On the Distance Distribution of Codes *

Gil Kalai † Nathan Linial ‡

Abstract

The *distance distribution* of a binary code C is the sequence $(G_i)_{i=0}^n$ defined as follows: Let $G_i(w)$ be the number of code words at distance i from the code word w, and let G_i be the average of $G_i(w)$ over all w in C.

In this paper we study the distance distribution for codes of length n and minimal distance δn, with $\delta > 0$ fixed and $n \to \infty$. Our main aim is to relate the size of the code with the distribution of distances near the minimal distance.

This is an early version of the paper which appeared which contains some more material.

*1991 Mathematics Subject Classification. Primary 94B65, Secondary 05B40, 52C17. Supported in part by the Binational Israel-US Science Foundation.

†Institute of Mathematics, Hebrew University, Jerusalem 91904, Israel.

‡Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel.

1 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 *code words.* The *minimum distance* of C is the least Hamming distance between two distinct code words. 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 [13, 15, 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 $0 \le \delta \le 1$ let

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

where $d_n = \delta n(1+o(1))$. (Here and elsewhere in the paper all $o(1)$ terms are taken for $n \to \infty$.) 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 [8]

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

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

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), [12] using Delsarte's linear programming method (see below). They showed:

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

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 [12].

The number of code words at distance i from a code word w is denoted $G_i(w)$ and $G_i(= G_i(C))$ is the average of $G_i(w)$ over all w in C. For $0 \leq s \leq 1$ we write $g_s(= g_s(C)) = n^{-1}\log(G_{[sn]}(C))$. We also define

$$R_s(\delta) = \limsup_{n\to\infty} \; g_{z_n}(C),$$

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 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 $g_s(C) = H(s) - H(\delta) + o(1)$, for every $s > \delta$, see e.g. [7]. The same distance distribution is obtained (almost surely) if one selects at random $2^n/(3\sum_{i=1}^{[\delta n]-1}\binom{n}{i})$ vectors in V_n and deletes all pairs of vectors of distance $< [\delta n]$. No family of codes is known, which meets the Gilbert bound and has asymptotically a different distance distribution.

Our main result is that if C is a code of length n and $g_s(C) \leq H(s) - H(\delta) + o(1)$ 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 1. $u_1(x)$ is smaller than $u_2(x)$ For $x < 0.082...$

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

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

Corollary 1.2

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

In particular,

Corollary 1.3 *If $R_s(\delta) \leq (H(s) - H(\delta))$ for every s, $\delta \leq s < u(s)$, 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 $s < 0.0198$. The equality $R(0.3) \leq 1 - H(0.3)$ would follow from $R_s(0.3) = H(s) - H(0.3)$ for $s < 0.375$.

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

- If $g_s(C) = 0$ 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 $g_s(C) - (H(s) - H(\delta)) \leq 0$ 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 $g_s(C) = 0$ for $s < \beta$ and $g_s(C) \leq H(s) - H(\beta) + o(1)$ 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 2. The proof for $u_2(\delta)$ is based on a variant of the linear programming method as applied in the proof of (2) and on some asymptotic analysis of Krawtchouk polynomials. This is done in Sections 3 and 4. Both proofs give, in fact, a slightly stronger statement, namely, that $R(C) \leq 1 - H(\delta) + o(1)$ already follows if $g_s(C) \leq H(s) - H(\delta) + w_\delta(s) + o(1)$ for every s in the interval $[\delta, u(\delta))$, where $w_\delta(s)$ is a certain nonnegative decreasing function of s. The actual function w_δ as obtained in the two proofs is given in Sections 2 and 5 respectively. (The asymptotic analysis of Krawtchouk polynomials in Section 5 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. These matters will be pursued in a subsequent paper.

In Section 6 we discuss possible ways to get upper bounds on the individual $G_i's$. We suggest a method to derive such bounds using a hypercontractive inequality of Beckner. (This inequality was first applied in extremal combinatorial problems by Kahn, Kalai and Linial [9, 10, 3].) Although currently the consequences of this method for codes are inferior to known results we feel that it may be found useful.

What remains a mystery is the behavior of the distance distribution of codes near the minimum. We conjecture, for example, that $R_\delta(\delta) = 0$ for every δ. Several open problems on the behavior of binary and spherical codes near the minimal distance are discussed in the final Section 7.

2 An averaging argument

Proposition 2.1 *For every binary code C of length n and every $s > 0$*

$$R(C) - (1 - H(s)) \leq \max\{g_t(C) - (H(t) - H(s)) - w_s(t) : 0 \leq t \leq 2s\} + o(1). \qquad (6)$$

where $w_s(t) = H(s) - [sH(\frac{t}{2s}) + (1-s)H(\frac{t}{2(1-s)})]$ is always nonnegative.

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

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

If $\text{dist}(x_1, x_2) = b$, 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 $|C| \cdot G_b(C)$ pairs of codewords (x_1, x_2) at distance b, and the inequality simplifies to:

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

Whence,

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

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

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

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 sub-interval $s \leq t \leq u_1(s)$. 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,

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

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 code words form (a translation by z of) a code of constant weight a and we need the following easy fact:

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 code words w in Γ with $w_i = 1$, then $\sum p_i = a$. Therefore, two randomly chosen code words 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 code words is $2\sum p_i(1-p_i) \leq 2a - 2a^2/n$, (since $\sum p_i = a$). □

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 $\Pr(X \leq (1 + \frac{1}{n^2})(2a - 2a^2/n)) = \Pr(X \leq 2a - 2a^2/n)$. Apply Markov's inequality with $c = 1 + \frac{1}{n^2}$ to conclude:

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

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 $(\delta + o(1))n$,*

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

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 $\delta_m \nearrow \delta$.

3 MacWilliams-Delsartes relations

In 1972 Delsarte [4, 5] 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 G_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:

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

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

We identify 0-1 vectors of length n with subsets of $[n] = \{1, 2, \cdots, n\}$ in the standard way. Let $f : V_n \to \mathbf{R}$ be a function and consider its Walsh-Fourier expansion

$$f = \sum \{\hat{f}(S) u_S : \quad S \subseteq [n]\}, \tag{8}$$

where, u_S is the function defined by $u_S(T) = (-1)^{|S \cap T|}$. Note that if $F_i = \sum\{f(S) : |S| = i\}$ then

$$\sum\{\hat{f}(S) : |S| = k\} = 2^n \sum_{i=0}^{n} K_k(i) F_i, \tag{9}$$

where $K_k(x)$ is the $k - th$ Krawtchouk polynomial.

In the context of harmonic analysis, it is convenient to view V_n as a probability space, and so given a function $f : V_n \to R$, its p-th norm is defined as $\|f\|_p = (2^{-n} \sum_{S \subseteq V_n} |f(S)|^p)^{1/p}$. Parseval's identity asserts that $\|f\|_2^2 = \sum_{S \subseteq V_n} \hat{f}^2(S)$. Also, the *convolution* $h = f * g$ of two functions is given by $h(S) = 2^{-n} \sum_T f(T) g(S \triangle T)$, where $S \triangle T$ is the symmetric difference between S and T. Recall that $\hat{h} = \hat{f} \cdot \hat{g}$.

The relevance of convolution in our work is that if f is the characteristic function of a binary code C, and if $g = f * f$, then $g(Z)$ is 2^{-n} times the number of pairs of codewords $S, T \in C$ with $S \triangle T = Z$. We recall that for a code word w, $G_i(w)$ is the number of code words of distance i from w and G_i is the average of $G_i(w)$ over all w in C. In other words, $G_i = \frac{2^n}{|C|} \sum\{g(S) : |S| = i\}$. Delsarte's inequalities, which for the special case of linear codes go back to MacWilliams, can be derived as follows: Since $g = f * f$, it follows that $\hat{g}(S) = \hat{f}^2(S) \geq 0$. Together with equation (9) we obtain:

$$\sum_{i=0}^{n} K_k(i) \cdot G_i = \frac{2^{2n}}{|C|} \sum\{\hat{g}(S) : |S| = k\} = \frac{2^{2n}}{|C|} \sum\{\hat{f}^2(S) : |S| = k\} \geq 0. \tag{10}$$

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

$$G_0 = 1 \tag{11}$$

$$G_i = 0, \text{ for } i = 1, 2, \ldots d - 1$$

$$G_i \geq 0, \text{ for } i = d, d + 1 \ldots n$$

$$\sum_{i=0}^{n} G_i K_k^n(i) \geq 0 \text{ for } k = 0, 1, \ldots, n$$

Delsarte's *linear programming method* is to derive an upper bound on the size of the code, by maximizing the sum of the G_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 $j = d, d+1, \cdots, n$,*

$$\sum_{i=0}^{n} G_i \leq \beta(0). \tag{12}$$

Remarks: 1. The optimum of this linear program equals the maximum of $\frac{\|f\|_1^2}{\|f\|_2^2}$ over all real functions f on V_n such that $f * f$ is non-negative and $f * f(S) = 0$ for $0 < |S| < d$. The fact that f itself, being a characteristic function of a code is non-negative, and even a 0-1 function is not used.

2. It was pointed out by Levenshtein that the MRRW bounds for $R(\delta)$ cannot be improved by selecting a function $\beta(x)$ which is non-positive for the entire interval $[d, n]$. (Levenshtein identified explicitly the best such $\beta(x)$. This led to improved bounds for $A(d, n)$ but not for $R(\delta)$.)

It may still be possible to get an improvement by choosing a function β which is non-positive on $\{d, d+1, ..., n\}$ but takes positive values elsewhere in the interval $[d, n]$.

4 A variant of the linear programming bound

In this section we discuss the effect of adding to the Delsarte's inequalities, upper bounds for the individual $G_i's$ and derive our main Theorem for $u_2(\delta)$.

Proposition 4.1 *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 $\beta_k \geq 0$ for $1 \leq k \leq n$, such that $\beta(j) \leq 0$ for $j = m, m+1, \cdots, n$,*

$$\sum_{i=0}^{n} G_i \leq (\beta_0)^{-1} \cdot \left[\beta(0) + \sum_{i=d}^{m-1} G_i \beta(i) \right]. \tag{13}$$

Proof: The coefficients β_k are a feasible solution to the dual of the linear program: $\max \sum G_i$ 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. □

Now define

$$k(a, b) = \limsup \{ \frac{1}{n} \log |K_j^n(x)| : j = (a + o(1))n \text{ and } x = (b + o(1))n \}. \tag{14}$$

(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 [12] 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 4.2 *For binary codes of length n, minimal distance δn and for every s,*

$$R(C)\leq \max\{H(\alpha(s)),\ \max\{g_x(C)+2k(\alpha(s),x)-H(\alpha(s)):\delta\leq x\leq s\}\}+o(1). \tag{15}$$

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

$$\beta(x)=(a-x)^{-1}(K_t(a)K_{t+1}(x)-K_{t+1}(a)K_t(x))^2. \tag{16}$$

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 [12], $\beta(x)$ 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 right hand side of (13), which we proceed to do. As shown in [12], (see also [15], 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: $n^{-1}\log(G_i\beta(i)\cdot\beta_0^{-1})=g_x(C)+2k(\alpha(s),x)-H(t)+o(1)$. By the previous proposition

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

Taking log on both sides and dividing by n we get the statement of the proposition. □

Theorem 4.3 *For C a binary code of length n, and minimal distance $(\delta+o(1))n$,*

$$R(C)\leq 1-H(\delta)+\max\{g_s(C)-(H(s)-H(\delta)-\bar{w}_\delta(s)):\delta\leq s\leq u_2(\delta)\}+o(1)), \tag{17}$$

where

$$\bar{w}_\delta(x)=2-H(x)-H(\delta)-2k(\alpha(u_2(\delta)),x) \tag{18}$$

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

$$R(C) - (1 - H(\delta)) \leq max\{0, g_x(C) - (H(x) - H(\delta)) +$$

$$+(H(x) - H(\delta)) - (1 - H(\delta)) + 2k(\alpha(u_2(\delta)), x) - H(\alpha(u_2(\delta))) : \delta \leq x \leq u_2(\delta)\}.$$

We get relation (17) with

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

$$1 - H(x) + H(\alpha(u_2(\delta))) - 2k(\alpha(u_2(\delta)), x),$$

which simplifies to relation (18). To show that $\bar{w}$ is nonegative we need

Proposition 4.4 *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., [15])

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

5 Asymptotics of Krawtchouk polynomials

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

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 (A.14 in [12]):

$$(n - x)K_j(x + 1) - (n - 2j)K_j(x) + xK_j(x - 1) = 0. \tag{19}$$

Recall also [12] that all j zeros of K_j are in the interval

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

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'd like 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:

$$\frac{K_j(x+1)K_j(x-1)}{K_j^2(x)} = \prod \frac{(x_i - x - 1)(x_i - x + 1)}{(x_i - x)^2} = 1 + O(\frac{1}{n}). \tag{20}$$

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 $\frac{K_j(x)}{K_j(x-1)} = (1 + O(\frac{1}{n}))z$), we may rewrite the basic identity (19) as a quadratic equation:

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

We still have to decide which of the two roots of the quadratic to select. Because of Equation (20) the choice of the sign is uniform throughout the region where x is bounded away from $\frac{n}{2} - \sqrt{j(n-j)}$. However, (equations A.8, A.9 in [12])

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

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

$$\frac{K_j(x+1)}{K_j(x)} = (1 + O(\frac{1}{n}))\frac{(n - 2j) + \sqrt{(n - 2j)^2 - 4x(n - x)}}{2(n - x)}. \tag{21}$$

We also know, of course that $K_j(0) = \binom{n}{j}$ and by multiplying appropriate instances of Equation (21) we get an approximate value for $K_j(x)$. How good is this approximationΓ 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 $\frac{1}{n} \log K_j(x)$, so we'll get our answer with an additive error of $O(1/n)$, which suits us just fine.

We get, then

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

By Euler-McLauren, 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 $\Delta(a,b) \geq 0$:

$$k(a,b) = H(a) + b\log(1 + \Delta(a,b) - 2a) + \tag{22}$$

$$0.5\log(1 + 2b(2 - 2b - \Delta(a,b))/(1 + \Delta(a,b) - 2a)) + a\log((1 + \Delta(a,b) - 2b)/(2 - 2a)).$$

6 Beckner's hypercontractive estimates

This section concerns some possible ways to obtain upper bounds for individual G_i's. An obvious upper bound on $G_i(C)$ for a code C of length n and minimal distance d is the maximal size $A(n,d,i)$, of a code of length n, constant weight i and minimal distance d. These bounds in conjunction with our main theorem cannot yield improvements for $A(n,d)$ since a theorem by Elias (proved by an easy averaging argument) asserts that $A(n,d) \leq A(n,d,i)\frac{2^n}{\binom{n}{i}}$. (See e.g., [12].) Upper bounds on $G_i(C)$ for codes of length n which go below $A(n,d,i)$ may lead to improved upper bounds for $R(\delta)$ via Theorem 1.1.

The following inequality may be useful in establishing upper bounds for $G_i(C)$.

Theorem 6.1 (Beckner [1]) *For $f : \{0,1\}^r \to \{0,1\}$, define $T_\epsilon(f) = \sum\{\hat{f}(S)\epsilon^{|S|}u_S : \quad S \subset [r]\}$. Then*

$$\|T_\epsilon f\|_2 \leq \|f\|_{1+\epsilon^2}. \tag{23}$$

Note that if $|S| = k$, then $T_\epsilon(1_\emptyset)(S) = (\frac{1-\epsilon}{2})^k(\frac{1+\epsilon}{2})^{n-k}$. Thus when f is the characteristic function of a subset C of V_n, inequality (23) reads:

$$\sum_{i=0}^{n} G_i(1+\epsilon)^{n-i}(1-\epsilon)^i \leq |C|^{\frac{1-\epsilon}{1+\epsilon}} \cdot 2^{\frac{2\epsilon}{1+\epsilon}n}. \tag{24}$$

Still another form obtained by expanding the terms in (24) in powers of ϵ is

$$\sum_{k=0}^{n}\sum_{i=0}^{n} G_i K_k^n(i)\epsilon^k \leq |C|^{\frac{1-\epsilon}{1+\epsilon}} \cdot 2^{\frac{2\epsilon}{1+\epsilon}n}. \tag{25}$$

In the range of interest to us, a direct application of the Beckner inequality yields nothing useful. Application of Equation (23) for functions 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 $G_i's$. So far, all the upper bounds on individual $G_i's$ 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 inequality (23) 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, (including the direct analogue of (23), see [2]). These may yield upper bounds for the distance distribution of spherical codes.

7 Open problems on the distance distributions near the minimal distance

In this paper we revealed relations between the distribution of distances at the vicinity of the least distance and the size of the whole code. The distance distribution near the minimum remains a great mystery. We list here several open problems on this *terra incognita.*

We start with a few problems on the number of occurrences of the *minimal distance* in a code and the analogous problem for packing of spheres.

Conjecture 2 *For every binary code of length n and minimal distance d, G_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, $G_d \leq (1+\epsilon)^n$.*

For linear codes Conjecture 2 simply reads

Conjecture 3 *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 4 *For a fixed $t > 0$, $m(n,t)$ is subexponential as n tends to infinity.*

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

Conjecture 5 *For every δ, $0 \leq \delta \leq 1$, $R_\delta(\delta) = 0$*

References

[1] W. Beckner, Inequalities in Fourier analysis, Annals of Math. 102(1975), 159-182.

[2] W. Beckner, Sobolev inequalities, the Poisson semigroup, and analysis on the sphere S^n, Proc. Nat. Acad. Sci. USA 89(1992), 4816-4819.

[3] J. Bourgain, J. Kahn, G. Kalai, I. Katznelson and N. Linial, The influence of variables in product spaces, Israel Jour. Math., 77(1992) 55-64.

[4] P. Delsarte, Bounds for unrestricted codes, by linear programming, Philips Research Reports, 27 (1972), 272-289.

[5] P. Delsarte, An algebraic approach to the association schemes of coding theory, Phillips Research Reports Supplements, 10(1973).

[6] P. Delsarete, J.M.Goethals and J. J. Seidel, Spherical codes and designs, Geom. Dedicata 6(1977), 363-388.

[7] R. G. Gallagher, Information Theory and Reliable Communication, Wiley, New-York, 1968.

[8] E. N. Gilbert, A comparison of signaling alphabet, Bell Syst. Tech. J. 31(1952), 504-522.

[9] Kahn, J. Kalai, G. and Linial, N., The influence of variables on Boolean functions, Proc. 29-th Ann. Symp. on Foundations of Comp. Sci., 68-80, Computer Society Press, 1988.

[10] Kahn, J. Kalai, G. and Linial, N., Collective coin flipping, the influence of variables on Boolean functions and harmonic analysis, to appear.

[11] G. A. Kabatiansky and V. I. Levenshtein, Bounds on packing on a sphere and in space, Problems of Information Transmission 14(1978), 1-17.

[12] 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. Infor. Th. 23(1977), 157-166.

[13] F. J. MacWilliams and N. J. A. Sloane, The theory of Error Correcting Codes, North-Holland, 1977.

[14] N. J. A. Sloane, Recent bounds for codes, sphere packings and related problems obtained by linear programming and other methods, Contemporary Math. 9(1982), 153-185.

[15] J. H. van Lint, Introduction to Coding Theory, Springer-Verlag, 1982.

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