z^2 - \left(1 + O\left(\frac{1}{n}\right)\right)(n-2j)z + x = 0.$$

We still have to decide which of the two roots of the quadratic to select. Because of (17), the choice of the sign is uniform throughout the region where x is bounded away from $\frac{n}{2} - \sqrt{j(n-j)}$. However ([11, eqs. A.8, A.9])

$$K_j(1) = \frac{n-2j}{n} K_j(0)$$

implies that the plus sign is the correct choice. Summing up, we already know that

$$\frac{K_j(x+1)}{K_j(x)} = \left(1 + O\left(\frac{1}{n}\right)\right) \frac{(n-2j) + \sqrt{(n-2j)^2 - 4x(n-x)}}{2(n-x)}. \tag{18}$$

We also know, of course, that $K_j(0) = \binom{n}{j}$ and by multiplying appropriate instances of (18) we get an approximate value for $K_j(x)$. How good is this approximation? Our only inaccuracy comes in from a product of $O(n)$ terms each of which equals $1 + O(1/n)$, so we get an answer that is correct up to a constant factor that is bounded away from zero. Our final goal is to obtain an expression for

$$\frac{1}{n} \log K_j(x)$$

and the above analysis will yield an answer with an additive error of $O(1/n)$.

We get, then

$$\begin{aligned} \frac{1}{n} \log K_j(x) &= H(j/n) \\ &+ \frac{1}{n} \sum_{i \leq x} \log \left(\frac{(n-2j) + \sqrt{(n-2j)^2 - 4t(n-t)}}{2(n-t)} \right) \\ &+ O(1/n). \end{aligned}$$

By Euler–McLaurin, this sum may be approximated by the appropriate integral.

It follows that

$$k(a, b) = H(a) + \int_0^b \log \left(\frac{1 - 2a + \sqrt{(1-2a)^2 - 4t(1-t)}}{2(1-t)} \right) dt.$$

Define

$$\Delta = \Delta(a, b) = \sqrt{(1-2a)^2 + (1-2b)^2} - 1.$$

Integrating (using Mathematica) we obtain that whenever $\Delta(a, b) \geq 0$

$$\begin{aligned} k(a, b) &= H(a) + b \log(1 + \Delta(a, b) - 2a) \\ &+ 0.5 \log(1 + 2b(2 - 2b - \Delta(a, b)) / (1 + \Delta(a, b) - 2a)) \\ &+ a \log((1 + \Delta(a, b) - 2b) / (2 - 2a)). \end{aligned} \tag{19}$$

V. FINAL REMARKS AND OPEN PROBLEMS

In this correspondence we revealed relations between the distribution of distances in the vicinity of the least distance and the size of the whole code. The distance distribution near the minimum distance remains a great mystery.

Conjecture 1: For every binary code of length n and minimal distance d , B_d is subexponential in n . In other words, for every ϵ there is $N = N(\epsilon)$ so that for every code of length $n > N$ and minimal distance d , $B_d \leq (1 + \epsilon)^n$.

For linear codes Conjecture 1 simply reads as follows:

Conjecture 2: The number of codewords of minimal weight in a linear code of length n is subexponential in n .

Remark: We cannot even show that for a binary linear code of exponential size, the number of codewords of minimal weight is exponentially smaller than the size of the code.

Here is the analogous (more general) conjecture for sphere packing. Let $m(n, t)$ be the smallest integer so that in any packing of spherical caps of radius t in the unit n -sphere there is a sphere which touches at most $m(n, t)$ other spheres.

Conjecture 3: For a fixed $t > 0$, $m(n, t)$ is subexponential as n tends to infinity.

By the results of this correspondence, slightly stronger forms of the above conjectures suffice to improve the known upper bound for codes.

Conjecture 4: For every δ , $0 \leq \delta \leq 1$, $R_\delta(\delta) = 0$

Finally, let us mention that the starting point of this research was our attempt to obtain upper bounds on individual B_i using a certain hypercontractive estimate by Beckner [1]. (For combinatorial application of this inequality see [8], [2].) Beckner's inequality is equivalent to the following relation:

$$\sum_{k=0}^n \sum_{i=0}^n B_i K_k^n(i) \epsilon^k \leq |C|^{\frac{1-\epsilon}{1+\epsilon}} \cdot 2^{\frac{2\epsilon}{1+\epsilon}n}. \tag{20}$$

Application of (20) for convolutions of the form $h = f * g$, where f is the characteristic function of a code and g is the characteristic function of a certain Hamming ball, do lead to nontrivial upper bounds on the B_i 's. So far, all the upper bounds on individual B_i 's we managed to derive this way, have been inferior to those obtained from the best known bounds for constant-weight codes together with Elias' Lemma. It is possible that by applying Beckner's or other hypercontractive estimates to other functions related to the original codes, or by finding sharper forms of Beckner's inequality for characteristic functions of sets of size β^n for $\beta < 2$ some progress can be made.

Analogues of Beckner's inequality for subsets of the Johnson Scheme (constant-weight codes) are not known and are of interest. There is a vast literature on hypercontractive estimates for certain operators on real functions on S^n . These may yield upper bounds for the distance distribution of spherical codes.

REFERENCES

- [1] W. Beckner, "Inequalities in Fourier analysis," *Annals Math.*, vol. 102, pp. 159–182, 1975.
- [2] J. Bourgain, J. Kahn, G. Kalai, I. Katznelson, and N. Linial, "The influence of variables in product spaces," *Israel J. Math.*, vol. 77, pp. 55–64, 1992.
- [3] P. Delsarte, "Bounds for unrestricted codes, by linear programming," *Philips Res. Rep.*, vol. 27, pp. 272–289, 1972.
- [4] ———, "An algebraic approach to the association schemes of coding theory," *Phillips Res. Rep., Suppl.*, vol. 10, 1973.
- [5] P. Delsarte, J. M. Goethals, and J. J. Seidel, "Spherical codes and designs," *Geom. Dedicata*, vol. 6, pp. 363–388, 1977.
- [6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

- [7] E. N. Gilbert, "A comparison of signaling alphabet," *Bell Syst. Tech. J.*, vol. 31, pp. 504–522, 1952.
- [8] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on Boolean functions," in *Proc. 29th Ann. Symp. on Foundations of Computer Science*. Los Alamitos, CA: Computer Soc. Press, 1988, pp. 68–80.
- [9] G. A. Kabatiansky and V. I. Levenshtein, "Bounds on packing on a sphere and in space," *Probl. Inform. Transmission*, vol. 14, pp. 1–17, 1978.
- [10] R. J. McEliece and H. C. Rumsey, "Sphere-packing in the Hamming metric," *Bull. Amer. Math. Soc.*, vol. 75, pp. 32–34, 1969.
- [11] R. J. McEliece, E. R. Rodemich, H. C. Rumsey, and L. R. Welch, "New upper bounds on the rate of codes via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 157–166, 1977.
- [12] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.
- [13] N. J. A. Sloane, "Recent bounds for codes, sphere packings and related problems obtained by linear programming and other methods," *Contemp. Math.*, vol. 9, pp. 153–185, 1982.
- [14] J. H. van Lint, *Introduction to Coding Theory*. New York: Springer-Verlag, 1982.

On the Accuracy of the Binomial Approximation to the Distance Distribution of Codes

Iliia Krasikov and Simon Litsyn

Abstract—The binomial distribution is a well-known approximation to the distance spectra of many classes of codes. We derive a lower estimate for the deviation from the binomial approximation.

Index Terms—Spectra of codes, Krawtchouk polynomials.

I. INTRODUCTION

The binomial distribution is a well-known approximation to the distance spectra of many classes of codes. For example, it is known to be tight for the weights of BCH codes (see, e.g. [7, sec. 9.10]). Several upper bounds for the error term of such approximation have been derived in [1], [2], [4], [8], [9]. These estimates show that, provided the dual distance is large enough, the spectrum of the code rapidly converges to the binomial distribution. How close can the real distribution be to the binomial one? In this correspondence we give a lower estimate for the deviation from the binomial approximation thus showing that it cannot be too sharp. We also establish an identity relating the error terms to the dual spectrum of a code.

II. RESULTS

We start with the following auxiliary lemma [3]. The proof is presented for self-completeness.

Manuscript received July 19, 1994; revised February 2, 1995. This research was partially supported by the Guastallo Fellowship and a Grant from the Israeli Ministry of Science and Technology.

I. Krasikov is with Tel-Aviv University, School of Mathematical Sciences, Ramat-Aviv 69978, Tel-Aviv, Israel, and Beit-Berl College, Kfar-Sava, Israel.

S. Litsyn is with Tel-Aviv University, Department of Electrical Engineering—Systems, Ramat-Aviv 69978, Tel-Aviv, Israel.

IEEE Log Number 9413879.

Lemma 1: Let \mathcal{F} be the set of real monic polynomials of degree c . Define

$$E(a, b, c) = \min_{f \in \mathcal{F}} \max_{x \in [-1, 1]} |(1-x)^a (1+x)^b f(x)|.$$

Then

$$\frac{4^{a+b+c}}{2a+2b+2c+1} \leq \binom{2a+2b+2c}{c} \binom{2a+2b+2c}{2a+c} (E(a, b, c))^2$$

provided $a, b > -1/2$, real.

Proof: Let f be an optimal polynomial. Expand f in the series of Jacobi polynomials (see, e.g., [10])

$$f(x) = \sum_{j=0}^c q_j P_j^{(2a, 2b)}(x)$$

where

$$(1-x)^\alpha (1+x)^\beta P_j^{(\alpha, \beta)}(x) = \frac{(-1)^j}{2^j j!} \frac{d^j}{dx^j} ((1-x)^{\alpha+j} (1+x)^{\beta+j}).$$

The leading coefficient of $P_j^{(\alpha, \beta)}(x)$ is $2^{-j} \binom{\alpha+\beta+2j}{j}$, and so

$$q_c = \frac{2^c}{\binom{2a+2b+2c}{c}}.$$

The orthogonality relation for Jacobi polynomials is given by

$$\begin{aligned} g_{jl}(\alpha, \beta) &= \int_{-1}^1 (1-x)^\alpha (1+x)^\beta P_j^{(\alpha, \beta)}(x) P_l^{(\alpha, \beta)}(x) dx \\ &= \frac{2^{\alpha+\beta+1} \Gamma(j+\alpha+1) \Gamma(j+\beta+1)}{(2j+\alpha+\beta+1) \Gamma(j+1) \Gamma(j+\alpha+\beta+1)} \delta_{jl} \end{aligned}$$

where δ_{jl} is the Kronecker delta.

Now we get

$$\begin{aligned} &\max_{x \in [-1, 1]} (1-x)^a (1+x)^b (f(x))^2 \\ &\geq \frac{1}{2} \int_{-1}^1 (1-x)^{2a} (1+x)^{2b} (f(x))^2 dx \\ &\geq \frac{1}{2} \sum_{j=0}^c \sum_{l=0}^c q_j q_l \int_{-1}^1 (1-x)^{2a} (1+x)^{2b} P_j^{(2a, 2b)}(x) \\ &\quad \cdot P_l^{(2a, 2b)}(x) dx \\ &\geq \frac{1}{2} q_c^2 g_{cc}(2a, 2b) \end{aligned}$$

and we are done. \square

The binary Krawtchouk polynomial $P_k^n(x)$ (of degree k in x) is defined by the following generating function:

$$\sum_{k=0}^{\infty} P_k^n(x) z^k = (1-z)^x (1+z)^{n-x}. \quad (1)$$

When it does not lead to confusion n is omitted, i.e., $P_k(x) = P_k^n(x)$. The following values are of importance for us:

$$P_i(0) = \binom{n}{i} \quad P_n(i) = (-1)^i \quad P_i(n) = (-1)^i \binom{n}{i}.$$

Let the distance distribution of a code C be $\underline{B} = (B_0, \dots, B_n)$, and $\underline{B}' = (B'_0, \dots, B'_n)$ stand for the dual spectrum, that is, \underline{B}' is determined by the MacWilliams transform of \underline{B}

$$B'_k = \frac{1}{|C|} \sum_{i=0}^n B_i P_k(i). \quad (2)$$