Note

An improved algorithm for quantitative group testing*

L. Gargano, V. Montuori, G. Setaro and U. Vaccaro

Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (SA), Italy

Received 29 January 1991
Revised 19 August 1991

Abstract


Given a set of \( n \) elements two of which are defective, the quantitative group testing problem asks for the search of the defectives under the constraint that the result of each test gives the number of defectives contained in the tested subset. We give an algorithm to solve the quantitative group testing problem that improves on the Fibonaccian algorithm by Christen and Aigner.

1. Introduction

Consider a set \( S \) of \( n \) coins two of which are defective. The problem is to identify the two defectives by weighing subsets of \( S \). Depending on the type of the weighing device several models are possible. Hwang [8] classified two-coins models, according to the receivable feedback of the weighing, into seven types and reviewed the best-to-date upper and lower bounds to the worst-case number of tests to find both defectives for each model. Progresses on the results quoted by Hwang are reported in [3, 6, 9, 10].

* Correspondence to: Professor U. Vaccaro, Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84081 Baronissi (SA), Italy.

* This work was supported in part by the Italian Ministry of the University and of the Scientific Research in the framework of the “Algoritmi e Sistemi di Calcolo” project.

0166-218X/92/$05.00 © 1992 — Elsevier Science Publishers B.V. All rights reserved
In this paper we study the quantitative group testing problem where the feedback one receives when a set of coins is weighed gives the number of defective coins contained in the tested subset. Denote by $T(n)$ the worst-case number of tests required by an optimal procedure to identify the two defectives among $n$ coins under the hypothesis of the quantitative group testing problem, and denote by $T(m, n)$ the analogous quantity when it is known that one defective is in a set $A$, $|A| = n$, and the other is in a set $B$, $|B| = m$, $A \cap B = \emptyset$. The information-theoretic bound gives $T(n) \geq \lceil \log_3(n^2) \rceil = 2 \log_3 n$ and no better lower bound is known. The best algorithm known so far is the Fibonaccian algorithm by Christen [4] and Aigner [1] which shows that $T(n) \leq \log_3 n = 2.28 \ldots \log_3 n$, where $\phi = (1 + \sqrt{5})/2$ is the golden ratio. In a recent paper Hao [7] considered the problem of evaluating the limit $\alpha = \lim_{n \to \infty} T(n) / \log_3 n$ and showed that $\alpha \leq 12/\log_3 330 = 2.27 \ldots$. In this paper we present a novel algorithm which improves on the Fibonaccian algorithm and shows that $T(n) < (7/\log_3 3L) \log_3 n = 2.18 \ldots \log_3 n$. Therefore, we improve on Hao’s estimation of $\alpha$ as well.

2. The result

From now on all log’s are of base 3. We first state the following result which is a consequence of [7, Lemma 21. We report here a proof for reader’s convenience.

**Corollary 2.1.** For any positive integers $m$ and $i$ one has

$$T(m^i, m^i) \leq iT(m, m).$$

**Proof.** Let one defective coin be in $A$, $|A| = m^i$ and the other defective be in $B$, $|B| = m^i$. Partition $A$ into $A_1, \ldots, A_m$ and $B$ into $B_1, \ldots, B_m$, with $|A_j| = |B_j| = m^{i-1}$ for each $j$. Since exactly one set among the $A_j$ and one among the $B_j$ contains a defective coin, the pair of sets containing the defectives can be identified in $T(m, m)$ tests. Therefore,

$$T(m^i, m^i) \leq T(m, m) + T(m^{i-1}, m^{i-1}) \leq \cdots \leq iT(m, m).$$

The next two lemmas represent the main step towards the derivation of our result.

**Lemma 2.2.** For any positive integers $m$ and $n$, $n > m$, one has

$$T(n) \leq \max_{0 < k \leq \lfloor \log_3 n \rfloor} \{k + T(m^{[\log n - k \log 2]}), \frac{n}{m}^{[\log n - k \log 2] / \log m}\}.$$

**Proof.** For each integer $a$, denote by $A_{a,a}$ a search procedure which finds the two defectives in two disjoint sets $A$ and $B$, $|A| = |B| = a$, with $T(a, a)$ tests. Moreover,

---

1 One of the referees pointed out that an unpublished result similar to ours was announced by Christen [5] in a talk given at the 1986 SIAM Conference on Discrete Mathematics.
for each set of coins $A$, denote by $F(A)$ the outcome of the test on $A$. Let $S$, $|S| = n$
be the set of coins. The following procedure finds the two defectives in $S$.

**Step 1.** Let $i = 1$, $A_1 = S$;

**Step 2.** [$A_i$ is the current solution space, i.e., both defectives belong to $A_i$]

- if $|A_i| = 2$ then stop;
- otherwise partition $A_i$ into $B$ and $C$ with $|B| = \lceil |A_i|/2 \rceil$ and $|C| = \lfloor |A_i|/2 \rfloor$;

**Step 3.** Test on $B$;
- if $F(B) = 0$ then set $i = i + 1$, $A_i = C$ and go to Step 2;
- if $F(B) = 2$ then set $i = i + 1$, $A_i = B$ and go to Step 2;
- if $F(B) = 1$ then set $j = \lfloor \log|B|/\log m \rfloor$ and apply $\mathcal{A}_{m',m'}$ to $(B,C)$.

Let us count the number of tests. Suppose that the first $k - 1$ tests gave feedback either equal to 0 or to 2 and the $k$th gave feedback 1. The procedure $\mathcal{A}_{m',m'}$ will be applied to sets of size not greater than $\lfloor n/2^k \rfloor \leq m \lfloor (\log n - k \log 2)/\log m \rfloor$. Therefore, the number of tests is at most

$$k + T(m \lfloor (\log n - k \log 2)/\log m \rfloor, m \lfloor (\log n - k \log 2)/\log m \rfloor).$$

**Lemma 2.3.** $T(32,32) = 7$.

**Proof.** The information-theoretic bound gives $T(32,32) \geq \lfloor \log 32^2 \rfloor = 7$. The opposite inequality $T(32,32) \leq 7$ is proved by the procedure given in the Appendix. $\square$

We can now prove our main result.

**Theorem 2.4.** $T(n) \leq 2.18\cdots \log n + O(1)$.

**Proof.** By Lemmas 2.2 and 2.3 and by Corollary 2.1 one gets

$$T(n) \leq \max_{0 < k \leq \lfloor \log_2 n \rfloor} \{k + T(32 \lfloor (\log n - k \log 2)/\log 32 \rfloor, 32 \lfloor (\log n - k \log 2)/\log 32 \rfloor)\}$$

$$= \max_{0 < k \leq \lfloor \log_2 n \rfloor} \{k + 7 \lfloor (\log n - k \log 2)/\log 32 \rfloor\} = 2.18\cdots \log n + O(1),$$

where the last equality follows since the quantity $k + 7 \lfloor (\log n - k \log 2)/\log 32 \rfloor$ is decreasing with $k$. $\square$

3. Concluding remarks

The search problem considered in this paper can be rephrased, using the language of [1], as a problem of searching an edge in a graph. More precisely, suppose
Fig. 2.
Fig. 3.
to have a graph $G = (V, E)$ and an unknown edge $e$. After every test on some $A \subseteq V$ we receive a feedback 2 if the edge $e$ we are looking for has both endvertices in $A$, feedback 1 if $e$ has one endvertex in $A$ and 0 otherwise. It is clear that the quantity $T(n)$ is the minimum number of tests to search for an edge in the complete graph $K_n$ and $T(m, n)$ corresponds to the analogous quantity for the complete bipartite graph $K_{m,n}$. Aigner [2] calls a graph $G = (V, E)$ 3-optimal if the minimum number of tests to search for an edge in $G$ is equal to the information-theoretic bound, i.e., $\lfloor \log_3 |E| \rfloor$. Very few graphs are known to be 3-optimal and it is conjectured that, for fixed $m$, only finitely-many complete bipartite graphs $K_{m,n}$ are 3-optimal (see [2]). Lemma 2.3 shows that $K_{32,32}$ is 3-optimal. An analysis of the algorithm given in the Appendix shows that also $K_{14,9}$ and $K_{21,17}$ are 3-optimal.

Finally, we mention that the technique developed in this paper can be fruitfully applied to search for more than two coins. In [11] are reported two algorithms to search for three and four defectives. The number of required tests is $3.78 \ldots \log_4 n$ and $5.23 \ldots \log_5 n$, respectively.

Appendix

Proof of Lemma 2.3. In Figs. 1, 2, and 3 is reported the tree which contains the tests to find both defectives, where one belongs to a set $A$, $|A| = 32$ and the other belongs to $B$, $|B| = 32$. $A \cap B = \emptyset$. For each node, the outgoing left branch corresponds to a test with feedback 2, the middle branch to a test with feedback 1 and the right branch to a test with feedback 0. To keep notation shorter, we have indicated only the cardinalities of the solution spaces and of the tested subsets. The cardinalities of the tested subset are in bold. For instance, a node in the tree of the form

$$(21, 15) \quad (11, 17)$$

$$(14, 9) \quad (2, 3)$$

means that, in the corresponding step of the algorithm, either one defective coin is in a set $A_1$ of cardinality 21 and the other defective is in a set $A_2$ of cardinality 15 or one defective coin is in a set $B_1$ of cardinality 11 and the other is in a set $B_2$ of cardinality 17. The set of coins which will be tested in the next step of the algorithm is constructed choosing (in arbitrary fashion) 14 coins from $A_1$, 9 coins from $A_2$, 2 coins from $B_1$ and 3 coins from $B_2$.

References


