Lower Bound for Non-Adaptive Estimation of the Number of Defective Items

Nader H. Bshouty
Department of Computer Science, Technion, Haifa, Israel
bshouty@cs.technion.ac.il

Abstract

We prove that to estimate within a constant factor the number of defective items in a non-adaptive randomized group testing algorithm we need at least $\tilde{\Omega}(\log n)$ tests. This solves the open problem posed by Damaschke and Sheikh Muhammad in [6, 7].

2012 ACM Subject Classification Mathematics of computing; Mathematics of computing → Discrete mathematics; Mathematics of computing → Probabilistic algorithms; Theory of computation → Probabilistic computation

Keywords and phrases Group Testing, Estimation, Defective Items

Digital Object Identifier 10.4230/LIPIcs.ISAAC.2019.2

Related Version `https://eccc.weizmann.ac.il/report/2018/053/`

1 Introduction

Let X be a set of *items* that contains *defective items* $I \subseteq X$. In Group testing, we *test (query)* a subset $Q \subset X$ of items and the answer to the query is 1 if Q contains at least one defective item, i.e., $Q \cap I \neq \emptyset$, and 0 otherwise. Group testing was originally introduced as a potential approach to the economical mass blood testing, [8]. However it has been proven to be applicable in a variety of problems, including DNA library screening, [18], quality control in product testing, [21], searching files in storage systems, [14], sequential screening of experimental variables, [16], efficient contention resolution algorithms for multiple-access communication, [14, 25], data compression, [12], and computation in the data stream model, [5]. See a brief history and other applications in [4, 9, 10, 13, 17, 18] and references therein.

Estimating the number of defective items to within a constant factor λ is the problem of finding an integer D that satisfies[1] $|I| \le D \le \lambda |I|$. This problem is extensively used in biological and medical applications [2, 22]. It is used to estimate the proportion of organisms capable of transmitting the aster-yellows virus in a natural population of leafhoppers [23], estimating the infection rate of yellow-fever virus in a mosquito population [24] and estimating the prevalence of a rare disease using grouped samples to preserve individual anonymity [15].

In *adaptive algorithms*, the queries can depend on the answers to the previous ones. In the *non-adaptive algorithms* they are independent of the previous one and; therefore, one can ask all the queries in one parallel step. In many applications in group testing non-adaptive algorithms are most desirable.

Estimating the number of defective items to within a constant factor with an *adaptive* deterministic, Las Vegas and Monte Carlo algorithms is studied in [1, 3, 6, 7, 11, 20]. For $|X| = n$ items and $|I| = d$ defective items the bounds are $\Theta(d \log(n/d))$ queries for Las Vegas and Deterministic algorithms and $\Theta(\log \log d + \log(1/\delta))$ queries for Monte Carlo algorithm [1, 11]. There are also polynomial time algorithms that achieve such bounds [1, 11].

[1] In all the applications in group testing the estimation $\lambda'|I| \le D \le \lambda |I|$ in not interesting.

30th International Symposium on Algorithms and Computation (ISAAC 2019).
Editors: Pinyan Lu and Guochuan Zhang; Article No. 2; pp. 2:1–2:9
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In this paper we study this problem in the non-adaptive setting. We first show that any deterministic and Las Vegas algorithm must ask at least $\Omega(n)$ queries. For randomized algorithm with any *constant* failure probability δ, Damaschke and Sheikh Muhammad give in [7] a non-adaptive randomized algorithm that asks $O(\log n)$ queries and with probability at least $1-\delta$ returns an integer D such that $D \ge d$ and $\mathbf{E}[D] = O(d)$. In this paper we give a polynomial time Monte Carlo algorithm that asks $O(\log(1/\delta)\log n)$ queries and with probability at least $1-\delta$ estimates the number of defective items to within a constant factor. They then prove in [6] the lower bound $\Omega(\log n)$ queries, but only for algorithms that choose each item in each query randomly and independently with some fixed probability. They conjecture that $\Omega(\log n)$ queries are needed for *any* randomized algorithm with constant failure probability. In this paper we prove this conjecture (up to $\log\log n$ factor). We show that for any non-adaptive randomized algorithm that with probability at least $3/4$ estimates the number of defective items to within a constant factor must ask at least

$$s = \Omega\left(\frac{\log n}{\log\log n}\right) = \tilde{\Omega}\left(\log n\right)$$

queries.

This paper is organised as follows: In Section 2 we give some preliminary results. In Section 3 we give the proof of the above lower bound. The lower bounds will be for an estimation within the factor of 1.5 and confidence 3/4, but it will be clear from the proof that this can be replaced by any constant factor λ and any constant confidence δ. In Section 4 we give the lower bound $\Omega(n)$ for any deterministic algorithm. In Section 5 we give the upper bound. The technique for the upper bound is standard and implicitly follows from [7, 11]. It is given for completeness.

2 Preliminary Results

In this section we give some definitions and then prove some preliminary results.

We will consider the set of *items* $X = [n] = \{1, 2, \ldots, n\}$ and the set of *defective items* $I \subseteq X$. The algorithm knows n and has an access to an oracle $\mathcal{O}_I$. The algorithm can ask the oracle $\mathcal{O}_I$ a *query* $Q \subset X$ and the oracle answers $\mathcal{O}_I(Q) := 1$ if $Q \cap I \neq \emptyset$ and $\mathcal{O}_I(Q) := 0$ otherwise. We say that algorithm A λ-*estimates* the number of defective items if for every $I \subseteq X$ it runs in polynomial time in n, asks queries to the oracle $\mathcal{O}_I$ and returns an integer D such that $|I| \le D \le \lambda|I|$. If λ is constant then we say that the algorithm *estimates the number of defective items to within a constant factor* . Our goal is to find such an algorithm that asks a minimum number of queries in the worst case.

For an algorithm A that asks queries we denote by $A(I)$ the output of A when it runs with the oracle $\mathcal{O}_I$. When the algorithm is randomized then we write $A(\sigma, I)$ where σ is the random seed of the algorithm.

We now prove two results that will be used for the lower bound:

▶ **Lemma 1.** *Let k be any real number. Let N' be a finite set of elements and s be an integer. Let S be a probability space of s-tuples $W = (w_1, w_2, \ldots, w_s) \in N'^s$. Let $N \subseteq N'$ and $N = N_1 \cup N_2 \cup \cdots \cup N_r$ be a partition of N to r disjoint sets. There is i_0 such that for a random $W \in S$, the probability that at least k of the elements (coordinates) of W are in N_{i_0}, is at most $s/(kr)$.*

Equivalently, there is i_0 such that, with probability at least $1 - s/(kr)$, the number of elements in W that are in N_{i_0} is less than k.

Proof. Define the random variables X_i, $i = 1, \ldots, r$, where $X_i(W) = 1$ if at least k of the elements of W are in N_i and 0 otherwise. Obviously, $k(X_1 + \ldots + X_r) \le s$ and therefore

$$\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_r] = \mathbf{E}[X_1 + \cdots + X_r] \le \frac{s}{k}.$$

Therefore there is i_0 such that $\mathbf{Pr}[X_{i_0} = 1] = \mathbf{E}[X_{i_0}] \le s/(kr)$. ◀

▶ **Lemma 2.** *Let $X' \subseteq X = [n]$. Let $\mathcal{D}$ be the probability space of random uniform subsets $I \subset X'$ of size d and $\mathcal{D}'$ be the probability space of random uniform and independent d chosen elements $I = \{x_1, \ldots, x_d\} \subseteq X'$ with replacement. Let A be any event and B be the event that $I \in \mathcal{D}'$ has size d, i.e., $x_1, \ldots, x_d$ are distinct. Then*

$$\boldsymbol{Pr}_{\mathcal{D}'}[A] + \boldsymbol{Pr}_{\mathcal{D}'}[\bar{B}] \ge \boldsymbol{Pr}_{\mathcal{D}}[A] \ge \boldsymbol{Pr}_{\mathcal{D}'}[A] - \boldsymbol{Pr}_{\mathcal{D}'}[\bar{B}].$$

Proof. Since

$$\begin{aligned} \mathbf{Pr}_{\mathcal{D}'}[A] &= \mathbf{Pr}_{\mathcal{D}'}[A|B]\mathbf{Pr}_{\mathcal{D}'}[B] + \mathbf{Pr}_{\mathcal{D}'}[A|\bar{B}]\mathbf{Pr}_{\mathcal{D}'}[\bar{B}] \\ &\le \mathbf{Pr}_{\mathcal{D}'}[A|B] + \mathbf{Pr}_{\mathcal{D}'}[\bar{B}] = \mathbf{Pr}_{\mathcal{D}}[A] + \mathbf{Pr}_{\mathcal{D}'}[\bar{B}], \end{aligned}$$

we have $\mathbf{Pr}_{\mathcal{D}}[A] \ge \mathbf{Pr}_{\mathcal{D}'}[A] - \mathbf{Pr}_{\mathcal{D}'}[\bar{B}]$. In the same way we have $\mathbf{Pr}_{\mathcal{D}}[\bar{A}] \ge \mathbf{Pr}_{\mathcal{D}'}[\bar{A}] - \mathbf{Pr}_{\mathcal{D}'}[\bar{B}]$ which implies the left-hand side inequality. ◀

3 Lower Bound for Randomized Algorithms

In this section we prove the lower bound for the number of queries in any non-adaptive randomized algorithm that λ-estimates the number of defective items. We give the proof for $\lambda = 1.5$ and confidence $\delta = 1/4$. The proof for any other constants λ and δ is similar.

The idea of the proof is the following. Suppose there is a randomized algorithm A that asks $\log n/(c \log \Delta)$ queries where $\Delta = \log n$ and c is a large constant. We partition the interval $[0, n]$ of all the possible sizes $|Q|$ of the queries Q into $\Theta(\log n/\log \Delta)$ disjoint sets $N_i = [n/\Delta^{4i+4}, n/\Delta^{4i}]$ for integers i. We then show, by Lemma 2, that with high probability, there is an interval N_{i_0} such that no query Q asked by the algorithm satisfies $|Q| \in N_{i_0}$. That is, with high probability, there is no query with a size that falls in $N_{i_0} = [n/\Delta^{4i_0+4}, n/\Delta^{4i_0}]$. We then show that if we choose a random uniform set of defective items I of size $d' := \Delta^{4i_0+2}$ or $2d' = 2\Delta^{4i_0+2}$ then, with high probability, all the queries of sizes more than n/Δ^{4i_0} will have answer 1 and all the queries of sizes less than n/Δ^{4i_0+4} will have answer 0. So the only useful queries are those that fall in N_{i_0} that, by Lemma 1, with high probability, there are none. Therefore, with high probability, the algorithm fails to distinguish between sets of defective sets of size d' and of size $2d'$. This implies that algorithm A cannot estimate, with high probability, the size of the defective sets within a factor of 1.5. Therefore any randomized algorithm that, with high probability, estimates the size of the defective sets within a factor of 1.5 must ask at least $\log n/(c \log \Delta) = \Omega(\log n/\log\log n)$ queries.

We now give the proof.

▶ **Theorem 3.** *Any non-adaptive Monte Carlo randomized algorithm that with probability at least 3/4, 1.5-estimates the number of defective items must ask at least*

$$s = \Omega\left(\frac{\log n}{\log\log n}\right)$$

queries.

Proof. Let c be a large enough constant. Suppose, for the contrary, there is a non-adaptive Monte Carlo algorithm $A(\sigma, I)$ that chooses a random sequence of queries $M := Q_1, \ldots, Q_s \subseteq X = [n]$ from some probability space where $s = \Delta/(c \log \Delta)$ and $\Delta = \log n$, asks queries to $\mathcal{O}_I$ and with probability at least 3/4, 1.5-estimates the number of defective items $|I|$. Let $r = \Delta/(16 \log \Delta)$ and let

$$N_i = [n/\Delta^{4i+4}, n/\Delta^{4i}] := \{x \mid n/\Delta^{4i+4} < x \le n/\Delta^{4i}\},$$

$i = 0, 1, \ldots, r-1$, be a partition of $N = [n^{3/4}, n]$. By Lemma 1, for $k = 1/16$ and the s-tuple $W := (|Q_1|, \ldots, |Q_s|)$, there is i_0 such that, with probability at least

$$1 - \frac{s}{kr} = 1 - \frac{256}{c} \ge \frac{15}{16}$$

the number of queries Q in M that satisfy $|Q| \in N_{i_0}$ is at most k. Therefore, with probability at least 15/16 there are no queries Q in M of size $|Q| \in N_{i_0}$. Let C be the event that there is no query Q in M of size $|Q| \in N_{i_0}$. Then

$$\mathbf{Pr}[\bar{C}] \le \frac{1}{16}.$$

Let $d' = \Delta^{4i_0+2}$. For a random uniform set $I \subset X$ of size $d = d'$, with probability at least 3/4, $A(\sigma, I)$ returns an integer in the interval $[d', 1.5d']$. For a random uniform set $I \subset X$ of size $d = 2d'$, with probability at least 3/4, $A(\sigma, I)$ returns an integer in the interval $[2d', 3d']$. Since both intervals are disjoint, algorithm A, with success probability at least 3/4, can distinguish between defective sets of size d' and $2d'$. We have constructed an algorithm, call it A', that distinguishes, with success probability 3/4, between defective sets of size d' and defective sets of size $2d'$. The probability that A' fails is at most 1/4.

Let $\mathcal{D}$, $\mathcal{D}'$ and $\{x_1, \ldots, x_d\}$ be as in Lemma 2. Here $d \in \{d', 2d'\}$. Let B be the event that $x_1, \ldots, x_d$ are distinct. Since $i_0 \le r$ we have

$$d \le 2d' = 2\Delta^{4i_0+2} \le 2\Delta^{4r+2} = 2n^{1/4} \log^2 n$$

and therefore, for large enough n,

$$\mathbf{Pr}_{\mathcal{D}'}[\bar{B}] = 1 - \prod_{i=1}^{d-1} \left(1 - \frac{i}{n}\right) \le \frac{d(d-1)}{2n} \le \frac{2\log^4 n}{n^{1/2}} \le \frac{1}{16}.$$

Now partition the queries in M to three sets of queries $M_1 \cup M_2 \cup M_3$ where M_1 are the queries that contain at most n/Δ^{4i_0+4} items, M_2 are the queries that contains at least n/Δ^{4i_0} items and $M_3 = M \backslash (M_1 \cup M_2)$, i.e., M_3 are the queries Q that satisfies $|Q| \in N_{i_0}$. Let $A_1(I)$ be the event that for $I \subseteq X$ all the queries in M_1 give answer 0. Then

$$\begin{aligned}
\mathbf{Pr}_{\mathcal{D}'}[\bar{A}_1] &= \mathbf{Pr}[(\exists Q \in M_1) Q \cap I \ne \emptyset] \\
&\le s\mathbf{Pr}[Q \cap I \ne \emptyset | Q \in M_1] \\
&= s(1 - \mathbf{Pr}[Q \cap I = \emptyset | Q \in M_1]) \\
&\le s\left(1 - \left(1 - \frac{1}{\Delta^{4i_0+4}}\right)^d\right) \\
&\le \frac{sd}{\Delta^{4i_0+4}} = \frac{2}{c\Delta \log \Delta} \le \frac{1}{16}.
\end{aligned} \qquad (1)$$

Then by Lemma 2, $\mathbf{Pr}_{\mathcal{D}}[\bar{A}_1] \le 2/16$. Let $A_2(I)$ be the event that for $I \subseteq X$ all the queries in M_2 give answer 1. Then

$$\begin{aligned}
\mathbf{Pr}_{\mathcal{D}'}[\bar{A}_2] &= \mathbf{Pr}[(\exists Q \in M_2) Q \cap I = \emptyset] \\
&\le s\mathbf{Pr}[Q \cap I = \emptyset | Q \in M_2] \\
&\le s\left(1 - \frac{1}{\Delta^{4i_0}}\right)^d \\
&\le se^{-\frac{d}{\Delta^{4i_0}}} = \frac{\Delta}{ce^{\Delta^2}\log \Delta} \le \frac{1}{16}.
\end{aligned}$$

Thus, by Lemma 2, $\mathbf{Pr}_{\mathcal{D}}[\bar{A}_2] \le 2/16$.

Now

$$\begin{aligned}
\mathbf{Pr}[A' \text{ fails}] &\ge \mathbf{Pr}[A_1 \wedge A_2 \wedge C] \\
&= 1 - \mathbf{Pr}[\bar{A}_1 \vee \bar{A}_2 \vee \bar{C}] \\
&\ge 1 - \mathbf{Pr}[\bar{A}_1] - \mathbf{Pr}[\bar{A}_2] - \mathbf{Pr}[\bar{C}] \\
&\ge \frac{1}{2}.
\end{aligned}$$

We got $\mathbf{Pr}[A' \text{ fails}] \ge 1/2$ which gives a contradiction. ◀

In the proof of Theorem 3, one cannot take smaller intervals for N_i (for example $[n/2^{4i+4}, n/2^{4i}]$). This is because, with the multiplicand s for the union bound in (1), the probability of $\bar{A}_1$ cannot then be bounded by 1/16.

The proof is also true for estimating the number of defective items to within a factor $\lambda = \Theta(\log n)$. In fact, for such λ the lower bound is tight.

4 Lower Bound for Deterministic Algorithms

In this section we prove

▶ **Theorem 4.** *Let $c > 1$ be any constant. Any non-adaptive deterministic algorithm that c-estimates the number of defective items must ask at least $\Omega(n)$ queries.*

Proof. Let A be a non-adaptive deterministic algorithm that c-estimates the number of defective items. Let $Q_1, \ldots, Q_s$ be the queries that A asks. Let $d = n/2c$. For possible answers $a_1, \ldots, a_s \in \{0, 1\}$ to the queries we define $S_{(a_1,\ldots,a_s)}$, the set of all defective sets of size d that give the answers $a_1, \ldots, a_s$ to the queries $Q_1, \ldots, Q_s$, respectively. That is, for every $I \in S_{(a_1,\ldots,a_s)}$ we have $|I| = d$ and for every $i = 1, \ldots, s$ we have $Q_i \cap I \ne \emptyset$ if $a_i = 1$ and $Q_i \cap I = \emptyset$ if $a_i = 0$. For $a = (a_1, \ldots, a_s) \in \{0, 1\}^s$ let $I_a = \cup_{I \in S_a} I$. We now prove two claims:

▷ Claim 5. If the defective set is I_a then the algorithm gets the answers a to the queries.

Proof. If $Q_i \cap I_a \ne \emptyset$ then there is $I \in S_a$ such that $Q_i \cap I \ne \emptyset$ and then $a_i = 1$. If $Q_i \cap I_a = \emptyset$ then for every $I \in S_a$ we have $Q_i \cap I = \emptyset$ and then $a_i = 0$. ◁

▷ Claim 6. $|I_a| \le cd$.

Proof. If $|I_a| > cd$ then the algorithm returns a value in $[cd + 1, c^2 d]$ and then for the sets in S_a, that are of size d, this answer is not a c-estimation. A contradiction. ◁

Since each $I \in S_a$ is of size d and is a subset of I_a we have

$$|S_a| \le \binom{cd}{d}.$$

Since there are $\binom{n}{d}$ sets of size d we get

$$\binom{n}{d} = \sum_{a\in\{0,1\}^s} |S_a| \le 2^s \binom{cd}{d}.$$

Since $d = n/(2c)$,

$$s \ge \log \binom{n}{\frac{n}{2c}} - \log \binom{\frac{n}{2}}{\frac{n}{2c}} = \Omega(n).$$ ◀

5 Upper Bounds

In this section is written for completeness. We use techniques similar to the ones in [7, 11] to prove

▶ **Theorem 7.** *Let c be any constant. There is a non-adaptive Monte Carlo randomized algorithm that asks*

$$s = O\left(\log \frac{1}{\delta} \log n\right)$$

queries and with probability at least $1-\delta$, c-estimates the number of defective items.

We recall the Chernoff Bound.

▶ **Lemma 8** (Chernoff Bound). *Let $X_1, \ldots, X_t$ be independent random variables that takes values in $\{0,1\}$. Let $X = (X_1 + \cdots + X_t)/t$ and $\boldsymbol{E}[X] \le \mu$. Then for any $\Delta \ge \mu$*

$$\boldsymbol{Pr}[X \ge \Delta] \le \left(\frac{e^{1-\frac{\mu}{\Delta}}\mu}{\Delta}\right)^{\Delta t} \quad (2)$$

$$\le \left(\frac{e\mu}{\Delta}\right)^{\Delta t}. \quad (3)$$

We will assume that $d \ge 6$. Otherwise, d can be estimated exactly in $O(\log n)$ more queries. Just run the algorithm that finds the defective items that asks $O(\log n)$ queries [19]. Here we give a 2-estimation algorithm. This can be extended in a straightforward manner to c-estimation for any constant c.

A p-query is a query Q that contains each item $i \in [n]$ randomly and independently with probability p. In the algorithm, $\mathcal{O}_I(Q) = 1$ if $Q \cap I \ne \emptyset$ and 0 otherwise.

Consider the following algorithm We now prove

▶ **Lemma 9.** *Let $|I| = d \ge 6$. If $u \le d \le w$ then with probability at least $1-\delta$, $d \le D \le 2d$. The algorithm asks $O(\log(1/\delta)\log(w/u))$ queries.*

In particular, for $u = 1$ and $w = n$, the algorithm asks

$$O\left(\log \frac{1}{\delta} \log n\right)$$

queries.

Algorithm 1 Estimate (u, w, δ).

Input: u and w such that $u \le d \le w$ and a failure probability δ

Output: D such that w.p. at least $1-\delta$, $d \le D \le 2d$.

1. For each $p_i = 1/(u \cdot 2^{i/4})$, $i = 0, 1, 2, 3, \cdots, 8\log(w/u)$,
2. For $t = O(\log(1/\delta))$ independent p_i-queries $Q_{i,1}, \ldots, Q_{i,t}$ do:
3. $q_i = (\mathcal{O}_I(Q_{i,1}) + \cdots + \mathcal{O}_I(Q_{i,t}))/t$.
4. Choose the first i_0 such that $q_{i_0} < 0.83$.
5. If no such i_0 exists then output("$d > w$").
6. Otherwise output($D := 2/p_{i_0}$).

Proof. Let i_1 be such that $p_{i_1-1} > 2/d$ and $p_{i_1} \le 2/d$. Then for $j = 0, 1, \cdots,$

$$2^{j/4}/d < p_{i_1+3-j} \le 2^{(j+1)/4}/d.$$

For every i, j we have

$$\mu_i := \mathbf{E}[q_i] = \mathbf{E}[\mathcal{O}_I(Q_{i,j})] = \mathbf{Pr}[I \cap Q_{i,j} \ne \emptyset] = 1 - (1-p_i)^d.$$

Since $d \ge 6$ we have $\mathbf{E}[q_{i_1+3}] = \mu_{i_1+3} \le 1 - (1 - 2^{1/4}/d)^d \le 0.74$ and

$$\begin{aligned} \mathbf{Pr}[D > 2d] &= \mathbf{Pr}[p_{i_0} < 1/d] = \mathbf{Pr}[i_0 > i_1 + 3] \\ &\le \mathbf{Pr}[q_{i_1+3} \ge 0.83] \le \delta/2. \end{aligned} \tag{4}$$

The first inequality in (4) follows from the fact that if $i_0 > i_1 + 3$ then $q_{i_1+3} \ge 0.83$. The second inequality follows from Chernoff bound (2) with $\mu = 0.74$ and $\Delta = 0.83$.

Now, since

$$\begin{aligned} \mathbf{E}[1 - q_{i_1+3-j}] &= 1 - \mu_{i_1+3-j} = (1 - p_{i_1+3-j})^d \\ &\le e^{-p_{i_1+3-j}d} < e^{-2^{j/4}}, \end{aligned}$$

we have $\mathbf{E}[1 - q_{i_1-2}] \le \mathbf{E}[1 - q_{i_1-1}] \le 0.136$ and

$$\begin{aligned} \mathbf{Pr}[D < d] &= \mathbf{Pr}[p_{i_0} > 2/d] = \mathbf{Pr}[i_0 \le i_1 - 1] \\ &= \sum_{i=0}^{i_1-1} \mathbf{Pr}[i_0 = i] \le \sum_{i=0}^{i_1-1} \mathbf{Pr}[q_i < 0.83] \\ &= \sum_{i=0}^{i_1-3} \mathbf{Pr}[1 - q_i > 0.17] + \sum_{i=i_1-2}^{i_1-1} \mathbf{Pr}[1 - q_i > 0.17] \\ &\le \sum_{i=0}^{i_1-3} \left(\frac{e \cdot e^{-2^{(i_1-i+3)/4}}}{0.17} \right)^{0.17 \cdot t} + \frac{\delta}{4} \\ &\le \sum_{k=0}^{\infty} \left(0.95 \cdot e^{-2^{k/4}} \right)^{0.17 \cdot t} + \frac{\delta}{4} \le \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2}. \end{aligned} \tag{5}$$

In the first summand of (5) we use Chernoff bound (3). In the second summand we use Chernoff bound (2) for $\mu = 0.136$ and $\Delta = 0.17$. ◀

References

1 Nader H. Bshouty, Vivian E. Bshouty-Hurani, George Haddad, Thomas Hashem, Fadi Khoury, and Omar Sharafy. Adaptive Group Testing Algorithms to Estimate the Number of Defectives. *ALT*, 2017. arXiv:1712.00615.

2 Chao L. Chen and William H. Swallow. Using Group Testing to Estimate a Proportion, and to Test the Binomial Model. *Biometrics.*, 46(4):1035–1046, 1990.

3 Yongxi Cheng and Yinfeng Xu. An efficient FPRAS type group testing procedure to approximate the number of defectives. *J. Comb. Optim.*, 27(2):302–314, 2014. doi:10.1007/s10878-012-9516-5.

4 Ferdinando Cicalese. *Fault-Tolerant Search Algorithms - Reliable Computation with Unreliable Information.* Monographs in Theoretical Computer Science. An EATCS Series. Springer, 2013. doi:10.1007/978-3-642-17327-1.

5 Graham Cormode and S. Muthukrishnan. What's hot and what's not: tracking most frequent items dynamically. *ACM Trans. Database Syst.*, 30(1):249–278, 2005. doi:10.1145/1061318.1061325.

6 Peter Damaschke and Azam Sheikh Muhammad. Bounds for Nonadaptive Group Tests to Estimate the Amount of Defectives. In *Combinatorial Optimization and Applications - 4th International Conference, COCOA 2010, Kailua-Kona, HI, USA, December 18-20, 2010, Proceedings, Part II*, pages 117–130, 2010. doi:10.1007/978-3-642-17461-2_10.

7 Peter Damaschke and Azam Sheikh Muhammad. Competitive Group Testing and Learning Hidden Vertex Covers with Minimum Adaptivity. *Discrete Math., Alg. and Appl.*, 2(3):291–312, 2010. doi:10.1142/S179383091000067X.

8 R. Dorfman. The detection of defective members of large populations. *Ann. Math. Statist.*, pages 436–440, 1943.

9 D. Du and F. K Hwang. Combinatorial group testing and its applications. *World Scientific Publishing Company.*, 2000.

10 D. Du and F. K Hwang. Pooling design and nonadaptive group testing: important tools for DNA sequencing. *World Scientific Publishing Company.*, 2006.

11 Moein Falahatgar, Ashkan Jafarpour, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Estimating the number of defectives with group testing. In *IEEE International Symposium on Information Theory, ISIT 2016, Barcelona, Spain, July 10-15, 2016*, pages 1376–1380, 2016. doi:10.1109/ISIT.2016.7541524.

12 Edwin S. Hong and Richard E. Ladner. Group testing for image compression. *IEEE Trans. Image Processing*, 11(8):901–911, 2002. doi:10.1109/TIP.2002.801124.

13 F. K. Hwang. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, 67:605—-608, 1972.

14 William H. Kautz and Richard C. Singleton. Nonrandom binary superimposed codes. *IEEE Trans. Information Theory*, 10(4):363–377, 1964. doi:10.1109/TIT.1964.1053689.

15 Joseph L.Gastwirth and Patricia A.Hammick. Estimation of the prevalence of a rare disease, preserving the anonymity of the subjects by group testing: application to estimating the prevalence of aids antibodies in blood donors. *Journal of Statistical Planning and Inference.*, 22(1):15–27, 1989.

16 C. H. Li. A sequential method for screening experimental variables. *J. Amer. Statist. Assoc.*, 57:455–477, 1962.

17 Anthony J. Macula and Leonard J. Popyack. A group testing method for finding patterns in data. *Discrete Applied Mathematics*, 144(1-2):149–157, 2004. doi:10.1016/j.dam.2003.07.009.

18 Hung Q. Ngo and Ding-Zhu Du. A survey on combinatorial group testing algorithms with applications to DNA Library Screening. In *Discrete Mathematical Problems with Medical Applications, Proceedings of a DIMACS Workshop, December 8-10, 1999*, pages 171–182, 1999. doi:10.1090/dimacs/055/13.

19 Ely Porat and Amir Rothschild. Explicit Nonadaptive Combinatorial Group Testing Schemes. *IEEE Trans. Information Theory*, 57(12):7982–7989, 2011. doi:10.1109/TIT.2011.2163296.

20 Dana Ron and Gilad Tsur. The Power of an Example: Hidden Set Size Approximation Using Group Queries and Conditional Sampling. *CoRR*, abs/1404.5568, 2014. `arXiv:1404.5568`.

21 M. Sobel and P. A. Groll. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Tech. J.*, 38:1179–1252, 1959.

22 William H. Swallow. Group Testing for Estimating Infection Rates and Probabilities of Disease Transmission. *Phytopathology*, 1985.

23 Keith H. Thompson. Estimation of the Proportion of Vectors in a Natural Population of Insects. *Biometrics*, 18(4):568–578, 1962.

24 S. D. Walter, S. W. Hildreth, and B. J. Beaty. Estimation of infection rates in population of organisms using pools of variable size. *Am J Epidemiol.*, 112(1):124–128, 1980.

25 Jack K. Wolf. Born again group testing: Multiaccess communications. *IEEE Trans. Information Theory*, 31(2):185–191, 1985. `doi:10.1109/TIT.1985.1057026`.