ON THE CORRECTING CAPABILITIES OF LINEAR CODES

1969

SovietRxiv

Abstract

Full Text

UDC 519.95

CYBERNETICS AND CONTROL THEORY

M. V. KOZLOV

ON THE CORRECTING CAPABILITIES OF LINEAR CODES

(Presented by Academician A. N. Kolmogorov on 1 XI 1968)

Let $GF(2)$ be the Galois field with 2 elements*; let V^n be the set of all ordered collections $\mathbf{v} = (v^1, \dots, v^n)$ with components $v^i \in GF(2)$. V^n forms an n-dimensional vector (coordinate) space over $GF(2)$**. A linear (k, n)-code is any homomorphism from V^k into V^n, $k < n$ (see [1]). The image of V^k under this mapping is called the code space, and its elements are called codewords. It is convenient to specify a linear (k, n)-code by means of a matrix $\|a_j^i,\ i = 1, \dots, k;\ j = 1, \dots, n\|$ over $GF(2)$ of rank k, whose rows $\mathbf{a}_1, \dots, \mathbf{a}_k$ form a basis of the code space.

The correcting capabilities of a linear (k, n)-code can be characterized by the minimum weight*** of nonzero codewords: if the minimum is equal to $2t + 1$, then any two codewords differ in at least $2t + 1$ positions, and consequently the distortion of any codeword in no more than t positions (replacement of 0 by 1 and 1 by 0) does not lead to loss of information. One of the problems of coding theory consists in finding theoretical bounds for the correcting capabilities of linear codes. There are a number of results here (see [1,2]), among which we note the following, due to Varshamov and Gilbert:

If

$$\sum_{0 \le i \le t-2} C_{n-1}^i < 2^{n-k}, \tag{1}$$

then there exists a (k, n)-code with minimum weight $\geq t$. However, the exact upper bound for the minimum weights of (k, n)-codes is still unknown. In the present note it is proved that the minimum weights of the majority of (k, n)-codes are grouped around the largest solution (in t) of inequality (1). We now pass to the precise formulation of the assertion.

On the set of all binary $k \times n$ matrices let us define the uniform probability measure, assuming that $a_j^i,\ i = 1, \dots, k;\ j = 1, \dots, n$, are mutually independent random variables taking the values 0 and 1 with equal probabilities. Introduce

the random variable η_n, equal to the minimum weight of the code space generated by the matrix $\|a_j^i\|$, and denote by $\beta_n(t) = P\{\eta_n > t\}$ its distribution function. We shall assume in what follows that $k = [nR]$****, where $0 < R < 1$ is fixed, and shall be interested in the asymptotic behavior of $\beta_n(t)$ as $n \to \infty$. For the largest solution t_n of inequality (1) one can write the asymptotic expression

$$t_n = np + \frac{1}{2}\left(\log_2 \frac{1-p}{p}\right)^{-1} \log_2 n + O(1), \tag{2}$$

where $p < 1/2$ is the root of the equation $1 - R = H(p)$*****. Taking (2) into account, we rewrite the Varshamov–Gilbert result in the following form.

---

* The elements of the field are 0 and 1; addition and multiplication are defined by the relations
$0 \oplus 0 = 1 \oplus 1 = 0,\ 0 \oplus 1 = 1,\ 0 \cdot 0 = 0 \cdot 1 = 0,\ 1 \cdot 1 = 1.$

** The operations in V^n are defined by the relations
$\mathbf{v} \oplus \mathbf{u} = (v^1, \dots, v^n) \oplus (u^1, \dots, u^n) =$
$(v^1 \oplus u^1, \dots, v^n \oplus u^n),\ w \cdot \mathbf{v} = (w \cdot v^1, \dots, w \cdot v^n)$, where $w, v^i, u^i \in GF(2)$.

*** The weight $w(\mathbf{a})$ of a vector $\mathbf{a}$ is the number of its nonzero coordinates.

**** $[a]$ is the integer part of the number a.

***** $H(p) = -p \log_2 p - (1-p)\log_2(1-p)$.

Theorem 1. (Varshamov, Gilbert). If the difference

$$\left[np + \frac{1}{2}\left(\log_2 \frac{1-p}{p}\right)^{-1} \log_2 n\right] - s_n \tag{3}$$

is bounded below by some constant c_p^*, then $\beta_n(s_n) > 0$.

For the distribution function $\beta_n(t)$ the following estimate is known (see [3,4]):

$$\beta_n(t) \geqslant 1 - 2^{k-n} \sum_{0 \leqslant i \leqslant t} C_n^i,$$

from which one can derive:

Theorem 2. (Gallager, Koshelev). If the difference (3) tends to $+\infty$, then $\beta_n(s_n) \to 1$ as $n \to \infty$.

The following complements Theorem 2.

Theorem 3. If the difference (3) tends to $-\infty$, then $\beta_n(s_n) \to 0$ as $n \to \infty$.

Theorem 3 is an immediate consequence of the following:

Theorem 4. Uniformly for

$$t \leqslant \tau_n = np + \left(\log_2 \frac{1-p}{p}\right)^{-1} \log_2 n - c'_p,$$

the relation

$$\beta_n(t) = \left[1 + O\left(2^{-\delta\sqrt{n}}\right)\right] \exp\left\{-2^{k-n} \sum_{i=0}^{t} C_n^i\right\}, \qquad n \to \infty. \tag{4}$$

Corollary. If the difference

$$\left[np + \left(\log_2 \frac{1-p}{p}\right)^{-1} \log_2 n\right] - s_n \geqslant c''_p,$$

then $\beta_n(s_n) > 0$. In other words, there exists a (k, n)-code with minimum weight

$$np + \left(\log_2 \frac{1-p}{p}\right)^{-1} \log_2 n + O(1)$$

or greater.

We proceed to the proof of Theorem 4. Introduce the notation D_v^t for the event

$$\{w(x_v^1 a_1 \oplus \cdots \oplus x_v^k a_k) > t\},$$

which consists in the fact that the weight of the linear combination of the rows a_i of the matrix $\|a_i^j,\ i = 1, \ldots, k;\ j = 1, \ldots, n\|$, with coefficients x_v^i, is greater than t; here $x_v^1 \ldots x_v^k$ is the binary representation of the number v, $1 \leqslant v \leqslant 2^k - 1$. Then

$$\beta_n(t) = P\{\eta_n > t\} = P\left\{\bigcap_{1 \leqslant v \leqslant 2^k - 1} D_v^t\right\}.$$

Lemma 1 ([5]). Let $G_v,\ v = 1, \ldots, N$, be an arbitrary collection of events, and

$$S_r = \sum_{1 \leqslant v_1 < v_2 < \cdots < v_r \leqslant N} P\{G_{v_1} G_{v_2} \ldots G_{v_r}\}. \tag{5}$$

Then

$$P\left\{\bigcup_{v=1}^{N} G_v\right\} = S_1 - S_2 + S_3 - \cdots + (-1)^{N-1} S_N.$$

Moreover,

$$P\left\{\bigcup_{v=1}^{N} G_v\right\} \geqslant S_1 - S_2 + \cdots - S_{2m}, \tag{6}$$

$$P\left\{\bigcup_{v=1}^{N} G_v\right\} \leqslant S_1 - S_2 + \cdots - S_{2m} + S_{2m+1}, \tag{7}$$

where m is any number such that $2m + 1 \leqslant N$.

Applying (6) and (7) to the events $G_v = \overline{D}_v^t$, $N = 2^k - 1$, and using Lemma 4, proved below, we arrive at the following expression for $\beta_n(t)$:

$$\beta_n(t) = \sum_{r=0}^{2m} \frac{[-u_n(t)]^r}{r!} + O\left\{\frac{[u_n(t)]^{2m+1}}{(2m+1)!} + 2^{-\varepsilon n} e^{u_n(t)}\right\}, \tag{8}$$

---

* c_p, c_p', c_p'' are constants depending only on p.

where

$$u_n(t) = 2^{k-n} \sum_{0 \leq i \leq t} C_n^i, \qquad m = O(\sqrt{n}).$$

If $u_n(\tau_n) \leq \frac{1}{2}m$, then the remainder term in (8) tends to zero uniformly for $t \leq \tau_n$ as $n \to \infty$. Further,

$$\log_2 u_n(\tau_n) = -\frac{1}{2}\log_2 n + n\left[R - 1 + H\left(\frac{\tau_n}{n}\right)\right] + O(1) =$$

$$= n\left[H\left(\frac{\tau_n}{n}\right) - H\left(\frac{t_n}{n}\right)\right] + O(1) = \frac{1}{2}\log_2 n - \left(\log_2 \frac{1-p}{p}\right) c_p' + O(1).$$

Choosing the constant c_p' sufficiently large, we ensure that the condition $u_n(\tau_n) \leq \frac{1}{2}m$ is fulfilled, whatever the m of the form $O(\sqrt{n})$. Thus it remains for us to prove Lemma 4, which is preceded by the following two propositions.

Lemma 2. If a system of vectors $\mathbf{x}_i = (x_i^1, \dots, x_i^k) \in V^k,\ i = 1, \dots, l;\ l \le k$, is linearly independent, then the random vectors

$$\mathbf{b}_i = x_i^1 \mathbf{a}_1 \oplus \cdots \oplus x_i^k \mathbf{a}_k, \qquad i = 1, \dots, l,$$

are mutually independent.*

Proof. Since all the quantities a_i^j are mutually independent, it suffices to verify that

$$x_i^1 a_1^1 \oplus \cdots \oplus x_i^k a_k^1, \qquad i = 1, \dots, l,$$

are mutually independent. Without loss of generality one may assume that $l = k$. We shall show that the probability of the intersection of the events

$$\{x_i^1 a_1^1 \oplus \cdots \oplus x_i^k a_k^1 = y_i\}, \qquad i = 1, \dots, k, \tag{9}$$

is equal to the product of the probabilities, i.e. 2^{-k}; here $y_i = 0$ or $1,\ i = 1, \dots, k$. Solving the system of linear equations (9) by elimination of the unknowns, we arrive at an event equivalent to (9),

$$\bigcap_{1 \le i \le k} \{a_i^k = \tilde{y}_i\}$$

(where $\tilde{y}_i = 0$ or $1,\ i = 1, \dots, k$), which obviously has probability 2^{-k}. This proves Lemma 2.

Lemma 3. Let $\mathbf{b}_1, \dots, \mathbf{b}_l;\ l \ge 2$, be independent random vectors from V^n. Then, uniformly in $t \le \tau_n$ and $l \le n$,

$$P\{w(\mathbf{b}_1) \le t, \dots, w(\mathbf{b}_l) \le t,\ w(\mathbf{b}_1 \oplus \cdots \oplus \mathbf{b}_l) \le t\} = O(2^{-\varepsilon_1 n}) \left[\sum_{i \le t} C_n^i 2^{-n} \right]^l, \tag{10}$$

where $\varepsilon_1 > 0$ does not depend on l.

Proof. Denote by α_l the probability (10), and by γ_l the probability of the same event as in (10), but now assuming that the components of the vectors $\mathbf{b}_1, \dots, \mathbf{b}_l$ are mutually independent random variables taking the values 1 and 0 with probabilities p_n and $1 - p_n$, respectively; here $p_n = (t_n/n)$. Each elementary outcome associated with the vectors $\mathbf{b}_1, \dots, \mathbf{b}_l$ is described by a binary $l \times n$ matrix. The number of such matrices that correspond to the event in (10) is equal to $\alpha_l \cdot 2^{ln}$, and each such matrix has no more than tl ones. Therefore,

$$\gamma_l \geq \alpha_l 2^{ln} p_n^{tl} (1-p_n)^{nl-tl}, \tag{11}$$

provided only that $p_n \leq \frac{1}{2}$. Obviously,

$$\gamma_l \leq \hat{P}\{w(\mathbf{b}_1 \oplus \cdots \oplus \mathbf{b}_l) \leq t\},$$

where the circumflex over P indicates the new distribution of the components of the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_l$. The vector $\mathbf{b}_1 \oplus \cdots \oplus \mathbf{b}_l$ has independent coordinates, each of which takes the value 1 with probability δ_l satisfying the recurrence relation

$$\delta_l = \delta_{l-1}(1-p_n) + (1-\delta_{l-1})p_n, \qquad \delta_1 = p_n.$$

It is not hard to derive from this that δ_l increases monotonically with l (toward $\frac{1}{2}$). Therefore, for $t \leq \delta_2 n$,

$$\gamma_l \leq \sum_{i \leq t} C_n^i \delta_2^i (1-\delta_2)^{n-i}, \qquad l \geq 2. \tag{12}$$

---

* Here and below, random vectors $\mathbf{b}_i$, $i = 1, \ldots, l$, are called mutually independent if the collection of all coordinates b_i^j, $i = 1, \ldots, l$; $j = 1, \ldots, n$, is mutually independent.

But $\delta_2 = 2p_n(1-p_n) \to 2p(1-p) > p$, whereas $\tau_n/n \to p$. Consequently, the right-hand side of (12) is $O(2^{-\varepsilon_1 n})$ uniformly in $t \leqslant \tau_n$. Combining (11) and (12) and noting that, uniformly in $t \leqslant \tau_n$,

$$\sum_{i \leqslant t} C_n^i = 2^{o(n)} [p_n^t (1-p_n)^{n-t}]^{-1},$$

we arrive at the required result.

Lemma 4. *For S_r^t, defined by (3), with $G_\nu = \overline{D}_\nu^t$, $N = 2^k - 1$, $r \leqslant 2m+1$, $m = O(\sqrt[7]{n})$, uniformly in $t \leqslant \tau_n$ the relation holds*

$$S_r^t = \frac{[u_n(t)]^r}{r!} + O(2^{-\varepsilon_2 n}) \sum_{i \leqslant r} \frac{[u_n(t)]^i}{i!}. \tag{13}$$

Proof. To simplify the exposition, we identify the indices ν_j in the sum (5) for S_r^t with the k-dimensional binary vectors corresponding to the k-digit binary notation of the numbers ν_j. We split the sum (5) for S_r^t into $r - [\log_2 r]$ parts according to the number of vectors in a maximal linearly independent subsystem of $\nu_1, \ldots, \nu_r$ (such subsystems contain at least $[\log_2 r] + 1$ vectors, since

all $\nu_1, \dots, \nu_r$ are distinct). For the sum $\Sigma^{(r)}$, corresponding to the case of linear independence of the vectors $\nu_1, \dots, \nu_r$, using the independence of the events $\overline{D}^t_{\nu_1}, \dots, \overline{D}^t_{\nu_r}$ (Lemma 2), we obtain

$$\sum^{(r)} P\{\overline{D}^t_{\nu_1} \dots \overline{D}^t_{\nu_r}\} = \frac{[u_n(t)]^r}{r!} \prod_{0 \leqslant j \leqslant r-1} (1 - 2^{-k+j})^*, \tag{14}$$

Let now the maximal linearly independent system consist of $l < r$ vectors, and, for definiteness, suppose that the independent vectors are $\nu_1, \dots, \nu_l$. Then

$$P\{\overline{D}^t_{\nu_1} \dots \overline{D}^t_{\nu_r}\} \leqslant P\{\overline{D}^t_{\nu_1} \dots \overline{D}^t_{\nu_l}, \overline{D}^t_{\nu_{l+1}}\} =$$

$$= P\{w(\mathbf{b}_1) \leqslant t, \dots, w(\mathbf{b}_l) \leqslant t,\ w(x^1\mathbf{b}_1 \oplus \dots \oplus x^l\mathbf{b}_l) \leqslant t\},$$

where the vectors $\mathbf{b}_1, \dots, \mathbf{b}_l$ are mutually independent, and x^i is equal to 0 or 1, with the number of x^i equal to 1 being at least 2. We estimate from above the number of terms in the sum $\Sigma^{(l)}$ by the quantity

$$\frac{1}{l!} \prod_{0 \leqslant i \leqslant l-1} (2^k - 2^i)\, 2^{l(r-l)} r! = \frac{1}{l!} 2^{kl} O(2^{r^2}) = \frac{1}{l!} 2^{kl} O(2^{m^2}).$$

Choosing $m = O(\sqrt[7]{n})$ so that $m^2 \leqslant \frac{\varepsilon_1}{2} n$, and applying Lemma 3, we arrive at the following estimate for $\Sigma^{(l)}$:

$$\sum^{(l)} P\{\overline{D}^t_{\nu_1} \dots \overline{D}^t_{\nu_r}\} = O(2^{-\varepsilon_2 n}) \frac{[u_n(t)]^l}{l!}, \qquad \varepsilon_2 = \frac{\varepsilon_1}{2}. \tag{15}$$

Summing (14) and (15) for $l = 1, 2, \dots, r-1$, we arrive at (13).

The author expresses his sincere gratitude to A. N. Kolmogorov for posing the problem and for his guidance.

Moscow State University
named after M. V. Lomonosov

Received
30 X 1968

REFERENCES

1. W. Peterson, *Error-Correcting Codes*. Moscow, 1964.

2. L. A. Bassalygo, *Some Problems in Coding Theory*, Dissertation, Moscow State University, 1967.

3. R. Gallager, *Low-Density Parity-Check Codes*, Moscow, 1966.

4. V. N. Koshelev, *Problems of Information Transmission*, 1, no. 4 (1965).

5. W. Feller, *An Introduction to Probability Theory and Its Applications*, Moscow, 1964.

6. D. Slepian, in: *Theory of Message Transmission*, collection, IL, 1967.

$$^{*}\prod_{0\leqslant j\leqslant r-1}(2^k-2^j)$$

is the number of $r\times k$ matrices of rank r (see (6)).

Note: Figure translations are in progress. See original paper for figures.