Fifteenth International Workshop on Algebraic and Combinatorial Coding Theory
June 18-24, 2016, Albena, Bulgaria *pp. 151–156*

Threshold Decoding for Disjunctive Group Testing [1]

A.G. D'yachkov agd-msu@yandex.ru
I.V. Vorobyev vorobyev.i.v@yandex.ru
N.A. Polyanskii nikitapolyansky@gmail.com
V.Yu. Shchukin vpike@mail.ru
Lomonosov Moscow State University, Moscow, Russia

Abstract. Let $1 \le s < t$, $N \ge 1$ be integers and a complex electronic circuit of size t is said to be an s-active, $s \ll t$, and can work as a system block if not more than s elements of the circuit are defective. Otherwise, the circuit is said to be an s-defective and should be replaced by a similar s-active circuit. Suppose that there exists a possibility to run N non-adaptive group tests to check the s-activity of the circuit. As usual, we say that a (disjunctive) group test yields the positive response if the group contains at least one defective element. Along with the conventional decoding algorithm based on disjunctive s-codes, we consider a threshold decision rule with the minimal possible decoding complexity, which is based on the simple comparison of a fixed threshold T, $1 \le T \le N-1$, with the number of positive responses p, $0 \le p \le N$. For the both of decoding algorithms we discuss upper bounds on the α-level of significance of the statistical test for the null hypothesis $\{H_0$: the circuit is s-active$\}$ verse the alternative hypothesis $\{H_1$: the circuit is s-defective$\}$.

1 Statement of Problem

Let $N \ge 2$, $t \ge 2$, s and T be integers, where $1 \le s < t$ and $1 \le T < N$. The symbol $\triangleq$ denote the equality by definition, $|A|$ – the size of the set A and $[N] \triangleq \{1, 2, \dots, N\}$ – the set of integers from 1 to N. A binary $(N \times t)$-matrix

$$X = \|x_i(j)\|, \quad x_i(j) = 0, 1, \quad i \in [N],\ j \in [t], \tag{1}$$

with t columns (*codewords*) $\boldsymbol{x}(j) \triangleq (x_1(j), x_2(j), \dots, x_N(j))$, $j \in [t]$, and N rows $\boldsymbol{x}_i \triangleq (x_i(1), x_i(2) \dots, x_i(t))$, $i \in [N]$, is called a *binary code of length N and size $t = \lfloor 2^{RN} \rfloor$*, where a fixed parameter $R > 0$ is called a *rate* of the code X. The number of 1's in a binary column $\boldsymbol{x} = (x_1, \dots, x_N) \in \{0,1\}^N$, i.e., $|\boldsymbol{x}| \triangleq \sum_{i=1}^N x_i$, is called a *weight* of $\boldsymbol{x}$. A code X is called a *constant weight binary code of weight w*, $1 \le w < N$, if for any $j \in [t]$, the weight $|\boldsymbol{x}(j)| = w$. The conventional symbol $\boldsymbol{u} \bigvee \boldsymbol{v}$ will be used to denote the disjunctive (Boolean)

[1] The research is supported in part by the Russian Foundation for Basic Research under Grant No. 16-01-00440.

sum of binary columns $\boldsymbol{u}, \boldsymbol{v} \in \{0,1\}^N$. We say that a column $\boldsymbol{u}$ *covers* a column $\boldsymbol{v}$ ($\boldsymbol{u} \succeq \boldsymbol{v}$) if $\boldsymbol{u} \bigvee \boldsymbol{v} = \boldsymbol{u}$.

1.1 Disjunctive and Threshold Disjunctive Codes

Definition 1. [1]. A code X (1) is called a *disjunctive s-code*, $s \in [t-1]$, if the disjunctive sum of any s-subset of codewords of X covers those and only those codewords of X which are the terms of the given disjunctive sum.

Let $\mathcal{S}$, $\mathcal{S} \subset [t]$, be an arbitrary fixed collection of defective elements of size $|\mathcal{S}|$. For a binary code X and collection $\mathcal{S}$, define the binary *response vector* of length N, namely:

$$\boldsymbol{x}(\mathcal{S}) \triangleq \bigvee_{j \in \mathcal{S}} \boldsymbol{x}(j), \quad \text{if} \quad \mathcal{S} \neq \emptyset \quad \text{and} \quad \boldsymbol{x}(\mathcal{S}) \triangleq (0,0,\ldots,0) \quad \text{if} \quad \mathcal{S} = \emptyset. \tag{2}$$

In the classical problem of *non-adaptive group testing*, we describe N tests as a binary $(N \times t)$-matrix $X = \|x_i(j)\|$, where a column $\boldsymbol{x}(j)$ corresponds to the j-th element, a row $\boldsymbol{x}_i$ corresponds to the i-th test and $x_i(j) \triangleq 1$ if and only if the j-th element is included into the i-th testing group. The result of each test equals 1 if at least one defective element is included into the testing group and 0 otherwise, so the column of results is exactly equal to the response vector $\boldsymbol{x}(\mathcal{S})$. Definition 1 of disjunctive s-code X gives the important sufficient condition for the evident identification of any unknown collection of defective elements $\mathcal{S}$ if the number of defective elements $|\mathcal{S}| \le s$. In this case, the identification of the unknown $\mathcal{S}$ is equivalent to discovery of all codewords of code X covered by $\boldsymbol{x}(\mathcal{S})$, and its complexity is equal to the code size t. Note that this algorithm also allows us to check s-activity of the circuit defined in the abstract. Moreover, it is easy to prove by contradiction that every code X which allows to check s-activity of the circuit without error is a disjunctive s-code.

Definition 2. Let s, $s \in [t-1]$, and T, $T \in [N-1]$, be arbitrary fixed integers. A disjunctive s-code X of length N and size t is said to be a a disjunctive s-code with *threshold* T (or, briefly, *s^T-code*) if the disjunctive sum of any $\le s$ codewords of X has weight $\le T$ and the disjunctive sum of any $\ge s+1$ codewords of X has weight $\ge T+1$.

Obviously, for any s and T, the definition of s^T-code gives a sufficient condition for code X applied to the group testing problem described in the abstract of our paper. In this case, only on the base of the known *number of positive responses* $|\boldsymbol{x}(\mathcal{S})|$, we decide that the controllable circuit identified by an unknown collection $\mathcal{S}$, $\mathcal{S} \subset [t]$, is s-active, i.e., the unknown size $|\mathcal{S}| \le s$ (s-defective, i.e., the unknown size $|\mathcal{S}| \ge s+1$) if $|\boldsymbol{x}(\mathcal{S})| \le T$ ($|\boldsymbol{x}(\mathcal{S})| \ge T+1$).

Remark 1. The concept of s^T-codes was motivated by troubleshooting in complex electronic circuits using a non-adaptive identification scheme which was considered in [2].

1.2 Hypothesis Test

Let a circuit of size t is identified by an unknown collection $\mathcal{S}_{un}$, $\mathcal{S}_{un} \subset [t]$, of defective elements of an unknown size $|\mathcal{S}_{un}|$ and X be a code (1) of size t and length N. Introduce the null hypothesis $\{H_0 : |\mathcal{S}_{un}| \le s\}$ (the circuit is s-active) verse the alternative $\{H_1 : |\mathcal{S}_{un}| \ge s+1\}$ (the circuit is s-defective). In this paper we focus on the testing of statistical hypotheses H_0 and H_1. The similar problem related to constructing of a confidence interval for $|\mathcal{S}_{un}|$ was considered in [3], where the authors construct the interval $[\hat{s}/c\,;\,\hat{s}]$, such that given a *random code* X, the statistic $\hat{s}$, i.e., a function of the random response vector $\boldsymbol{x}(\mathcal{S}_{un})$, satisfies the following properties: $\Pr\{\hat{s} < |\mathcal{S}_{un}|\}$ is upper bounded by a small parameter $\epsilon \ll 1$ and the expected value of $\hat{s}/|\mathcal{S}_{un}|$ is upper bounded by a number $c > 1$.

For fixed parameters s, $1 \le s < t$, and T, $1 \le T < N$, consider the following *threshold decision rule* motivated by Definition 2, namely:

$$\begin{cases} \text{accept } \{H_0 : |\mathcal{S}_{un}| \le s\} & \text{if } |\boldsymbol{x}(\mathcal{S}_{un})| \le T, \\ \text{accept } \{H_1 : |\mathcal{S}_{un}| \ge s+1\} & \text{if } |\boldsymbol{x}(\mathcal{S}_{un})| \ge T+1. \end{cases} \tag{3}$$

For the conventional *statistical* interpretation of the decision rule (3), it is reasonable to assume that the different collections of defective elements of the same size are *equiprobable*. That is why, we set that the *probability distribution* of the random collection $\mathcal{S}_{un}$, $\mathcal{S}_{un} \subset [t]$, is identified by an unknown probability vector $\boldsymbol{p} \triangleq (p_0, p_1, \ldots, p_t)$, $p_k \ge 0$, $k = 0, 1, \ldots, t$, $\sum_{k=0}^{t} p_k = 1$, as follows:

$$\Pr\{\mathcal{S}_{un} = \mathcal{S}\} \triangleq \frac{p_{|\mathcal{S}|}}{\binom{t}{|\mathcal{S}|}} \quad \text{for any subset} \qquad \mathcal{S} \subseteq [t]. \tag{4}$$

Introduce a *maximal error probability* of the decision rule (3) :

$$\varepsilon_s(T, \boldsymbol{p}, X) \triangleq \max\left\{\Pr\{\text{accept } H_1 \big| H_0\}, \Pr\{\text{accept } H_0 \big| H_1\}\right\} \tag{5}$$

where the conditional probabilities in the right-hand side of (5) are identified by (3)-(4). Note that the number $\varepsilon_s(T, \boldsymbol{p}, X) = 0$ if and only if the code X is an s^T-code. Denote by $t_s(N, T)$ the maximal size of s^T-codes of length N. For a parameter τ, $0 < \tau < 1$, introduce the rate of $s^{\lfloor \tau N \rfloor}$-codes:

$$R_s(\tau) \triangleq \varlimsup_{N \to \infty} \frac{\log_2 t_s(N, \lfloor \tau N \rfloor)}{N} \ge 0, \quad 0 < \tau < 1. \tag{6}$$

Definition 3. Let τ, $0 < \tau < 1$, and a parameter R, $R > R_s(\tau)$, be fixed. For the maximal error probability $\varepsilon_s(T, \boldsymbol{p}, X)$, defined by (3)-(5), consider the function

$$\varepsilon_s^N(\tau, R) \triangleq \max_{\boldsymbol{p}} \left\{\min_{X} \varepsilon_s(\lfloor \tau N \rfloor, \boldsymbol{p}, X)\right\}, \quad R > R_s(\tau), \tag{7}$$

where the minimum is taken over all codes X of length N and size $t = \lfloor 2^{RN} \rfloor$ with parameter $R > R_s(\tau)$. The number $\varepsilon_s^N(\tau, R) > 0$ does not depend on the unknown probability vector $\boldsymbol{p}$ and can be called the *universal* error probability of the decision rule (3). The corresponding *error exponent*

$$E_s(\tau, R) \triangleq \overline{\lim_{N\to\infty}} \frac{-\log_2 \varepsilon_s^N(\tau, R)}{N}, \quad s \geq 1, \quad 0 < \tau < 1, \quad R > R_s(\tau) \tag{8}$$

identifies the asymptotic behavior of α*-level of significance* for the decision rule (3), i.e.,

$$\alpha \triangleq \exp_2\{-N\,[E_s(\tau, R) + o(1)]\}, \quad \text{if} \quad E_s(\tau, R) > 0, \quad N \to \infty. \tag{9}$$

Along with (3) we introduce the *disjunctive decision rule* based on the conventional algorithm:

$$\begin{cases} \text{accept } H_0 & \text{if } \boldsymbol{x}(\mathcal{S}_{un}) \text{ covers } \leq s \text{ codewords of } X, \\ \text{accept } H_1 & \text{if } \boldsymbol{x}(\mathcal{S}_{un}) \text{ covers } \geq s+1 \text{ codewords of } X. \end{cases} \tag{10}$$

For a fixed code rate R, $R > 0$, the error exponent for disjunctive decision rule (10) $E_s(R)$ is defined similarly to (5)-(8). The function $E_s(R)$ was firstly introduced in our paper [1], where we proved

Theorem 1. [1]. *The function $E_s(R) = 0$ if $R \geq 1/s$.*

2 Lower Bounds on Error Exponents

In this Section, we formulate and compare the random coding lower bounds for the both of error exponents $E_s(R)$ and $E_s(\tau, R)$. These bounds were proved applying the random coding method based on the ensemble of constant-weight codes. A parameter Q in formulations of theorems 2-3 means the relative weight of codewords of constant-weight codes. Introduce the standard notations

$$h(Q) \triangleq -Q\log_2 Q - (1-Q)\log_2[1-Q], \quad [x]^+ \triangleq \max\{x, 0\}.$$

In [1], we established

Theorem 2. [1]. *The error exponent $E_s(R) \geq \underline{E}_s(R)$ where the random coding lower bound*

$$\underline{E}_s(R) \triangleq \max_{0<Q<1} \min_{Q \leq q < \min\{1, sQ\}} \left\{\mathcal{A}(s, Q, q) + [h(Q) - qh(Q/q) - R]^+\right\},$$

$$\mathcal{A}(s, Q, q) \triangleq (1-q)\log_2(1-q) + q\log_2\left[\frac{Qy^s}{1-y}\right] + sQ\log_2\frac{1-y}{y} + sh(Q),$$

and y is the unique root of the equation

$$q = Q\frac{1-y^s}{1-y}, \quad 0 < y < 1.$$

In addition, as $s \to \infty$ and $R \le \frac{\ln 2}{s}(1+o(1))$, the lower bound $\underline{E}_s(R) > 0$.

Theorem 3. 1. *The error exponent $E_s(\tau, R) \ge \underline{E}_s(\tau, R)$ where the random coding bound $\underline{E}_s(\tau, R)$ does not depend on $R > 0$ and has the form:*

$$\underline{E}_s(\tau, R) \triangleq \max_{1-(1-\tau)^{1/(s+1)} < Q < 1-(1-\tau)^{1/s}} \min\left\{\mathcal{A}'(s, Q, \tau), \mathcal{A}(s+1, Q, \tau)\right\},$$

$$\mathcal{A}'(s, Q, q) \triangleq \begin{cases} \mathcal{A}(s, Q, q), & \text{if } Q \le q \le sQ, \\ \infty, & \text{otherwise.} \end{cases}$$

2. *As $s \to \infty$ the optimal value of $\underline{E}_s(\tau, R)$ satisfies the inequality:*

$$\underline{E}_{Thr}(s) \triangleq \max_{0<\tau<1} \underline{E}_s(\tau, R) \ge \frac{\log_2 e}{4s^2}(1+o(1)), \quad s \to \infty.$$

It is possible to use decision rule (3) with any value of parameter T. The numerical values of the optimal error exponent $\underline{E}_{\mathrm{Thr}}(s)$ along with the corresponding optimal values of threshold parameter τ and the constant-weight code ensemble parameter Q are presented in Table 1. Besides in Table 1 the values of $\underline{E}_s(0)$ and $R_{\mathrm{cr}} \triangleq \sup\{R : \underline{E}_s(R) > \underline{E}_{\mathrm{Thr}}(s)\}$ are shown. Theorems 1-3 show that, for large values of R, the threshold decision rule (3) has an advantage over the disjunctive decision rule (10) as $N \to \infty$.

Table 1: The numerical values of $\underline{E}_{\mathrm{Thr}}(s)$

s	2	3	4	5	6	7	8
$\underline{E}_{\mathrm{Thr}}(s)$	0.1380	0.0570	0.0311	0.0196	0.0135	0.0098	0.0075
τ	0.2065	0.1365	0.1021	0.0816	0.0679	0.0582	0.0509
Q	0.1033	0.0455	0.0255	0.0163	0.0113	0.0083	0.0064
$\underline{E}_s(0)$	0.3651	0.2362	0.1754	0.1397	0.1161	0.0994	0.0869
R_{cr}	0.2271	0.1792	0.1443	0.1201	0.1027	0.0896	0.0794

3 Simulation for finite code parameters

For finite N and t, we carried out a simulation as follows. The probability distribution vector $\boldsymbol{p}$ (4) is defined by

$$p_k \triangleq \binom{t}{k} p^k (1-p)^{t-k}, \quad p \triangleq \frac{s+1/2}{t}, \quad 0 \le k \le t,$$

i.e. the number of defective elements $|\mathcal{S}_{un}|$ is binomially distributed and has the expected value $s+1/2$. A code X is generated randomly from the ensemble of constant-weight codes, i.e. for some weight parameter w, every codeword of X is chosen independently and equalprobably from the set of all $\binom{t}{w}$ codewords. For every weight w and every decision rule, we repeat generation 1000 times and choose the code with minimal error probability. Note that for disjunctive decision rule $\Pr\{\text{accept } H_0|H_1\} = 0$. In Table 2 the best values of maximal error probability for fixed parameters s, t and N are shown in bold.

Table 2: Results of simulation

	Threshold decision rule				Disjunctive decision rule	
N	$\Pr\{\text{acc. } H_1\|H_0\}$	$\Pr\{\text{acc. } H_0\|H_1\}$	w	T	$\Pr\{\text{acc. } H_1\|H_0\}$	w
			$s=2,\quad t=15$			
5	**0.1366**	0.1355	2	3	0.4780	2
8	0.0732	**0.0824**	3	5	0.3610	2
10	0	**0.0744**	1	2	0.2390	3
12	0	**0.0440**	1	2	0.1220	3
14	0	**0.0349**	2	4	0.0537	3
15	0	0.0258	2	4	**0.0195**	3
			$s=2,\quad t=20$			
5	**0.1398**	0.1365	2	3	0.5356	2
8	**0.0897**	0.0890	3	5	0.4169	2
10	**0.0897**	0.0858	3	5	0.3008	3
12	**0.0580**	0.0576	4	7	0.1979	3
15	0	**0.0324**	2	4	0.0792	4

References

[1] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Almost Disjunctive List-Decoding Codes, *Problems of Information Transmission*, vol. 51, no. 2, pp. 110-131, 2015.

[2] *Zubashich V.F.*, *Lysyansky A.V.*, *Malyutov M.B.*, Blockrandomized distributed troubleshooting construction in large circuits with redundancy. *Izvestia of the USSR Acad. of Sci., Technical Cybernetics*, vol. 6, 1976.

[3] *Damaschke P.*, *Muhammad A.S.*, Competitive group testing and learning hidden vertex covers with minimum adaptivity, *Discrete Math. Algorithm. Appl.*, vol. 2, no. 3, pp. 291-311, 2010.