POISSON GROUP TESTING: A PROBABILISTIC MODEL FOR NONADAPTIVE STREAMING BOOLEAN COMPRESSED SENSING

Amin Emad and Olgica Milenkovic

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, IL, USA

ABSTRACT

We introduce a novel probabilistic group testing framework, termed Poisson group testing, in which the number of defectives follows a right-truncated Poisson distribution. The Poisson model applies to a number of biological testing scenarios, where the subjects are assumed to be ordered based on their arrival times and where the probability of being defective decreases with time. Our main result is an information-theoretic upper bound on the minimum number of tests required to achieve an average probability of detection error asymptotically converging to zero.

Index Terms— Boolean compressed sensing, Dynamical group testing, Information-theoretic bounds, Poisson and Binomial probabilistic group testing.

1. INTRODUCTION

Group testing (GT) is a method for identifying a group of subjects with some distinguishable characteristic, frequently referred to as defectives, among a large group of subjects [1], [2]. The gist of the GT approach is that for a small number of defectives, one can reduce the required number of experiments by testing subgroups of subjects. If a subgroup tests negatively, all items in the subgroup are eliminated from future screening batches. Otherwise, additional tests are performed on subgroups of items in order to narrow down the options for the defectives. Given its simple working principles and the potential for reducing the cost of component screening, GT has found many applications in communication theory, signal processing, bioinformatics and mathematics [3]-[7].

The test model of the GT framework varies depending on the application at hand. The original setup, also known as conventional GT or CGT, was proposed by Dorfman [8] and includes logical OR computations of the test signatures. More precisely, in CGT, the result of a test is positive if there exists *at least one* defective in the test pool, and negative otherwise. Many other models have been proposed in the literature, such as the adder channel, also known as quantitative GT [7], threshold GT [9], and symmetric GT [10]. More recent developments include the semi-quantitative group testing (SQGT) paradigm, which provides a unifying framework for a number of GT models and generalizes the notion of GT to non-binary test matrices and non-binary test outcomes [11], [12]. Note that GT is also closely related to compressed sensing (CS) [13], [14] and integer compressed sensing [15]; the main differences lie in the structure of the alphabet used ($\mathbb{R}$ or $\mathbb{C}$ in CS, $\{0, 1\}$ or a discrete set of integers for GT) and the operations used to perform dimensionality reduction (addition in CS, Boolean OR in CGT).

The group testing literature may be divided into two categories based on how the number of defectives is modeled. In combinatorial GT, the number of defectives or an upper bound on the number of defectives is fixed and assumed to be known in advance [7]. On the other hand, in probabilistic GT the number of defectives is a random variable with a given probability distribution, see for example [8]. With very few exceptions, the probabilistic GT literature focuses on a Binomial(n, p_0) distribution for the number of defectives. Such a model arises when each of the n subjects is defective with a fixed probability p_0, independent of all other subjects. Binomial models are not necessarily sparse models, given that p_0 may be taken to be relatively large and given the probabilistic nature of the defective selection process.

Here, we propose the novel paradigm of *Poisson GT* that has a number of useful properties that distinguishes it from classical binomial models. Although a binomial GT distribution with $p_0 \ll 1$ and large n, such that $\lambda = np_0$ is a constant, converges to a Poisson distribution [16], our model allows for the mean of the (truncated) Poisson variable to grow with n, i.e. $\lambda(n) = o(n)$. Such a model is suitable in settings were test subjects are assumed to arrive at different times, and were tests are performed when a sufficient number of subjects is present. Furthermore, the assumption $\lambda = o(n)$ ensures that the longer the waiting time, the smaller the average *relative* fraction of defectives. In other words, the rate of defectives diminishes with time. Our motivation for these assumptions comes from clinical testing, where one is trying to identify infected individuals under the assumption that the infection will gradually die out. A similar scenario is encountered in screening DNA clones for the presence of certain DNA sub-

This work was supported in parts by NSF grants CCF 0809895, CCF 1218764 and the Emerging Frontiers for Science of Information Center, CCF 0939370.

strings, where the clones are test subjects and defectives are clones that contain the given substrings. The distribution of clones containing a given DNA pattern is frequently modeled as Poisson [7]. Other applications include identifying faulty items generated by a process, the quality of which improves with time.

We would like to point out that a number of papers have considered a Poisson model to capture the dynamics of the *arrivals of subjects* to the test center [17], [18] in a streaming GT scenario. In contrast, our model assumes that the number of defectives follows a right-truncated Poisson distribution. In addition, the focus in the aforementioned papers was on the total amount of time (delay) required to test a batch of subjects arriving at the test center at random times. The focus of this work is on finding bounds on the smallest number of tests that ensure that misidentification probability converges to zero. Other related results include [19]-[24], pertaining to two-stage (i.e. adaptive) Binomial group testing; in this paper, we consider *nonadaptive* GT and extensions of the proposed Poisson paradigm to adaptive GT and including noisy measurements will be described in the full version of the paper.

The paper is organized as follows. Section 2 introduces the Poisson GT model, while Section 3 outlines the main result of the paper: an information theoretic upper bound on the minimum number of tests required to achieve asymptotic zero-error average probability.

2. THE POISSON GROUP TESTING MODEL

Let $\mathcal{S}$ denote a set of n test subjects, among which a subset of subjects $\mathcal{D}$ are defective. We assume that the number of defectives follows a right-truncated Poisson distribution with parameters $\lambda(n)$ and n, i.e.,

$$P(D=d) = \begin{cases} c(n)\frac{\lambda(n)^d}{d!}\,\mathrm{e}^{-\lambda(n)}, & 1 \leq d \leq n \\ 0, & \text{otherwise} \end{cases}, \qquad (1)$$

where $D = |\mathcal{D}|$ denote the number of defectives, $\lambda(n)$ is proportional to the expected number of defectives[1], while $c(n)$ is a normalization coefficient such that $\lim_{n\to\infty} c(n) = 1$. Note that the parameter $\lambda(n)$ is assumed to be a function of n, so that the normalization function $c(n)$ that depends on both n and $\lambda(n)$ reduces to a function of n only.

A right-truncated Poisson distribution is closely related to a *finite support version* of the non-uniform Bernoulli model on the set of test subjects, in which the i^{th} subject is defective with probability p_i, $0 \leq p_i \leq 1$, independent of all other test subjects. From Le Cam's theorem [25], it is known that the number of defectives in this model satisfies

$$\sum_{k=0}^{\infty} \left| P\{D=d\} - e^{-\lambda}\,\frac{\lambda^d}{d!} \right| \leq 2\sum_{i=1}^{n} p_i^2, \qquad (2)$$

where $\lambda(n) = \sum_i^n p_i$. As an example, one can choose $p_i = \frac{\beta}{i}$, $\beta > 0$, to obtain a model where individual subjects have decreasing probabilities of being defective as i increases, so that $\lambda(n) = O(\log n)$. The approximation error to the Poisson distribution scales as $2\,\beta\,\zeta(2)$, where $\zeta(\cdot)$ denotes the Riemann zeta function. By choosing β sufficiently small, the approximation error can be reduced to a desired level. For a discussion of adaptive and other classes of non-uniform Bernoulli models, see [19], [20].

Each test includes a subset of the subjects; let m denote the number of tests [2]. The assignment of subjects to different tests are usually specified via a binary matrix termed the *test matrix*, $\mathbf{C} \in \{0,1\}^{m\times n}$. If $\mathbf{C}(i,j) = 1$, for $1 \leq i \leq m$ and $1 \leq j \leq n$, the j^{th} subject is present in the i^{th} test; on the other hand, $\mathbf{C}(i,j) = 0$ implies that the j^{th} subject is excluded from the i^{th} test. The output of a test is equal to 1 if at least one defective was included in the test, and 0 otherwise. In other words, the vector of test results $\mathbf{y} \in \{0,1\}^m$ is equal to the Boolean OR of columns of $\mathbf{C}$ corresponding to the defectives. Fig. 1 demonstrates an example of a test matrix, the set of defectives, and the resulting vector of test results.

$$\mathbf{C} = \begin{array}{c} \begin{array}{cccccccccc} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 & S_7 & S_8 & S_9 & S_{10} \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} \mathbf{y} \\ \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \end{array}$$

Fig. 1. Example of a test matrix and the test results where $n = 10$, $m = 4$, and the set of defectives is $\mathcal{D} = \{S_4, S_8\}$.

We assume that the test matrix is constructed in a probabilistic manner, with each entry of $\mathbf{C}$ being an independent, identically distributed (i.i.d.) Bernoulli(p) random variable; in other words, each entry of the test matrix is equal to 1 with probability p, and 0 with probability $1-p$. Given the vector of test results and the test matrix, the set of defectives is identified using a maximum likelihood decoder,

$$\hat{\mathcal{D}} = \arg\max P(\mathbf{y}|\mathbf{C}, \mathcal{D}'), \qquad (3)$$

where $P(\mathbf{y}|\mathbf{C}, \mathcal{D}')$ is the conditional distribution of observing $\mathbf{y}$ given the test matrix $\mathbf{C}$ and set of defectives $\mathcal{D}'$. Our goal is to find an upper bound on the minimum number of tests m required to ensure that the average probability of error converges to zero as $n \to \infty$. The average is taken over all realizations of the set of defectives and test matrices. Given that our derivations are information-theoretic, the complexity of the decoder is not addressed. Practical structured testing approaches and efficient decoders will be described in the full version of the paper.

3. MAIN RESULTS

The main result of our work is that for $\lambda(n) = o(n)$, a proper choice of p ensures that $m \geq (2+\delta)\lambda^{2+\alpha} \log n$ measurements

[1] It is straightforward to show that the expected value of the right-truncated Poisson distribution equals $\lambda(n)(1 + \frac{\partial}{\partial\lambda}\log c(n))$ which converges to $\lambda(n)$ as $n \to \infty$.

[2] In the model considered in this paper m is deterministic.

suffice to achieve an average probability of error converging to zero, for any $\delta > 0$ and $\alpha > 0$.

Assume that $\mathcal{D}_t$ is the true set of defectives, and let D be the random variable equal to the cardinality of $\mathcal{D}_t$, denoted by $|\mathcal{D}_t|$. Let $\mathcal{E}$ denote the event that there exists a set of subjects $\mathcal{D} \neq \mathcal{D}_t$ such that $P(\mathbf{y}|\mathbf{C}, \mathcal{D}) \geq P(\mathbf{y}|\mathbf{C}, \mathcal{D}_t)$. Given $D = d$, $1 \leq d \leq n$, let $\mathcal{E}_i$, $1 \leq i \leq d$, denote the event that there exists a set of subjects with cardinality d, differing from the true defective set in i items, that is at least as likely as the true defective given the decoder. Let $P_E(i, d, n)$ denote the probability of $\mathcal{E}_i$. From these definitions, one has

$$P_e = \mathbb{E}\left[P(\mathcal{E}|D)\right] = \sum_{d=1}^{n} c(n)\, \frac{\lambda(n)^d}{d!}\, \mathrm{e}^{-\lambda(n)}\, P\big(\cup_{i=1}^{d} \mathcal{E}_i\big)$$

$$\leq \sum_{d=1}^{n}\sum_{i=1}^{d} c(n)\, \frac{\lambda^d}{d!}\, \mathrm{e}^{-\lambda}\, P_E(i, d, n), \quad (4)$$

where the last inequality follows from the union bound [16]. At first glance, it may seem that a bound on P_e may be obtained using an upper bound on $P_E(i, d, n)$ for a *fixed* value of d (such as in [26]) and subsequent averaging; however, there are two subtle, yet important issues that prohibit us from using this approach. First, in (4) the value of d, and hence i, may be as large as n. Since we are interested in asymptotic results as $n \to \infty$, a bound on $P_E(i, d, n)$ should account for the growth of d and i with respect to n. Second, the bounds obtained in [26] rely on a test matrix $\mathbf{C}$ with i.i.d. Bernoulli$(1/d)$ entries. However, in Poisson PGT, the true value of d is unknown (D is a random variable) and cannot be used as a design parameter.

In order to overcome the aforementioned problems, we use different functions to bound $P_E(i, d, n)$ for different ranges of d, using new bounds that do not employ the value of d as a design parameter. In [26], it was shown that for $d = o(n)$ and for all ρ, $0 \leq \rho \leq 1$, one has

$$P_E(i, d, n) \leq 2^{-m\left(E_o(\rho,i,d,n) - \frac{\rho \log \binom{n-d}{i}\binom{d}{i}}{m}\right)} \quad \text{(5a)}$$

where the error exponent E_o satisfies

$$E_o(\rho, i, d, n) = \quad \text{(5b)}$$

$$-\log \sum_{Y\in\{0,1\}} \sum_{T_2} \left(\sum_{T_1} P(\mathbf{t}_1) P(y, \mathbf{t}_2|\mathbf{t}_1, \mathcal{D}_t)^{\frac{1}{1+\rho}} \right)^{1+\rho}.$$

In these equations, Y is a random variable corresponding to the result of a single test. Let $(\mathcal{D}_1, \mathcal{D}_2)$ be a partition of $\mathcal{D}_t$ into disjoint sets with cardinalities $|\mathcal{D}_1| = i$ and $|\mathcal{D}_2| = d - i$, respectively. The vectors T_1 and T_2 are binary-valued row-vectors of length i and $d - i$, indicating which subjects in $\mathcal{D}_1$ and $\mathcal{D}_2$ are present in a given test, respectively.

In order to prove the main results of this section, we need the following lemma.

Lemma 1. *Let $f(n) : \mathbb{N} \mapsto \mathbb{R}^+$. Assume that each entry of the binary test matrix is an i.i.d. Bernoulli(p) random variable, such that $\lceil f(n) \rceil\, p = o(n)$. Then $\forall i, d$ such that $1 \leq i \leq d \leq \lceil f(n) \rceil$, and $\forall \rho$ such that $0 < \rho < 1$, one has the following bound on the error exponent:*

$$E_o(\rho, i, d, n) \geq \rho(1-p)^d ip \left(1 - \frac{\rho}{2}\log^2(ip) + o(1)\right).$$

Proof. The proof of the lemma is rather technical and omitted due to space limitations. □

Note that this lemma is a generalization of a lower bound on $E_o(\rho, i, d, n)$ in [26]. However, the bound in [26] does not apply for the model considered here, since it addressed a combinatorial GT setting where the number of defectives d is assumed to be known and was used as a design parameter.

The next theorem presents the main results for the asymptotic regime where $\log n \leq \lambda = o(n)$. Note that we do not require that $\lambda \geq \log n$, $\forall n > 0$, but only a bound in the limit of large n; in other words, we require that $\exists n' > 0$, such that $\forall n > n'$, $\lambda(n) \geq \log n$.

Theorem 1. *Let $\lambda(n) = o(n)$, where for sufficiently large n, $\lambda(n) > \log n$. Then $m \geq (2+\delta)\lambda^{2+\alpha} \log n$ tests ensure that $P_e = o(1)$, for any $\delta > 0$ and $\alpha > 0$.*

Proof. Since $\lambda = o(n)$, there exists a fixed $\epsilon > 0$ small enough such that $f(n) := \lambda^{(1+\epsilon)} = o(n)$. Choose $p = \lceil f(n) \rceil^{-(1+\gamma)}$, for some $0 < \gamma < 1$. The probability of error in formula (4) can be written as $P_e = P_{e_1} + P_{e_2}$, where

$$P_{e_1} = \sum_{d=1}^{\lceil f(n) \rceil} \sum_{i=1}^{d} c(n)\, \frac{\lambda^d}{d!}\, \mathrm{e}^{-\lambda}\, P_E(i, d, n),$$

$$P_{e_2} = \sum_{d=\lceil f(n) \rceil + 1}^{n} \sum_{i=1}^{d} c(n)\, \frac{\lambda^d}{d!}\, \mathrm{e}^{-\lambda}\, P_E(i, d, n).$$

The idea is to bound these probabilities by finding a tight upper bound on $P_E(i, d, n)$, independent of i and d, for $1 \leq d \leq \lceil f(n) \rceil$, while using the upper bound $P_E(i, d, n) \leq 1$ for $\lceil f(n) \rceil + 1 \leq d \leq n$.

First, note that the modes of a standard Poisson distribution with parameter λ are equal to $\lfloor \lambda \rfloor$ and $\lceil \lambda \rceil - 1$. Hence,

$$\arg\max_{d} \frac{\lambda^{(d-1)}}{(d-1)!} \leq \lfloor \lambda \rfloor + 1 \leq \lceil \lambda \rceil + 1,$$

and for any $d \geq \lceil \lambda \rceil + 1$, the aforementioned function is monotonically decreasing. Since $\lceil f(n) \rceil + 1 \geq \lceil \lambda \rceil + 1$, then

$$\max_{d:\lceil f(n) \rceil + 1 \leq d \leq n} \frac{\lambda^d}{(d-1)!} = \frac{\lambda^{\lceil f(n) \rceil + 1}}{(\lceil f(n) \rceil)!}.$$

Now, one has

$$P_{e_2} = \sum_{d=\lceil f(n) \rceil + 1}^{n} \sum_{i=1}^{d} c(n)\, \frac{\lambda^d}{d!}\, \mathrm{e}^{-\lambda}\, P_E(i, d, n)$$

$$\leq \sum_{d=\lceil f(n) \rceil + 1}^{n} c(n)\, \frac{\lambda^d}{(d-1)!}\, \mathrm{e}^{-\lambda}$$

$$\leq n\, c(n) \max_{d:\lceil f(n) \rceil + 1 \leq d \leq n} \frac{\lambda^d}{(d-1)!}\, \mathrm{e}^{-\lambda}$$

$$\leq \exp\left(-\epsilon \lceil f(n) \rceil \log \lambda + o(\lceil f(n) \rceil \log \lambda)\right) = o(1).$$

Since we have chosen $p = \lceil f(n) \rceil^{-(1+\gamma)}$, for some $0 < \gamma < 1$, using Lemma 1 one can show that $\forall i, d,\ 1 \le i \le d \le \lceil f(n) \rceil$ and $\forall \rho,\ 0 < \rho < 1$,

$$\begin{aligned} E_o(\rho, i, d, n) &\ge \rho(1-p)^d ip \left(1 - \frac{\rho}{2}\log^2(ip) + o(1)\right) \\ &\ge \rho\, p(1-p)^{\lceil f(n) \rceil} \left(1 - \frac{\rho}{2}\log^2(p) + o(1)\right) \\ &\ge \frac{\rho\,(1+o(1))}{(f(n)+1)^{1+\gamma}} \left(1 - \frac{\rho}{2}(1+\gamma)^2 \log^2(f(n)+1)\right). \end{aligned}$$

Now, by choosing $\rho = \frac{1}{(1+\gamma)^2 \log^2(f(n)+1)}$, one arrives at

$$E_o(\rho, i, d, n) \ge \frac{1}{2} \frac{\rho}{(f(n)+1)^{1+\gamma}}\,(1+o(1)),$$

for any i, d such that $1 \le i \le d \le \lceil f(n) \rceil$. In addition, using the inequality $\binom{d}{i} \le \left(\frac{de}{i}\right)^i$, it can be easily shown that for $1 \le i, d \le \lceil f(n) \rceil$, it holds that

$$\log \binom{n-d}{i}\binom{d}{i} \le f(n) \log n\,(1+o(1)).$$

As a result, if

$$m > \frac{\rho f(n) \log n}{\frac{1}{2}\frac{\rho}{(f(n)+1)^{1+\gamma}}}(1+o(1)) = 2 f(n)^{2+\gamma} \log n\,(1+o(1)),$$

then the exponent in (5) is positive. Therefore, by using $m \ge (2+\delta) f(n)^{2+\gamma} \log n$ for any fixed $\delta > 0$, we may write

$$\begin{aligned} P_E(i, d, n) &\le 2^{-m\left(E_o(\rho,i,d,n) - \frac{\rho \log \binom{n-d}{i}\binom{d}{i}}{m}\right)} \\ &\le 2^{-m\rho\left(\frac{1}{2}\frac{1+o(1)}{(f(n)+1)^{1+\gamma}} - \frac{f(n)\log n(1+o(1))}{m}\right)} \\ &= 2^{-\frac{\delta \log n f(n)}{(1+\gamma)^2 \log^2 f(n)}(1+o(1))} := P_1(n). \end{aligned}$$

Since

$$\begin{aligned} P_{e_1} &= \sum_{d=1}^{\lceil f(n) \rceil} \sum_{i=1}^{d} c(n)\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda}\,P_E(i,d,n) \\ &\le c(n)\,P_1(n) \sum_{d=1}^{\lceil f(n) \rceil} d\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda} \\ &\le c(n)\,P_1(n) \sum_{d=1}^{\infty} d\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda} = c(n) P_1(n) \lambda(n), \end{aligned}$$

it follows that

$$\begin{aligned} P_{e_1} &\le c(n)\lambda(n)P_1(n) \\ &= 2^{-\frac{\delta \log n f(n)}{(1+\gamma)^2 \log^2 f(n)}(1+o(1))} = o(1). \end{aligned}$$

Consequently, the probability of error converges to zero, i.e., $P_e = o(1)$ if $m \ge (2+\delta) f(n)^{2+\gamma} \log n$, for any fixed $\delta > 0$ and $\gamma > 0$. Substituting $f(n) = \lambda^{1+\epsilon}$ in the previous expression, and performing some straightforward simplifications reduces the bound to $m \ge (2+\delta)\lambda^{2+\alpha} \log n$, for $\delta > 0$ and $\alpha > 0$. $\square$

Let $\log^{(K)} n := \underbrace{\log \log \cdots \log}_{K \text{ times}} n$, for an integer $K \ge 1$, and let $\log^{(0)} n := n$. The next theorem states the main results for the case when asymptotically, $\lambda(n) < \log n$. The idea is to confine the growth of $\lambda(n) = o(n)$ between $\log^{(K)} n$ and $\log^{(K-1)} n$ for some $K > 1$, and prove the results for such bounded values of $\lambda(n)$.

Theorem 2. *Let $\log^{(K)} n \le \lambda < \log^{(K-1)} n$, for some $K > 1$ and for sufficiently large n. Then $m \ge (2+\delta)\lambda^{2+\alpha} \log n$ tests suffice to ensure $P_{\mathrm{e}} = o(1)$, for any $\delta > 0$ and $\alpha > 0$.*

Proof. The proof of this theorem is similar to the proof of Theorem 1. As a result, and due to space restrictions, we only describe the sketch of the proof. First, we write a bound on the probability of error as

$$P_e \le \sum_{d=1}^{n} \sum_{i=1}^{d} c(n)\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda}\,P_E(i,d,n) = \sum_{k=1}^{K+1} P_{e_k},$$

where

$$\begin{aligned} P_{e_1} &= \sum_{d=1}^{\lceil f(n) \rceil} \sum_{i=1}^{d} c(n)\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda}\,P_E(i,d,n), \\ P_{e_2} &= \sum_{d=\lceil f(n) \rceil + 1}^{\lceil \log^{(K-1)} n \rceil} \sum_{i=1}^{d} c(n)\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda}\,P_E(i,d,n), \\ P_{e_k} &= \sum_{d=\lceil \log^{(K-k+2)} n \rceil + 1}^{\lceil \log^{(K-k+1)} n \rceil} \sum_{i=1}^{d} c(n)\,\frac{\lambda^d}{d!}\,\mathrm{e}^{-\lambda}\,P_E(i,d,n), \end{aligned}$$

where $k \in \{3, \ldots, K+1\}$. In our proof, we first show that $P_{e_k} = o(1), \forall k \in \{2, 3, \ldots, K+1\}$. Then, we use Lemma 1 to show that by choosing $p = \lceil f(n) \rceil^{-(1+\gamma)}$ for some $0 < \gamma < 1$, $m \ge (2+\delta)\lambda^{2+\alpha} \log n$, we can also ensure that $P_{e_1} = o(1)$, for any $\delta > 0$ and $\alpha > 0$. $\square$

In [26], information-theoretic arguments were used to show that for nonadaptive combinatorial GT, and $d = o(n)$, $m = O(d \log^2 d \log n)$ measurements ensure that the probability of error converges to zero. In our setup, even though the *average number of measurements* is asymptotically equal to $\lambda(n)$, we need $O(\lambda^{2+\alpha}(n)\, \log\, n)$ measurements due to the probabilistic nature of the number of defectives. In the well-studied adaptive Binomial(n, p_0) group testing scenario of [24], it was shown that when $p_0 = n^{-\beta}$, $\beta \in (0, \frac{1}{2})$, the expected number of measurements needed for asymptotically accurate detection equals $\mathbb{E}(m) = O(np_0 |\log p_0|)$. Although the Poisson GT model considered in this paper represents a generalization of the model described in [24], we would like to point out that setting $p_0 = \lambda/n$ we obtain an approximation of $\mathbb{E}(m) = O(\lambda(n) \log n)$ tests needed for accurate Poisson GT. This reduction in the cost of measurements stems from the adaptive nature of the GT scheme in [24] as compared to the non-adaptive approach studied in this manuscript.

4. REFERENCES

[1] D. -Z. Du and F. K. Hwang, *Pooling designs and non-adaptive group testing*, NJ: World Scientific, 2006.

[2] A. G. D'yachkov, "Lectures on designing screening experiments," *Lecture Note Series 10*, Feb. 2004.

[3] J. K. Wolf, "Born again group testing: multi-access communications," *IEEE Trans. Inf. Theory*, vol. 31, pp. 185-191.

[4] P. -O. Anderson, "Superimposed codes for the Euclidean channel," Ph.D. thesis no. 342, Linköping Univ., Linköping, Sweden, 1994.

[5] P. Z. Fan, M. Darnell, and B. Honary, "Superimposed codes for the multiaccess binary adder channel," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1178-1182, 1995.

[6] A. Barg, G. R. Blakley, and G. A. Kabatiansky, "Digital fingerprinting codes: problem statements, constructions, identification of traitors," *IEEE Trans. Inf. Theory*, vol. 49, pp. 852-865, 2003.

[7] D. -Z. Du and F. K. Hwang, *Combinatorial group testing and its applications*, NJ: World Scientific, 2000.

[8] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Stat.*, vol. 14, pp. 436-440, 1943.

[9] P. Damaschke, "Threshold group testing," *General Theory of Information Transfer and Combinatorics* in: LNCS, vol. 4123, pp. 707-718, 2006.

[10] A. Emad and O. Milenkovic, "Symmetric Group Testing and Superimposed Codes," *Proc. Inf. Theory Workshop (ITW11)*, Oct. 2011, pp. 20-24.

[11] A. Emad and O. Milenkovic, "Semi-quantitative Group Testing," *Proc. IEEE Int. Symp. Inf. Theory (ISIT'12)*, July 2012, pp. 1847-1851.

[12] A. Emad and O. Milenkovic, "Semi-Quantitative Group Testing: A Unifying Framework for Group Testing with Applications in Genotyping," *arXiv:1210.2484v2*, pp. 1-69, Oct. 2012.

[13] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489-509, 2006.

[14] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289-1306, 2006.

[15] W. Dai and O. Milenkovic, "Weighted superimposed codes and constrained integer compressed sensing," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2215-2229, 2009.

[16] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes*, 4th ed., NY: McGraw-Hill Education, 2002.

[17] S. Bar-Lev, M. Parlar, D. Perry, W. Stadje, and F. V. der, "Applications of bulk queues to group testing models with incomplete identification," *Eur. J. Oper. Res.*, vol. 183, pp. 226-237, 2007.

[18] D. Claeys, J. Walraevens, K. Laevens, and H. Bruneel, "A queueing model for general group screening policies and dynamic item arrivals," *Eur. J. Oper. Res.*, vol. 207, pp. 827-835, 2010.

[19] M. Garey, and F. K. Hwang, "Isolating a single defective using group testing," *J. Amer. Stat. Assoc.*, vol. 69, pp. 151-153, Mar. 1974.

[20] F. K. Hwang, "A generalized binomial group testing problem," *J. Amer. Stat. Assoc.*, vol. 70, pp. 923-926, Dec. 1975.

[21] E. Knill, "Lower bounds for identifying subset members with subset queries," *Proc. 6th Ann. ACM-SIAM Symp. discr. Algorithms*, San Francisco, Jan. 1995, pp. 369-377.

[22] T. Berger and V. I. Levenshtein, "Asymptotic efficiency of two-stage disjunctive testing," *IEEE Trans. Inf.Theory*, vol. 48, pp. 1741-1749, 2002.

[23] L. A. De Bonis and U. Vaccaro, "Optimal two-stage algorithms for group testing problems," *SIAM J. Computing*, vol. 34, pp. 1253-1270, 2005.

[24] M. Mézard and C. Toninelli, "Group testing with random pools: optimal two-stage algorithms," *IEEE Trans. Inf. Theory* , vol. 57, pp. 1736-1745, March 2011.

[25] L. Le Cam, "An approximation theorem for the Poisson Binomial distribution," *Pacific J. Math.*, vol. 10, pp. 1181-1197, 1960.

[26] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, pp. 1880-1901, 2012.