On a Hypergraph Approach to Multistage Group Testing Problems.

D'yachkov A.G., Vorobyev I.V., Polyanskii N.A., Shchukin V.Yu.

Vorobyev I.V.

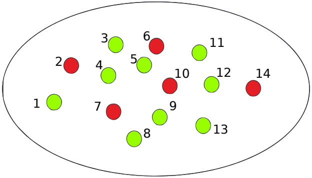
Lomonosov Moscow State University, Faculty of Mechanics and Mathematics

ACCT 2016

Problem Statement

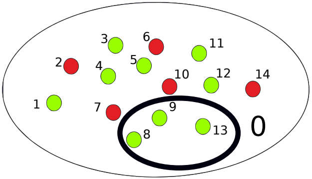
Let $T = \{1, 2, \ldots, t\}$ be a set of objects and $S_{un} \subset T$, $|S_{un}| \leq s$, be a set of defective elements. Our goal is to find the set S_{un} by performing the minimal number of tests on a chosen subsets of T. The answer to the test $S \subset T$ is positive iff $S \cap S_{un} \neq \varnothing$.

Combinatorial Group Testing



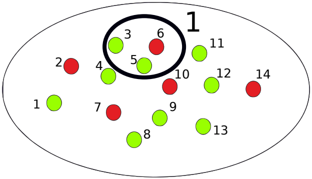
Example:
$t = 14$,
$S_{un} = \{2, 6, 7, 10, 14\}$.

Combinatorial Group Testing



Example:
$t = 14$,
$S_{un} = \{2, 6, 7, 10, 14\}$.
$S = \{8, 9, 13\}$.
$S \cap S_{un} = \varnothing$, thus the test result is negative(0).

Combinatorial Group Testing



Example:
$t = 14$,
$S_{un} = \{2, 6, 7, 10, 14\}$.
$S = \{3, 5, 6\}$.
$S \cap S_{un} = \{6\}$, thus the test result is positive(1).

Different types of search algorithms

- adaptive - later tests depend on the results of previous tests
- nonadaptive - all tests are carried out in parallel.
  Example: disjunctive codes
- multistage - algorithm consists of the several stages, where tests of stage i depend on the results of tests from stages $1, 2, \ldots i-1$.
  Example: list decoding disjunctive codes for 2 stages.

Matrix representation

Any non-adaptive algorithm consisting of N tests can be represented by a binary $N \times t$ matrix X such that each test corresponds to the row, and each element stands for the column.

We put $x_i(j) = 1$ if the j-th element is included in i-th test;
otherwise, $x_i(j) = 0$.

Outcomes of tests can be represented by a binary vector $r(X, S_{un})$.

By $N^p(t, s)$ we denote minimal number of tests in an algorithm, which finds s defects among t elements using p stages.

Hypergraph

Definition

A *hypergraph is a pair* $H = (V, E)$ *such that* $E \subset 2^V \setminus \varnothing$, *where* V *is a set of vertices and* E *is a set of hyperedges.*

Chromatic number

Definition

A *coloring* of H is a map $\varphi : V \to \mathbb{N}$ such that each hyperedge $e \in E$ contains at least two vertices $u, v \in e$ of distinct colors $\varphi(u) \neq \varphi(v)$. The corresponding *chromatic number* $\chi_s(H)$ is the least number of colors for which H has a proper coloring.

Strong Coloring

Definition

A *strong coloring* of H is a map $\varphi : V \to \mathbb{N}$ such that whenever $u, v \in e$ for some $e \in E$, we have that $\varphi(u) \neq \varphi(v)$. The corresponding *strong chromatic number* $\chi_s(H)$ is the least number of colors for which H has a proper strong coloring.

Suppose that we have already performed some set of tests X.

The set of vertices of hypergraph $H(X, S_{un}) = (T, E)$ is equal to the set of elements T. The set of edges E equals to the set of all possible sets of defects, i.e, all subsets $S \subset T$, $|S| \leq s$, such that $r(X, S_{un}) = r(X, S)$.

Previous Results

Bounds from disjunctive codes

$$c_1 \frac{s^2}{\log_2 s} \log_2 t \leq N^1(t, s) \leq c_2 s^2 \log_2 t$$

If X is a matrix of tests of non-adaptive algorithm, then hypergraph $H(X, S_{un})$ has only one hyperedge.

Previous Results

Bounds from list-decoding disjunctive codes

$$s \log_2 t(1 + o(1)) \le N^2(t, s) \le cs \log_2 t(1 + o(1)), t \to \infty$$

If X is a matrix of tests corresponding to the list-decoding disjunctive code, then hypergraph $H(X, S_{un})$ has only constant (independent of t) number of hyperedges.

Previous Results

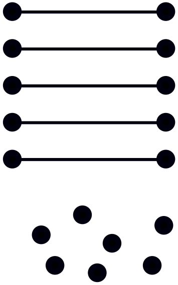
Specific two stage algorithm for s=2

There exists a $N \times t$ matrix X, $N = 2\log_2 t$, such that the graph $H(X, S_{un})$ has $\sqrt{t}$ vertices with degree 1 and $t - \sqrt{t}$ isolated vertices.

Previous Results

Specific two stage algorithm for s=2

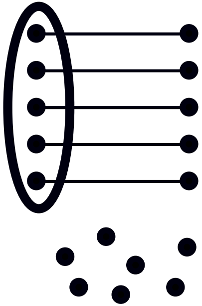
There exists a $N \times t$ matrix X, $N = 2\log_2 t$, such that the graph $H(X, S_{un})$ has $\sqrt{t}$ vertices with degree 1 and $t - \sqrt{t}$ isolated vertices.



Previous Results

Specific two stage algorithm for s=2

There exists a $N \times t$ matrix X, $N = 2\log_2 t$, such that the graph $H(X, S_{un})$ has $\sqrt{t}$ vertices with degree 1 and $t - \sqrt{t}$ isolated vertices.



We can find one defective element among $\sqrt{t}/2$ objects using $0.5\log_2 t$ tests.

Previous Results

Specific two stage algorithm for s=2

There exists a $N \times t$ matrix X, $N = 2\log_2 t$, such that the graph $H(X, S_{un})$ has $\sqrt{t}$ vertices with degree 1 and $t - \sqrt{t}$ isolated vertices.

Corollary

$$N^2(t, 2) \leq 2.5\log_2 t(1 + o(1)), t \to \infty$$

Previous Results

Specific two stage algorithm for s=2

There exists a $N \times t$ matrix X, $N = 2\log_2 t$, such that the graph $H(X, S_{un})$ has $\sqrt{t}$ vertices with degree 1 and $t - \sqrt{t}$ isolated vertices.

Corollary

$$N^2(t,2) \le 2.5\log_2 t(1 + o(1)), t \to \infty$$

For $s = 2$ list decoding disjunctive code give algorithm with the number of tests

$$N^2_{LD}(t,2) \approx 3.11\log_2 t(1 + o(1)).$$

Goal

Information theory bound

$$N^p(t, s) \geq s \log_2 t(1 + o(1)), t \to \infty$$

Adaptive algorithm

$$N^\infty(t, s) = s \log_2 t(1 + o(1)), t \to \infty$$

Goal

Find p such that

$$N^p(t, s) = s \log_2 t(1 + o(1)), t \to \infty$$

4 stage procedure

First stage

Let X be a $N \times t$ matrix of tests of the first stage and $H(X, S_{un}) = (V, E)$ is a corresponding hypergraph. Find the strong chromatic number $\chi_s(H)$ such that there exist disjoint sets $V_1, V_2, \ldots, V_k$, $V = V_1 \bigcup V_2 \bigcup \ldots \bigcup V_k$, $|V_i \bigcap e| \leq 1$ for all $e \in E$.

Note that each set V_i has at most one defective element.

Second stage

Test each set V_i individually.

Here we find the cardinality of the set S_{un} and the set $\{V_{i_1}, V_{i_2}, \ldots, V_{i_{|S_{un}|}}\}$, each of which contains one defective element.

Third stage

Find defective element in the set V_{i_1} by carrying out $\lceil \log_2 |V_{i_1}| \rceil$ tests.

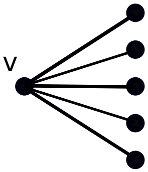
Observe that actually by performing $\sum\limits_{j=1}^{\mathcal{S}_{un}} \lceil \log_2 |V_{i_j}| \rceil$ tests we could identify all defects $\mathcal{S}_{un}$ on this stage.

Fourth stage

Consider all hyperedges $e \in E$, such that e contains the found vertex v and consists of vertices of $v \cup V_{i_2} \cup \ldots \cup V_{i_{|\mathcal{S}_{un}|}}$. For each such e test the set $T \setminus e$.

Fourth stage

Consider all hyperedges $e \in E$, such that e contains the found vertex v and consists of vertices of $v \cup V_{i_2} \cup \ldots \cup V_{i_{|\mathcal{S}_{un}|}}$. For each such e test the set $T \backslash e$.

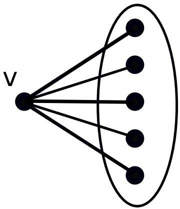


Case s=2

We can use $\lceil \log_2 \deg(v) \rceil$ tests instead of $\deg(v)$.

Fourth stage

Consider all hyperedges $e \in E$, such that e contains the found vertex v and consists of vertices of $v \cup V_{i_2} \cup \ldots \cup V_{i_{|\mathcal{S}_{un}|}}$. For each such e test the set $T \backslash e$.



Case s=2

We can use $\lceil \log_2 \deg(v) \rceil$ tests instead of $\deg(v)$.

Total number of tests

Let t' be a number of non-isolated vertices in hypergraph H, and d be a maximal degree of vertex from V. Then the total number of tests can be bounded by

$$N + \chi_s(H) + \lceil \log_2 t' \rceil + d.$$

Total number of tests for s=2

$$N + \chi_s(H) + \lceil \log_2 t' \rceil + \lceil \log_2 d \rceil .$$

Construction for s=2

Let D be the set of all binary words with length N_1 such that the number of ones in each codeword is fixed and equals wN_1. Let C be the q-ary code, $q = \binom{N_1}{wN_1}$, consisting of all q-ary words of length N_1 and having size $t = q^{N_2}$. Let X be a binary $N_1 N_2 \times q^{N_1}$ matrix of a concatenated code with inner code D and outer code C.

Theorem

Chromatic number $\chi(H(X, S_{un}))$ is less or equal to q. For $s = 2$, the product of the maximal degree and the number of non-isolated vertices of H is estimated as follows

$$t' \cdot d \le \max_{w \le \hat{w} \le 2w} \left(\binom{\hat{w} N_1}{w N_1} \cdot \binom{w N_1}{(2w - \hat{w}) N_1} \right)^{N_2}.$$

The optimal choice of parameters gives the algorithm with total number of tests

$$T = N_1 N_2 + q + \lceil \log_2 t' \rceil + \lceil \log_2 d \rceil \sim 2 \log_2 t.$$

Theorem

Chromatic number $\chi(H(X, S_{un}))$ is less or equal to q. For $s = 2$, the product of the maximal degree and the number of non-isolated vertices of H is estimated as follows

$$t' \cdot d \leq \max_{w \leq \hat{w} \leq 2w} \left(\binom{\hat{w} N_1}{w N_1} \cdot \binom{w N_1}{(2w - \hat{w}) N_1} \right)^{N_2}.$$

The optimal choice of parameters gives the algorithm with total number of tests

$$T = N_1 N_2 + q + \lceil \log_2 t' \rceil + \lceil \log_2 d \rceil \sim 2 \log_2 t.$$

Corollary

$$N^4(t, 2) = 2 \log_2 t (1 + o(1)).$$

Construction for s>2

Theorem

$$N^{2s+1}(t, s) \leq (2s - 1)\log_2 t(1 + o(1)).$$

Number of tests for $t \leq 1000$

t	tests	t	tests	t	tests
8-9	8	29-36	14	126-256	20
10-16	10	37-64	15	257-441	22
17-27	12	65-81	16	442-784	24
28	13	82-125	18	785-1000	25

Number of tests for $t = 10^k$

$t = q^{N_1}$	tests	information bound	tests / $\log_2 t$
10^3	26	19	2.609
10^4	33	26	2.483
10^5	41	33	2.468
10^6	48	39	2.408
10^7	56	46	2.408
10^8	64	53	2.408
10^9	71	59	2.375
10^{10}	79	66	2.378
10^{11}	86	73	2.354
10^{12}	94	79	2.358
10^{13}	102	86	2.362
10^{14}	109	93	2.344
10^{15}	117	99	2.348
10^{16}	124	106	2.333
10^{17}	132	112	2.337
10^{18}	139	119	2.325

Number of tests for t with small ratio tests / $\log_2 t$

$q^{N_1} = t$	tests	information bound	tests / $\log_2 t$
$28^2 = 784$	24	19	2.496
$15^3 = 3375$	29	23	2.474
$21^3 = 9261$	32	26	2.428
$28^3 = 21952$	35	28	2.427
$15^4 = 50625$	37	31	2.368
$21^4 = 194481$	41	35	2.334
$21^5 = 4084101$	51	43	2.322
$15^6 = 11390625$	54	46	2.304
$21^6 = 85766121$	60	52	2.277
$21^9 = 794280046581$	89	79	2.251
$21^{11} \approx 3.5 \cdot 10^{14}$	108	96	2.235

Thank you for your attention!