Univ. u Novom Sadu
Zb. Rad. Prirod.–Mat. Fak.
Ser. Mat. 24, 2 (1994), 137-141

Review of Research
Faculty of Science
Mathematics Series

GROUP TESTING WITH THREE DEFECTIVES

Ivica Bošnjak
Institute of Mathematics, University of Novi Sad
Trg Dositeja Obradovića 4, 21000 Novi Sad, Yugoslavia

Abstract

The problem of identifying three defective elements on a set of n elements is considered. An algorithm is constructed which can differ from the optimal one by at most one test.

AMS Mathematics Subject Classification (1991): 90B40, 62C20, 90C39
Key words and phrases: group testing, defective element, optimal algorithm.

1. Introduction

Consider the following problem. There are exactly three defective elements in the set $X = \{x_1, x_2, ..., x_n\}$, all possibilities occurring with equal probabilities. We want to identify defective elements by testing some subset A of X. The test of an individual subset A informs us whether A contains some defective elements or not. Such test we denote by T_A. We use the expresion "bad element" for defective elements, and "good element" for other elements. A subset A of X is said to be "bad" if it contains at least one bad element. Otherwise, it is said to be "good". Our aim is to minimize the maximum number of tests required to find all bad elements. We denote by $P_n^m(l)$ any algorithm which enables us to find all bad elements, if there are m of them in a set of n elements, l being the number of tests to be required.

We write $c_m(n) = l$, if $P_n^m(l)$ is optimal. It is important to notice that the function $c_m(n)$ is non-decreasing.
It follows by information-theoretical reasoning that

$$(1) \qquad c_m(n) \geq \left\lceil \log_2 \binom{n}{m} \right\rceil$$

where $\lceil x \rceil$ denotes the least integer $\geq x$.
It is well known that

$$(2) \qquad c_1(n) = \lceil \log_2 n \rceil$$

The case $m = 2$ was investigated in [1]- [4]. It was proved that

$$(3) \qquad c_2(n) \leq \left\lceil \log_2 \binom{n}{2} \right\rceil + 1$$

for all $n \geq 2$.
Chang, Hwang and Lin in [1] proved that

$$(4) \qquad n \leq \left\lceil 43 \cdot 2^{\frac{t}{2}-5} \right\rceil - 1 \Rightarrow c_2(n) \leq t$$

for all even $t \geq 4$.

$$(5) \qquad n \leq \left\lceil 31 \cdot 2^{\frac{t-1}{2}-4} \right\rceil - 1 \Rightarrow c_2(n) \leq t$$

for all odd $t \geq 3$.

2. Results

It is obvious that we can find all bad elements in at most $n - 1$ steps, by testing the subsets of X which contain exactly one element. It is easy to prove that the strategy "element-by-element" is optimal for all $n \leq 8$.
In the case $n = 10$ we must test the subset A of X such that $|A| = 2$.
If A is good, we can apply the algorithm $P_8^3(7)$.
If A is bad we can find bad element in A using only one test. Remaining two bad elements can be identified in additional 6 tests, for $c_2(9) \leq 6$, according to (4). Therefore, $c_3(10) \leq 8$.
In the case $n = 12$ we can apply the following algorithm. The first test is T_A, where $|A| = 2$.
If A is good, we can apply the algorithm $P_{10}^3(8)$.
If A is bad, we can apply the algorithms $P_2^1(1)$ and $P_{11}^2(7)$. Therefore, $c_3(12) \leq 9$.

Lemma 1. $c_3(16) \leq 10$

Proof. The first test is T_A, where $|A| = 4$.
If A is bad, we can find bad element from A using 2 tests. Then, we can identify the remaining bad elements using at most 7 tests, because $c_2(15) \leq 7$ according to (5).
If A is good, we can apply the algorithm $P^3_{12}(9)$.

Lemma 2. *Let $|X| = n$ where $n = 2^k + s$, $0 < s \leq 2^k$ and X contains at least one bad element. Then we can identify one bad element from X using k tests, or using $k+1$ tests and identifying simultaneously at least $2^k - s$ good elements.*

Proof. It is easy to check that the statement holds for $k = 1$.
Suppose that the statement is true for all $l < k$. We must consider two cases.
(a) $s \leq 2^{k-1}$
The first test is T_A where $|A| = 2^{k-1}$.
If A is bad, we can find bad element from A using $k-1$ additional tests.
If A is good, the set $X - A$ is bad and $|X - A| = 2^{k-1} + s$. Therefore, by the induction hypothesis, we can identify one bad element and $2^{k-1} - s$ good elements in additional k steps. The statement is now proved because we have already identified 2^{k-1} good elements from A.
(b) $s > 2^{k-1}$
The first test is A, where $|A| = s$. If A is bad, then there is a bad element in A, and by the induction hypothesis we can find it in $k-1$ steps or in k steps identifying at the same time $2^{k-1} - (s - 2^{k-1}) = 2^k - s$ good elements.
If A is good, then A contains $s > 2^k - s$ good elements, and the set $X - A$ is bad, where $|X - A| = 2^k$. Thus, we can identify one bad element by the halving algorithm using additional k tests.

Theorem 1. *For $k \geq 0$, the following inequalities hold:*

$$c_3(16 \cdot 2^k) \leq 10 + 3k \tag{6}$$

$$c_3(20 \cdot 2^k) \leq 11 + 3k \tag{7}$$

$$c_3(25 \cdot 2^k) \leq 12 + 3k \tag{8}$$

Proof. The proof is by induction. Suppose that $k > 1$ and the statement is true for all $l < k$.

(a) Let $n = 16 \cdot 2^k = 32 \cdot 2^{k-1}$. The first test is T_A where $|A| = 7 \cdot 2^{k-1}$.

If A is good, then, by the induction hypothesis we can apply the algorithm $P^3_{25 \cdot 2^{k-1}}(9+3k)$.

If A is bad, then, by Lemma 2. we can identify a bad element from A using additional $k+2$ tests, identifying 2^{k-1} good elements too. The remaining bad elements are in the set of $31 \cdot 2^{k-1} - 1$ elements and can be identified in $2k+7$ additional tests according to (5).

(b) Let $n = 20 \cdot 2^k$. The first test is T_A, where $|A| = 4 \cdot 2^k$.

If A is good, we can apply the algorithm $P^3_{16 \cdot 2^k}(10+3k)$ from (a).

If A is bad, then, a bad element from A can be identified using additional $k+2$ tests. Then, we can apply the algorithm $P^2_{20 \cdot 2^k - 1}(2k+8)$ which exists according to (4).

(c) Let $n = 25 \cdot 2^k$ and A,B,C,D are disjoint subsets of X such that $|A| = 3 \cdot 2^k$, $|B| = |C| = |D| = 2 \cdot 2^k$. The first test is $T_{A \cup B}$.

If $A \cup B$ is good, we can apply the algorithm $P^3_{20 \cdot 2^k}(11+3k)$ from (b).

If $A \cup B$ is bad, the second test is $T_{B \cup C}$.

If $B \cup C$ is good, there is a bad element in A and we can find it using additional $k+2$ tests. According to Lemma 2., the remaining bad elements are in the set of $20 \cdot 2^k - 1$ elements. Thus, we can identify them using at most $2k+8$ tests.

If $B \cup C$ is bad, the next test is $T_{C \cup D}$.

If $C \cup D$ is good, there is a bad element in B and we can find it using $k+1$ additional tests. Then, we can apply the algorithm $P^2_{21 \cdot 2^k - 1}(2k+8)$ which exists according to (4).

If $C \cup D$ is bad, the fourth test is T_B.

If B is good, A and C are bad, and we can identify bad elements from A and C using additional $2k+3$ tests. Then, we can find the third bad element using $k+5$ tests.

Finally, if B is bad, the second bad element is in $C \cup D$, and we can find them using additional $2k+3$ tests. The remaining bad element can be found in $k+5$ tests.

A proof of inequalities $c_3(20) \leq 11$ and $c_3(25) \leq 12$ is contained in (b) and (c). Therefore, the statement is true for $k = 0$ and we have completed the proof of Theorem 1 by induction.

Theorem 2.

$$c_3(n) \le \left\lceil \log_2 \binom{n}{3} \right\rceil + 1 \tag{9}$$

Proof. We have already proved that $c_3(8) \le 7$, $c_3(10) \le 8$, $c_3(12) \le 9$, $c_3(16) \le 10$ and according to the monotonicity of the function $c_3(n)$, it is easy to prove that (9) is true for all $n \le 16$.
It is not difficult to verify that for all $k \ge 0$

$$n > 16 \cdot 2^k \Rightarrow \left\lceil \log_2 \binom{n}{3} \right\rceil > 3k + 10 \tag{10}$$

$$n > 20 \cdot 2^k \Rightarrow \left\lceil \log_2 \binom{n}{3} \right\rceil > 3k + 11 \tag{11}$$

$$n > 25 \cdot 2^k \Rightarrow \left\lceil \log_2 \binom{n}{3} \right\rceil > 3k + 12 \tag{12}$$

On the basis of these statements we conclude that the algorithm constructed in Theorem 1 is optimal for all $k \ge 0$. Also, for other values of n, (9) is true, because of the monotonicity of the function $c_3(n)$.

References

[1] Chang, G., Hwang, F., Lin, S., Group Testing with Two Defectives, Discrete Applied Mathematics 4, 97-102 (1982).

[2] Hwang, F., A Tale of Two Coins, American Mathematical Monthly 2, 121-129 (1987).

[3] Tošić, R., An Optimal Search Procedure, Journal of Statistical Planning and Inference 4, 169-171 (1980).

[4] Tošić, R., An Optimal Group-Testing Procedure, Studia Scientarium Mathematicarum Hungarica 17, 319-323 (1982).

Received by the editors March 18, 1993.