Density-Dependent Group Testing

Rahil Morjaria
University of Bristol

Saikiran Bulusu
The Ohio State University

Venkata Gandikota
Syracuse University

Sidharth Jaggi
University of Bristol

Abstract

Group testing is the problem of identifying a small subset of defectives from a large set using as few binary tests as possible. In most current literature on group testing the binary test outcome is 1 if the pool contains at least one defective, and 0 otherwise. In this work we initiate the study of a generalized model of group testing that accommodates the physical effects of dilution of infected samples in large pools. In this model the binary test outcome is 1 with probability $f(\rho)$, where ρ is the density of the defectives in the test, and $f : [0,1] \to [0,1]$ is a given "test function" that models this dilution process. For a large class of test functions our results establish near-optimal sample complexity bounds, by providing information-theoretic lower bounds on the number of tests necessary to recover the set of defective items, and providing computationally efficient algorithms with sample complexities that match these lower bounds up to constant or logarithmic factors. Furthermore, using tools from real analysis, we extend our results to any "sufficiently well-behaved function" $f : [0,1] \to [0,1]$.

1 INTRODUCTION

Group testing is a method used to efficiently identify a small subset $\mathcal{K}$ ($|\mathcal{K}| = k$) of defective items within a larger population $\mathcal{N}$ ($|\mathcal{N}| = n \gg k$) by pooling multiple items into a single test. Instead of testing each item individually, group testing combines items into groups and conducts binary tests. This approach significantly reduces the number of tests required, especially when the number of defective items is small relative to the population.

Proceedings of the 28[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

Originally introduced by Dorfman during World War II to screen for diseases in soldiers (Dorfman, 1943), group testing has since found applications in diverse areas such as industrial quality assurance (Sobel and Groll, 1959), DNA sequencing (Pevzner and Lipshutz, 1994), molecular biology (Macula, 1999; Schliep et al., 2003), wireless communication (Wolf, 1985; Berger et al., 1984; Luo and Guo, 2008), data compression (Hong and Ladner, 2002; Emad and Milenkovic, 2015), pattern matching (Clifford et al., 2007), secure key distribution (Stinson et al., 2000), network tomography (Ma et al., 2014; Cheraghchi et al., 2012), efficient data analysis (Engels et al., 2021; Cormode and Muthukrishnan, 2005) and more recently in COVID-19 testing protocols (Wang et al., 2023; Ghosh et al., 2021).

The central question in group testing is to investigate the number of binary tests necessary and sufficient to recover $\mathcal{K}$ with a vanishing error probability. The algorithmic question revolves around efficiently recovering the set of defectives $\mathcal{K}$ given sufficient test outcomes. Decades of research, motivated by various applications, have answered these questions in a variety of settings and recovery guarantees. See (D'yachkov, 2003; Malyutov, 2013; Aldridge et al., 2019; Ngo and Du, 2000; Chen and Hwang, 2008) for excellent surveys on Group Testing and its variants.

This work focuses on the problem of exact recovery of $\mathcal{K}$ in the non-adaptive setting (where the tests are fixed in advance which allows for parallelization of the testing process). This is in contrast with the adaptive setting where tests may be designed based on the previous test outcomes.

A large proportion of the literature is focused on classical group testing, where the outcome is positive if at least one defective item is present in the group and negative otherwise. In the classical setting of non-adaptive probabilistic group testing, assuming an i.i.d. prior on the defective items, it has been established that $\Theta(k\log(n/k))$ tests are necessary and sufficient for efficient recovery (Scarlett and Cevher, 2017) of the defectives $\mathcal{K}$. These algorithms are also known to be robust to errors (Guruswami and Wang, 2023; Cheraghchi et al., 2011; Cai et al., 2017).

In this work, we study the group testing problem under the influence of dilution[1], where defective items in a pooled test may become harder to detect as their proportion decreases. Damaschke (2006) introduced this line of research by using threshold test functions to model dilution effects. In their model, a test outcome is positive if the pool contains more than u defective items, negative if it contains fewer than ℓ and arbitrary otherwise. Following the adaptive schemes of Damaschke (2006), several non-adaptive (Chen and Fu, 2009; D'yachkov et al., 2013; Bui et al., 2019; Gerbner et al., 2013) schemes have been proposed. Efficient algorithms that are robust to test errors have also been studied in this model (Bui et al., 2024). Furthermore, Ahlswede et al. (2013) extended these results to slightly more general threshold test functions, where a test outcome is positive if the number of defectives exceeds a function $f_1(k)$ and negative if it falls below a function $f_2(k)$, with $f_1 > f_2$.

The above-mentioned threshold test functions and various other variants were generalized by Cheng et al. (2022, 2023), who consider general monotonic stochastic test functions $f : [k] \to [0,1]$ such that $f(\ell)$ is the probability of a positive test outcome when a pool has ℓ defectives. The authors characterize the properties of f that allow for efficient recovery with near-optimal number of tests. In spirit, our paper is a natural extension of the work done in Cheng et al. (2022, 2023).

The effect of dilution is more accurately captured by the defective density, defined as $\rho = \ell/g$, where g and ℓ are respectively the size of the pool and the number of defectives in it. This distinction is critical because the likelihood of detecting a defective item decreases as the pool size increases relative to the number of defectives. For instance, a single defective item in a pool of 100 is far less likely to be detected than in a pool of 10, highlighting the limitations of traditional threshold-based models that treat pools of different sizes similarly.

Our Contributions: In this work, we build upon this observation and propose a more general framework by introducing *test functions* $f : [0,1] \to [0,1]$, where $f(\rho)$ models the probability of a positive test outcome as a function of the defective density ρ. This density-based approach allows us to better capture the physical effects of dilution, providing a more accurate representation of real-world group testing scenarios, such as medical testing or signal processing, where sample concentration can drastically affect detection probabilities.

Our contributions go beyond extending previous models; we establish new theoretical guarantees for efficient testing under this generalized framework. Specifically, we characterize the properties of the test function $f(\rho)$ that enable accurate recovery of defectives with an optimal number of tests, even when the dilution effect is strong.

We consider successively broader classes of density-dependent test functions $f(\rho)$. For each class of test functions, we present information-theoretic lower bounds on the number of tests required to obtain the defective set with error probability bounded above by some $\varepsilon > 0$ for *any* non-adaptive group testing algorithm, and matching such lower bounds with corresponding computationally efficient non-adaptive group testing algorithms that meet these sample complexity lower bounds up to small factors.

In this model of density-dependent test functions, the sample complexity scales quite differently from classical group testing. In classical group testing, the sample complexity required to ensure vanishing probability of error scales as $\Theta(k\log n)$ whereas in the model considered in this paper

[1] Certain other works such as Hwang (1976); Atia and Saligrama (2012); Arpino et al. (2021) have also considered a *dilution noise* model. However, the model in these works is quite different to the one considered here; for each defective item, it is assumed that in an i.i.d. manner its presence will be missed due to a dilution effect. Such models do not explicitly model the density of defectives in tests and rather assign a fixed dilution probability, and hence arguably a less precise way to measure dilution. As a measure of the difference from the model in this paper, the information-theoretic limits on sample complexities in those works are typically $\Theta(k\log(n))$, arguably overly optimistic, as opposed to the $\Omega((n/k)^{\alpha}k\log(n))$ or $\Omega((n/k)^{2\alpha}k\log(n))$ scaling (for some positive constant α) in the more physically motivated Density-Dependent model in this paper.

the corresponding sample complexity scales quite differently. In particular, roughly speaking, if the "shape" of the test function $f(\rho)$ for ρ in the neighbourhood of 0 scales roughly proportionally to ρ^α, then the sample complexity scales essentially as one of $\mathcal{O}(k\left(\frac{n}{k}\right)^\alpha \log n)$, $\mathcal{O}(k\left(\frac{n}{k}\right)^{2\alpha} \log n)$, or $\Theta(n \log n)$, depending on the "shape" of the test function $f(\rho)$.

As an example, in Theorem 1 we consider test functions of the form $f(\rho) = a_1\rho^\alpha$, for $a_1 \in (0,1]$ and $\alpha \in (0,\infty]$, our choice of algorithm depends on the value of α. For $\alpha \geq 1$ our algorithm is simply "repetition testing", wherein each of the n items is tested individually repeatedly $\Theta(\log n)$ times, for a total of $\Theta(n \log n)$ tests. On the other hand, for $\alpha \in (0,1)$, our groups are chosen using a Bern$(1/k)$ "random test design", which requires a total of $\mathcal{O}(k\left(\frac{n}{k}\right)^\alpha \log n)$ tests. For both test designs the items are declared to be defective or non-defective depending on whether the fraction of observed positive test outcomes in which the item participates is skewed towards one of two values corresponding to the two hypotheses that the item is defective or non-defective. This scaling behaviour is explained by noting that the expected group density is quite low, about $\frac{k}{n}$, hence the probability of test outcome positivity is commensurately low, scaling as $a_1\left(\frac{k}{n}\right)^\alpha$, and hence each test conveys relatively little information about $\mathcal{K}$.

In Theorem 2 test functions of the form $f(\rho) = a_0 + a_1\rho^\alpha$, for $a_0, a_1 \in (0,1]$ and $\alpha \in (0,\infty)$ are considered. Compared to Theorem 1, the story is broadly similar, except that the sample complexity is significantly larger for some parameter regimes, scaling as $\mathcal{O}\left(k\left(\frac{n}{k}\right)^{2\alpha} \log n\right)$ for $\alpha \in \left(0, \frac{1}{2}\right)$ and as $\mathcal{O}(n \log n)$ for $\alpha \geq \frac{1}{2}$. The difference in scaling behaviour relative to that seen in the scenario in Theorem 1 can be explained via the following observations. The expected density is still $\frac{k}{n}$, hence test outcome positivity probability scales to first order as a_0 *regardless* of the number of defectives in the test (since $a_0 \gg a_1\left(\frac{k}{n}\right)^\alpha$), therefore it is even more challenging to extract meaningful information from test outcomes.

In Theorems 3 and 4, where we again obtain near-optimal sample complexities for a much more general class of test functions in the *set of α-power series functions* $\mathcal{A}(\alpha)$ defined as

$$\begin{aligned}\mathcal{A}(\alpha) := \Bigg\{ f : [0,1] \to [0,1]; f(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha}\ , \\ c_f := \sum_{i=0}^{\infty} |a_i| < \infty, a_1 > 0 \Bigg\},\end{aligned} \tag{1}$$

where the condition $a_1 > 0$ arises from the physically motivated requirement that $f(\rho)$ be increasing at $\rho = 0$. In Theorem 3, we analyse test functions in $\mathcal{A}(\alpha)$ where $f(0) = 0$, where we see the same scaling behaviour as Theorem 1. In Theorem 4 we analyse test functions where $f(0) > 0$, where we see the same scaling behaviour as Theorem 2. Intuitively this behaviour follows for the same reasons as before, with the "lower order" terms contributing minimally.

The choice of analysing test functions in $\mathcal{A}(\alpha)$ is motivated by Section 6, where we show that the Stone-Weierstrass theorem (Weierstrass, 1885; Stone, 1937, 1948) implies that for any $\alpha > 0$ the set of (fractional) power series in ρ^α are dense in the set of continuous real-valued functions on $[0,1]$. Theorem 11 complements this by giving a method to compute, for "sufficiently smooth" $f(\cdot)$, an absolutely convergent (fractional) power series via an appropriately defined fractional derivative defined in Katugampola (2014).

In Theorems 5 and 6 we extend our results to obtain near-optimal sample-complexity results for a large class of functions outside $\mathcal{A}(\alpha)$ if such a function can be appropriately "sandwiched" between two functions in $\mathcal{A}(\alpha)$ for any $\alpha > 0$.

While not our main focus, in Theorems 7 and 8 we extend the analysis of our algorithms for even more general test functions that are *not* density-dependent, such as $f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ – in these cases our upper and lower bounds, while non-trivial, are not in general order-optimal.

2 MODEL DESIGN

2.1 MODEL DEFINITIONS

Definition 1 (Number of Items). *We denote by $\mathcal{N}$ the set $\{1,\ldots,n\}$ of n items.*

Definition 2 (Defective Set). *The defective set $\mathcal{K} \subseteq \mathcal{N}$ is defined by the length-n defectivity vector $K := (k_i)_{i\in\mathcal{N}}$. Each component of the defectivity vector is drawn i.i.d. with each $k_i \sim$ Bern$\left(\frac{k}{n}\right)$, $i \in$*

$\mathcal{N}$. If $k_i = 1$ then $i \in \mathcal{K}$, if $k_i = 0$ then $i \notin \mathcal{K}$. Here k denotes the expected number of defective items. In this work we focus on the sparse scenario $k = n^\theta$, where $\theta \in (0,1)$ is the sparsity parameter.

Definition 3 (**Test Function**). *A test function $f(\ell, g) : [0,1] \to [0,1]$ outputs the probability of a positive test outcome given inputs ℓ which represents the number of defective items in the test and g which is the total number of items in the test. Test functions are assumed to be increasing functions, with our main focus being on 'density-dependent' inputs of the form $\rho := \frac{\ell}{g}$.*

Definition 4 (**Non-Adaptive Test Design**). *A non-adaptive testing scheme with n items and T tests is represented by the binary $T \times n$ test design matrix $\mathbf{X}$ where $x_{ij} = 1$ denotes the inclusion of item $i \in \mathcal{N}$ in the j'th test and $x_{ij} = 0$ denotes the exclusion of item $i \in \mathcal{N}$ from the j'th test.*

The T test results are denoted by the test outcome vector $Y \in \{0,1\}^T$, where 0 denotes a negative test outcome, and 1 denotes a positive test outcome.

The corresponding decoder is a function taking as inputs the test outcome vector Y and the test design matrix $\mathbf{X}$, and outputs the estimated defective set $\hat{\mathcal{K}} \subset \mathcal{N}$. The probability of error of the scheme is defined as the probability (over the set $\mathcal{K}$, any randomness in the test design matrix $\mathbf{X}$, test function $f(\cdot)$, and decoder) $\Pr(\hat{\mathcal{K}} \neq \mathcal{K})$.

2.2 MODEL OVERVIEW

Given a set of items $\mathcal{N} = [n]$ of size $n \in \mathbb{N}$. We assume that each item is defective i.i.d. $\mathrm{Bern}\left(\frac{k}{n}\right)$ for some *expected number of defective items* $k = n^\theta$ where $\theta \in (0,1)$ is the *sparsity parameter*. In our model we assume that the testing apparatus generates the binary test outcomes corresponding to a given *test function* $f(\ell, g)$ where $\ell \in \mathbb{N}_0$ denotes the number of defective items in a given test and $0 < g \leq n$ corresponds to the total number of items in a given test. Here the *test function* $f(\ell, g)$ outputs the probability of a positive test outcome.

Our objective is to minimize the *total number of tests* $T \in \mathbb{N}$ needed to obtain the defective set $\mathcal{K}$ with a probability of error bounded above by $\varepsilon \in (0,1)$ using a non-adaptive test design. In this paper we obtain lower and upper bounds on T as a function of n, k, ε and $f(\cdot)$.

3 ALGORITHM DEFINITIONS

To derive our upper bounds we consider two algorithms. Algorithms considered in this work are a naïve "Repetition Testing + Midpoint Decoder" that has near-optimal sample complexity for some parameter regimes, and a "Random Test Design + NCOMP Decoder" that has near-optimal sample complexity in other regimes.

3.1 REPETITION TESTING

Algorithm 1 (**Repetition Testing + Midpoint Decoding**).

Encoder: *Repetition testing is a non-adaptive test design, where each $i \in \mathcal{N}$ is tested individually $c_I \log n$ times for some $c_I > 0$.*

Decoder: *Given any test function f, the Midpoint Decoder is a decoder with computational complexity $\mathcal{O}(nT)$ which outputs the estimation for the defective set $\hat{\mathcal{K}}$,*

$$\hat{\mathcal{K}} = \left\{ i \in \mathcal{N} : \frac{p_i}{c_I \log n} > \frac{f(1) + f(0)}{2} \right\},$$

where for each $i \in \mathcal{N}$, p_i denotes the total number of positive tests i participates in.

3.2 RANDOM TEST DESIGN

As we shall see, random test designs allow for recovery via significantly fewer tests than via repetition testing for a large range of test functions. To define the NCOMP Decoder, which is a generalisation of the Midpoint Decoder, we require a preliminary definition.

Definition 5 (**Item-Included Test-Positivity Probability**). *Given test function $f(\cdot)$ and some $g \in \mathbb{N}$ two important quantities used by our decoding rule are*

$$\mu_g^+(f) = \mathbb{E}\left[f\left(\ell + 1, g\right)\right] \text{ and } \mu_g^-(f) = \mathbb{E}\left[f\left(\ell, g\right)\right].$$

Here the expectations are with respect to the number of defectives in a group ℓ being distributed as a binomial distribution $Bin(g-1, k/n)$. The quantities $\mu_g^+(f)$ and $\mu_g^-(f)$ respectively denote the probability of a positive test outcome conditioned on a specific item in the test (containing g items) being defective or non-defective, with there being g items in the test. For notational convenience we sometimes abbreviate these as μ_g^+ and μ_g^-.

Algorithm 2 (Random Test Design + NCOMP Decoding).

Encoder: *A random test design is a non-adaptive test design, where we choose each entry of the test design matrix* $\mathbf{X}$ *in an i.i.d. manner, with each entry* $x_{ij} \sim Bern(1/k)$.

Decoder: *Given any test function* f*, the NCOMP Decoder is a decoder with computational complexity* $\mathcal{O}(nT)$ *which outputs the estimation for the defective set* $\hat{\mathcal{K}}$,

$$\hat{\mathcal{K}} = \left\{ i \in \mathcal{N} : \frac{p_i}{t_i} > \frac{\mu_g^+(f) + \mu_g^-(f)}{2} \right\},$$

where for each $i \in \mathcal{N}$, p_i *and* t_i *denote the total number of positive tests and the total number of tests* i *participates in respectively. In this algorithm we apply the decoder at* $g = \frac{n}{k}$.

4 OVERVIEW OF TECHNIQUES

Fn Class	**Regime**	**Lower Bound on T**	**Upper Bound on T**
$f(\rho) = a_1\rho^\alpha$ $a_1 \in (0,1]$ [Theorem 1]	$\alpha \in (0,1)$	$\Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$
	$\alpha \in [1,\infty)$	$\Omega(n)$	$\mathcal{O}(n\log n)$
$f(\rho) = a_1\rho^\alpha + a_0$, $a_1, a_0 \in (0,1]$ $a_1 + a_0 \leq 1$ [Theorem 2]	$\alpha \in \left(0, \frac{1}{2}\right)$	$\Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$
	$\alpha \in \left[\frac{1}{2},\infty\right)$	$\Omega(n\log n)$	$\mathcal{O}(n\log n)$
$f(\rho) \in \mathcal{A}(\alpha)$, $f(0) = 0$ [Theorem 3]	$\alpha \in (0,1)$	$\Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$
	$\alpha \in [1,\infty)$	$\Omega(n)$	$\mathcal{O}(n\log n)$
$f(\rho) \in \mathcal{A}(\alpha)$, $f(0) > 0$ [Theorem 4]	$\alpha \in \left(0, \frac{1}{2}\right)$	$\Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$
	$\alpha \in \left[\frac{1}{2},\infty\right)$	$\Omega(n\log n)$	$\mathcal{O}(n\log n)$
$f(\rho)$ s.t. $f(0) = 0$ $f(\rho)$ sandwiched between functions in $\mathcal{A}(\alpha)$. [Theorem 5]	$\alpha \in (0,1)$	$\Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$
	$\alpha \in [1,\infty)$	$\Omega(n)$	$\mathcal{O}(n\log n)$
$f(\rho)$ s.t. $f(0) > 0$ $f(\rho)$ sandwiched between functions in $\mathcal{A}(\alpha)$. [Theorem 6]	$\alpha \in \left(0, \frac{1}{2}\right)$	$\Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$
	$\alpha \in \left[\frac{1}{2},\infty\right)$	$\Omega(n\log n)$	$\mathcal{O}(n\log n)$
$f(\ell, g) = a_1 \frac{\ell^\alpha}{g^\beta}$ $a_1 \in (0,1]$, $0 \leq \alpha \leq \beta$ [Theorem 7]	$\alpha \in (0,1)$	$\Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$
	$\alpha \in [1,\infty)$	$\Omega(n)$	$\mathcal{O}(n\log n)$
$f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ $a_0 + a_1 \leq 1$ $0 < \alpha \leq \beta$ [Theorem 8]	$\alpha \in \left(0, \frac{1}{2}\right)$	$\Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\right)$	$\mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$
	$\alpha \in \left[\frac{1}{2},\infty\right)$	$\Omega(n\log n)$	$\mathcal{O}(n\log n)$

Table 1: Summary of Sample Complexity Results.

We present here a high-level overview of the techniques used in proving our results, with details deferred to the appendix.

▶ **Theorems 1-8, lower bounds:** To prove our information-theoretic sample-complexity lower bounds we use two methods.

- The first method takes the ratio of the entropy of the defective set $\mathcal{K}$ and the entropy of test outcomes to obtain a lower bound on the sample complexity. This technique provides us with reasonably good lower bounds when the probability of a positive test outcome is small enough (for instance in Theorem 1 when $f(\rho) = a_1\rho^\alpha$, for the parameter regime $\alpha \in (0,1)$, this probability scales as $a_1\left(\frac{k}{n}\right)^\alpha$).

- In other scenarios, for instance in Theorem 2 when $f(\rho) = a_0 + a_1\rho^\alpha$, the presence of the non-zero constant a_0 means that the first method would result in too loose a lower bound. Instead we extend a technique developed in Cheng et al. (2022, 2023), where a novel lower bound is developed in terms of the mean and variance of the random variable corresponding to positive test outcomes.

▶ **Theorems 1-8, upper bounds:** Our sample complexity upper bounds are all obtained by the two algorithms in Section 3. Where the tests are designed either using repetition testing (for instance for functions of the form $f(\rho) = a_1\rho^\alpha$ in Theorem 1, with $\alpha \geq 1$) or via random test designs (for instance in Theorem 1 with $\alpha \in [0,1)$).

- Repetition testing involves testing each item individually $c_I \log(n)$ times, and then applying Midpoint Decoding. We use standard concentration arguments to show that for a test function f (Definition 3) if $c_I \geq \frac{4}{(f(1)-f(0))^2}$ the likelihood an item is misclassified is bounded above by ε/n, hence a union bound over all items results in at most ε error.

- A random test design involves testing groups of items, where each item $i \in \mathcal{N}$ is included in a test with probability $1/k$ and then applying NCOMP Decoding. Using concentration arguments we show that for the test functions explored in Section 5, if T is chosen large enough the likelihood an item is misclassified is bounded above by ε/n, hence a union bound over all items results in at most ε error. We give upper bounds on the constants $u_{i,R}$ in the appendix. Complementing the similar phenomenon in our lower bounds, the difference in the sample complexity required for settings where $f(0) = 0$ versus when $f(0) > 0$ follows from the looser level of concentration of the probability of misclassification error in the latter case.

▶ **Theorem 11**: For general test functions, the formalism of fractional power series is useful.

The fractional derivative defined in Katugampola (2014) comes equipped with a Fractional Rolle's Theorem (see Katugampola (2014) for details) and is relatively tractable with $D^{\alpha}(f)(x) = x^{1-\alpha} f'(x)$. By applying the Fractional Rolle's Theorem iteratively, we are able to compute the closed form of the remainder.

5 MAIN RESULTS

We now present our main results (with proofs deferred to the appendix) in a sequence of theorems characterizing the sample complexity for density-dependent group testing, each focusing on a successively broader class of test functions $f(\rho)$. The lower bounds are information-theoretic in nature and hold for *any* non-adaptive group testing scheme, and the upper bounds follow by analyzing the performance of the two algorithms presented in Section 3, both of which have computational complexity that is polynomial in n and k, and in particular $\mathcal{O}(nT)$.

First, in Section 5.1 we consider test functions that scale polynomially (with fractional powers) in ρ, to functions in $\mathcal{A}(\alpha)$. Then in Section 5.2 we generalize this to test functions outside this class, which can be bounded by functions in $\mathcal{A}(\alpha)$. Finally, while this is not the main thrust of this work, in Section 5.3 we show the power of our analytical techniques by generalizing beyond density-dependent test functions, by providing results even for test functions of the form $f(\ell, g) = a_0 + a_1 \frac{\ell^{\alpha}}{g^{\beta}}$ (recall that ℓ denotes the number of defectives and g the number of items in a test) – when $\alpha \neq \beta$ this cannot be written as a density-dependent test function. Note that all logs are base two unless otherwise specified.

5.1 POLYNOMIAL TEST FUNCTIONS

A natural class of test functions are functions of the form $f(\rho) = a_0 + a_1\rho^{\alpha}$ where $\alpha \in (0, \infty)$, $a_1, a_0 \in (0, 1]$ and $a_0 + a_1 \leq 1$ (these constraints ensure that $f(\rho) \in [0, 1]$). We consider separately in Theorems 1 and 2 below the cases corresponding to $a_0 = 0$ and $a_0 > 0$, since the results are structurally different.

Theorem 1. *Given any test function $f(\rho) = a_1\rho^{\alpha}$, where $a_1 \in (0, 1]$ and $\alpha \in (0, \infty)$,*

1a **Lower bound:** *Any non-adaptive group testing algorithm ensuring a probability error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{1,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{\alpha}, & \alpha \in (0,1) \\ \ell_{1,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases}.$$

1b **Upper bound:** *There exist non-adaptive group testing algorithms with probability of error at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{1,R}k\left(\frac{n}{k}\right)^{\alpha}\log\frac{n}{\varepsilon}, & \alpha \in (0,1) \\ u_{1,I}n\log\frac{n}{\varepsilon}, & \alpha \in [1,\infty) \end{cases}.$$

Here $\ell_{1,1}$, $\ell_{1,2}$, $u_{1,R}$ and $u_{1,I}$ are universal constants that depend only on f.

Theorem 2. *Given any test function $f(\rho) = a_0 + a_1\rho^{\alpha}$ where $a_1, a_0 \in (0, 1]$, $a_0 + a_1 \leq 1$ and $\alpha \in (0, \infty)$,*

2a **Lower bound:** *Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{2,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log n, & \alpha \in (0,\frac{1}{2}) \\ \ell_{2,2}(1-\varepsilon)n\log n, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

2b **Upper bound:** *There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{2,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}, & \alpha \in (0,\frac{1}{2}) \\ u_{2,I}n\log\frac{n}{\varepsilon}, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

Here $\ell_{2,1}$, $\ell_{2,2}$, $u_{2,R}$ and $u_{2,I}$ are universal constants that depend only on f.

We broaden our results to *test functions* in the *set of α-power series functions* $\mathcal{A}(\alpha)$,

$$\mathcal{A}(\alpha) := \left\{ f : [0,1] \to [0,1]; f(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha}, \right. \\ \left. c_f := \sum_{i=0}^{\infty} |a_i| < \infty, a_1 > 0 \right\}.$$

By analogy to the types of functions considered in Theorems 1 and 2, the analogous generalizations in Theorems 3 and 4 below consider respectively the scenario where $f(0) = 0$ and $f(0) > 0$.

Theorem 3. *Given any test function $f(\rho) \in \mathcal{A}(\alpha)$, where $f(0) = 0$ and $\alpha \in (0, \infty)$,*

3a **Lower bound:** *Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{3,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{\alpha}, & \alpha \in (0,1) \\ \ell_{3,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases}.$$

3b **Upper bound:** *There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{3,R}k\left(\frac{n}{k}\right)^{\alpha} \log\frac{n}{\varepsilon}, & \alpha \in (0,1) \\ u_{3,I}n\log\frac{n}{\varepsilon}, & \alpha \in [1,\infty) \end{cases}.$$

Here $\ell_{3,1}$, $\ell_{3,2}$, $u_{3,R}$ and $u_{3,I}$ are universal constants that depend only on f.

Theorem 4. *Given any test function $f(\rho) \in \mathcal{A}(\alpha)$, where $f(0) \in (0, 1)$ and $\alpha \in (0, \infty)$,*

4a **Lower bound:** *Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{4,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log n, & \alpha \in (0,\frac{1}{2}) \\ \ell_{4,2}(1-\varepsilon)n\log n, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

4b **Upper bound:** *There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{4,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}, & \alpha \in (0,\frac{1}{2}) \\ u_{4,I}n\log\frac{n}{\varepsilon}, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

Here $\ell_{4,1}$, $\ell_{4,2}$, $u_{4,R}$ and $u_{4,I}$ are universal constants that depend only on f.

5.2 GENERAL TEST FUNCTIONS

We extend our results to general density-dependent test functions $f(\rho) : [0, 1] \to [0, 1]$ by "sandwiching" these functions by functions in $\mathcal{A}(\alpha)$ for some $\alpha \in (0, \infty)$, obtaining lower and upper bounds for the sample complexity of f.

Theorem 5. *Given any test function $f(\rho) \in [0, 1]$ where $f(0) = 0$,*

5a **Lower bound:** *For any $\alpha \in (0, \infty)$, if there exists $d \in \mathcal{A}(\alpha)$ where $d(0) = 0$ and $f(\rho) \leq d(\rho)$ for $\rho \in (0, 1]$. Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{5,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{\alpha}, & \alpha \in (0,1) \\ \ell_{5,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases}.$$

5b **Upper bound:** *For any $\alpha \in (0, \infty)$, if there exists $h, d \in \mathcal{A}(\alpha)$ where $h(0) = d(0) = 0$ and $h(\rho) \leq f(\rho) \leq d(\rho)$ for $\rho \in \left(0, \frac{k}{n}\log\frac{n^2}{k}\right]$. There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{5,R}k\left(\frac{n}{k}\right)^{\alpha}\log\frac{n}{\varepsilon}, & \alpha \in (0,1) \\ u_{5,I}n\log\frac{n}{\varepsilon}, & \alpha \in [1,\infty) \end{cases}.$$

Here $\ell_{5,1}$, $\ell_{5,2}$, $u_{5,R}$ and $u_{5,I}$ are universal constants that depend only on f.

Theorem 6. *Given any test function $f(\rho) \in [0, 1]$ where $f(0) \in (0, 1)$,*

6a **Lower bound:** *For any $\alpha \in (0, \infty)$, if there exists $d \in \mathcal{A}(\alpha)$ such that $f(0) = d(0) \leq f(\rho) \leq d(\rho)$ for $\rho \in [0, 1]$. Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{6,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log n, & \alpha \in (0,\frac{1}{2}) \\ \ell_{6,2}(1-\varepsilon)n\log n, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

6b **Upper bound:** *For any $\alpha \in (0, \infty)$, if there exists $h \in \mathcal{A}(\alpha)$ where $f(0) = h(0)$ and $f(\rho) \geq h(\rho)$ for $\rho \in \left(0, \frac{k}{n}\log\frac{n^2}{k}\right]$. There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{6,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}, & \alpha \in (0,\frac{1}{2}) \\ u_{6,I}n\log\frac{n}{\varepsilon}, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

Here $\ell_{6,1}$, $\ell_{6,2}$, $u_{6,R}$ and $u_{6,I}$ are universal constants that depend only on f.

5.3 OTHER TEST FUNCTIONS OF INTEREST

Finally, we go beyond the class of density-dependent test functions and use our techniques to provide non-trivial sample complexity bounds for *test functions* of the form $f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ where $a_1 > 0$, $a_0 \in [0,1]$ such that $a_0 + a_1 \leq 1$ and $0 < \alpha \leq \beta$ (these constraints ensure that $f(\ell, g) \in [0,1]$),

Theorem 7. *Given any test function $f(\ell, g) = a_1 \frac{\ell^\alpha}{g^\beta}$ where $a_1 \in (0,1]$ and $0 \leq \alpha \leq \beta$,*

7a **Lower bound:** *Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{7,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^\alpha, & \alpha \in (0,1) \\ \ell_{7,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases}.$$

7b **Upper bound:** *There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{7,R}k\left(\frac{n}{k}\right)^\beta \log\frac{n}{\varepsilon}, & \beta \in (0,1) \\ u_{7,I}n\log\frac{n}{\varepsilon}, & \beta \in [1,\infty) \end{cases}.$$

Here $\ell_{7,1}$, $\ell_{7,2}$, $u_{7,R}$ and $u_{7,I}$ are universal constants that depend only on f.

Theorem 8. *Given any test function $f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ where $a_1, a_0 \in (0,1]$ and $0 < \alpha \leq \beta$,*

8a **Lower bound:** *Any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{8,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log n, & \alpha \in (0,\frac{1}{2}) \\ \ell_{8,2}(1-\varepsilon)n\log n, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

8b **Upper bound:** *There exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{8,R}k\left(\frac{n}{k}\right)^{2\beta}\log\frac{n}{\varepsilon}, & \beta \in (0,\frac{1}{2}) \\ u_{8,I}n\log\frac{n}{\varepsilon}, & \beta \in [\frac{1}{2},\infty) \end{cases}.$$

Here $\ell_{8,1}$, $\ell_{8,2}$, $u_{8,R}$ and $u_{8,I}$ are universal constants that depend only on f.

6 APPROXIMABILITY BY $\mathcal{A}(\alpha)$

In Section 6.1 below we note that due to the Stone-Weierstrass theorem (Weierstrass, 1885; Stone, 1937, 1948), for any $\alpha > 0$, the set of (fractional) power series in ρ^α are dense in the set of continuous real-valued functions on $[0,1]$, equipped with the supremum norm. In Section 6.2 we complement this with an appropriate computational method to compute a (fractional) power series for sufficiently smooth $f(\cdot)$. These motivate our focus on functions in $\mathcal{A}(\alpha)$.

6.1 STONE-WEIERSTRASS THEOREM

Theorem 9 (Stone-Weierstrass Theorem (Weierstrass, 1885; Stone, 1937, 1948))**.** *Let K be a compact metric space and $\mathcal{C}(K,\mathbb{R})$ denote the set of real-valued continuous functions on K with the topology of uniform convergence. Let $\mathcal{B} \subset \mathcal{C}(K,\mathbb{R})$ a unital sub-algebra ($\mathcal{B}$ is closed under multiplication and real linear combinations of pairs of elements in $\mathcal{B}$) which separates points of K (for every $x, y \in K$, with $x \neq y$, there exists $f \in \mathcal{B}$ such that $f(x) \neq f(y)$). Then $\mathcal{B}$ is dense in $\mathcal{C}(K,\mathbb{R})$.*

Theorem 10. *For $\alpha \in (0,\infty)$ the set $\mathcal{B}(\alpha)$ defined below is dense in $\mathcal{C}([0,1],\mathbb{R})$,*

$$\mathcal{B}(\alpha) := \left\{ \sum_{i=0}^{\infty} a_i x^{i\alpha}; |a_i| < \infty \right\}.$$

Proof. We have that for any $f, g \in \mathcal{B}(\alpha)$, $fg \in \mathcal{B}(\alpha)$ and for any $a, b \in \mathbb{R}$, $af + bg \in \mathcal{B}(\alpha)$. Hence $\mathcal{B}(\alpha)$ is a unital sub-algebra. It is also separating as for any distinct $x, y \in [0,1]$, $x^\alpha \neq y^\alpha$. Hence by Theorem 9, $\mathcal{B}(\alpha)$ is dense in $\mathcal{C}([0,1],\mathbb{R})$. ☐

6.2 FRACTIONAL TAYLOR SERIES

Using results from Katugampola (2014) we give a method to approximate "α-smooth" functions as an α power series.

Definition 6 (Fractional Derivative (Katugampola (2014)))**.** *Let $f : [0,\infty] \to \mathbb{R}$, the $\alpha \in (0,1)$ derivative is*

$$D^\alpha(f)(x) := \lim_{\varepsilon \to 0} \frac{f\left(xe^{\varepsilon x^{-\alpha}}\right) - f(x)}{\varepsilon} = f'(x)x^{1-\alpha}$$

for $x > 0$, where $D^\alpha(f)(0) := \lim_{a\to 0} D^\alpha(f)(a)$. We denote i applications of the D^α function as $D^{i\alpha}$. A function is α-smooth if for $i \in \mathbb{N}$, $|D^{i\alpha}(f)| < \infty$ for the domain of the function.

We now provide an analogous Taylor Series type expansion for fraction power series, with proof deferred to Appendix E.

Theorem 11. *For some $\alpha \in (0,1)$, an α-smooth function $f : [0,1] \to \mathbb{R}$ can be approximated around $c \in [0,1]$ as*

$$H(x) := \sum_{i=0}^{\infty} a_i (x-c)^{i\alpha}, \text{ where } a_i = \frac{D^{i\alpha}(f)(c)}{i!\alpha^i}.$$

Defining the sum truncated at n as $H_n(x) = \sum_{i=0}^{n} a_i (x-c)^{i\alpha}$ and the remainder term at n as $R_n(x) = f(x) - H_n(x)$ we have that for all $x \in [0,1]$,

$$R_n(x) = \frac{(x-c)^{(n+1)\alpha} D^{(n+1)\alpha}(f)(z)}{\alpha^{n+1}(n+1)!},$$

for some $z \in [0,1]$. Therefore, $\lim_{n\to\infty} R_n(x) \to 0$.

7 SIMULATION RESULTS

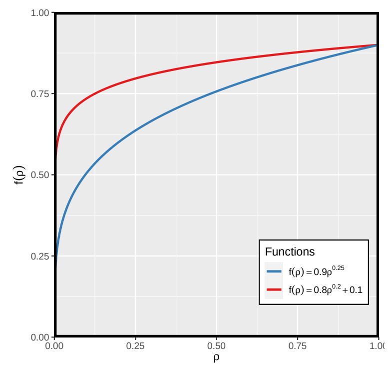
Here we present our simulation results shown in Figure 1 to evaluate the performance of our proposed schemes.

Testing Phase: For a given (n, k, T) and the test function $f(\rho) = a_0 + a_1\rho^{\alpha}$, we randomly generate a binary matrix $\mathbf{X}$ of size $T \times n$ where each entry is i.i.d. Bern$(\frac{1}{k})$. Note that each row of $\mathbf{X}$ corresponds to a distinct test and each column corresponds to a distinct item.

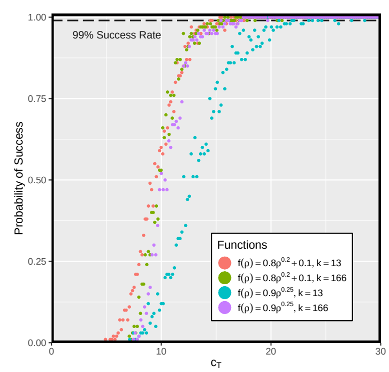
Decoding Phase: We utilize the "NCOMP Decoder". Our results are provided as the following.

Simulations: For the simulations, we analyse two functions $f(\rho) = 0.9\rho^{0.25}$ and $f(\rho) = 0.1 + 0.8\rho^{0.2}$. For $f(\rho) = 0.9\rho^{0.25}$, we choose T as given in Theorem 1, $T = c_T k \left(\frac{n}{k}\right)^{0.25} \log \frac{n}{0.01}$. For $f(\rho) = 0.1 + 0.8\rho^{0.2}$ we choose T as given in Theorem 2, $T = c_T k \left(\frac{n}{k}\right)^{0.4} \log \frac{n}{0.01}$. Here c_T is between 1 and 30 with step 29/200. We simulate this for $n = 5000$ and $k = 13$ and 166 (corresponding respectively to $\theta = 0.3$ and 0.6), repeating each simulation 100 times.

Simulation values for c_T differ significantly from the analytical bounds, suggesting room for refining our analysis. For $f(\rho) = 0.9\rho^{0.25}$, we have $u_{1,R} = \frac{\ln(2)2^{3+\alpha}}{a_1(2^\alpha-1)^2} = 220$ (Theorem 1b) and $\ell_{1,1} = \frac{1-\varepsilon}{2a_1\alpha} = 2.2$ (Theorem 1a). From Fig. 1(b), we obtain $c_T = 24$ for 99% success rate ($k = 13$), confirming $\ell_{1,1} \le c_T \le u_{1,R}$. For $f(\rho) = 0.1 + 0.8\rho^{0.2}$, we have $u_{2,R} = \frac{a_0 8\ln(2)}{a_1^2(2^\alpha-1)^2} = 39.0625$ (Theorem 2b) and $\ell_{2,1} = \frac{2a_0(1-a_0)(1-\varepsilon)}{a_1^2 \log e} = 0.64$ (Theorem 2a). From Fig. 1(b), we obtain $c_T = 14$ for 99% success rate ($k = 13$), confirming $\ell_{2,1} \le c_T \le u_{2,R}$.



(a) Test functions (3) utilized in simulations, showing the relationship between density of defectives versus probability of positive test outcomes.



(b) Number of tests (as parametrized by c_T) vs probability of successful decoding for same test functions as Figure 1(a).

Figure 1: Simulation Results

8 ACKNOWLEDGEMENTS

The authors would like to acknowledge helpful discussions with Xiwei Cheng and Daniel McMorrow, and also insightful feedback from the AISTATS meta reviewer and anonymous reviewers which helped improve the presentation in this paper. The work of S. Jaggi and that of R. Morjaria was supported in part by the Engineering and Physical Sciences Research Council (respectively grant refs: EP/Y028732/1 and EP/S023569/1). V. Gandikota's work was partially supported by NSF CCF-2217058.

References

Ahlswede, R., Deppe, C., and Lebedev, V. (2013). Threshold and majority group testing. *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, pages 488–508.

Aldridge, M., Johnson, O., and Scarlett, J. (2019). Group testing: an information theory perspective. *Foundations and Trends in Communications and Information Theory*, 15(3-4):196–392.

Arpino, G., Grometto, N., and Bandeira, A. S. (2021). Group testing in the high dilution regime. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1955–1960. IEEE.

Atia, G. K. and Saligrama, V. (2012). Boolean compressed sensing and noisy group testing. *IEEE Transactions on Information Theory*, 58(3):1880–1901.

Berger, T., Mehravari, N., Towsley, D., and Wolf, J. (1984). Random multiple-access communication and group testing. *IEEE Transactions on Communications*, 32(7):769–779.

Bui, T. V., Chee, Y. M., and Vu, V. K. (2024). Efficient designs for threshold group testing without gap. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 3005–3010. IEEE.

Bui, T. V., Kuribayashi, M., Cheraghchi, M., and Echizen, I. (2019). Efficiently decodable non-adaptive threshold group testing. *IEEE Transactions on Information Theory*, 65(9):5519–5528.

Cai, S., Jahangoshahi, M., Bakshi, M., and Jaggi, S. (2017). Efficient algorithms for noisy group testing. *IEEE Transactions on Information Theory*, 63(4):2113–2136.

Chen, H.-B. and Fu, H.-L. (2009). Nonadaptive algorithms for threshold group testing. *Discrete Applied Mathematics*, 157(7):1581–1585.

Chen, H.-B. and Hwang, F. K. (2008). A survey on nonadaptive group testing algorithms through the angle of decoding. *Journal of Combinatorial Optimization*, 15:49–59.

Cheng, X., Jaggi, S., and Zhou, Q. (2022). Generalized group testing. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 10777–10835. PMLR.

Cheng, X., Jaggi, S., and Zhou, Q. (2023). Generalized group testing. *IEEE Transactions on Information Theory*, 69(3):1413–1451.

Cheraghchi, M., Hormati, A., Karbasi, A., and Vetterli, M. (2011). Group testing with probabilistic tests: Theory, design and application. *IEEE Transactions on Information Theory*, 57(10):7057–7067.

Cheraghchi, M., Karbasi, A., Mohajer, S., and Saligrama, V. (2012). Graph-constrained group testing. *IEEE Transactions on Information Theory*, 58(1):248–262.

Clifford, R., Efremenko, K., Porat, E., and Rothschild, A. (2007). k-mismatch with don't cares. In *European Symposium on Algorithms*, pages 151–162. Springer.

Cormode, G. and Muthukrishnan, S. (2005). What's hot and what's not: Tracking most frequent items dynamically. *ACM Transactions on Database Systems (TODS)*, 30(1):249–278.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA.

Damaschke, P. (2006). Threshold group testing. In *General theory of information transfer and combinatorics*, pages 707–718. Springer.

Dorfman, R. (1943). The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440.

D'yachkov, A. G. (2003). Lectures on designing screening experiments. *Lecture Note Series 10, Combinatorial and Computational Mathematics Center, Pohang University of Science and Technology (POSTECH)*, pages 12–78.

D'yachkov, A., Rykov, V., Deppe, C., and Lebedev, V. (2013). Superimposed codes and threshold group testing. *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, pages 509–533.

Emad, A. and Milenkovic, O. (2015). Poisson group testing: A probabilistic model for Boolean compressed sensing. *IEEE Transactions on Signal Processing*, 63(16):4396–4410.

Engels, J., Coleman, B., and Shrivastava, A. (2021). Practical near neighbor search via group testing. *Advances in Neural Information Processing Systems*, 34:9950–9962.

Gerbner, D., Keszegh, B., Pálvölgyi, D., and Wiener, G. (2013). Density-based group testing. *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, pages 543–556.

Ghosh, S., Agarwal, R., Rehan, M. A., Pathak, S., Agarwal, P., Gupta, Y., Consul, S., Gupta, N., Ritika, Goenka, R., Rajwade, A., and Gopalkrishnan, M. (2021). A compressed sensing approach to pooled RT-PCR testing for covid-19 detection. *IEEE Open Journal of Signal Processing*, 2:248–264.

Guruswami, V. and Wang, H.-P. (2023). Noise-resilient group testing with order-optimal tests and fast-and-reliable decoding. *arXiv preprint arXiv:2311.08283*.

Hong, E. S. and Ladner, R. E. (2002). Group testing for image compression. *IEEE Transactions on Image Processing*, 11(8):901–911.

Hwang, F. K. (1976). Group testing with a dilution effect. *Biometrika*, 63(3):671–680.

Katugampola, U. N. (2014). A new fractional derivative with classical properties. *arXiv preprint arXiv:1410.6535*.

Luo, J. and Guo, D. (2008). Neighbor discovery in wireless ad-hoc networks based on group testing. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 791–797. IEEE.

Ma, L., He, T., Swami, A., Towsley, D., Leung, K. K., and Lowe, J. (2014). Node failure localization via network tomography. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 195–208.

Macula, A. J. (1999). Probabilistic nonadaptive group testing in the presence of errors and DNA library screening. *Annals of Combinatorics*, 3:61–69.

Malyutov, M. (2013). Search for sparse active inputs: A review. *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, pages 609–647.

Ngo, H. Q. and Du, D.-Z. (2000). A survey on combinatorial group testing algorithms with applications to DNA library screening. *Discrete mathematical problems with medical applications*, 55:171–182.

Pevzner, P. A. and Lipshutz, R. J. (1994). Towards DNA sequencing chips. In *International Symposium on Mathematical Foundations of Computer Science*, pages 143–158. Springer.

Scarlett, J. and Cevher, V. (2017). Limits on support recovery with probabilistic models: An information-theoretic framework. *IEEE Transactions on Information Theory*, 63(1):593–620.

Schliep, A., Torney, D. C., and Rahmann, S. (2003). Group testing with DNA chips: generating designs and decoding experiments. In *Computational Systems Bioinformatics. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, pages 84–91. IEEE.

Sobel, M. and Groll, P. A. (1959). Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Technical Journal*, 38(5):1179–1252.

Stinson, D. R., Van Trung, T., and Wei, R. (2000). Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *Journal of Statistical Planning and Inference*, 86(2):595–617.

Stone, M. H. (1937). Applications of the theory of Boolean rings to general topology. *Transactions of the American Mathematical Society*, 41(3):375–481.

Stone, M. H. (1948). The generalized Weierstrass approximation theorem. *Mathematics Magazine*, 21(5):237–254.

Wang, H.-P., Gabrys, R., and Vardy, A. (2023). Tropical Group Testing. *IEEE Transactions on Information Theory*, 69(9):6098–6120.

Weierstrass, K. (1885). Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 2(633-639):364.

Wolf, J. (1985). Born-again group testing: Multiaccess communications. *IEEE Transactions on Information Theory*, 31(2):185–191.

Checklist

1. For all models and algorithms presented, check if you include:
    (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] (see Sections 2 and 3)
    (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] (see Sections 4 and 5)
    (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]
2. For any theoretical claim, check if you include:
    (a) Statements of the full set of assumptions of all theoretical results. [Yes] (see Section 2)
    (b) Complete proofs of all theoretical results. [Yes] (see Appendix)
    (c) Clear explanations of any assumptions. [Yes] (see Section 2)
3. For all figures and tables that present empirical results, check if you include:
    (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [No]
    (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] (see Section 7)
    (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] (see Section 7)
    (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] (see Section 7)
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
    (a) Citations of the creator If your work uses existing assets. [No]
    (b) The license information of the assets, if applicable. [Not Applicable]
    (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
    (d) Information about consent from data providers/curators. [Not Applicable]
    (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
    (a) The full text of instructions given to participants and screenshots. [Not Applicable]
    (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
    (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A APPENDIX TABLE OF CONTENTS

This appendix is segmented into the following sections:

- In Section C, we introduce and prove necessary pre-requisites and prove the lower bound results for Theorems 1-8.
- In Section D, we introduce and prove necessary pre-requisites and prove the upper bound results for Theorems 1-8.
- In Section E we introduce necessary pre-requisites and prove Theorem 11.

B DEFINITIONS

For the reader's convenience we restate here our basic definitions, and also present definitions useful in our proofs.

We first restate the definition of the *set of α power series functions* as

$$\mathcal{A}(\alpha) := \left\{ f : [0,1] \to [0,1]; f(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha}, c_f := \sum_{i=0}^{\infty} |a_i| < \infty, a_1 > 0 \right\}, \tag{2}$$

where the physically motivated requirement for the function to be increasing at 0 requires $a_1 > 0$.

Claim 1. *For $\alpha \in (0,1)$, $f \in \mathcal{A}(\alpha)$ such that f is increasing at 0, assuming $a_1 \neq 0$ we have that $a_1 > 0$.*

Proof. Since $f(\rho) > 0$ and increasing at 0, therefore $0 \leq f(\rho) - f(0)$. Substituting $f(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha}$, we have $a_1 \rho^{\alpha} + \sum_{i=2}^{\infty} a_i \rho^{i\alpha} \geq 0, \forall \rho > 0$ which by the triangular inequality, absolute convergence of $\{a_i\}$, and $\rho < 1$ implies $a_1 \rho^{\alpha} + \sum_{i=2}^{\infty} |a_i| \rho^{2\alpha} \geq 0, \forall \rho \in [0,1]$. The previous inequality implies $a_1 \rho^{\alpha} + (c_f - a_1 - a_0)\rho^{2\alpha} > 0, \forall \rho \in [0,1]$. Finally, taking the limit $\rho \to 0^+$, the $a_1 \rho^{\alpha}$ term dominates, hence $a_1 > 0$.

□

B.1 INFORMATION-THEORETIC DEFINITIONS

In this subsection we introduce standard information-theoretic definitions that are useful for our analysis.

Definition 7. *The entropy of random variable X with distribution p_X is defined as*

$$H(X) := \mathbb{E}_X \left(\log \frac{1}{p_X(x)} \right).$$

For some $p \in (0,1)$, if $X \sim Bern(p)$ then we can write $H(X)$ as the binary entropy function $H_2(p)$,

$$H_2(p) := p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}.$$

Definition 8. *The joint entropy of random variables (X,Y) with joint distribution $p_{X,Y}$ is defined as*

$$H(X,Y) := \mathbb{E}_{X,Y} \left(\log \frac{1}{p_{X,Y}(x,y)} \right).$$

The conditional entropy is defined as

$$H(X|Y) := \mathbb{E}_{X,Y} \left(\log \frac{1}{p_{X|Y}(x|y)} \right).$$

Definition 9. *The mutual information of random variables (X,Y) with joint distribution $p_{X,Y}$ and marginals p_X and p_Y is defined as*

$$I(X;Y) := \mathbb{E}_{X,Y} \left(\log \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} \right) = H(X) - H(X|Y).$$

Definition 10 (KL Divergence for Bernoulli Random Variables)**.** *Given distributions Bern(p) and Bern(q) the Kullback-Leibler Divergence is defined as*

$$D_{KL}(p||q) = p \log \frac{p}{q} + (1-p) \log \left(\frac{1-p}{1-q} \right).$$

C ANALYSIS OF LOWER BOUNDS

In this section, we present the prerequisites and proofs of the lower bounds in Theorems 1-8.

C.1 LOWER BOUND METHODS

In this subsection, we prove the main theorems used in our lower bound analysis. Theorem 12 is a standard information-theoretic result, while Theorem 13 is on extension of Theorem 2 from Cheng et al. (2022, 2023).

C.1.1 PREREQUISITES FOR PROOFS OF LOWER BOUNDS

Definition 11. *For any test function $f\left(\frac{\ell}{g}\right) : [0,1] \to [0,1]$ we define $\mu_g(f)$ and $\sigma_g^2(f)$ as*

$$\mu_g(f) = \mathbb{E}\left[f\left(\frac{\ell}{g}\right)\right] \text{ and } \sigma_g^2(f) = \mathbb{E}\left[\left(f\left(\frac{\ell}{g}\right) - \mu_g\right)^2\right],$$

where $\ell \sim Bin(g, \frac{k}{n})$. If the test function f and the group size g are clear from the content, then we simply denote these quantities by μ and σ^2.

Lemma 1 (Data-Processing Inequality Cover and Thomas (2006))**.** *If $\mathcal{K} \leftrightarrow Y \leftrightarrow \hat{\mathcal{K}}$ form a Markov chain, then*

$$I(\mathcal{K}; \hat{\mathcal{K}}) \leq I(\mathcal{K}; Y).$$

Lemma 2 (Fano's Inequality Cover and Thomas (2006))**.** *Let $\mathcal{K} \to Y \to \hat{\mathcal{K}}$ be a Markov chain, letting $\varepsilon := \Pr\left(\mathcal{K} \neq \hat{\mathcal{K}}\right)$ then*

$$H(\mathcal{K}|\hat{\mathcal{K}}) \leq 1 + \varepsilon H(\mathcal{K}).$$

Lemma 3 (Lemma 19 from Cheng et al. (2022, 2023))**.** *For any $b \in (0,1)$ and $c \in (-b, 1-b)$, it follows that*

$$-b \ln b + (b+c)\ln(b+c) \leq c(1 + \ln b) + \frac{c^2}{b}.$$

Proof. Since (i) $b + c > 0$; (ii) $\frac{c}{b} > -1$; and (iii) $x - \ln(1+x) > 0, \forall x > -1$, therefore we have,

$$\begin{aligned} 0 &\leq (b+c)\left(\frac{c}{b} - \ln\left(1 + \frac{c}{b}\right)\right) \\ &= c + \frac{c^2}{b} - (b+c)\ln(b+c) + (b+c)\ln b. \end{aligned}$$

Rearranging we have the desired inequality. □

Lemma 4 (Cheng et al. (2022, 2023))**.** *Given any test function $f(\ell, g)$ we have that*

$$\begin{aligned} &-\mu_g \ln \mu_g - (mu_g - \mu_g)\ln(1-\mu_g) + f(\ell, G_i)\ln f(\ell, G_i) + (1 - f(\ell, G_i))\ln(1 - f(\ell, G_i)) \\ &\qquad \leq (f(\ell, G_i) - \mu_g)(\ln \mu_g - \ln(1-\mu_g)) + \frac{f(\ell, G_i) - \mu_g)^2}{\mu_g(1-\mu_g)}. \end{aligned}$$

Proof. The proof is very similar to Lemma 19 from Cheng et al. (2022, 2023).

- When $f(\ell, g) = 0$,

$$\text{L.H.S} - \text{R.H.S} = 1 - \frac{1}{1-\mu_g} \leq 0, \text{ if } \mu_g \in (0,1).$$

- When $f(\ell, g) = 1$,

$$\text{L.H.S} - \text{R.H.S} = 1 - \frac{1}{\mu_g} - \ln \mu_g \leq 0, \text{ if } \mu_g \in (0,1).$$

- When $f(\ell, g) \in (0,1)$, we first apply Lemma 3 with $b = \mu_g$ and $c = f(\ell, g) - \mu_g$ to obtain

$$-\mu_g \ln \mu_g + f(\ell, g) \ln f(\ell, g) \leq (f(\ell, g) - \mu_g)(1 + \ln \mu_g) + \frac{(f(\ell, g) - \mu_g)^2}{\mu_g}.$$

  Applying Lemma 3 again with $b = 1 - \mu_g$ and $c = \mu_g - f(\ell, g)$ we obtain,

$$-(1-\mu_g)\ln(1-\mu_g) + (1 - f(\ell, g))\ln(1 - f(\ell, g)) \leq (\mu_g - f(\ell, g))(1 + \ln(1-\mu_g)) + \frac{(f(\ell, g) - \mu_g)^2}{1-\mu_g}.$$

  Summing both sides of the inequalities we obtain the required inequality.

□

Lemma 5. *For $p \in (0, 1/2)$, the binary entropy function can be bounded as*

$$H_2(p) < 2p \log\left(\frac{1}{p}\right).$$

Proof. By the symmetry of the binary entropy function $H_2(p)$ around $p = 1/2$, we have that for $p \in (0, 1/2)$, $p \log\left(\frac{1}{p}\right) > (1-p)\log\left(\frac{1}{1-p}\right)$. Hence, we have $H_2(p) < 2p \log\left(\frac{1}{p}\right)$. □

C.1.2 PROOFS OF LOWER BOUNDS

Theorem 12. *Given any test function $f(\ell, g) : [0,1] \to [0,1]$, and any non-adaptive test design with a probability of error of at most ε, we have that*

$$T \geq \min_{0 < g \leq n} \left\{\frac{1}{H_2(\mu_g)}\right\} (1-\varepsilon) k \log\left(\frac{n}{k}\right)(1 - o(1)).$$

Proof. Let $Y \in \{0,1\}^T$ be the outcomes of the T non-adaptive tests, and $G := (G_1, G_2, \ldots, G_T) \in \mathbb{N}^T$, denote the group sizes (number of items pooled together) of each test.

We follow a standard information-theoretic approach to establish the lower bounds on the number of non-adaptive tests that are required to achieve an error probability of at most ε. Consider the Markov chain $\mathcal{K} \leftrightarrow Y \leftrightarrow \hat{\mathcal{K}}$, where $\hat{\mathcal{K}}$ is the output of the decoder with input Y with probability of error at most ε. Note that $H_2(\mathcal{K}) = nH_2(\frac{k}{n}) = k \log\left(\frac{n}{k}\right)(1 + o(1))$. Moreover, via standard information-theoretic definitions, we have

$$\begin{aligned}
H(\mathcal{K}) &= H(\mathcal{K}|\hat{\mathcal{K}}) + I(\mathcal{K}; \hat{\mathcal{K}}) \\
&\overset{(a)}{\leq} 1 + \varepsilon H(\mathcal{K}) + I(\mathcal{K}; \hat{\mathcal{K}}) \\
&\overset{(b)}{\leq} 1 + \varepsilon H(\mathcal{K}) + I(\mathcal{K}; Y) \\
&\overset{(c)}{\leq} 1 + \varepsilon H(\mathcal{K}) + \sum_{i=1}^{T} H(Y_i) \\
&\leq 1 + \varepsilon H(\mathcal{K}) + T \max_i H(Y_i),
\end{aligned}$$

where (a) follows from Fano's inequality, (b) follows from the Data Processing Inequality, (c) follows from [Lemma 8.9.2 Cover and Thomas (2006)] and the fact that $I(K_i; Y_i) \leq H(Y_i)$.

Rearranging the terms and using the fact that $H_2(\mathcal{K}) = k \log\left(\frac{n}{k}\right)(1 - o(1))$, we obtain

$$T \geq \frac{(1-\varepsilon)k \log \frac{n}{k}}{\max_i H(Y_i)}. \tag{3}$$

As $Y_i \in \{0, 1\}$ we can write this in terms of the binary entropy function $H(Y_i) = H_2(\Pr(Y_i = 1)) = H_2(\mu_g)$ for μ_g from Definition 11. □

Theorem 13. *For any test function $f(\ell, g) : [0, 1] \to [0, 1]$, for any non-adaptive test design with a probability of error ε we have that*

$$T \geq \min_{0<g\leq n} \left\{\frac{\mu_g(1-\mu_g)}{\sigma_g^2 \log e}\right\}(1-\varepsilon)k \log \frac{n}{k}. \tag{4}$$

Proof. Consider the Markov chain $(\mathcal{K}, L^{i-1}, Y^{i-1}) \leftrightarrow L_i \leftrightarrow Y_i \leftrightarrow \hat{\mathcal{K}}$, where $\hat{\mathcal{K}}$ is the output of the decoder with input Y with probability of error at most ε and L_i is the number of defectives in test i. Note that $H(\mathcal{K}) = nH_2(\frac{k}{n})(1 - o(1))$.

Via standard information-theoretic definitions we have

$$\begin{aligned} H(\mathcal{K}) &= H(\mathcal{K}|\hat{\mathcal{K}}) + I(\mathcal{K}; \hat{\mathcal{K}}) \\ &\stackrel{(a)}{\leq} 1 + \varepsilon H(\mathcal{K}) + I(\mathcal{K}; \hat{\mathcal{K}}) \\ &\stackrel{(b)}{\leq} 1 + \varepsilon H(\mathcal{K}) + I(\mathcal{K}; Y), \end{aligned} \tag{5}$$

where (a) follows from Fano's inequality (2) and (b) follows from the Data Processing Inequality.

Let $Y^{i-1} = (Y_1, \ldots, Y_{i-1})$ and $L^{i-1} = (L_1, \ldots, L_{i-1})$. From standard set of inequalities we have

$$\begin{aligned} I(\mathcal{K}; Y) &= \sum_{i=1}^{T} H(Y_i|Y^{i-1}) - H(Y_i|\mathcal{K}, Y^{i-1}) \\ &\leq \sum_{i=1}^{T} H(Y_i) - H(Y_i|\mathcal{K}, Y^{i-1}) \\ &\leq \sum_{i=1}^{T} H(Y_i) - H(Y_i|\mathcal{K}, Y^{i-1}, L_i) \\ &\leq \sum_{i=1}^{T} H(Y_i) - H(Y_i|L_i). \end{aligned} \tag{6}$$

Note that $H(Y_i) - H(Y_i|L_i) = \sum_{\ell=0}^{G_i} \Pr(L_i = \ell)\left[H(Y_i) - H(Y_i|L_i = \ell)\right]$ where $\Pr(L_i = \ell) = \binom{G_i}{\ell}\left(\frac{k}{n}\right)^{\ell}\left(1 - \frac{k}{n}\right)^{G_i - \ell}$ with $G_i \in \mathbb{N}$ is the number of items pooled in test i. Moreover, using Definition 11 we have that

$$\mu(G_i) = \sum_{\ell=0}^{G_i} \Pr(L_i = \ell) f(\ell, G_i) \text{ and } \sigma^2(G_i) = \sum_{\ell=0}^{G_i} \Pr(L_i = \ell)\left(f(\ell, G_i) - \mu(G_i)\right)^2.$$

From the definition of our test function we have that $\Pr(Y_i = 1|L_i = \ell) = f(\ell, G_i)$. Therefore, we can write the probability of a positive outcome in terms of the mean

$$\Pr(Y_i = 1) = \sum_{\ell=0}^{G_i} \Pr(L_i = \ell) P(Y_i = 1|L_i = \ell) = \mu(G_i).$$

Hence, we have

$$\begin{aligned}H(Y_i) &= -\mu(G_i)\log\mu(G_i) - (1-\mu(G_i))\log(1-\mu(G_i)) \\ &= [-\mu(G_i)\ln\mu(G_i) - (1-\mu(G_i))\ln(1-\mu(G_i))]\log e,\end{aligned}$$

and

$$\begin{aligned}H(Y_i|L_i=\ell) &= -f(\ell,G_i)\log f(\ell,G_i) - (1-f(\ell,G_i))\log(1-f(\ell,G_i)) \\ &= [-f(\ell,G_i)\ln f(\ell,G_i) - (1-f(\ell,G_i))\ln(1-f(\ell,G_i))]\log e.\end{aligned}$$

Further, from Lemma 4, we have

$$\begin{aligned}&-\mu(G_i)\ln\mu(G_i) - (1-\mu(G_i))\ln(1-\mu(G_i)) + f(\ell,G_i)\ln f(\ell,G_i) + (1-f(\ell,G_i))\ln(1-f(\ell,G_i)) \\ &\qquad\le (f(\ell,G_i)-\mu(G_i))(\ln\mu(G_i) - \ln(1-\mu(G_i))) + \frac{f(\ell,G_i)-\mu(G_i))^2}{\mu(G_i)(1-\mu(G_i))}.\end{aligned}$$

Therefore, we get

$$\begin{aligned}H(Y_i) - H(Y_i|L_i) &\le \sum_{\ell=0}^{G_i}\Pr(L_i=\ell)\left[(f(\ell,G_i)-\mu(G_i))(\ln\mu(G_i)-\ln(1-\mu(G_i))) + \frac{(f(\ell,G_i)-\mu(G_i))^2}{\mu(G_i)(1-\mu(G_i))}\right] \\ &= \frac{\sigma^2(G_i)\log e}{\mu(G_i)(1-\mu(G_i))}. \qquad (7)\end{aligned}$$

Substituting eq. (7) in eq. (6), we obtain

$$\begin{aligned}I(\mathcal{K};Y) &\le \sum_{i=1}^{T}\frac{\sigma^2(G_i)\log e}{\mu(G_i)(1-\mu(G_i))} \\ &\le T\frac{\sigma^2(G_i^*)\log e}{\mu(G_i^*)(1-\mu(G_i^*))}, \qquad (8)\end{aligned}$$

where G_i^* is the value for G_i that maximises $\frac{\sigma^2(G_i^*)\log e}{\mu(G_i^*)(1-\mu(G_i^*))}$.

Next, substituting the bound obtained in Equation (8) into Equation (5) yields

$$H(\mathcal{K}) \le 1 + \varepsilon H(\mathcal{K}) + T\frac{\sigma^2(G_i^*)\log e}{\mu(G_i^*)(1-\mu(G_i^*))}. \qquad (9)$$

Rearranging the terms we get

$$T \ge \frac{\mu(G_i^*)(1-\mu(G_i^*))}{\sigma^2(G_i^*)\log e}\left((1-\varepsilon)H(K) - 1\right).$$

As $H(K) = nH_2(\frac{k}{n}) \ge k\log\frac{n}{k}$ we get the final result as

$$T \ge \frac{\mu(G_i^*)(1-\mu(G_i^*))}{\sigma^2(G_i^*)\log e}(1-\varepsilon)k\log\frac{n}{k}.$$

□

C.2 BOUNDS ON MEANS AND VARIANCES

In this section, we state and prove Lemmas 6-9 which provide bounds on $\mu_g(f)$ and $\sigma_g^2(f)$ and are required for our sample complexity lower bounds. We begin by proving a claim used in the proofs of the lemmas.

Claim 2. *For $\beta \geq 0$ we have that for $\ell \sim Bin\left(g, \frac{k}{n}\right)$, $\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{1+\beta}\right] \leq \frac{k}{n}$, with equality when $\beta = 0$.*

Proof. Note that

$$\begin{aligned}
\mathbb{E}\left[\frac{\ell}{g}\right] &= \sum_{\ell=1}^{g} \left(\frac{\ell}{g}\right) \binom{g}{\ell} \left(\frac{k}{n}\right)^{\ell} \left(1 - \frac{k}{n}\right)^{g-\ell} \\
&= \sum_{\ell=1}^{g} \binom{g-1}{\ell-1} \left(\frac{k}{n}\right)^{\ell} \left(1 - \frac{k}{n}\right)^{g-\ell} \\
&= \left(\frac{k}{n}\right) \sum_{\ell'=0}^{g-1} \binom{g-1}{\ell'} \left(\frac{k}{n}\right)^{\ell'} \left(1 - \frac{k}{n}\right)^{g-1-\ell'} \\
&= \left(\frac{k}{n}\right) \left(\frac{k}{n} + \left(1 - \frac{k}{n}\right)\right)^{g-1} = \frac{k}{n},
\end{aligned}$$

where the last equality follows using the binomial expansion of $(x+y)^{g-1} = \sum_{i=0}^{g-1} \binom{g-1}{i} x^i y^{g-1-i}$. Since $\ell/g \leq 1$ and $\beta \geq 0$, therefore, $\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{1+\beta}\right] \leq \mathbb{E}\left(\frac{\ell}{g}\right) = \frac{k}{n}$ the claim is proven. □

We now compute bounds on the quantities from Definition 11 required for application in Theorem 12 and 13. For application in Theorem 12 we require upper bounds for $\mu_g(f)$, to apply Theorem 13 we require lower bounds on $\mu_g(f)$ but upper bounds on $\sigma_g(f)$.

Lemma 6. *Given any test function $f(\rho) = a_0 + a_1\rho^\alpha$ where $a_1 \in (0,1]$, $a_0 \in [0,1]$, $a_0 + a_1 \leq 1$ and $\alpha \in (0,\infty)$, for any group size $0 < g \leq n$, we have*

(i)(a)	$a_0 + a_1\frac{k}{n} \leq \mu_g(f) \leq a_0 + a_1\left(\frac{k}{n}\right)^{\alpha}$,	*for* $\alpha \in (0,1)$,
(i)(b)	$a_0 + a_1\left(\frac{k}{n}\right)^{\alpha} \leq \mu_g(f) \leq a_0 + a_1\frac{k}{n}$,	*for* $\alpha \in [1,\infty)$,
(ii)(a)	$\sigma_g^2(f) \leq \mathbb{E}\left[f^2\left(\frac{\ell}{g}\right)\right] \leq a_1^2\left(\frac{k}{n}\right)^{2\alpha}$,	*for* $\alpha \in \left(0, \frac{1}{2}\right)$,
(ii)(b)	$\sigma_g^2(f) \leq \mathbb{E}\left[f^2\left(\frac{\ell}{g}\right)\right] \leq a_1^2\frac{k}{n}$,	*for* $\alpha \in \left[\frac{1}{2},\infty\right)$,

where the expectation is taken over $\ell \sim Bin(g, \frac{k}{n})$.

Proof. We first bound $\mu_g(f)$ from below and above as the following.

(i)(a) For $\alpha \in (0,1)$, we use the concavity of $f(\rho)$ to apply Jensen's Inequality to obtain $\mathbb{E}\left[a_0 + a_1\left(\frac{\ell}{g}\right)^{\alpha}\right] < a_0 + a_1\left(\frac{k}{n}\right)^{\alpha}$. For the lower bound we note that since $\alpha < 1$, we have

$$\mathbb{E}\left[a_0 + a_1\left(\frac{\ell}{g}\right)^{\alpha}\right] > \mathbb{E}\left[a_0 + a_1\frac{\ell}{g}\right] = a_0 + a_1\frac{k}{n}.$$

(i)(b) For $\alpha \in [1,\infty)$, we use the convexity of $f(\rho)$ to apply Jensen's Inequality $\mathbb{E}\left[a_0 + a_1\left(\frac{\ell}{g}\right)^{\alpha}\right] \geq a_0 + a_1\left(\frac{k}{n}\right)^{\alpha}$. For the upper bound we apply Claim 2 to obtain $\mathbb{E}\left[a_0 + a_1\left(\frac{\ell}{g}\right)^{\alpha}\right] \leq a_0 + a_1\frac{k}{n}$.

To bound $\sigma_g^2(f)$ we first note that

$$\sigma_g^2(a_0 + a_1\rho^\alpha) = \sigma_g^2(a_1\rho^\alpha) \leq \mathbb{E}\left[a_1^2\left(\frac{\ell}{g}\right)^{2\alpha}\right].$$

(ii)(a) If $\alpha \in (0, \frac{1}{2})$, we use Jensen's Inequality to obtain $\mathbb{E}\left[a_1^2\left(\frac{\ell}{g}\right)^{2\alpha}\right] < a_1^2\left(\frac{k}{n}\right)^{2\alpha}$.

(ii)(b) For $\alpha \in [\frac{1}{2}, \infty)$, we apply Claim 2 to obtain $\mathbb{E}\left[a_1^2\left(\frac{\ell}{g}\right)^{2\alpha}\right] < a_1^2\left(\frac{k}{n}\right)$. □

Lemma 7. *Given any test function $f(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha} \in \mathcal{A}(\alpha)$, for $\alpha \in (0, \infty)$, for any group size $0 < g \leq n$, we have*

$$\begin{array}{lll} \textit{(i)(a)} & a_0 \leq \mu_g(f) < a_0 + (c_f - a_0)\left(\dfrac{k}{n}\right)^{\alpha}, & \textit{for } \alpha \in (0,1), \\ \textit{(i)(b)} & a_0 \leq \mu_g(f) \leq a_0 + (c_f - a_0)\dfrac{k}{n}, & \textit{for } \alpha \in [1,\infty), \\ \textit{(ii)(a)} & \sigma_g^2(f) \leq \mathbb{E}\left[f^2\left(\dfrac{\ell}{g}\right)\right] \leq c_f^2\left(\dfrac{k}{n}\right)^{2\alpha}, & \textit{for } \alpha \in \left(0,\tfrac{1}{2}\right), \\ \textit{(ii)(b)} & \sigma_g^2(f) \leq \mathbb{E}\left[f^2\left(\dfrac{\ell}{g}\right)\right] \leq c_f^2\dfrac{k}{n}, & \textit{for } \alpha \in \left[\tfrac{1}{2},\infty\right), \end{array}$$

where the expectation is taken over $\ell \sim Bin\left(g, \frac{k}{n}\right)$.

Proof. First, to bound $\mu_g(f)$, we note that

$$\begin{aligned} \mu(f) &= a_0 + \mathbb{E}\left[\sum_{i=1}^{\infty} a_i \left(\frac{\ell}{g}\right)^{i\alpha}\right] \\ &\overset{(a)}{\leq} a_0 + \sum_{i=1}^{\infty} |a_i| \mathbb{E}\left[\left(\frac{\ell}{g}\right)^{i\alpha}\right] \\ &\overset{(b)}{\leq} a_0 + \sum_{i=1}^{\infty} |a_i| \mathbb{E}\left[\left(\frac{\ell}{g}\right)^{\alpha}\right] \\ &\leq a_0 + (c_f - a_0)\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{\alpha}\right], \end{aligned}$$

where $c_f = \sum_{i=0}^{\infty} |a_i|$, (a) follows from the triangle inequality, and (b) follows by noting that $\frac{\ell}{g} \leq 1$.

(i)(a) If $\alpha \in (0,1)$ using Jensen's Inequality we have that $\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{\alpha}\right] \leq \left(\frac{k}{n}\right)^{\alpha}$.

(i)(b) If $\alpha \in [1,\infty)$, Claim 2 shows that $\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{\alpha}\right] \leq \frac{k}{n}$.

To bound $\sigma_g^2(f)$, we note that

$$\begin{aligned}
\sigma_G^2\left(\sum_{i=0}^{\infty} a_i \rho^{i\alpha}\right) &= \sigma_g^2\left(\sum_{i=1}^{\infty} a_i \rho^{i\alpha}\right) \\
&\leq \mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_i a_j \left(\frac{\ell}{g}\right)^{\alpha(i+j)}\right] \\
&\stackrel{(a)}{\leq} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_i a_j \mathbb{E}\left[\left(\frac{\ell}{g}\right)^{i\alpha}\right] \\
&\leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |a_i a_j| \mathbb{E}\left[\left(\frac{\ell}{g}\right)^{i\alpha}\right] \\
&\stackrel{(b)}{\leq} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |a_i a_j| \mathbb{E}\left[\left(\frac{\ell}{g}\right)^{2\alpha}\right] \\
&\leq c_f^2 \mathbb{E}\left[\left(\frac{\ell}{g}\right)^{2\alpha}\right],
\end{aligned}$$

where (a) and (b) follow from the fact that $\frac{\ell}{g} \leq 1$.

(ii)(a) For $\alpha \in (0, \frac{1}{2})$, using Jensen's Inequality we have $\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{2\alpha}\right] < \left(\frac{k}{n}\right)^{2\alpha}$.

(ii)(b) For $\alpha \in [\frac{1}{2}, \infty)$, we apply Claim 2 to obtain $\mathbb{E}\left[\left(\frac{\ell}{g}\right)^{2\alpha}\right] < \frac{k}{n}$. □

Lemma 8. *Given any test function $f : [0,1] \to [0,1]$, if there exists some $h(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha} \in \mathcal{A}(\alpha)$ such that $f(0) = h(0)$ and $f(0) \leq f(\rho) \leq h(\rho)$ for $\rho \in (0,1]$, for any group size $0 < g \leq n$, we have*

(i)(a)	$a_0 \leq \mu_g(f) < a_0 + (c_h - a_0)\left(\frac{k}{n}\right)^{\alpha}$,	*for* $\alpha \in (0,1)$,
(i)(b)	$a_0 \leq \mu_g(f) \leq a_0 + (c_h - a_0)\frac{k}{n}$,	*for* $\alpha \in [1,\infty)$,
(ii)(a)	$\sigma_g^2(f) \leq c_h^2 \left(\frac{k}{n}\right)^{2\alpha}$,	*for* $\alpha \in (0,\frac{1}{2})$,
(ii)(b)	$\sigma_g^2(f) \leq c_h^2 \frac{k}{n}$,	*for* $\alpha \in [\frac{1}{2},\infty)$,

where the expectation is taken over $\ell \sim Bin(g, \frac{k}{n})$.

Proof. (i)(a) and (i)(b).

Defining $\tilde{f}(\rho)$ as $f(\rho) - f(0)$ and $\tilde{h}(\rho)$ as $h(\rho) - h(0)$ by assumption we have that $\tilde{f} \geq 0$ and that $\tilde{f} \leq \tilde{h}$, therefore $\mu_g(\tilde{f}) \leq \mu_g(\tilde{h})$. Combined with Lemma 7 we get our bounds for $\mu_g(f)$.

(ii)(a) and (ii)(b).

Similarly for $\sigma_g^2(f)$ we note

$$\sigma_g^2(f) = \sigma_g^2(\tilde{f}) < \mathbb{E}\left[\tilde{f}\left(\frac{\ell}{g}\right)^2\right] \leq \mathbb{E}\left[\tilde{h}\left(\frac{\ell}{g}\right)^2\right].$$

Therefore from Lemma 7 we have that

$$\sigma_g^2(f) \leq \begin{cases} c_h^2 \left(\frac{k}{n}\right)^{2\alpha}, & \alpha \in (0, \frac{1}{2}) \\ c_h^2 \frac{k}{n}, & \alpha \in [\frac{1}{2}, \infty) \end{cases}.$$

□

Lemma 9. *Given any test function $f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ where $0 < \alpha \leq \beta < \infty$, $a_1 \in (0, 1]$, $a_0 \in [0, 1]$ and $a_0 + a_1 \leq 1$, for any group size $0 < g \leq n$, we have*

$$\text{(i)(a)} \qquad a_0 \leq \mu_g(f) < a_0 + a_1 \left(\frac{k}{n}\right)^\alpha, \qquad \text{for } \alpha \in (0, 1),$$

$$\text{(i)(b)} \qquad a_0 \leq \mu_g(f) \leq a_0 + a_1 \frac{k}{n}, \qquad \text{for } \alpha \in [1, \infty),$$

$$\text{(ii)(a)} \qquad \sigma_g^2(f) \leq a_1^2 \left(\frac{k}{n}\right)^{2\alpha}, \qquad \text{for } \alpha \in (0, \tfrac{1}{2}),$$

$$\text{(ii)(b)} \qquad \sigma_g^2(f) \leq a_1^2 \frac{k}{n}, \qquad \text{for } \alpha \in [\tfrac{1}{2}, \infty),$$

where the expectation is taken over $\ell \sim Bin(g, \frac{k}{n})$.

Proof. First we bound $\mu_g(f)$. Since $a_i \geq 0$ by assumption, therefore $\mu_g(f) \geq a_0$.
(i)(a) To bound $\mu_g(f)$ from above, we have that for $\alpha \in (0, 1)$,

$$\mu_g(f) = \mathbb{E}\left[a_0 + a_1 \frac{\ell^\alpha}{g^\beta}\right] < a_0 + a_1 \left(\frac{k}{n}\right)^\alpha g^{\alpha - \beta} < a_0 + a_1 \left(\frac{k}{n}\right)^\alpha,$$

where the last inequality follows from Jensen's Inequality.
(i)(b) For $\alpha \in [1, \infty)$,

$$\mu_g(f) = \mathbb{E}\left[a_0 + a_1 \frac{\ell^\alpha}{g^\beta}\right] \leq \mathbb{E}\left[a_0 + a_1 \frac{\ell^\alpha}{g^\alpha}\right] < a_0 + a_1 \frac{k}{n},$$

where the first inequality follows from the fact $\alpha \leq \beta$ and the last inequality follows from Claim 2.

To bound $\sigma_g^2(f)$ we note

$$\sigma_g^2(f) = \sigma_g^2\left(a_1 \frac{\ell^\alpha}{g^\beta}\right) < \mathbb{E}\left[a_1^2 \frac{\ell^{2\alpha}}{g^{2\beta}}\right].$$

(ii)(a) If $\alpha \in \left(0, \frac{1}{2}\right)$ using Jensen's Inequality we have

$$\mathbb{E}\left[a_1^2 \frac{\ell^{2\alpha}}{g^{2\beta}}\right] < a_1^2 \left(\frac{k}{n}\right)^{2\alpha} g^{2(\alpha - \beta)} \leq a_1^2 \left(\frac{k}{n}\right)^{2\alpha}.$$

(ii)(b) If $\alpha \in \left[\frac{1}{2}, 1\right)$ we have

$$\mathbb{E}\left[a_1^2 \frac{\ell^{2\alpha}}{g^{2\beta}}\right] \leq \mathbb{E}\left[a_1^2 \frac{\ell^{2\beta}}{g^{2\beta}}\right] \leq \mathbb{E}\left[a_1^2 \frac{\ell}{g}\right] \leq a_1^2 \frac{k}{n}.$$

□

C.3 PROOFS OF LOWER BOUNDS OF THEOREMS 1-8

In this section we prove the sample complexity lower bounds in Theorems 1-8.

Function Class	μ	σ^2	Theorem 12	Theorem 13
			$T \in \Omega\left(\frac{k\log n}{H_2(\mu)}\right)$	$T \in \Omega\left(\frac{\mu(1-\mu)}{\sigma^2}k\log n\right)$
$f(\rho) = a_1\rho^\alpha$ $a_1 \in (0,1]$	For $\alpha \in (0,1)$: $\mu \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$, $\Omega\left(\frac{k}{n}\right)$. For $\alpha \in [1,\infty)$: $\mu \in \mathcal{O}\left(\frac{k}{n}\right)$, $\Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$ For $\alpha \in [1,\infty)$: $T \in \Omega(n)$	For $\alpha \in (0,1/2)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{2\alpha-1} k\log n\right)$ For $\alpha \in [1/2,1)$: $T \in \Omega(k\log n)$ For $\alpha \in [1,\infty)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{1-\alpha} k\log n\right)$
$f(\rho) = a_1\rho^\alpha + a_0$, $a_1, a_0 \in (0,1]$ $a_1 + a_0 \leq 1$	For $\alpha \in (0,\infty)$: $\mu \in \theta(1)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1/2)$: $T \in \Omega(k\log n)$ For $\alpha \in [1/2, \infty)$: $T \in \Omega(k\log n)$	For $\alpha \in (0,1/2)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\log n\right)$ For $\alpha \in [1/2, \infty)$: $T \in \Omega(n\log n)$
$f(\rho) \in \mathcal{A}(\alpha)$, $f(0) = 0$	For $\alpha \in (0,1)$: $\mu \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$ For $\alpha \in [1,\infty)$: $\mu \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$ For $\alpha \in [1,\infty)$: $T \in \Omega(n)$	N/A
$f(\rho) \in \mathcal{A}(\alpha)$, $f(0) > 0$	For $\alpha \in (0,\infty)$: $\mu \in \theta(1)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1/2)$: $T \in \Omega(k\log n)$ For $\alpha \in [1/2, \infty)$: $T \in \Omega(k\log n)$	For $\alpha \in (0,1/2)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\log n\right)$ For $\alpha \in [1/2,\infty)$: $T \in \Omega(n\log n)$
$f(\rho)$ s.t. $f(0) = 0$ $f(\rho)$ is sandwiched	For $\alpha \in (0,1)$: $\mu \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$, $\Omega\left(\frac{k}{n}\right)$ For $\alpha \in [1,\infty)$: $\mu \in \mathcal{O}\left(\frac{k}{n}\right)$, $\Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$ For $\alpha \in [1,\infty)$: $T \in \Omega(n)$	For $\alpha \in (0,1/2)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{2\alpha-1} k\log n\right)$ For $\alpha \in [1/2,1)$: $T \in \Omega(k\log n)$ For $\alpha \in [1,\infty)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{1-\alpha} k\log n\right)$
$f(\rho)$ s.t. $f(0) > 0$ $f(\rho)$ is sandwiched	For $\alpha \in (0,\infty)$: $\mu \in \theta(1)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1/2)$: $T \in \Omega(k\log n)$ For $\alpha \in [1/2, \infty)$: $T \in \Omega(k\log n)$	For $\alpha \in (0,1/2)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\log n\right)$ For $\alpha \in [1/2, \infty)$: $T \in \Omega(n\log n)$
$f(\ell,g) = a_1\frac{\ell^\alpha}{g^\beta}$ $a_1 \in (0,1]$, $0 \leq \alpha \leq \beta$	For $\alpha \in (0,1)$: $\mu \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$ For $\alpha \in [1,\infty)$: $\mu \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^\alpha k\right)$ For $\alpha \in [1,\infty)$: $T \in \Omega(n)$	N/A
$f(\ell,g) = a_0 + a_1\frac{\ell^\alpha}{g^\beta}$ $a_0 + a_1 \leq 1$ $0 < \alpha \leq \beta$	For $\alpha \in (0,\infty)$: $\mu \in \theta(1)$	For $\alpha \in (0,1/2)$: $\sigma^2 \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right)$ For $\alpha \in [1/2,\infty)$: $\sigma^2 \in \mathcal{O}\left(\frac{k}{n}\right)$	For $\alpha \in (0,1/2)$: $T \in \Omega(k\log n)$ For $\alpha \in [1/2,\infty)$: $T \in \Omega(k\log n)$	For $\alpha \in (0,1/2)$: $T \in \Omega\left(\left(\frac{n}{k}\right)^{2\alpha} k\log n\right)$ For $\alpha \in [1/2,\infty)$: $T \in \Omega(n\log n)$

Table 2: Summary of Converse Results (Theorem 1 to Theorem 4). The entries in the first column distinguish between successively richer function classes of $f(\rho)$. The second and third columns respectively presents bounds obtained in Lemmas 6,7,8,9. The fourth and the fifth columns present the concomitant lower bounds obtained via Theorems 12 and 13, with text in green indicating the tighter of the two bounds, and text in red the looser bound.

C.3.1 PROOF OUTLINES

The proofs of the sample complexity lower bounds for each of Theorems 1-8 broadly follows the same structure, by first bounding the means and variances for each class of test functions and then applying either Theorem 12 or Theorem 13. To apply Theorem 12 we only require upper bounds on μ, while for Theorem 13 we require lower bounds on μ and upper bounds for σ^2. Hence, our decision on whether to use Theorem 12 or 13 depends in the first instance on our ability to analytically compute tight bounds on required quantities. If both Theorems can be applied, we then choose to use the one that offers the tighter bound.

C.3.2 THEOREM 1 LOWER BOUND

Theorem 1a. *Given any test function $f(\rho) = a_1\rho^\alpha$, where $a_1 \in (0,1]$ and $\alpha \in (0,\infty)$, any non-adaptive group testing algorithm ensuring a probability error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{1,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^\alpha, & \alpha \in (0,1) \\ \ell_{1,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases},$$

where $\ell_{1,1}$ and $\ell_{1,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g shown in Lemma 6. For $\alpha \in (0,\infty)$, we obtain the tightest bounds when we apply Theorem 12.

For $\alpha \in (0,1)$, we have that $\mu_g \leq a_1\left(\frac{k}{n}\right)^\alpha$. From Lemma 5, we have $H_2(p) < 2p\log\left(\frac{1}{p}\right)$, therefore $H_2\left(a_1\left(\frac{k}{n}\right)^\alpha\right) \leq 2a_1\alpha\left(\frac{k}{n}\right)^\alpha\log(\frac{n}{k})$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2a_1\alpha}\left((1-\varepsilon)\left(\frac{n}{k}\right)^\alpha k\right).$$

Similarly for $\alpha \in [1,\infty)$, we have that $\mu_g \leq a_1\frac{k}{n}$, from Lemma 5 we have that $H_2\left(a_1\left(\frac{k}{n}\right)\right) \leq 2a_1\left(\frac{k}{n}\right)\log(\frac{n}{k})$ for large enough n. Inserting these bounds into Theorem 12 we have that

$$T \geq \frac{1}{2a_1}\left((1-\varepsilon)n\right).$$

□

C.3.3 THEOREM 2 LOWER BOUND

Theorem 2a. *Given any test function $f(\rho) = a_0 + a_1\rho^\alpha$, where $a_1, a_0 \in (0,1]$, $a_0 + a_1 \leq 1$ and $\alpha \in (0,\infty)$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{2,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log(\frac{n}{k}), & \alpha \in (0,\frac{1}{2}) \\ \ell_{2,2}(1-\varepsilon)n\log(\frac{n}{k}), & \alpha \in [\frac{1}{2},\infty) \end{cases},$$

where $\ell_{2,1}$ and $\ell_{2,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 6. For $\alpha \in (0,\infty)$ we obtain the tightest bounds when we apply Theorem 13.
For $\alpha \in \left(0,\frac{1}{2}\right)$, we have $\sigma_g^2 < a_1^2\left(\frac{k}{n}\right)^{2\alpha}$ and we have that $a_0 < \mu_g < a_0(1+o(1))$ from Lemma 7 . Inserting into Theorem 13 we have that

$$T \geq \frac{a_0(1-a_0)}{a_1^2\log e}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log\left(\frac{n}{k}\right)(1-o(1)).$$

For $\alpha \in \left[\frac{1}{2}, \infty\right)$, we have that $\sigma_g^2 < a_1^2 \frac{k}{n}$ and $\mu_g > a_0$. Inserting this into Theorem 13 we have that

$$T \geq \frac{a_0(1-a_0)}{a_1^2 \log e}(1-\varepsilon)n \log\left(\frac{n}{k}\right)(1-o(1)).$$

□

C.3.4 THEOREM 3 LOWER BOUND

Theorem 3a. *Given any test function $f(\rho) \in \mathcal{A}(\alpha)$, where $f(0) = 0$ and $\alpha \in (0, \infty)$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{3,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{\alpha}, & \alpha \in (0,1) \\ \ell_{3,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases},$$

where $\ell_{3,1}$ and $\ell_{3,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 7. For $\alpha \in (0, \infty)$ we obtain the tightest bounds when we apply Theorem 12.
For $\alpha \in (0,1)$, we have that $\mu_g < c_f\left(\frac{k}{n}\right)^{\alpha}$, then from Lemma 5 we have that $H_2\left(c_f\left(\frac{k}{n}\right)^{\alpha}\right) \leq 2c_f\alpha\left(\frac{k}{n}\right)^{\alpha}\log(\frac{n}{k})$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2c_f\alpha}(1-\varepsilon)\left(\frac{n}{k}\right)^{\alpha} k.$$

For $\alpha \in [1, \infty)$, we have that $\mu_g < c_f\left(\frac{k}{n}\right)$. We then use that $H_2\left(c_f\left(\frac{k}{n}\right)\right) \leq 2\left(\frac{k}{n}\right)\log(\frac{n}{k})$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2c_f}(1-\varepsilon)n.$$

□

C.3.5 THEOREM 4 LOWER BOUND

Theorem 4a. *Given any test function $f(\rho) \in \mathcal{A}(\alpha)$, where $f(0) \in (0,1)$ and $\alpha \in (0, \infty)$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{4,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log(\frac{n}{k}), & \alpha \in (0,\frac{1}{2}) \\ \ell_{4,2}(1-\varepsilon)n\log(\frac{n}{k}), & \alpha \in [\frac{1}{2},\infty) \end{cases},$$

where $\ell_{4,1}$ and $\ell_{4,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 7. For $\alpha \in (0, \infty)$ we obtain the tightest bounds using Theorem 13.
For $\alpha \in \left(0, \frac{1}{2}\right)$, we have $\sigma_g^2 < c_f^2\left(\frac{k}{n}\right)^{2\alpha}$ and we have that $a_0 < \mu_g < a_0(1+o(1))$. Inserting into Theorem 13, we obtain

$$T \geq \frac{a_0(1-a_0)}{c_f^2 \log e}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log\left(\frac{n}{k}\right)(1-o(1)).$$

For $\alpha \in \left[\frac{1}{2}, \infty\right)$, we have that $\sigma_g^2 < c_f^2 \frac{k}{n}$ and $\mu_g > a_0$. Inserting it into Theorem 13 we obtain

$$T \geq \frac{a_0(1-a_0)}{c_f^2 \log e}(1-\varepsilon)n \log\left(\frac{n}{k}\right)(1-o(1)).$$

□

C.3.6 THEOREM 5 LOWER BOUND

Theorem 5a. *Given any test function $f(\rho) \in [0,1]$ where $f(0) = 0$. For any $\alpha \in (0,\infty)$, if there exists $d \in \mathcal{A}(\alpha)$ where $d(0) = 0$ and $f(\rho) \leq d(\rho)$ for $\rho \in (0,1]$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{5,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{\alpha}, & \alpha \in (0,1) \\ \ell_{5,2}(1-\varepsilon)n, & \alpha \in [1,\infty) \end{cases},$$

where $\ell_{5,1}$ and $\ell_{5,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 8. For $\alpha \in (0,\infty)$, we obtain the tightest bounds when we apply Theorem 12.
For $\alpha \in (0,1)$, we have that $\mu_g < c_d\left(\frac{k}{n}\right)^{\alpha}$, from Lemma 5, we have that $H_2\left(c_d\left(\frac{k}{n}\right)^{\alpha}\right) \leq 2c_d\alpha\left(\frac{k}{n}\right)^{\alpha}\log(\frac{n}{k})$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2c_d\alpha}(1-\varepsilon)\left(\frac{n}{k}\right)^{\alpha}k.$$

For $\alpha \in [1,\infty)$, we have that $\mu_g < c_d\left(\frac{k}{n}\right)$. We then use that $H_2\left(c_d\left(\frac{k}{n}\right)\right) \leq 2\left(\frac{k}{n}\right)\log(\frac{n}{k})$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2c_d}(1-\varepsilon)n.$$

□

C.3.7 THEOREM 6 LOWER BOUND

Theorem 6a. *Given any test function $f(\rho) \in [0,1]$ where $f(0) \in (0,1)$. For any $\alpha \in (0,\infty)$, if there exists $d \in \mathcal{A}(\alpha)$ such that $f(0) = d(0) \leq f(\rho) \leq d(\rho)$ for $\rho \in [0,1]$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{6,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log(\frac{n}{k}), & \alpha \in (0,\frac{1}{2}) \\ \ell_{6,2}(1-\varepsilon)n\log(\frac{n}{k}), & \alpha \in [\frac{1}{2},\infty) \end{cases},$$

where $\ell_{6,1}$ and $\ell_{6,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 8. For $\alpha \in (0,\infty)$, we obtain the tightest bounds using Theorem 13. Writing $d(\rho) = \sum_{i=0}^{\infty} a_i\rho^{i\alpha} \in \mathcal{A}(\alpha)$.
For $\alpha \in (0,\frac{1}{2})$, we have $\sigma_g^2 < c_d^2\left(\frac{k}{n}\right)^{2\alpha}$ and we have that $a_0(1+o(1)) > \mu_g > a_0 = f(0) \neq 0$. Inserting into the equation above we have

$$T \geq \frac{a_0(1-a_0)}{c_d^2\log e}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log\left(\frac{n}{k}\right)(1-o(1)).$$

For $\alpha \in [\frac{1}{2},\infty)$, we have that $\sigma_g^2 < c_d^2\frac{k}{n}$ and $a_0(1+o(1)) > \mu_g > a_0$. Inserting these bounds into Theorem 13 we have

$$T \geq \frac{a_0(1-a_0)}{c_d^2\log e}(1-\varepsilon)n\log\left(\frac{n}{k}\right)(1-o(1)).$$

□

C.3.8 THEOREM 7 LOWER BOUND

Theorem 7a. *Given any test function $f(\ell, g) = a_1 \frac{\ell^\alpha}{g^\beta}$ where $a_1 \in (0,1]$ and $0 < \alpha \leq \beta$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least*

$$T \geq \begin{cases} \ell_{7,1}\left((1-\varepsilon)k\left(\frac{n}{k}\right)^\alpha\right), & \alpha \in (0,1) \\ \ell_{7,2}\left((1-\varepsilon)n\right), & \alpha \in [1,\infty) \end{cases},$$

where $\ell_{7,1}$ and $\ell_{7,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 9. For $\alpha \in (0,\infty)$, we obtain the tightest bounds when we use Theorem 12.
For $\alpha \in (0,1)$, we have that $\mu_g < a_1\left(\frac{k}{n}\right)^\alpha$, from Lemma 5, we have that $H_2\left(a_1\left(\frac{k}{n}\right)^\alpha\right) \leq 2a_1\alpha\left(\frac{k}{n}\right)^\alpha \log(\frac{n}{k})$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2a_1\alpha}\left((1-\varepsilon)\left(\frac{n}{k}\right)^\alpha k\right).$$

For $\alpha \in [1,\infty)$, we have that $\mu < a_1\left(\frac{k}{n}\right)$, from Lemma 5, we have that $H_2\left(a_1\left(\frac{k}{n}\right)\right) \leq 2a_1\left(\frac{k}{n}\right)\log\left(\frac{n}{k}\right)$ for large enough n. Inserting into Theorem 12 we have that

$$T \geq \frac{1}{2a_1}\left((1-\varepsilon)n\right).$$

□

C.3.9 THEOREM 8 LOWER BOUND

Theorem 8a. *Given any test function $f(\ell, g) = a_0 + a_1\frac{\ell^\alpha}{g^\beta}$, where $a_0, a_1 \in (0,1]$ and $0 < \alpha \leq \beta$, any non-adaptive group testing algorithm that ensures a probability of error of at most ε requires a number of tests T that is at least,*

$$T \geq \begin{cases} \ell_{8,1}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log\left(\frac{n}{k}\right), & \alpha \in (0,\frac{1}{2}) \\ \ell_{8,2}(1-\varepsilon)n\log\left(\frac{n}{k}\right), & \alpha \in [\frac{1}{2},\infty) \end{cases},$$

where $\ell_{8,1}$ and $\ell_{8,2}$ are universal constants independent of f.

Proof. To prove this Theorem we use the bounds on μ_g and σ_g^2 shown in Lemma 9. For $\alpha \in (0,\infty)$, we obtain the tightest bounds using Theorem 13.
For $\alpha \in \left(0,\frac{1}{2}\right)$, we have $\sigma_g^2 < a_1^2\left(\frac{k}{n}\right)^{2\alpha}$ and we have that $a_0(1+o(1)) > \mu_g > a_0$. Inserting into Theorem 13 we have

$$T \geq \frac{a_0(1-a_0)}{a_1^2\log e}(1-\varepsilon)k\left(\frac{n}{k}\right)^{2\alpha}\log\left(\frac{n}{k}\right)(1-o(1)).$$

For $\alpha \in \left[\frac{1}{2},\infty\right)$, we have that $\sigma_g^2 < a_1^2\frac{k}{n}$ and $\mu_g > a_0$. Inserting this into Theorem 13 we have

$$T \geq \frac{a_0(1-a_0)}{a_1^2\log e}(1-\varepsilon)n\log\left(\frac{n}{k}\right)(1-o(1)).$$

□

D ANALYSIS OF UPPER BOUNDS

D.1 PREREQUISITES FOR ANALYSIS OF UPPER BOUNDS

Definition 12 (Item Test Sensitivity). *Given test function $f : [0,1] \to [0,1]$ the item test sensitivity is defined as,*

$$\Delta(f) = \frac{\mu_g^+(f) - \mu_g^-(f)}{2}.$$

For notational convenience we sometimes abbreviate this as Δ.

Definition 13 (**Insufficient Test Result Error**). *For the algorithms restated in Section 3, we obtain an error if an item* $j \in \mathcal{N}$ *is not in sufficiently many tests. This insufficient test result error is defined as* $E_j^{<} = \mathbb{1}(t_j < \delta)$. *This* δ *will be defined in our analysis. Note that for repetition testing* $E_j^{<} = 0$ *for all* $j \in \mathcal{N}$ *as there is no randomness in the test design.*

Definition 14 (**Misclassification Error**). *For both algorithms in Section 3, we define the misclassification errors for some* $j \in \mathcal{N}$ *as,*

$$E_j^{+|-} = \mathbb{1}\left[j \in \hat{\mathcal{K}} | \, (j \notin \mathcal{K}) \cap \left(E_j^{<}\right)^c\right], \tag{10}$$

$$E_j^{-|+} = \mathbb{1}\left[j \notin \hat{\mathcal{K}} | \, (j \in \mathcal{K}) \cap \left(E_j^{<}\right)^c\right]. \tag{11}$$

Here $E_j^{+|-}$ *denotes the false alarm event and* $E_j^{-|+}$ *denotes a missed detection event where both are conditioned on the complement of* $E_j^{<}$.

Definition 15 (**Group Size Deviation Error**). *When using a Bern*$(1/k)$ *random test design, we obtain a group size deviation error* E_g *when the number of items in any of the tests is not in the neighbourhood* $\left[\frac{n}{k} - 6\sqrt{\frac{n \ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n \ln n}{k}}\right]$.

D.2 CHERNOFF BOUNDS

To bound the probability of error we use concentration inequalities, specifically the Chernoff Type Bounds below.

Theorem 14 (Additive Chernoff–Hoeffding Theorem Cover and Thomas (2006)). *Suppose* $X_1, X_2, \ldots, X_n$ *are i.i.d random variables taking values is* $\{0,1\}$. *Defining* $p := \mathbb{E}[X_1]$, *for* $\Delta > 0$ *we have,*

$$\Pr\left(\frac{1}{n}\sum X_i \geq p + \Delta\right) \leq \left(\left(\frac{p}{p+\Delta}\right)^{p+\Delta}\left(\frac{1-p}{1-p-\Delta}\right)^{1-p-\Delta}\right)^n = 2^{-D_{KL}(p+\Delta||p)n},$$

$$\Pr\left(\frac{1}{n}\sum X_i \leq p - \Delta\right) \leq \left(\left(\frac{p}{p-\Delta}\right)^{p-\Delta}\left(\frac{1-p}{1-p+\Delta}\right)^{1-p+\Delta}\right)^n = 2^{-D_{KL}(p-\Delta||p)n}.$$

Theorem 15 (Multiplicative Chernoff Bound Cover and Thomas (2006)). *Suppose* $X_1, X_2, \ldots, X_n$ *are i.i.d random variables taking values is* $\{0,1\}$. *Defining* $X := \sum_{i=1}^n X_i$ *and* $\mu = \mathbb{E}[X]$, *for* $\delta > 0$ *we have,*

$$\Pr\left(X \geq (1+\delta)\mu\right) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \leq e^{-\frac{\delta^2 \mu}{2+\delta}},$$

$$\Pr\left(X \leq (1-\delta)\mu\right) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \leq e^{-\frac{\delta^2 \mu}{2}},$$

$$\Pr\left(|X-\mu| \geq \delta\mu\right) \leq 2e^{\frac{-\delta^2 \mu}{3}}.$$

D.3 BOUNDS ON KL DIVERGENCE.

We use some well-known bounds on the Kullback-Liebler Divergence (Definition 10). These will allow us to bound the probability of misclassification (Definition 14) in the next section.

Claim 3. *Given distributions Bern*$(p-\Delta)$ *and Bern*(p) *we have that* $D_{KL}\left(p-\Delta||p\right) > \frac{\Delta^2}{2p\ln(2)}$.

Proof. First we can write the KL-divergence as

$$D_{\mathrm{KL}}\left(p-\Delta||p\right)=\frac{1}{\ln 2}\left[\underbrace{(p-\Delta)\ln\left(1-\frac{\Delta}{p}\right)}_{T_1}+\underbrace{(1-p+\Delta)\ln\left(1+\frac{\Delta}{1-p}\right)}_{T_2}\right].$$

Then using Taylor series expansion of $\ln(1+x)$ we get

$$T_1=\sum\frac{\Delta^{k+1}}{kp^k}-\frac{\Delta^k}{kp^{k-1}}=-\Delta+\sum\frac{\Delta^{k+1}}{k(k+1)p^k},$$

where the second equality results from a change of variables.

Similarly for T_2,

$$T_2=\Delta+\sum(-1)^{k+1}\left(\frac{\Delta^{k+1}}{k(1-p)^k}-\frac{\Delta^{k+1}}{(k+1)(1-p)^k}\right)=\Delta+\underbrace{\sum(-1)^{k+1}\left(\frac{\Delta^{k+1}}{k(k+1)(1-p)^k}\right)}_{T_{22}}.$$

Note that $T_{22}\geq 0$ as every negative term has a larger positive term before it. Therefore,

$$D_{\mathrm{KL}}\left(p-\Delta||p\right)=\frac{1}{\ln 2}\left(T_1+T_2\right)>\frac{\Delta^2}{2p\ln 2}.$$

□

Claim 4. *Given distributions Bern*$(p+\Delta)$ *and Bern*(p) *we have that* $D_{KL}\left(p+\Delta||p\right)>\frac{2}{3}\frac{\Delta^2}{2p\ln 2}$.

Proof. First we can write the KL-divergence as

$$D_{\mathrm{KL}}\left(p+\Delta||p\right)=\frac{1}{\ln 2}\left[\underbrace{(p+\Delta)\ln\left(1+\frac{\Delta}{p}\right)}_{T_3}+\underbrace{(1-p-\Delta)\ln\left(1-\frac{\Delta}{1-p}\right)}_{T_4}\right].$$

Then using the Taylor series expansion of $\ln(1+x)$ terms we get

$$T_3=\sum(-1)^{k+1}\left(\frac{\Delta^{k+1}}{kp^k}+\frac{\Delta^k}{kp^{k-1}}\right)=\Delta+\underbrace{\sum(-1)^{k+1}\frac{\Delta^{k+1}}{k(k+1)p^k}}_{T_{32}},$$

where the second equality results from a change of variables.

Similarly for T_4,

$$T_4=-\Delta+\sum\left(\frac{\Delta^{k+1}}{k(1-p)^k}-\frac{\Delta^{k+1}}{(k+1)(1-p)^k}\right)=-\Delta+\sum\frac{\Delta^{k+1}}{k(k+1)(1-p)^k}.$$

Note that $T_{32}\geq 0$ as every negative term has a larger positive term before it. Therefore,

$$D_{\mathrm{KL}}\left(p+\Delta||p\right)=\frac{1}{\ln 2}\left(T_1+T_2\right)>\frac{1}{\ln 2}\left(\frac{\Delta^2}{2p}-\frac{\Delta^3}{6p^2}\right)=\frac{1}{\ln 2}\frac{\Delta^2}{2p}\left(1-\frac{\Delta}{3p}\right)>\frac{2}{3}\frac{\Delta^2}{2p\ln 2}.$$

□

Claim 5 (Pinkser's Inequality Cover and Thomas (2006), Lemma 11.6.1)**.** *Given distributions Bern*$(p\pm\Delta)$ *and Bern*(p) *Pinkser's Inequality states that* $D_{KL}\left(p\pm\Delta||p\right)\geq\frac{2\Delta^2}{\ln(2)}\geq\frac{\Delta^2}{\ln(2)}$.

D.4 BOUNDS ON ERROR EVENTS

Here we bound the probabilities of the error events for our algorithms. The main error event is misclassification error (Definition 14) which we can bound using lower bounds for Δ (Definition 12) and upper bounds for μ_g^+ and μ_g^- (Definition 5). For each class of test functions, bounds for Δ, μ_g^- and μ_g^+ are computed in Section D.5.

Theorem 16. *For an $i \in \mathcal{K}$, using the NCOMP Decoder, if we have that $p_i \sim Bin\left(t_i, \mu_g^+\right)$ we can write the probability of missed detection (11) as,*

$$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{KL}(\mu_g^+ - \Delta || \mu_g^+) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}} \leq 2^{-\frac{t_i \Delta^2}{\ln(2)}}.$$

Similarly if $i \notin \mathcal{K}$ we have that $p_i \sim Bin\left(t_i, \mu_g^-\right)$ then we can write the probability of false alarm (10) as,

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{KL}\left(\mu_g^- + \Delta || \mu_g^-\right) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^-}} \leq 2^{-\frac{t_i \Delta^2}{\ln(2)}}.$$

Proof. The first inequality comes directly from Theorem 14 with the final inequalities coming from Claim 3, Claim 4 and Claim 5. $\square$

Theorem 16 requires an item $i \in \mathcal{N}$ to be included in 'enough' tests to bound the probability of misclassification. Defining the minimum number of tests for a certain error ε as t_ε, we can see that for repetition testing we can guarantee each $i \in \mathcal{N}$ inclusion in a certain $t_i > t_\varepsilon$ as there is no randomness in our test design.

However for a Bern$(1/k)$ random test design, if for some $i \in \mathcal{N}$, $t_i < t_\varepsilon$ we obtain an insufficient test result error (Definition 13). We bound the probability of this event in Theorem 17.

Theorem 17. *Using a Bern$(\frac{1}{k})$ random test design an item $i \in \mathcal{N}$ partakes in $t_i \sim Bin\left(T, \frac{1}{k}\right)$ tests. For some $c_\delta, u > 0$ such that $c_\delta^2 u \geq 6\ln 2$, if we have $T \geq u\left(\frac{n}{k}\right)^\alpha k \log\frac{n}{\varepsilon}$ then,*

$$\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right) u \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2.$$

If $T \geq u\left(\frac{n}{k}\right)^{2\alpha} k \log\frac{n}{\varepsilon}$ then,

$$\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha}\right) u \left(\frac{n}{k}\right)^{2\alpha} \log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2.$$

Proof. The Multiplicative Chernoff Bound (Theorem 15) states that,

$$\Pr\left(t_i < (1-\delta)\mu\right) < e^{-\frac{\delta^2}{2+\delta}}.$$

For the first case, we have $t_i \sim \text{Bin}\left(T, \frac{1}{k}\right)$ where $T = u\left(\frac{n}{k}\right)^\alpha k \log\frac{n}{\varepsilon}$, therefore $\mathbb{E}(t_i) = u\left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}$ inserting into Theorem 15,

$$\begin{aligned}\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right) u\left(\frac{n}{k}\right)^\alpha \log\frac{3n}{\varepsilon}\right) &< e^{-\frac{c_\delta^2 u}{2 + c_\delta\left(\frac{k}{n}\right)^{\alpha/2}}\log\frac{n}{\varepsilon}} < e^{-\frac{c_\delta^2 u}{3}\log\frac{n}{\varepsilon}} \\ &= e^{-2\ln(2)\log\frac{n}{\varepsilon}} = \left(\frac{\varepsilon}{n}\right)^2.\end{aligned}$$

For the second case, we have $t_i \sim \text{Bin}(T, 1/k)$ where $t = u\left(\frac{n}{k}\right)^{2\alpha} k \log \frac{n}{\varepsilon}$ therefore $\mathbb{E}(t_i) = u\left(\frac{n}{k}\right)^{2\alpha} \log \frac{n}{\varepsilon}$,

$$\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^\alpha\right) u \left(\frac{n}{k}\right)^{2\alpha} \log \frac{n}{\varepsilon}\right) < e^{-\frac{c_\delta^2 u}{2 + c_\delta\left(\frac{k}{n}\right)^\alpha} \log \frac{n}{\varepsilon}} < e^{-\frac{c_\delta^2 u}{3} \log \frac{n}{\varepsilon}}$$
$$= e^{-2\ln(2)\log\frac{n}{\varepsilon}} = \left(\frac{\varepsilon}{n}\right)^2.$$

□

When applying the Bern$\left(\frac{1}{k}\right)$ random test design many of our lemmas require that each group size is in an interval $\left[\frac{n}{k} - 6\sqrt{\frac{n\ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n\ln n}{k}}\right]$. In the following theorem we show that the probability of the group size deviation (15) is minimal.

Theorem 18. *Using a Bern$\left(\frac{1}{k}\right)$ random test design, if $|\mathcal{T}| \in o(n^3)$, the probability of a group size deviation error (Definition 15) occurring is bounded from above by $\frac{1}{n^9}$.*

Proof. For each test $t \in \mathcal{T}$ we define G_t as the number of participants in a test t. For a test designed *Bern* $\left(\frac{1}{k}\right)$ we have that $G_t \sim \text{Bin}\left(n, \frac{1}{k}\right)$. To bound the probability of group size deviation we use Theorem 15,

$$\Pr\left(|X - \mu| \geq \delta\mu\right) \leq 2e^{\frac{-\delta^2\mu}{3}}.$$

Defining δ as $6\sqrt{\frac{k\log n}{n}}$ and inserting into Theorem 15 we get

$$\bigcup_{t\in\mathcal{T}} \Pr\left(\left|G_t - \frac{n}{k}\right| \geq 6\sqrt{\frac{n\log n}{k}}\right) \leq \frac{2}{n^{12}} \leq \frac{1}{n^9},$$

where the final inequality comes from the fact that $|\mathcal{T}| \in o(n^3)$. □

D.5 BOUNDS ON Δ, μ_g^- AND μ_g^+

We now provide bounds on the quantities μ_g^-, μ_g^+ and Δ as defined in Definition 5 and Definition 12. This allows us to bound the misclassification error (Definition 14) using Theorem 16.

D.5.1 REPETITION TESTING

Using repetition testing (algorithm 1) we have that $g = 1$ which leads to the first bounds. This also shows that the NCOMP Decoder is just a specific case of the Midpoint Decoder.

Lemma 10. *For a test function $f(\rho)$ where $f(0) < f(1)$, using repetition testing we have that,*

$$\mu_g^- = f(0),\ \mu_g^+ = f(1),\ \textit{and}\ \Delta = \frac{f(1) - f(0)}{2}.$$

Proof. We have $\mu_g^- = \mathbb{E}\left[f\left(\frac{\ell}{g}\right)\right]$ and $\mu_g^+ = \mathbb{E}\left[f\left(\frac{\ell+1}{g}\right)\right]$ where $\ell \sim \text{Bin}\left(g-1, \frac{k}{n}\right)$. As $g = 1$ we have trivially that $\mu_g^- = f(0)$ and $\mu_g^+ = f(1)$. By definition of Δ we have that $\Delta = \frac{f(1)-f(0)}{2}$. □

D.5.2 FUNCTIONS IN $\mathcal{A}(\alpha)$

We now provide bounds for test functions in the large class of α power series test functions in $\mathcal{A}(\alpha)$ – recall that this was defined in (2) as,

$$\mathcal{A}(\alpha) := \left\{f : [0,1] \to [0,1]; f(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha}, c_f := \sum_{i=0}^{\infty} |a_i| < \infty, a_1 > 0\right\}.$$

In Lemma 11 below we first provide bounds for the special case of test functions of the form $a_0 + a_1\rho^\alpha$, before generalizing these bounds to all test functions in $\mathcal{A}(\alpha)$.

Lemma 11. *Given any test function $f(\rho) \in \mathcal{A}(\alpha)$ for $\alpha \in (0,1)$, if $g \in \left[\frac{n}{k} - 6\sqrt{\frac{n\ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n\ln n}{k}}\right]$, we have the following bounds for μ_g^-, μ_g^+ and Δ,*

$$\mu_g^- < \mu_g^+ < a_0 + a_1 \left(\frac{2k}{n}\right)^\alpha (1 + o(1)),$$

and

$$\Delta = \frac{\mu_g^+ - \mu_g^-}{2} > \frac{a_1(2^\alpha - 1)}{2} \left(\frac{k}{n}\right)^\alpha (1 + o(1)).$$

Proof. Noting that $\ell \sim \mathrm{Bin}\left(g-1, \frac{k}{n}\right)$ we can write μ_g^+ as,

$$\begin{aligned} \mu_g^+ &= \mathbb{E}\left[\sum_{i=0}^{\infty} a_i \left(\frac{\ell+1}{g}\right)^{i\alpha}\right] \\ &= a_0 + \mathbb{E}\left[a_1\left(\frac{\ell+1}{g}\right)^{\alpha}\right] + \mathbb{E}\left[\sum_{i=2}^{\infty} a_i \left(\frac{\ell+1}{g}\right)^{i\alpha}\right]. \end{aligned} \tag{12}$$

Note that there is no third term (12) for $f(\rho) = a_0 + a_1\rho^\alpha \in \mathcal{A}(\alpha)$.

Defining $\delta_g := 6\sqrt{\frac{n\ln n}{k}}$ we bound the second term of (12) of μ_g^+ using Jensen's Inequality as $\alpha \in (0,1)$,

$$\begin{aligned} \mathbb{E}\left[a_1\left(\frac{\ell+1}{g}\right)^{\alpha}\right] < a_1\left(\frac{\mathbb{E}[\ell]+1}{g}\right)^{\alpha} &= a_1\left(\frac{k}{n} - \frac{k}{ng} + \frac{1}{g}\right)^{\alpha} \\ &< a_1\left(\frac{k}{n} + \frac{k}{n}\left(1 \pm \frac{1}{\delta_g}\right)\right)^{\alpha} = a_1 2^\alpha \left(\frac{k}{n}\right)^\alpha \left(1 \pm \mathcal{O}\left(\sqrt{\frac{k}{n\ln n}}\right)\right). \end{aligned}$$

To bound the third term (12) we follow a standard set of inequalities using that $\left(\frac{\ell+1}{g}\right) \in (0,1]$,

$$\left|\mathbb{E}\left[\sum_{i=2}^{\infty} a_i \left(\frac{\ell+1}{g}\right)^{i\alpha}\right]\right| < \mathbb{E}\left[\sum_{i=2}^{\infty} |a_i| \left(\frac{\ell+1}{g}\right)^{2\alpha}\right] < \mathbb{E}\left[\left(\frac{\ell+1}{g}\right)^{2\alpha}\right]\sum_{i=2}^{\infty}|a_i|,$$

where $\sum_{i=2}^{\infty}|a_i|$ is finite as $\sum_{i=0}^{\infty}|a_i| \leq \infty$.

Defining $\alpha_\varepsilon = \min\{2\alpha, 1\}$, if $\alpha_\varepsilon = 2\alpha$ then $\alpha \in \left(0, \frac{1}{2}\right)$ and we can apply Jensen's Inequality,

$$\mathbb{E}\left[\left(\frac{\ell+1}{g}\right)^{2\alpha}\right] < \left(\frac{\mathbb{E}[\ell]+1}{g}\right)^{2\alpha} = \left(\frac{\left(\frac{k}{n}\right)(g-1)+1}{g}\right)^{2\alpha} = \left(\frac{k}{n} - \frac{k}{ng} + \frac{1}{g}\right)^{2\alpha} \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{2\alpha}\right).$$

If $\alpha_\varepsilon = 1$ then $\alpha \in \left[\frac{1}{2}, 1\right)$ and we have that,

$$\mathbb{E}\left[\left(\frac{\ell+1}{g}\right)^{2\alpha}\right] \leq \mathbb{E}\left[\frac{\ell+1}{g}\right] = \frac{\left(\frac{k}{n}\right)(g-1)+1}{g} = \frac{k}{n} - \frac{k}{ng} + \frac{1}{g} \in \mathcal{O}\left(\frac{k}{n}\right).$$

Therefore for any test function $f(\rho) \in \mathcal{A}(\alpha)$, with $\alpha \in (0,1)$,

$$\mu_g^- < \mu_g^+ \leq a_0 + a_1 2^\alpha \left(\frac{k}{n}\right)^\alpha (1 + o(1)).$$

To bound Δ for any test function $f(\rho) \in \mathcal{A}(\alpha)$ where $\alpha \in (0,1)$ we first note that $2\Delta = \mu_g^+ - \mu_g^-$, which we can split into two terms,

$$\mu_g^+ - \mu_g^- = \mathbb{E}\left[a_1 \left(\frac{1}{g}\right)^{\alpha} ((\ell+1)^{\alpha} - \ell^{\alpha})\right] + \mathbb{E}\left[\sum_{i=2}^{\infty} a_i \left(\frac{1}{g}\right)^{i\alpha} ((\ell+1)^{i\alpha} - \ell^{i\alpha})\right]. \tag{13}$$

Note that for test functions of the form $f(\rho) = a_0 + a_1\rho^{\alpha} \in \mathcal{A}(\alpha)$ for $\alpha \in (0,1)$, the second term (13) is zero.

To bound the first term on the RHS of (13), we use that $(\ell+1)^{\alpha} - \ell^{\alpha}$ is convex in ℓ for $\alpha \in (0,1)$, hence applying Jensen's Inequality we have that

$$\begin{aligned}
a_1 \mathbb{E}\left[\left(\frac{\ell+1}{g}\right)^{\alpha} - \left(\frac{\ell}{g}\right)^{\alpha}\right] &> a_1 \left(\frac{\mathbb{E}[\ell]+1}{g}\right)^{\alpha} - a_1 \left(\frac{\mathbb{E}[\ell]}{g}\right)^{\alpha} \\
&= a_1 \left(\frac{k}{n} - \frac{k}{ng} + \frac{1}{g}\right)^{\alpha} - a_1 \left(\frac{k}{n} - \frac{k}{ng}\right)^{\alpha} \\
&= a_1 \left(\frac{k}{n} - \frac{k}{n\left(\frac{n}{k} \pm \delta_g\right)} + \frac{k}{n}\left(1 \pm \frac{1}{\delta_g}\right)\right)^{\alpha} - a_1 \left(\frac{k}{n} - \frac{k}{n\left(\frac{n}{k} \pm \delta_g\right)}\right)^{\alpha} \\
&= a_1 \left(2^{\alpha}\left(1 + \mathcal{O}\left(\sqrt{\frac{k}{n \ln n}}\right)\right) - 1\right)\left(\frac{k}{n}\right)^{\alpha} \\
&= a_1 \left(2^{\alpha} - 1\right)\left(\frac{k}{n}\right)^{\alpha}\left(1 \pm \mathcal{O}\left(\sqrt{\frac{k}{n \ln n}}\right)\right).
\end{aligned}$$

To bound the second term on the RHS of (13),

$$\mathbb{E}\left[\sum_{i=2}^{\infty} a_i \left(\frac{1}{g}\right)^{i\alpha} ((\ell+1)^{i\alpha} - \ell^{i\alpha})\right] < \mathbb{E}\left[\sum_{i=2}^{\infty} |a_i| \left(\frac{1}{g}\right)^{i\alpha} (\ell+1)^{i\alpha}\right] \in \mathcal{O}\left(\left(\frac{k}{n}\right)^{\alpha_{\varepsilon}}\right),$$

which we bounded previously in this theorem. Therefore for any test function $f(\rho) \in \mathcal{A}(\alpha)$, with $\alpha \in (0,1)$,

$$\Delta > \frac{a_1}{2}\left(2^{\alpha} - 1\right)\left(\frac{k}{n}\right)^{\alpha}(1 + o(1)).$$

□

D.5.3 BOUNDED FUNCTIONS

In this section we use a version Poisson Limit Theorem to compute the bounds for test functions $f(\rho)$ bounded by functions in $\mathcal{A}(\alpha)$.

Theorem 19 (Poisson Limit Theorem)**.** *Let p_g be a sequence of real numbers in $[0,1]$ such that the gp_g converges to a finite limit λ then*

$$\lim_{g\to\infty} \binom{g}{\ell} p_g^{\ell}(1-p_g)^{g-\ell} = e^{-\lambda}\frac{\lambda^{\ell}}{\ell!}.$$

For our purposes we need a version of the theorem with finer-grained understanding of the rate of convergence, but only for certain parameter regimes.

Claim 6. *Let $p_g = \frac{k}{n}$ and $g \in \left[\frac{n}{k} - 6\sqrt{\frac{n \ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n \ln n}{k}}\right]$, if $\ell < \ln \frac{n^2}{k} + 1$, we have that*

$$\binom{g}{\ell} p_g^{\ell}(1-p_g)^{g-\ell} = \frac{e^{-1}}{\ell!}\left(1 + \mathcal{O}\left(\frac{k \ln^2 n}{n}\right)\right).$$

Proof. Paralleling the proof for the Poisson Limit Theorem,

$$\binom{g}{\ell} p_g^\ell (1-p_g)^{g-\ell} = \frac{\sqrt{2\pi g}\left(\frac{g}{e}\right)^g}{\sqrt{2\pi(g-\ell)}\left(\frac{g-\ell}{e}\right)^{g-\ell}\ell!} p_g^\ell(1-p_g)^{g-\ell}\left(1+\mathcal{O}\left(\frac{1}{g-\ell}\right)\right) \tag{14}$$

$$= \sqrt{\frac{g}{g-\ell}}\frac{g^g e^{-l}}{(g-\ell)^{g-\ell}\ell!} p_g^\ell(1-p_g)^{g-\ell}\left(1+\mathcal{O}\left(\frac{1}{g-\ell}\right)\right) \tag{15}$$

$$= \frac{\left(1-\frac{1}{g}\right)^{g-\ell} e^{-\ell}}{\left(1-\frac{\ell}{g}\right)^{g-\ell}\ell!}\left(1+\mathcal{O}\left(\frac{\ell}{g}\right)\right). \tag{16}$$

Here (14) follows via Stirling's approximation, and (16) by a Taylor series approximation. By taking the ratio of $\frac{(1-\frac{1}{g})^{g-\ell}e^{-\ell}}{(1-\frac{\ell}{g})^{g-\ell}\ell!}$ (16) with $\frac{e^{-1}}{\ell!}$ and taking natural logarithms we compute the error term,

$$(g-\ell)\ln\left(1-\frac{1}{g}\right) - (g-\ell)\ln\left(1-\frac{\ell}{g}\right) - \ell + 1 \tag{17}$$

$$= (g-\ell)\left[\frac{-1}{g} - \frac{1}{2g^2} + \mathcal{O}\left(\frac{1}{g^3}\right)\right] - (g-\ell)\left[\frac{-\ell}{g} - \frac{1}{2}\left(\frac{\ell}{g}\right)^2 + \mathcal{O}\left(\frac{\ell^3}{g^3}\right)\right] - \ell + 1 \tag{18}$$

$$= \frac{\ell - 1/2}{g} - \frac{3}{2}\frac{\ell^2}{g} + \mathcal{O}\left(\frac{\ell^3}{g^2}\right), \tag{19}$$

where (18) comes from taking a Taylor Series expansion. Hence the overall error term is $e^{\frac{k}{n}\left(\ell-\ell^2+\mathcal{O}\left(\frac{k\ell^3}{n}\right)\right)}\left(1+\mathcal{O}\left(\frac{\ell}{g}\right)\right)$. Taking a Taylor Series expansion we can see that $1+\mathcal{O}(\frac{k\ell^2}{n})$ is the dominant error term. Using the fact that $\ell < \ln\frac{n^2}{k}+1$ we have shown the result. ∎

Claim 7. *For some test function $f(\rho)$, if there exists $d \in \mathcal{A}(\alpha)$ where $f \leq d$ for $\rho \in \left(0, \frac{k}{n}\ln\frac{n^2}{k}\right]$, then for $g \in \left[\frac{n}{k} - 6\sqrt{\frac{n\ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n \ln n}{k}}\right]$, we have that $\mu_g^+(f) < \mu_g^+(d)$ and $\mu_g^-(f) < \mu_g^-(d)$.*

Proof. From Theorem 15 we know that $\Pr\left(\ell > \ln\frac{n^2}{k}+1\right) \leq \frac{k}{n^2}$,

$$\begin{aligned}
\mu_g^+(f) &= \sum_{\ell=0}^{g-1}\binom{g-1}{\ell}\left(\frac{k}{n}\right)^\ell\left(1-\frac{k}{n}\right)^{g-1-\ell} f\left(\frac{\ell+1}{g}\right)\\
&= \sum_{\ell=0}^{\log\frac{n^2}{k}}\binom{g-1}{\ell}\left(\frac{k}{n}\right)^\ell\left(1-\frac{k}{n}\right)^{g-1-\ell} f\left(\frac{\ell+1}{g}\right) + \frac{1}{n}\\
&< \sum_{\ell=0}^{\log\frac{n^2}{k}}\binom{g-1}{\ell}\left(\frac{k}{n}\right)^\ell\left(1-\frac{k}{n}\right)^{g-1-\ell} d\left(\frac{\ell+1}{g}\right) + \frac{1}{n} = \mu_g^+(d),
\end{aligned}$$

$$\begin{aligned}
\mu_g^-(f) &= \sum_{\ell=0}^{g-1}\binom{g-1}{\ell}\left(\frac{k}{n}\right)^\ell\left(1-\frac{k}{n}\right)^{g-1-\ell} f\left(\frac{\ell}{g}\right)\\
&= \sum_{\ell=0}^{\log\frac{n^2}{k}}\binom{g-1}{\ell}\left(\frac{k}{n}\right)^\ell\left(1-\frac{k}{n}\right)^{g-1-\ell} f\left(\frac{\ell}{g}\right) + \frac{1}{n}\\
&< \sum_{\ell=0}^{\log\frac{n^2}{k}}\binom{g-1}{\ell}\left(\frac{k}{n}\right)^\ell\left(1-\frac{k}{n}\right)^{g-1-\ell} d\left(\frac{\ell}{g}\right) + \frac{1}{n} = \mu_g^-(d).
\end{aligned}$$

$\square$

Claim 8. *Given any test function $f(\rho)$. If there exists $h \in \mathcal{A}(\alpha)$ such that $h(0) = f(0)$ and $f \geq h$ for $\rho \in \left(0, \frac{k}{n} \ln \frac{n^2}{k}\right]$. If $g \in \left[\frac{n}{k} - 6\sqrt{\frac{n \ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n \ln n}{k}}\right]$, we have that $\Delta(h) < \Delta(f)$.*

Proof. From Theorem 15 we know that $\Pr\left(\ell > \ln \frac{n^2}{k} + 1\right) \leq \frac{k}{n^2}$, combining with Corollary 6 we can write,

$$\begin{aligned}
\mu_g^+(f) - \mu_g^-(f) &= \sum_{\ell=0}^{g-1} \binom{g-1}{\ell} \left(\frac{k}{n}\right)^{\ell} \left(1 - \frac{k}{n}\right)^{g-1-\ell} \left[f\left(\frac{\ell+1}{g}\right) - f\left(\frac{\ell}{g}\right)\right] \\
&> \left[1 + \mathcal{O}\left(\frac{k \ln^2 n}{n}\right)\right] e^{-1} \sum_{\ell=0}^{\ln \frac{n^2}{k}} \frac{\left(f\left(\frac{\ell+1}{g}\right) - f\left(\frac{\ell}{g}\right)\right)}{\ell!} \\
&= \left[1 + \mathcal{O}\left(\frac{k \ln^2 n}{n}\right)\right] \left(-e^{-1} f(0) + e^{-1} \sum_{i=0}^{\ln \frac{n^2}{k}} \frac{i f(i/g)}{(i+1)!}\right) \\
&> \left[1 + \mathcal{O}\left(\frac{k \ln^2 n}{n}\right)\right] \left(-e^{-1} h(0) + e^{-1} \sum_{i=0}^{\ln \frac{n^2}{k}} \frac{i h(i/g)}{(i+1)!}\right) \\
&= \mu_g^+(h) - \mu_g^-(h),
\end{aligned}$$

where the second equality comes from Claim 6 and the Chernoff Bounds (Section D.2), and the third equality comes from rewriting the sum as a telescopic sum. Hence we have $\Delta(f) > \Delta(h)$. $\square$

D.5.4 SOME OTHER INTERESTING TEST FUNCTIONS

We now provide the bounds for test functions of the form $f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ where $\alpha \in (0,1)$, $\alpha \leq \beta$, $a_1 \in (0,1]$, $a_0 \in [0,1]$ and $a_0 + a_1 \leq 1$.

Lemma 12. *Given any test function $f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ where $\alpha \in (0,1)$, $\alpha \leq \beta$, $a_1 \in (0,1]$, $a_0 \in [0,1]$ and $a_0 + a_1 \leq 1$, if $g \in \left[\frac{n}{k} - 6\sqrt{\frac{n \ln n}{k}}, \frac{n}{k} + 6\sqrt{\frac{n \ln n}{k}}\right]$, we have the following bounds for μ_g^+, μ_g^- and Δ*

$$\begin{aligned}
\mu_g^- < \mu_g^+ &< a_0 + a_1 2^\alpha \left(\frac{k}{n}\right)^\beta (1 + o(1)), \\
\Delta &> \frac{a_1}{2} (2^\alpha - 1) \left(\frac{k}{n}\right)^\beta (1 + o(1)).
\end{aligned}$$

Proof. We can write μ_g^+ and Δ as an expectation with $\ell \sim \text{Bin}\big(g-1, \frac{k}{n}\big)$, combining with the fact that $\mathbb{E}[\ell] = 1 + \mathcal{O}\left(\sqrt{\frac{k}{n \ln n}}\right)$,

$$\mu_g^- < \mu_g^+ = a_0 + \frac{a_1}{g^\beta} \mathbb{E}\left[(\ell+1)^\alpha\right] < a_0 + a_1 2^\alpha \left(\frac{k}{n}\right)^\beta (1 + o(1)),$$

$$\Delta = \frac{a_1}{2g^\beta} \mathbb{E}\left[(\ell+1)^\alpha - \ell^\alpha\right] > \frac{a_1}{2} (2^\alpha - 1) \left(\frac{k}{n}\right)^\beta (1 + o(1)).$$

Where the final inequalities come from applying Jensen's as $(\ell+1)^\alpha$ is concave and $(\ell+1)^\alpha - \ell^\alpha$ is convex for $\alpha \in (0,1)$. $\square$

D.6 PROOFS OF UPPER BOUNDS OF THEOREMS 1-8

Function Class	μ_g^+	Δ	Random Design Achievability	Repetition Test Achievability
			$T \in \mathcal{O}\left(\frac{\mu_g^+}{\Delta^2} k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\rho) = a_1\rho^\alpha$ $a_1 \in (0,1]$	For $\alpha \in (0,1)$: $\mu_g^+ \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1)$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1)$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\rho) = a_1\rho^\alpha + a_0$, $a_1, a_0 \in (0,1]$ $a_1 + a_0 \leq 1$	For $\alpha \in (0,1/2)$: $\mu_g^+ \in \mathcal{O}(1)$	For $\alpha \in (0,1/2)$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1/2)$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\rho) \in \mathcal{A}(\alpha)$, $f(0) = 0$	For $\alpha \in (0,1)$: $\mu_g^+ \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1)$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1)$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\rho) \in \mathcal{A}(\alpha)$, $f(0) > 0$	For $\alpha \in (0,1/2)$: $\mu_g^+ \in \mathcal{O}(1)$	For $\alpha \in (0,1/2)$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1/2)$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\rho)$ s.t. $f(0) = 0$ $f(\rho)$ is sandwiched	For $\alpha \in (0,1)$: $\mu_g^+ \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1)$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\alpha\right)$	For $\alpha \in (0,1)$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^\alpha k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\rho)$ s.t. $f(0) > 0$ $f(\rho)$ is sandwiched	For $\alpha \in (0,1/2)$: $\mu_g^+ \in \mathcal{O}(1)$	For $\alpha \in (0,1/2)$: $\mu_g^+ \in \mathcal{O}(1)$	For $\alpha \in (0,1/2)$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^{2\alpha} k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\ell, g) = a_1 \frac{\ell^\alpha}{g^\beta}$ $a_1 \in (0,1]$, $0 \leq \alpha \leq \beta$	For $0 < \alpha \leq \beta \leq 1$: $\mu_g^+ \in \mathcal{O}\left(\left(\frac{k}{n}\right)^\beta\right)$	For $0 < \alpha \leq \beta \leq 1$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\beta\right)$	For $0 < \alpha \leq \beta \leq 1$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^\beta k \log n\right)$	$T \in \mathcal{O}(n \log n)$
$f(\ell, g) = a_0 + a_1 \frac{\ell^\alpha}{g^\beta}$ $a_0 + a_1 \leq 1$ $0 < \alpha \leq \beta$	For $0 < \alpha \leq \beta \leq 1/2$: $\mu_g^+ \in \mathcal{O}(1)$	For $0 < \alpha \leq \beta \leq 1/2$: $\Delta \in \Omega\left(\left(\frac{k}{n}\right)^\beta\right)$	For $0 < \alpha \leq \beta \leq 1/2$: $T \in \mathcal{O}\left(\left(\frac{n}{k}\right)^{2\beta} k \log n\right)$	$T \in \mathcal{O}(n \log n)$

Table 3: Summary of Achievability Results. Column 2 and Column 3 list the bounds on "item-included test-positivity probability" and the "item-test sensitivity" that are used to compute the bounds on the number of tests sufficient to recover $\mathcal{K}$ using the random-test design (Algorithm 2). This provides a near-optimal bound in the regimes of α (and β) as mentioned in Column 4. For α (and β) values beyond the range, we use the repetition testing (Algorithm 1) whose upper bounds are listed in Column 5.

D.6.1 PROOF OUTLINES FOR UPPER BOUNDS

Our proofs for the upper bounds for Theorem 1-8 are obtained by applying the algorithms in Section 3. Depending on the 'shape' of the test function we apply either repetition testing (1) or a Bern$(1/k)$ random test design (2).

When the optimal test design is repetition testing the proof follows from Theorem 0 [Repetition Testing Upper Bound] - in which we bound the only error event, as we have no randomness in our test design, the misclassification error using Pinsker's Inequality (5).

However, when the optimal test design is a Bern$(1/k)$ random test design we have to bound three possible error events:

1. As many of our pre-requisites (bounds on μ_g^-, μ_g^+ (5) and Δ (12)) require the number of items in each test to be in a neighbourhood, we bound the probability that there exists any group test outside that

neighbourhood. This is shown in Theorem 18, where we show that the probability of group size deviation error (15) is 'significantly' vanishing.

2. Another possible error event, due to the random test design, is the insufficient test result error (13). Where an item isn't included in enough tests for us to 'sufficiently' bound the misclassification error. In Theorem 17 we show that the probability of this event is also 'significantly' vanishing.

3. The final (and dominant) error event is the misclassification error (14), we bound the probability of this event from above, using our bounds for μ_g^-, μ_g^+ and Δ that we computed in Section D.4, completing the analysis.

D.6.2 REPETITION TESTING

Our upper bound for repetition testing holds for any 'reasonable' test function $f(\rho)$ we include it as a theorem.

Theorem 0 [Repetition Testing Upper Bound]. *Given any test function $f(\rho)$ where $f(0) < f(1)$, using repetition testing with the Midpoint Decoder we have an error probability bounded above by ε when the total number of tests T is at most*

$$T \leq u_I n \log \frac{n}{\varepsilon}.$$

For any $u_I > \frac{4\ln(2)}{(f(1)-f(0))^2}$.

Proof. For repetition testing we have from Lemma 10 that $\mu_g^- = f(0)$, $\mu_g^+ = f(1)$ and $\Delta = \frac{f(1)-f(0)}{2}$.

If we assume all $i \in \mathcal{N}$ are in $t_i = u_I \log \frac{n}{\varepsilon}$ tests we have the probability of misclassification is bounded above by ε/n if $u_I > \ln(2)/\Delta^2$. To prove this we use Theorem 16 which states

If $i \in \mathcal{K}$ has been in t_i tests the probability of a *missed detection* using midpoint decoding is

$$\Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) = 2^{-t_i D_{\mathrm{KL}}(\mu_g^+ - \Delta || \mu_g^+)} \leq 2^{-\frac{\Delta^2 t_i}{\ln(2)}} \leq 2^{-\log \frac{n}{\varepsilon}} = \frac{\varepsilon}{n}.$$

If $i \notin \mathcal{K}$ which has been in t_i tests, then the probability of *false alarm* is

$$\Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) = 2^{-t_i D_{\mathrm{KL}}(\mu_g^- + \Delta || \mu_g^-)} \leq 2^{-\frac{\Delta^2 t_i}{\ln(2)}} \leq 2^{-\log \frac{n}{\varepsilon}} = \frac{\varepsilon}{n}.$$

Taking union bounds over $\mathcal{N}$, we have that the probability of misclassification is bounded above by ε. $\square$

D.6.3 THEOREM 1 UPPER BOUND

Theorem 1b. *Given any test function $f(\rho) = a_1 \rho^\alpha$, where $a_1 \in (0,1]$ and $\alpha \in (0, \infty)$, there exist non-adaptive group testing algorithms with probability of error at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{1,R} k \left(\frac{n}{k}\right)^\alpha \log \frac{n}{\varepsilon}, & \alpha \in (0,1) \\ u_{1,I} n \log \frac{n}{\varepsilon}, & \alpha \in [1, \infty) \end{cases}.$$

Here $u_{1,R}$ and $u_{1,I}$ are universal constants that depend only on f.

Proof.

- For $\alpha \in (0,1)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{1,R}$ (that depends only on f) such that

$$T \leq u_{1,R} k \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}$$

test suffice to obtain a probability of error bounded above by ε.

To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error. From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g) \in o(n^{-9})$.

For an item $i \in \mathcal{N}$ we have it partakes in $t_i \sim \text{Bin}(T, 1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

$$\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) u_{1,R} \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{1,R} \geq 6 \ln 2$.

We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) u_{1,R} \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}$. We apply Theorem 16 which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ

$$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^+ - \Delta || \mu_g^+) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}},$$

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^- + \Delta || \mu_g^-) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}}.$$

From Lemma 11, we have that $\Delta > \frac{a_1(2^\alpha - 1)}{2} \left(\frac{k}{n}\right)^{\alpha} (1 + o(1))$ and $\mu_g^+ < a_1 2^\alpha \left(\frac{k}{n}\right)^{\alpha} (1 + o(1))$ which gives us a lower bound

$$\frac{\Delta^2}{2\ln(2)\mu_g^+} > a_1 \frac{(2^\alpha - 1)^2}{\ln(2) 2^{3+\alpha}} \left(\frac{k}{n}\right)^{\alpha} (1 + o(1)) = c_0 \left(\frac{k}{n}\right)^{\alpha}.$$

Inserting into Theorem 16, if $u_{1,R} > \frac{1}{c_0\left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right)}$ our misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{1,R}}$ we choose some constant $u_{1,R}\left(1 - \sqrt{\frac{6\ln(2)}{u_{1,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{1,R}$, therefore there exists a $u_{1,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error

$$P_e < \bigcup_{i \in \mathcal{N}} \left[\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) u_{1,R} \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}\right) \cup \Pr(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}) \cup \Pr(i \notin \hat{\mathcal{K}} | i \in \mathcal{K})\right] < \varepsilon.$$

Remark: As $\frac{k}{n} \in o(1)$ any $u_{1,R} > 1/c_0$ is sufficient for at most ε error.

- For $\alpha \in [1, \infty)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{1,I} > \frac{4\ln(2)}{(f(1) - f(0))^2}$ then,

$$T \leq u_{1,I} n \log \frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

D.6.4 THEOREM 2 UPPER BOUND

Theorem 2b. *Given any test function $f(\rho) = a_0 + a_1\rho^\alpha$ where $a_0, a_1 \in (0,1]$, $a_0 + a_1 \leq 1$ and $\alpha \in (0,\infty)$, there exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{2,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}, & \alpha \in (0,\frac{1}{2}) \\ u_{2,I}n\log\frac{n}{\varepsilon}, & \alpha \in [\frac{1}{2},\infty) \end{cases}.$$

Here $u_{2,R}$ and $u_{2,I}$ are universal constants that depend only on f.

Proof.

- For $\alpha \in \left(0,\frac{1}{2}\right)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{2,R}$ (that depends only on f) such that,

  $$T \leq u_{2,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}$$

  test suffice to obtain a probability of error bounded above by ε.

  To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error. From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g) \in o\left(n^{-9}\right)$.

  For an item $i \in \mathcal{N}$ we have it partakes in $t_i \sim \text{Bin}(T, 1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

  $$\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^\alpha\right)u_{2,R}\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

  for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{2,R} \geq 6\ln 2$.

  We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta\left(\frac{k}{n}\right)^\alpha\right)u_{2,R}\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ we have,

  $$\Pr\left(i \notin \hat{\mathcal{K}}|i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\text{KL}}(\mu_g^+ - \Delta||\mu_g^+)t_i} \leq 2^{-\frac{t_i\Delta^2}{2\ln(2)\mu_g^+}},$$

  $$\Pr\left(i \in \hat{\mathcal{K}}|i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\text{KL}}(\mu_g^- + \Delta||\mu_g^-)t_i} \leq 2^{-\frac{t_i\Delta^2}{2\ln(2)\mu_g^+}}.$$

  From Lemma 11, we have that $\Delta > \frac{a_1(2^\alpha - 1)}{2}\left(\frac{k}{n}\right)^\alpha(1 + o(1))$ and $\mu_g^+ < a_0 + a_1 2^\alpha\left(\frac{k}{n}\right)^\alpha(1 + o(1))$ which gives us a lower bound,

  $$\frac{\Delta^2}{2\ln(2)\mu_g^+} > a_1^2\frac{(2^\alpha - 1)^2}{8\ln(2)a_0}\left(\frac{k}{n}\right)^{2\alpha}(1 + o(1)) = c_0\left(\frac{k}{n}\right)^{2\alpha}.$$

  Inserting into Theorem 16 we have that if $u_{2,R} > \frac{1}{c_0\left(1 - c_\delta\left(\frac{k}{n}\right)^\alpha\right)}$ the probability of misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{2,R}}$ we choose some constant $u_{2,R}\left(1 - \sqrt{\frac{6\ln(2)}{u_{2,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{2,R}$, therefore there exists a $u_{2,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

  $$P_e < \bigcup_{i\in\mathcal{N}}\left[\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right)u_{2,R}\left(\frac{n}{k}\right)^\alpha\log\frac{n}{\varepsilon}\right) \cup \Pr(i \in \hat{\mathcal{K}}|i \notin \mathcal{K}) \cup \Pr(i \notin \hat{\mathcal{K}}|i \in \mathcal{K})\right] < \varepsilon.$$

  Remark: As $\frac{k}{n} \in o(1)$ any $u_{2,R} > 1/c_0$ is sufficient for at most ε error.

- For $\alpha \in \left[\frac{1}{2}, \infty\right)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{2,I} > \frac{4\ln(2)}{(f(1)-f(0))^2}$ then,

$$T \leq u_{2,I} n \log \frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

D.6.5 THEOREM 3 UPPER BOUND

Theorem 3b. *Given any test function $f(\rho) \in \mathcal{A}(\alpha)$, where $f(0) = 0$ and $\alpha \in (0, \infty)$, there exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{3,R} k \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}, & \alpha \in (0,1) \\ u_{3,I} n \log \frac{n}{\varepsilon}, & \alpha \in [1, \infty) \end{cases}.$$

Here $u_{3,R}$ and $u_{3,I}$ are universal constants that depend only on f.

Proof.
- For $\alpha \in (0,1)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{3,R}$ (that depends only on f) such that,

$$T \leq u_{3,R} k \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}$$

test suffice to obtain a probability of error bounded above by ε.

To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error. From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g) \in o\left(n^{-9}\right)$.

For an item $i \in \mathcal{N}$ we have it partakes in $t_i \sim$ Bin$(T, 1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

$$\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) c_T \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{3,R} \geq 6 \ln 2$.

We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) u_{3,R} \left(\frac{n}{k}\right)^{\alpha} \log \frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ, we have

$$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^+ - \Delta || \mu_g^+) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}},$$

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^- + \Delta || \mu_g^-) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}}.$$

From Lemma 11, we have that $\Delta > \frac{a_1(2^\alpha - 1)}{2} \left(\frac{k}{n}\right)^{\alpha} (1 + o(1))$ and $\mu_g^+ < a_1 2^\alpha \left(\frac{k}{n}\right)^{\alpha} (1 + o(1))$ which gives us a lower bound

$$\frac{\Delta^2}{2\ln(2)\mu_g^+} > a_1 \frac{(2^\alpha - 1)^2}{\ln(2) 2^{3+\alpha}} \left(\frac{k}{n}\right)^{\alpha} (1 + o(1)) = c_0 \left(\frac{k}{n}\right)^{\alpha}.$$

Inserting into Theorem 16, we have that if $u_{3,R} > \frac{1}{c_0\left(1-c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right)}$ the probability of misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{3,R}}$ we choose some constant $u_{3,R}\left(1-\sqrt{\frac{6\ln(2)}{u_{3,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{3,R}$, therefore there exists a $u_{3,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

$$P_e < \bigcup_{i\in\mathcal{N}}\left[\Pr\left(t_i < \left(1-c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right)c_T\left(\frac{n}{k}\right)^{\alpha}\log\frac{n}{\varepsilon}\right)\cup\Pr(i\in\hat{\mathcal{K}}|i\notin\mathcal{K})\cup\Pr(i\notin\hat{\mathcal{K}}|i\in\mathcal{K})\right] < \varepsilon.$$

Remark: As $\frac{k}{n}\in o(1)$ any $u_{3,R} > 1/c_0$ is sufficient for at most ε error.

- For $\alpha\in[1,\infty)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{3,I} > \frac{4\ln(2)}{(f(1)-f(0))^2}$ then,

$$T \leq u_{3,I}n\log\frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

D.6.6 THEOREM 4 UPPER BOUND

Theorem 4b. *Given any test function $f(\rho)\in\mathcal{A}(\alpha)$, where $f(0)\in(0,1)$ and $\alpha\in(0,\infty)$, there exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{4,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}, & \alpha\in(0,\frac{1}{2}) \\ u_{4,I}n\log\frac{n}{\varepsilon}, & \alpha\in[\frac{1}{2},\infty) \end{cases}.$$

Here $u_{4,R}$ and $u_{4,I}$ are universal constants that depend only on f.

Proof.

- For $\alpha\in\left(0,\frac{1}{2}\right)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{4,R}$ (that depends only on f) such that,

$$T \leq u_{4,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}$$

test suffice to obtain a probability of error bounded above by ε.

To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error. From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g)\in o\left(n^{-9}\right)$.

For an item $i\in\mathcal{N}$ we have it partakes in $t_i\sim\mathrm{Bin}(T,1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

$$\Pr\left(t_i < \left(1-c_\delta\left(\frac{k}{n}\right)^{\alpha}\right)u_{4,R}\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

for some large enough $c_\delta > 0$, so that $c_\delta^2 u_{4,R} \geq 6\ln 2$.

We can now bound the probability of misclassification error assuming $t_i > \left(1-c_\delta\left(\frac{k}{n}\right)^{\alpha}\right)u_{4,R}\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+,μ_g^- and Δ, we have that

$$\Pr\left(i\notin\hat{\mathcal{K}}|i\in\mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^+-\Delta||\mu_g^+)t_i} \leq 2^{-\frac{t_i\Delta^2}{2\ln(2)\mu_g^+}},$$

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^- + \Delta || \mu_g^-) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}}.$$

From Lemma 11, we have that $\Delta > \frac{a_1(2^\alpha - 1)}{2}\left(\frac{k}{n}\right)^\alpha (1 + o(1))$ and $\mu_g^+ < a_0 + a_1 2^\alpha \left(\frac{k}{n}\right)^\alpha (1 + o(1))$ which gives us a lower bound

$$\frac{\Delta^2}{2\ln(2)\mu_g^+} > a_1^2 \frac{(2^\alpha - 1)^2}{8\ln(2) a_0} \left(\frac{k}{n}\right)^{2\alpha} (1 + o(1)) = c_0 \left(\frac{k}{n}\right)^{2\alpha}.$$

Inserting into Theorem 16, we have that if $u_{4,R} > 1/c_0 (1 - c_\delta \left(\frac{k}{n}\right)^\alpha)$ the probability of misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{4,R}}$ we choose some constant $u_{4,R}\left(1 - \sqrt{\frac{6\ln(2)}{u_{4,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{4,R}$, therefore there exists a $u_{4,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

$$P_e < \bigcup_{i \in \mathcal{N}} \left[\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) c_T \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}\right) \cup \Pr(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}) \cup \Pr(i \notin \hat{\mathcal{K}} | i \in \mathcal{K})\right] < \varepsilon.$$

Remark: As $\frac{k}{n} \in o(1)$ any $u_{4,R} > 1/c_0$ is sufficient for at most ε error.

- For $\alpha \in \left[\frac{1}{2}, \infty\right)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{4,I} > \frac{4\ln(2)}{(f(1) - f(0))^2}$ then,

$$T \leq u_{4,I} n \log\frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

D.6.7 THEOREM 5 UPPER BOUND

Theorem 5b. *Given any test function $f(\rho) \in [0,1]$ where $f(0) = 0$. For any $\alpha \in (0,\infty)$, if there exists $h, d \in \mathcal{A}(\alpha)$ where $h(0) = d(0) = 0$ and $h(\rho) \leq f(\rho) \leq d(\rho)$ for $\rho \in \left(0, \frac{k}{n}\ln\frac{n^2}{k}\right]$, there exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{5,R} k \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}, & \alpha \in (0,1) \\ u_{5,I} n \log\frac{n}{\varepsilon}, & \alpha \in [1,\infty) \end{cases}.$$

Here $u_{5,R}$ and $u_{5,I}$ are universal constants that depend only on f.

Proof.
- For $\alpha \in (0,1)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{5,R}$ (that depends only on f) such that,

$$T \leq u_{5,R} k \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}.$$

test suffice to obtain a probability of error bounded above by ε.

To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error. From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g) \in o\left(n^{-9}\right)$.

For an item $i \in \mathcal{N}$ we have it partakes in $t_i \sim \text{Bin}(T, 1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

$$\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) u_{5,R} \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{5,R} \geq 6\ln 2$.

We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta \left(\frac{k}{n}\right)^{\alpha/2}\right) u_{5,R} \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ, we have

$$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\text{KL}}(\mu_g^+ - \Delta || \mu_g^+) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}},$$

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\text{KL}}(\mu_g^- + \Delta || \mu_g^-) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}}.$$

From Claim 8 we have that $\Delta(f) > \Delta(h)$, we also have from Claim 7 that $\mu_g^+(f) < \mu_g^+(d)$.

Writing $h, d \in \mathcal{A}(\alpha)$ as $h(\rho) = \sum_{i=0}^{\infty} a_i \rho^{i\alpha}$ and $d(\rho) = \sum_{i=0}^{\infty} b_i \rho^{i\alpha}$, from Lemma 11 we have that $\Delta(h) > a_1 \frac{2^\alpha - 1}{2} \left(\frac{k}{n}\right)^\alpha (1 + o(1)) = c_\Delta \left(\frac{k}{n}\right)^\alpha$, we also have that $\mu_g^+(d) < b_1 2^\alpha \left(\frac{k}{n}\right)^\alpha (1 + o(1)) = c_p \left(\frac{k}{n}\right)^\alpha$ we can write

$$\frac{\Delta(f)^2}{2\ln(2)\mu_g^+(f)} > \frac{\Delta(h)^2}{2\ln(2)\mu_g^+(d)} > \left(\frac{k}{n}\right)^\alpha \frac{c_\Delta^2}{2\ln(2)c_p} = c_0 \left(\frac{k}{n}\right)^\alpha.$$

Inserting into Theorem 16, we have that if $u_{5,R} > \frac{1}{c_0\left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right)}$ the misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{5,R}}$ we choose some constant $u_{5,R}\left(1 - \sqrt{\frac{6\ln(2)}{u_{5,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{5,R}$, therefore there exists a $u_{5,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

$$P_e < \bigcup_{i \in \mathcal{N}} \left[\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right) c_T \left(\frac{n}{k}\right)^\alpha \log\frac{n}{\varepsilon}\right) \cup \Pr(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}) \cup \Pr(i \notin \hat{\mathcal{K}} | i \in \mathcal{K})\right] < \varepsilon.$$

Remark: As $\frac{k}{n} \in o(1)$ any $u_{5,R} > 1/c_0$ is sufficient for at most ε error.

- For $\alpha \in [1, \infty)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{5,I} > \frac{4\ln(2)}{(f(1) - f(0))^2}$ then,

$$T \leq u_{5,I} n \log\frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

$\square$

D.6.8 THEOREM 6 UPPER BOUND

Theorem 6b. *Given any test function $f(\rho) \in [0,1]$ where $f(0) \in (0,1)$. For any $\alpha \in (0,\infty)$, if there exists $h \in \mathcal{A}(\alpha)$ where $f(0) = h(0)$ and $f(\rho) \geq h(\rho)$ for $\rho \in \left(0, \frac{k}{n}\ln\frac{n^2}{k}\right]$, there exist non-adaptive group testing* *algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{6,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}, & \alpha \in \left(0,\frac{1}{2}\right) \\ u_{6,I}n\log\frac{n}{\varepsilon}, & \alpha \in \left[\frac{1}{2},\infty\right) \end{cases}.$$

Here $u_{6,R}$ and $u_{6,I}$ are universal constants that depend only on f.

Proof.

- For $\alpha \in \left(0,\frac{1}{2}\right)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{6,R}$ (that depends only on f) such that,

  $$T \leq u_{6,R}k\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}.$$

  test suffice to obtain a probability of error bounded above by ε.

  To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error. From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g) \in o\left(n^{-9}\right)$.

  For an item $i \in \mathcal{N}$ we have it partakes in $t_i \sim \text{Bin}(T, 1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

  $$\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha}\right)u_{6,R}\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

  for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{6,R} \geq 6\ln 2$.
  We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha}\right)u_{6,R}\left(\frac{n}{k}\right)^{2\alpha}\log\frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ, we have that

  $$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^+ - \Delta || \mu_g^+)t_i} \leq 2^{-\frac{t_i\Delta^2}{2\ln(2)\mu_g^+}},$$

  $$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^- + \Delta || \mu_g^-)t_i} \leq 2^{-\frac{t_i\Delta^2}{2\ln(2)\mu_g^+}}.$$

  From Claim 8 we have that $\Delta(f) > \Delta(h)$, writing $h \in \mathcal{A}(\alpha)$ as $h(\rho) = \sum_{i=0}^{\infty} a_i\rho^{i\alpha}$, from Lemma 11 we have that $\Delta(h) > a_1\frac{2^\alpha - 1}{2}\left(\frac{k}{n}\right)^{\alpha}(1 + o(1)) = c_\Delta\left(\frac{k}{n}\right)^{\alpha}$, we also have that $\mu_g^+(f) \leq 1$ therefore

  $$\frac{\Delta(f)^2}{2\mu_g^+(f)} > \frac{\Delta(h)^2}{2} > \left(\frac{k}{n}\right)^{\alpha}\frac{c_\Delta^2}{2} = c_0\left(\frac{k}{n}\right)^{\alpha}.$$

  Inserting into Theorem 16, if $u_{6,R} > \frac{1}{c_0\left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha}\right)}$ we have that the probability of misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{6,R}}$ we choose some constant $u_{6,R}\left(1 - \sqrt{\frac{6\ln(2)}{u_{6,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{6,R}$, therefore there exists a $u_{6,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

  $$P_e < \bigcup_{i \in \mathcal{N}}\left[\Pr\left(t_i < \left(1 - c_\delta\left(\frac{k}{n}\right)^{\alpha/2}\right)c_T\left(\frac{n}{k}\right)^{\alpha}\log\frac{n}{\varepsilon}\right) \cup \Pr(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}) \cup \Pr(i \notin \hat{\mathcal{K}} | i \in \mathcal{K})\right] < \varepsilon.$$

  Remark: As $\frac{k}{n} \in o(1)$ any $u_{6,R} > 1/c_0$ is sufficient for at most ε error.

- For $\alpha \in \left[\frac{1}{2}, \infty\right)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{6,I} > \frac{4\ln(2)}{(f(1)-f(0))^2}$ then,

$$T \leq u_{6,I} n \log \frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

D.6.9 THEOREM 7 UPPER BOUND

Theorem 7b. *Given any test function $f(\ell, g) = a_1 \frac{\ell^\alpha}{g^\beta}$ where $a_1 \in (0,1]$ and $0 \leq \alpha \leq \beta$, there exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \leq \begin{cases} u_{7,R} k \left(\frac{n}{k}\right)^\beta \log \frac{n}{\varepsilon}, & \beta \in (0,1) \\ u_{7,I} n \log \frac{n}{\varepsilon}, & \beta \in [1, \infty) \end{cases}.$$

Here $u_{7,R}$ and $u_{7,I}$ are universal constants that depend only on f.

Proof.

- For $\beta \in (0,1)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{7,R}$ (that depends only on f) such that,

$$T \geq u_{7,R} k \left(\frac{n}{k}\right)^\beta \log \frac{n}{\varepsilon}.$$

test suffice to obtain a probability of error bounded above by ε.

To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error . From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g) \in o\left(n^{-9}\right)$.

For an item $i \in \mathcal{N}$ we have it partakes in $t_i \sim \text{Bin}(T, 1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

$$\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^{\beta/2}\right) u_{7,R} \left(\frac{n}{k}\right)^\beta \log \frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{7,R} \geq 6 \ln 2$.

We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta \left(\frac{k}{n}\right)^{\beta/2}\right) u_{7,R} \left(\frac{n}{k}\right)^\beta \log \frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ, we have

$$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\text{KL}}(\mu_g^+ - \Delta || \mu_g^+) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}},$$

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\text{KL}}(\mu_g^- + \Delta || \mu_g^-) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}}.$$

From Lemma 12 we have that $\mu_g^+ < a_1 2^\alpha \left(\frac{k}{n}\right)^\beta (1 + o(1)) = c_p \left(\frac{k}{n}\right)^\beta$ and $\Delta(f) > \frac{a_1(2^\alpha - 1)}{2} \left(\frac{k}{n}\right)^\beta (1 + o(1)) = c_\Delta \left(\frac{k}{n}\right)^\beta$. Therefore we can write

$$\frac{\Delta(f)^2}{2\ln(2)\mu_g^+(f)} > \left(\frac{k}{n}\right)^\beta \frac{c_\Delta^2}{2\ln(2)c_p} = c_0 \left(\frac{k}{n}\right)^\beta.$$

Inserting into Theorem 16, if $u_{7,R} > \frac{1}{c_0\left(1-c_\delta\left(\frac{k}{n}\right)^{\beta/2}\right)}$ we have that the probability of misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{7,R}}$ we choose some constant $u_{6,R}\left(1-\sqrt{\frac{6\ln(2)}{u_{7,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{7,R}$, therefore there exists a $u_{7,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

$$P_e < \bigcup_{i\in\mathcal{N}}\left[\Pr\left(t_i < \left(1-c_\delta\left(\frac{k}{n}\right)^{\beta/2}\right)c_T\left(\frac{n}{k}\right)^{\beta}\log\frac{n}{\varepsilon}\right)\cup\Pr(i\in\hat{\mathcal{K}}|i\notin\mathcal{K})\cup\Pr(i\notin\hat{\mathcal{K}}|i\in\mathcal{K})\right] < \varepsilon.$$

Remark: As $\frac{k}{n}\in o(1)$ any $u_{7,R} > 1/c_0$ is sufficient for at most ε error.

- For $\beta\in[1,\infty)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{7,I} > \frac{4\ln(2)}{(f(1)-f(0))^2}$ then,

$$T \leq u_{7,I}n\log\frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

D.6.10 THEOREM 8 UPPER BOUND

Theorem 8b. *Given any test function $f(\ell,g) = a_0 + a_1\frac{\ell^\alpha}{g^\beta}$ where $a_0, a_1 \in (0,1]$ and $0<\alpha\leq\beta$, there exist non-adaptive group testing algorithms with probability of error of at most ε and computational complexity $\mathcal{O}(nT)$ as long as the total number of tests T is at most*

$$T \geq \begin{cases} u_{8,R}k\left(\frac{n}{k}\right)^{2\beta}\log\frac{n}{\varepsilon}, & \beta\in(0,\frac{1}{2}) \\ u_{8,I}n\log\frac{n}{\varepsilon}, & \beta\in[\frac{1}{2},\infty)\end{cases}.$$

Here $u_{8,R}$ and $u_{8,I}$ are universal constants that depend only on f.

Proof.

- For $\beta\in\left(0,\frac{1}{2}\right)$, applying a Bern$(1/k)$ random test design with the NCOMP Decoder, we show that there exists some $u_{8,R}$ (that depends only on f) such that,

$$T \geq u_{8,R}k\left(\frac{n}{k}\right)^{2\beta}\log\frac{n}{\varepsilon}.$$

test suffice to obtain a probability of error bounded above by ε.

To bound the probability of misclassification error, we first must bound the probabilities of the group size deviation error and insufficient test error . From Theorem 18, as $T = o\left(n^3\right)$ we have that the probability of group size deviation is very small $\Pr(E_g)\in o\left(n^{-9}\right)$.

For an item $i\in\mathcal{N}$ we have it partakes in $t_i\sim\text{Bin}(T,1/k)$. To bound the probability of insufficient test error we use Theorem 17 which states that,

$$\Pr\left(t_i < \left(1-c_\delta\left(\frac{k}{n}\right)^{\beta}\right)u_{8,R}\left(\frac{n}{k}\right)^{2\beta}\log\frac{n}{\varepsilon}\right) < \left(\frac{\varepsilon}{n}\right)^2$$

for some large enough $c_\delta > 0$, such that $c_\delta^2 u_{8,R} \geq 6\ln 2$.

We can now bound the probability of misclassification error assuming $t_i > \left(1 - c_\delta \left(\frac{k}{n}\right)^\beta\right) u_{8,R} \left(\frac{n}{k}\right)^{2\beta} \log \frac{n}{\varepsilon}$. From Theorem 16, which bounds the probability of misclassification error in terms of μ_g^+, μ_g^- and Δ, we have that

$$\Pr\left(i \notin \hat{\mathcal{K}} | i \in \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} < \mu_g^+ - \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^+ - \Delta || \mu_g^+) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}},$$

$$\Pr\left(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}\right) = \Pr\left(\frac{p_i}{t_i} > \mu_g^- + \Delta\right) \leq 2^{-D_{\mathrm{KL}}(\mu_g^- + \Delta || \mu_g^-) t_i} \leq 2^{-\frac{t_i \Delta^2}{2\ln(2)\mu_g^+}}.$$

From Lemma 12 we have that $\mu_g^+ < a_0 + a_1 2^\alpha \left(\frac{k}{n}\right)^\beta (1 + o(1))$ and $\Delta(f) > \frac{a_1(2^\alpha - 1)}{2} \left(\frac{k}{n}\right)^\beta (1 + o(1)) = c_\Delta \left(\frac{k}{n}\right)^\beta$. Therefore we can write

$$\frac{\Delta(f)^2}{2\ln(2)\mu_g^+(f)} > \left(\frac{k}{n}\right)^{2\beta} \frac{c_\Delta^2}{2\ln(2)a_0}(1 + o(1)) = c_0 \left(\frac{k}{n}\right)^{2\beta}.$$

Inserting into Theorem 16, if $u_{8,R} > \frac{1}{c_0\left(1 - c_\delta\left(\frac{k}{n}\right)^\beta\right)}$ we have that the probability of misclassification error is bounded above by $\frac{\varepsilon}{n}$. As $c_\delta^2 > \frac{6\ln(2)}{u_{8,R}}$ we choose some constant $u_{8,R}\left(1 - \sqrt{\frac{6\ln(2)}{u_{8,R}}}\right) \geq \frac{1}{c_0}$ which exists as the left hand side is an increasing function in $u_{8,R}$, therefore there exists a $u_{8,R}$ that only depends on f. Taking the union bound over $\mathcal{N}$ we have the probability of error,

$$P_e < \bigcup_{i \in \mathcal{N}} \left[\Pr\left(t_i < \left(1 - c_\delta \left(\frac{k}{n}\right)^\beta\right) c_T \left(\frac{n}{k}\right)^{2\beta} \log \frac{n}{\varepsilon}\right) \cup \Pr(i \in \hat{\mathcal{K}} | i \notin \mathcal{K}) \cup \Pr(i \notin \hat{\mathcal{K}} | i \in \mathcal{K})\right] < \varepsilon.$$

Remark: As $\frac{k}{n} \in o(1)$ any $u_{8,R} > 1/c_0$ is sufficient for at most ε error.

- For $\beta \in \left[\frac{1}{2}, \infty\right)$, applying repetition testing using the Midpoint Decoder, from Theorem 0 [Repetition Testing Upper Bound] we have that if $u_{8,I} > \frac{4\ln(2)}{(f(1) - f(0))^2}$ then,

$$T \leq u_{8,I} n \log \frac{n}{\varepsilon}$$

tests suffice for a probability of error bounded above by ε.

□

E FRACTIONAL TAYLOR SERIES

In this section we prove Theorem 11 using tools from real analysis and provide insight on our choice to analyse functions in $\mathcal{A}(\alpha)$.

Theorem 9 (Stone-Weierstrass Theorem (Weierstrass, 1885; Stone, 1937, 1948))**.** *Let K be a compact metric space and $\mathcal{C}(K, \mathbb{R})$ denote the set of real-valued continuous functions on K with the topology of uniform convergence. Let $\mathcal{B} \subset \mathcal{C}(K, \mathbb{R})$ a unital sub-algebra ($\mathcal{B}$ is closed under multiplication and real linear combinations of pairs of elements in $\mathcal{B}$) which separates points of K (for every $x, y \in K$, with $x \neq y$, there exists $f \in \mathcal{B}$ such that $f(x) \neq f(y)$). Then $\mathcal{B}$ is dense in $\mathcal{C}(K, \mathbb{R})$.*

Theorem 10. *For $\alpha \in (0, \infty)$ the set $\mathcal{B}(\alpha)$ defined below is dense in $\mathcal{C}([0, 1], \mathbb{R})$,*

$$\mathcal{B}(\alpha) := \left\{\sum_{i=0}^{\infty} a_i x^{i\alpha}; |a_i| < \infty\right\}.$$

Proof. We have that for any $f, g \in \mathcal{B}$, $fg \in \mathcal{B}$ and for any $a, b \in \mathbb{R}$ $af + bg \in \mathcal{B}$. Hence it is a unital sub-algebra. It is also separating as for any $x, y \in [0,1]$, there exists $f \in \mathcal{B}$ such that $f(x) \neq f(y)$. Hence by Theorem 9 it is dense in C. □

Remark: We then have the reduced set $\mathcal{B}_{[0,1]} = \{f : [0,1] \to [0,1]; f \in \mathcal{B}\} \subseteq \mathcal{B}$ is dense in $\mathcal{C}([0,1],[0,1])$,this combined with Claim 1 motivate our exploration of functions $\mathcal{A}(\alpha)$.

Definition 6 (Fractional Derivative (Katugampola (2014)))**.** *Let $f : [0,\infty] \to \mathbb{R}$, the $\alpha \in (0,1)$ derivative is*

$$D^\alpha(f)(x) := \lim_{\varepsilon \to 0} \frac{f\left(xe^{\varepsilon x^{-\alpha}}\right) - f(x)}{\varepsilon} = f'(x)x^{1-\alpha}$$

for $x > 0$, where $D^\alpha(f)(0) := \lim_{a\to 0} D^\alpha(f)(a)$. We denote i applications of the D^α function as $D^{i\alpha}$. A function is α-smooth if for $i \in \mathbb{N}$, $|D^{i\alpha}(f)| < \infty$ for the domain of the function.

Many properties of this fractional derivative are given in Katugampola (2014), we state two that are used in the proof of Theorem 11:

- $D^\alpha(C) = 0$ where C is a constant function.
- $D^\alpha\left(x^\beta\right) = \beta x^{\beta-\alpha}$.

The final prerequisite for Theorem 11 is an analagous Rolle's Theorem for this fractional derivative.

Theorem 20 (Fractional Rolle's Theorem (Katugampola (2014)))**.** *Let $a > 0$ and $f : [a,b] \to \mathbb{R}$ be a function with the properties that*

1. *f is continuous on $[a,b]$.*
2. *f is α-differentiable on (a,b) for some $\alpha \in (0,1)$.*
3. *$f(a) = f(b)$.*

Then there exists $c \in (a,b)$ such that $f^\alpha(c) = 0$.

This allows us to define a Taylor Series type expansion for α power series.

Theorem 11. *For some $\alpha \in (0,1)$, an α-smooth function $f : [0,1] \to \mathbb{R}$ can be approximated around $c \in [0,1]$ as*

$$H(x) := \sum_{i=0}^{\infty} a_i(x-c)^{i\alpha}, \text{ where } a_i = \frac{D^{i\alpha}(f)(c)}{i!\alpha^i}.$$

Defining the sum truncated at n as $H_n(x) = \sum_{i=0}^{n} a_i(x-c)^{i\alpha}$ and the remainder term at n as $R_n(x) = f(x) - H_n(x)$ we have that for all $x \in [0,1]$,

$$R_n(x) = \frac{(x-c)^{(n+1)\alpha} D^{(n+1)\alpha}(f)(z)}{\alpha^{n+1}(n+1)!},$$

for some $z \in [0,1]$. Therefore, $\lim_{n\to\infty} R_n(x) \to 0$.

Proof. For some $c \in [0,1]$ and $b \in (0,1]$ we define

$$h(x) = f(x) - H_n(x) - s(x-c)^{(n+1)\alpha} \text{ where } s = \frac{R_n(b)}{(b-c)^{(n+1)\alpha}}.$$

First we note that $h(c) = 0$ as $h(c) = f(c) - a_0 = f(c) - f(c) = 0$. We also have that $h(b) = 0$

$$h(b) = f(b) - H_n(b) - \frac{R_n(b)}{(b-c)^{(n+1)\alpha}}(b-c)^{(n+1)\alpha} = f(b) - H_n(b) - R_n(b) = 0.$$

And that $D^{j\alpha}(h)(c) = 0$ for $j < n+1$

$$D^{j\alpha}(h)(x) = D^{j\alpha}(f)(x) - \sum_{i=j}^{n} a_i \alpha^j \frac{i!}{(i-1-j)!}(x-c)^{(i-j)\alpha} - \alpha^j \frac{(n+1)!}{(n-j)!}(x-c)^{(n+1-j)\alpha},$$

$$D^{j\alpha}(h)(c) = D^{j\alpha}(f)(c) - j! a_j \alpha^j = D^{j\alpha}(f)(c) - D^{j\alpha}(f)(c) = 0.$$

First for $b > c$ applying Fractional Rolle's Theorem iteratively:

1. $\exists z_1 \in (c, b)$ where $D^\alpha(h)(z_1) = 0$, as $h(c) = h(b) = 0$.
2. $\exists z_2 \in (c, z_1)$ where $D^{2\alpha}(h)(z_2) = 0$, as $D^\alpha(h)(c) = D^\alpha(h)(z_1) = 0$.
3. $\exists z_3 \in (c, z_2)$ where $D^{3\alpha}(h)(z_3) = 0$, as $D^{2\alpha}(h)(c) = D^{2\alpha}(h)(z_2) = 0$.
4. ...
5. $\exists z \in (c, z_n)$ where $D^{(n+1)\alpha}(h)(z) = 0$, as $D^{n\alpha}(h)(c) = D^{n\alpha}(h)(z_n) = 0$.

Thus we have

$$D^{(n+1)\alpha}(h)(z) = D^{(n+1)\alpha}(f)(z) - s\alpha^{n+1}(n+1)! = 0,$$

$$s = \frac{D^{(n+1)\alpha}(f)(z)}{\alpha^{n+1}(n+1)!} = \frac{R_n(b)}{(b-c)^{(n+1)\alpha}},$$

$$R_n(b) = \frac{(b-c)^{(n+1)\alpha} D^{(n+1)\alpha}(f)(z)}{\alpha^{n+1}(n+1)!}.$$

As all α derivatives of f are bounded we have for all $b \in (c, 1]$, $\lim_{n\to\infty} R_n(b) \to 0$.

Second for $b < c$ applying Fractional Rolle's Theorem iteratively:

1. $\exists z_1 \in (b, c)$ where $D^\alpha(h)(z_1) = 0$, as $h(c) = h(b) = 0$.
2. $\exists z_2 \in (z_1, c)$ where $D^{2\alpha}(h)(z_2) = 0$, as $D^\alpha(h)(c) = D^\alpha(h)(z_1) = 0$.
3. ...
4. $\exists z \in (z_n, c)$ where $D^{(n+1)\alpha}(h)(z) = 0$, as $D^{n\alpha}(h)(c) = D^{n\alpha}(h)(z_n) = 0$.

Thus we have

$$D^{(n+1)\alpha}(h)(z) = D^{(n+1)\alpha}(f)(z) - s\alpha^{n+1}(n+1)! = 0,$$

$$s = \frac{D^{(n+1)\alpha}(f)(z)}{\alpha^{n+1}(n+1)!} = \frac{R_n(b)}{(b-c)^{(n+1)\alpha}},$$

$$R_n(b) = \frac{(b-c)^{(n+1)\alpha} D^{(n+1)\alpha}(f)(z)}{\alpha^{n+1}(n+1)!}.$$

As all α derivatives of f are bounded we have for all $b \in [0, c)$, $\lim_{n\to\infty} R_n(b) \to 0$.

Therefore we have that for $b \in [0, 1]$ $\lim_{n\to\infty} R_n(b) \to 0$. $\square$