UNIVERSITY OF LEEDS

This is a repository copy of *On the All-or-Nothing Behavior of Bernoulli Group Testing*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/218506/

Version: Accepted Version

Article:
Truong, L.V., Aldridge, M. orcid.org/0000-0002-9347-1586 and Scarlett, J. (2020) On the All-or-Nothing Behavior of Bernoulli Group Testing. IEEE Journal on Selected Areas in Information Theory, 1 (3). pp. 669-680. ISSN 2641-8770

https://doi.org/10.1109/jsait.2020.3039790

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

On the All-Or-Nothing Behavior of Bernoulli Group Testing

Lan V. Truong, Matthew Aldridge, and Jonathan Scarlett

***Abstract*—In this paper, we study the problem of non-adaptive group testing, in which one seeks to identify which items are defective given a set of suitably-designed tests whose outcomes indicate whether or not at least one defective item was included in the test. The most widespread recovery criterion seeks to exactly recover the entire defective set, and relaxed criteria such as approximate recovery and list decoding have also been considered. In this paper, we study the fundamental limits of group testing under the significantly relaxed *weak recovery* criterion, which only seeks to identify a small fraction (e.g., 0.01) of the defective items. Given the near-optimality of i.i.d. Bernoulli testing for exact recovery in sufficiently sparse scaling regimes, it is natural to ask whether this design additionally succeeds with much fewer tests under weak recovery. Our main negative result shows that this is not the case, and in fact, under i.i.d. Bernoulli random testing in the sufficiently sparse regime, an *all-or-nothing* phenomenon occurs: When the number of tests is slightly below a threshold, weak recovery is impossible, whereas when the number of tests is slightly above the same threshold, high-probability exact recovery is possible. In establishing this result, we additionally prove similar negative results under Bernoulli designs for the weak detection problem (distinguishing between the group testing model vs. completely random outcomes) and the problem of identifying a single item that is definitely defective. On the positive side, we show that all three relaxed recovery criteria can be attained using considerably fewer tests under suitably-chosen non-Bernoulli designs. Thus, our results collectively indicate that when too few tests are available, naively applying i.i.d. Bernoulli testing can lead to catastrophic failure, whereas "cutting one's losses" and adopting a more carefully-chosen design can still succeed in attaining these less stringent criteria.**

***Index Terms*—Group testing, hypothesis testing, approximate recovery, phase transitions.**

I. Introduction

The group testing problem has recently regained significant attention following new applications and connections with compressive sensing; see [1] for a recent survey. Briefly, the idea of group testing is to identify a small subset of defective items within a larger subset of items, based on a number of tests whose binary outcomes indicate whether or not at least one defective item was included in the test.

The standard recovery goal in group testing is to exactly identify the entire defective set. In combinatorial group testing [2], [3] a single test design is required to succeed for all defective sets up to a certain size, whereas in probabilistic group testing [1], [4] only high-probability recovery is required with respect to a random defective set (and/or a random test design). Various relaxed recovery criteria have also appeared, including list decoding recovery [5]–[11] and approximate recovery criteria that allow a small number of false positives and/or false negatives in the reconstruction [11], [11]–[13].

In this paper, focusing on probabilistic group testing, our goal is to better understand the fundamental limits of what can be achieved in the group testing problem under *significantly weaker* recovery criteria. In particular, instead of asking when it is possible to recover most of the defectives, we seek to understand when it is possible just to *recover a small fraction*. For general non-adaptive designs, we show that this goal can be obtained with much fewer tests, and identify an exact threshold on the number required. On the other hand, for the widely-adopted i.i.d.[1] Bernoulli test matrix design, we identify scenarios under which an *all-or-nothing* phenomenon occurs: When the number of tests is slightly above a certain threshold, high-probability exact recovery is possible, whereas slightly below the same threshold, essentially nothing can be learned from the tests. Thus, while Bernoulli designs can be near-optimal under standard recovery criteria, they are also prone to complete failure when there are too few tests, and accordingly can be highly suboptimal under relaxed recovery criteria. Along the way, we additionally provide analogous results for the problems of weak detection and identifying a definite defective, which are formally defined in Section I-C.

A. Problem Setup

We consider a population of p items indexed as $\{1,\dots,p\}$, and we let k denote the number of defective items. The set of defective items is denoted by S, and is assumed to be uniform over the $\binom{p}{k}$ possibilities.

A group testing procedure performs a sequence of n tests, with $X^{(i)} \in \{0,1\}^p$ indicating which item is in the i-th test. The resulting outcomes are $Y^{(i)} = \bigvee_{j\in S} X_j^{(i)}$ for $i = 1,\dots,n$, where n is the number of tests. That is, the test outcome is 1 if there is any defective item in the test, and 0 otherwise. The tests can be represented by a matrix $\mathbf{X} \in \{0,1\}^{n\times p}$, whose i-th row is $X^{(i)}$. Similarly, the outcomes can be represented by a vector $\mathbf{Y} \in \{0,1\}^n$, whose i-th entry is $Y^{(i)}$.

In general, group testing procedures may be *adaptive* (i.e., $X^{(i)}$ may be chosen as a function of the previous outcomes) or *non-adaptive* (i.e., all $X^{(i)}$ must be selected prior to observing any outcomes). We focus on the non-adaptive setting,

L. V. Truong is with the Department of Engineering, University of Cambridge. M. Aldridge is with the School of Mathematics, University of Leeds. J. Scarlett is with the Department of Computer Science, Department of Mathematics, and Institute of Data Science, National University of Singapore (NUS). e-mail: lt407@cam.ac.uk; m.aldridge@leeds.ac.uk; scarlett@comp.nus.edu.sg

This work was supported by the Singapore National Research Foundation (NRF) under grant number R-252-000-A74-281.

[1] We write i.i.d. as a shorthand for *independent and identically distributed*.

which is often preferable in practice due to permitting highly parallelized tests. In particular, except where stated otherwise, we consider the widely-adopted *(i.i.d.) Bernoulli random test design* [1, Sec. 2.1], in which every item is independently placed in each test with probability $\frac{\nu}{k}$ for some $\nu > 0$, and we choose ν to be such that

$$\left(1 - \frac{\nu}{k}\right)^k = \frac{1}{2}. \tag{1}$$

This choice ensures that the probability of a positive test is exactly $\frac{1}{2}$, which maximizes the entropy of each test outcome. More importantly, this choice of ν leads to a provably optimal number of tests in broad scaling regimes, as we survey in Section I-B. A simple asymptotic analysis gives $\nu = (\log 2)(1 + o(1))$ as $k \to \infty$, which behaves similarly to the choice $\nu = \log 2$, but the exact choice described by (1) will be more convenient to work with.

B. Related Work

There have recently been numerous developments on theory and algorithms for probabilistic group testing [13]–[20] (see [1] for a survey); here we focus only on those most relevant to the present paper.

The most relevant works to us are those attaining upper and/or lower bounds on the number of tests of the form $\big(k \log_2 \frac{p}{k}\big)(1 + o(1))$. The most straightforward way that this quantity arises is that with $\binom{p}{k}$ possible defective sets and 2^n possible sequences of outcomes, we require $n \geq \log_2 \binom{p}{k}$ for each defective set to produce different outcomes. In the sub-linear regime $k = o(p)$, this simplifies to $n \geq \big(k \log_2 \frac{p}{k}\big)(1 + o(1))$. Building on this intuition, Fano's inequality was used in [4], [21] to show that $n \geq (1 - \delta) \log_2 \binom{p}{k}$ is required to attain an error probability of at most δ, and a refined bound $n \geq \log_2 \big(\binom{p}{k}(1 - \delta)\big)$ was established in [22].

More recently, various results showed that $\big(k \log_2 \frac{p}{k}\big)(1 + o(1))$ tests are *sufficient* for certain recovery guarantees under broad scaling regimes on k as a function of p. In [13], high-probability exact recovery was shown to be possible under Bernoulli random testing when $k = O(p^{1/3})$ and $n = \big(k \log_2 \frac{p}{k}\big)(1 + o(1))$, and in addition, this result was extended to all $k = o(p)$ when the exact recovery criterion is replaced by the following approximate recovery criterion (see also [11]): Output a set $\widehat{S} \subseteq \{1, \ldots, p\}$ of cardinality k such that

$$\max\big\{|\widehat{S} \setminus S|, |S \setminus \widehat{S}|\big\} \leq \alpha k \tag{2}$$

for some $\alpha \in (0, 1)$. The above-mentioned result holds for arbitrarily small $\alpha > 0$, as long as it is bounded away from zero as $p \to \infty$.

On the other hand, the lower bounds for approximate recovery in [11], [13] only state that in order to attain (2) for fixed $\alpha \in (0, 1)$, it is necessary that $n \geq (1-\alpha)\big(k \log_2 \frac{p}{k}\big)(1+o(1))$. This suggests that as α increases, the constant factor in the number of tests could be reduced. We will show that the preceding lower bound can in fact be matched with an upper bound using a suitably-chosen non-Bernoulli design, whereas Bernoulli designs can fail even for α close to one when n is slightly smaller than $k \log_2 \frac{p}{k}$.

While our discussion thus far focuses on Bernoulli designs, in the case of exact recovery, improved bounds have been shown for a different random test design based on *near-constant tests-per-item* [19], [20], [23], in particular permitting $n = \big(k \log_2 \frac{p}{k}\big)(1 + o(1))$ for all $k = O(p^{0.409})$ (improving on $O(p^{1/3})$). However, upon moving to approximate recovery with parameter α, both designs attain the threshold $n = \big(k \log_2 \frac{p}{k}\big)(1 + o(1))$ in the limit as $\alpha \to 0$ [11], [24], suggesting that there is less to be gained via the near-constant tests-per-item design under relaxed recovery criteria. Nevertheless, extending our results to this design may be an interesting direction for future work.

Our work is inspired by recent studies of the all-or-nothing behavior of sparse linear regression under i.i.d. Gaussian measurements; see [25] for a study of the maximum-likelihood estimator, and [26], [27] for general estimators. While group testing can be viewed as a non-linear Boolean counterpart to sparse linear regression [28], [29], and our work will adopt the same high-level approach as [26], the details will be very different.

C. Overview of the Paper

As hinted above, in this paper, our main goal is to investigate the question of when the following mild recovery requirement is possible:

- **(Weak recovery)** Can we find a set $\hat{S}$ of size k such that $|S \cap \hat{S}| \geq \delta k$ for small $\delta > 0$ with some non-zero constant probability?

The study of this goal essentially asks whether we can learn even a small amount of information from the test outcomes. As a result, any hardness result (lower bound on the number of tests) under this criterion serves as a much stronger claim compared to a hardness result for exact recovery.

While our main focus is on weak recovery, we will additionally address the following two recovery goals that are also much milder than exact recovery:

- **(Weak detection)** Can we perform a hypothesis test on $(\mathbf{X}, \mathbf{Y})$ to distinguish between the above group testing model and the "null model" in which $\mathbf{Y}$ is independent of $\mathbf{X}$?
- **(Identify a definite defective)** Can we identify just a single defective item, i.e., output a single index $\mathcal{I} \in \{1, \ldots, p\}$ with certainty that $\mathcal{I} \in S$? (Here we also allow "detected errors", in which the decoder declares that it is uncertain.)

We show in Section II that these goals can be achieved with very few tests using suitably-chosen non-Bernoulli test designs. On the other hand, for Bernoulli designs (studied in Section III), we will see that in sufficiently sparse regimes, these criteria all require essentially the same number of tests as exact recovery.

The main reason that we consider weak detection is that is serves as a useful stepping stone to establishing our negative result for weak recovery under Bernoulli designs. Identifying a definite defective is also not our central focus, but its negative result comes almost for free via our analysis.

Before proceeding, we briefly pause to discuss our emphasis on Bernoulli designs, despite our results demonstrating that they can be inferior to alternative designs under the relaxed recovery criteria. The justification for doing so is that Bernoulli designs (and other related unstructured random designs) are widespread and extensively studied in the literature [1], and thus serve as a standard "go-to" design. As a result, it is essential to not only identify the cases that they succeed, but also understand their limitations.

II. Positive Results for Bernoulli and non-Bernoulli Designs

To set the stage for our negative results on Bernoulli designs, we start by providing several positive (i.e., achievability) results for weak recovery, weak detection, and identifying a definite defective, focusing primarily on non-Bernoulli designs.

A. Asymptotically Optimal Approximate Recovery

As discussed in Section I-B, the lower bounds for approximate recovery in [11], [13] state that in order to attain

$$\max\left\{|\hat{S}\setminus S|, |S\setminus\hat{S}|\right\} \le \alpha k \tag{3}$$

for fixed $\alpha \in (0,1)$, it is necessary that $n \ge (1-\alpha)(k\log_2\frac{p}{k})(1-o(1))$. The following result shows that one can in fact attain a matching upper bound for general test designs.

Theorem 1. (Positive Result for Approximate Recovery) *Consider the probabilistic group testing problem, and for fixed $\alpha \in (0,1)$, suppose that $n \ge (1+\eta)(1-\alpha)\log_2\binom{p}{k}$ for some $\eta > 0$. Then, when $k \to \infty$ with $k = o(p)$ as $p \to \infty$, there exists a non-adaptive test design and decoder that outputs an estimate $\hat{S}$ of S satisfying the following as $p \to \infty$:*

$$\mathbb{P}\left[\max\left\{|\hat{S}\setminus S|, |S\setminus\hat{S}|\right\} \le \alpha k\right] \to 1. \tag{4}$$

Moreover, it can be assumed that $\hat{S}$ has cardinality exactly k.

By letting α be arbitrarily close to one, this result establishes the following corollary.

Corollary 1. (Positive Result for Weak Detection) *Under the preceding setup, when $k \to \infty$ with $k = o(p)$ as $p \to \infty$, as long as $k = \Omega\big(k\log\frac{p}{k}\big)$ (with any implied constant), there exists a non-adaptive test design and decoder that outputs an estimate $\hat{S}$ of cardinality k satisfying $|\hat{S}\cap S| = \Omega(k)$ with probability approaching one.*

To prove Theorem 1, we will use the previous best-known positive result on approximate recovery as a stepping stone. This is stated in the following lemma, whose main statement comes from [11], [13], with the second part regarding approximately-known k instead coming from [30, App. B].

Lemma 1. (Existing Positive Results for Approximate Recovery [11], [13], [30]) *Consider the probabilistic group testing problem with Bernoulli random testing using the choice of ν in (1), and suppose that $n \ge (1+\eta)\log_2\binom{p}{k}$ for some $\eta > 0$. Then, for any fixed $\alpha \in (0,1)$, when $k = o(p)$ as $p \to \infty$, there exists a decoder that outputs an estimate $\hat{S}$ of S satisfying the following as $p \to \infty$:*

$$\mathbb{P}\left[\max\left\{|\hat{S}\setminus S|, |S\setminus\hat{S}|\right\} \le \alpha k\right] \to 1. \tag{5}$$

Furthermore, this result remains true even when the decoder does not know the exact value of k but instead only knows some quantity $\bar{k}$ satisfying $\bar{k} = k(1+o(1))$, and the ν/k test-inclusion probability is replaced by $\nu/\bar{k}$.

Theorem 1 reduces the number of tests in Lemma 1 by a multiplicative factor of $1-\alpha$, and provides an asymptotically optimal result (including the constant). In addition, this result demonstrates that weak recovery is possible whenever $n = \Omega\big(k\log_2\frac{p}{k}\big)$ with an arbitrarily small implied constant. In contrast, we will show in Theorem 6 below that Bernoulli testing requires the implied constant to be one, and hence, the two differ by an arbitrarily large constant factor.

Proof of Theorem 1. The idea of the proof is straightforward: We ignore slightly less than a fraction α of the items, and use Bernoulli testing to achieve approximate recovery on the items that were not ignored.

More formally, fix $\alpha' \in (0,\alpha)$, and consider discarding $\alpha' p$ items chosen uniformly at random, leaving $p' = (1-\alpha')p$ items remaining. By Hoeffding's inequality for the Hypergeometric distribution [31] and the assumption $k \to \infty$, we have with probability $1-o(1)$ that the number of remaining defectives k' satisfies

$$k' = (1-\alpha')k\cdot(1+o(1)). \tag{6}$$

We apply the second part of Lemma 1 on this reduced problem, with $\bar{k} = (1-\alpha')k$ and an approximate recovery parameter α'' to be selected shortly. While the number of defectives k' in the reduced problem is random, we see from (6) that it is known up to a multiplicative factor of $1+o(1)$, as required in Lemma 1. Then, the number of tests n' required in Lemma 1 satisfies the following:

$$n' = (1+\eta')\log_2\binom{p'}{k'} \tag{7}$$

$$= (1+\eta')\log_2\binom{(1-\alpha')p'}{(1-\alpha')k'(1+o(1))} \tag{8}$$

$$= \big(1+\eta'+o(1)\big)(1-\alpha')\log_2\binom{p}{k} \tag{9}$$

for arbitrarily small $\eta' > 0$, where we used $\log\binom{p}{k} = \big(k\log\frac{p}{k}\big)(1+o(1))$ for $k = o(p)$. In addition, in accordance with the lemma, the returned set $\hat{S}'$ of size k' contains at least

$$(1-\alpha'')k' = (1-\alpha'')(1-\alpha')k\cdot(1+o(1)) \tag{10}$$

defective items, with probability approaching one.

It suffices to let the final estimate $\hat{S}$ equal $\hat{S}'$; alternatively, if an estimate of size k is sought, one can add $k-k'$ arbitrary items to $\hat{S}'$ to form $\hat{S}$. In either case, taking α' arbitrarily close to α and α'' arbitrarily close to 0, (10) ensures that $\hat{S}$ contains at least $(1-\alpha)k$ defective items, which implies $\max\left\{|\hat{S}\setminus S|, |S\setminus\hat{S}|\right\} \le \alpha k$ due to the fact that $|S| = k$. In addition, since $\eta' > 0$ is arbitrarily small, the number of tests

in (9) simplifies to $(1+\eta)(1-\alpha)\log_2\binom{p}{k}$ for arbitrarily small $\eta > 0$.

□

We note that while the preceding analysis still uses Bernoulli testing as a subroutine via Lemma 1, the full $n \times p$ test matrix is not i.i.d. Bernoulli, as a fraction α' of its columns are set to zero. Hence, we still consider this to be a non-Bernoulli test design.

Discussion. Since the proof of Theorem 1 is based on ignoring a fraction of the items, it amounts to a technique that immediately gives up on exact recovery, or "cuts its losses" from an early stage. This raises the interesting question as to whether such an approach is actually necessary to obtain the bound in Theorem 1.

To appreciate this distinction, note that Hwang's *adaptive* generalized binary splitting algorithm [32] works by repeatedly identifying a single defective using at most $\big(\log_2\frac{p}{k}\big)(1+o(1))$ tests, and then removing it from further consideration. Hence, exact recovery is guaranteed with $n = \big(k\log_2\frac{p}{k}\big)(1+o(1))$, but even if the procedure is stopped after $(1-\alpha)\big(k\log_2\frac{p}{k}\big)(1+o(1))$ tests, one will have already identified $(1-\alpha)k$ defective items. In this sense, Hwang's adaptive algorithm is *universally optimal* with respect to the approximate recovery parameter α,[2] and the algorithm degrades gracefully as the number of tests decreases below the exact recovery threshold.

In contrast, the non-adaptive designs that we have considered do not enjoy such universality. Under a Bernoulli design with $k = o(p)$, we can achieve approximate recovery with arbitrarily small α when $n = \big(k\log_2\frac{p}{k}\big)(1+o(1))$ (Lemma 1), or even exact recovery if $k = O(p^{1/3})$ [13], but we are prone to complete failure for smaller n, at least in sufficiently sparse regimes (see Theorem 6 below). Alternatively, ignoring roughly a fraction α of the items leads to α-approximate recovery when $n = (1-\alpha)\big(k\log_2\frac{p}{k}\big)(1+o(1))$, but one retains this guarantee *and no better* regardless of how much n is increased. Hence, it remains an interesting question for future work as to whether there exists a gracefully degrading (and ideally universally optimal) test matrix design in the non-adaptive setting.

B. A Trivial Strategy for Weak Detection

We first describe the weak detection problem in more detail. The goal is to distinguish between two joint distributions on the pair $(\mathbf{X},\mathbf{Y})$ for some specified distribution $P(\mathbf{X})$:

- Under distribution P, the X-marginal is $P(\mathbf{X})$, and the joint distribution $P(\mathbf{X},\mathbf{Y})$ is deduced by deterministically computing $\mathbf{Y}$ from $\mathbf{X}$ via the group testing model.
- Under distribution Q, the $\mathbf{X}$ and $\mathbf{Y}$ marginals match those of P, but $\mathbf{X}$ and $\mathbf{Y}$ are independent, i.e., $Q(\mathbf{X},\mathbf{Y}) = P(\mathbf{X})P(\mathbf{Y})$.

This is a binary hypothesis testing problem. The distribution Q corresponds to "completely uninformative outcomes", so intuitively, if we cannot reliably distinguish between P and Q, then we can view the group tests (under the distribution P) as being highly uninformative.

As hinted above, for Bernoulli designs, the main reason that we consider weak detection is as a stepping stone to providing a negative result for weak recovery (see Section III). It turns out that if we consider weak detection as a standalone problem with general designs, then a trivial strategy succeeds with very few tests. Formally, we have the following.

Proposition 1. (Trivial Strategy for Weak Detection) *For any number of tests n, there exists some distribution $P(\mathbf{X})$ such that given $(\mathbf{X},\mathbf{Y})$, the group testing joint distribution $P(\mathbf{X},\mathbf{Y})$ can be distinguished from $Q(\mathbf{X},\mathbf{Y}) := P(\mathbf{X})P(\mathbf{Y})$ with zero error probability under P and $1-2^{-n}$ error probability under Q.*

Proof. Consider letting each test independently contain all items with probability $\frac{1}{2}$ and no items with probability $\frac{1}{2}$. Hence, each outcome is independent and equiprobable on $\{0,1\}$. Consider a detection algorithm that declares P whenever the ones in $\mathbf{Y}$ exactly match the rows of ones in $\mathbf{X}$, and declares Q otherwise. Then, under P, success is guaranteed due to the group testing model being noiseless. Under Q, since the test outcomes are uniformly random, the probability of producing the correct outcomes is 2^{-n}, proving the proposition. □

This result suggests that weak detection is of limited interest as a standalone recovery criterion in general. Nevertheless, this finding arguably strengthens our subsequent negative results for Bernoulli designs (Section III), in the sense of showing that weak detection is not attained despite being an "extremely easy" goal.

C. Weak Detection with Bernoulli Designs

Our negative results for Bernoulli testing in Section III will demonstrate failure in sufficiently sparse regimes (e.g., $k = \mathrm{poly}(\log p)$) when the number of tests is slightly below $\log_2\binom{p}{k}$. On the other hand, fairly simple detection strategy can be used to attain the following positive result under Bernoulli testing when $k \gg p^{1/3}$ and $k = o(p)$.

Theorem 2. (Positive Result for Weak Detection with Bernoulli Designs) *Consider the probabilistic group testing problem with Bernoulli random testing using the choice of ν in (1), and suppose that $n = (1-\eta)\log_2\binom{p}{k}$ for some $\eta \in (0,1)$. Then, under the condition that $k = o(p)$ and $k = \Omega\big(p^{\frac{1+\eta}{3+\eta}+\delta}\big)$ for some (arbitrarily small) fixed constant $\delta \in (0,1)$, there exists a binary hypothesis testing scheme that succeeds in distinguishing $P(\mathbf{X},\mathbf{Y})$ from $Q(\mathbf{X},\mathbf{Y}) := P(\mathbf{X})P(\mathbf{Y})$ with probability approaching one as $p \to \infty$.*

Proof. The idea of is to perform weak detection based on the number of columns of $\mathbf{X}$ that are *covered* by the test outcomes (i.e., the test outcome is one whenever the column entry is one). The number of such columns is characterized by a binomial distribution under both the group testing model and the null model, and the distributions are shown to be statistically distinguishable under the conditions given. The details are given in Appendix A. □

[2]Note that the lower bound stated following (3) also holds for adaptive algorithms.

D. A Simple Strategy for Identifying a Definite Defective

We first describe the problem of *identifying a definite defective* in more detail. We suppose that the decoder either outputs a single index $\mathcal{I} \in \{1, \dots, p\}$ believed to be defective, or declares "I don't know" by outputting $\mathcal{I} = 0$. In the former scenario, we insist that $\mathcal{I}$ must be defective (i.e., $\mathcal{I} \in S$) with probability one, meaning that the only errors allowed are *detected errors* corresponding to $\mathcal{I} = 0$. This setup is partly motivated by the definite defectives algorithm for recovering the defective set [17], [23], as well as the notion of *zero undetected error capacity* in information theory [33].

The following result shows that under a suitably-chosen non-Bernoulli test design, a single definitely-defective item can be reliably identified using only $O(\log \frac{p}{k})$ tests, representing a reduction by a factor of k compared to the usual $O(k \log \frac{p}{k})$ scaling.

Theorem 3. (Positive Result for Identifying a Definite Defective) *Consider the probabilistic group testing problem, and suppose that $n \geq (1+\eta)2c \log_2 \frac{p}{k}$ for some $\eta > 0$ and some positive integer $c > 0$. Then there exists a non-adaptive test design and decoder that outputs an estimate $\mathcal{I} \in \{0, 1, \dots, p\}$ of a definite defective (with 0 representing "I don't know") satisfying*

$$\mathbb{P}[\mathcal{I} > 0 \cap \mathcal{I} \notin S] = 0 \quad \text{and} \quad \mathbb{P}[\mathcal{I} = 0] \leq (1 - e^{-1} + o(1))^c \tag{11}$$

as $k \to \infty$ and $p \to \infty$.

Proof. We first consider the case $c = 1$. Let $\mathcal{A}$ be a uniformly random set of $\frac{p}{k}$ items.[3] By the assumption that $k \to \infty$ and $k = o(p)$, it is straightforward to show that

$$\mathbb{P}[|\mathcal{A} \cap S| = 1] = e^{-1} + o(1). \tag{12}$$

Indeed, the analogous claim is standard when each item is included in $\mathcal{A}$ with probability $\frac{1}{k}$ [1, Sec. 2.3], and (12) can then by understood by approximating the Hypergeometric distribution by the binomial distribution [34].

We proceed by describing a procedure from the SAFFRON algorithm of [12] that is guaranteed to identify the single defective item in $\mathcal{A}$ whenever $|\mathcal{A} \cap S| = 1$, while also being able to identify with certainty whether $|\mathcal{A} \cap S| = 1$ or $|\mathcal{A} \cap S| \neq 1$. This procedure uses $2v$ tests, where $v = \lceil \log_2 \frac{p}{k} \rceil = (1 + o(1)) \log_2 \frac{p}{k}$.

Number the items in $\mathcal{A}$ from 0 to $\frac{p}{k} - 1$ in a fixed manner (e.g., maintaining the order that they take as items in $\{1, \dots, n\}$). For $i \in \{0, 1, \dots, \frac{p}{k} - 1\}$, let $\mathbf{b}_i \in \{0,1\}^{2v}$ be a binary vector of length $2v$ and weight v constructed as follows: The first v entries are the number i written in binary, and the last v entries are the same, but with the 0s and 1s swapped. We then construct $2v$ tests, where test j contains exactly the items corresponding to $i \in \mathcal{A}$ for which the j-th entry of $\mathbf{b}_i$ equals 1.

We now consider the following cases:

- If $\mathcal{A}$ contains no defective items, then all $2v$ tests will be negative. When this is observed, we set $\mathcal{I} = 0$.
- If $\mathcal{A}$ contains exactly one defective, then exactly v of the tests will be positive. When this is observed, we set $\mathcal{I}$ to be item whose value $i \in \mathcal{A}$ is spelled out (in binary) by the first v test results.
- If $\mathcal{A}$ contains two or more defectives, then more than v of the tests will be positive. When this is observed, we set $\mathcal{I} = 0$.

The first and third cases ensure that we never erroneously set $\mathcal{I} \neq 0$. In addition, we correctly identify a defective item in the second case, which occurs with probability $e^{-1} + o(1)$ due to (12). This proves the theorem for $c = 1$.

To handle $c > 1$, we simply repeat the preceding process c times, drawing $\mathcal{A}$ independently each time. By doing so, we only fail if none of the sets $\mathcal{A}$ contain exactly one defective item, which occurs with probability $(1 - e^{-1} + o(1))^c$. □

We briefly mention that the $O\big(\log \frac{p}{k}\big)$ scaling in Theorem 3 cannot be improved further. To see this, suppose by contradiction that a definite defective could be found with constant (non-zero) probability using $o\big(\log \frac{p}{k}\big)$ tests. By repeating this procedure with uniformly random shuffling of the items, ignoring any $\mathcal{I} = 0$ outcomes and removing any defectives identified, an adaptive group testing algorithm could recover the full defective set using $o\big(k \log \frac{p}{k}\big)$ tests on average. However, this would violate the standard $\Omega\big(k \log \frac{p}{k}\big)$ lower bound [1, Sec 1.4], which holds even in the adaptive setting.

III. All-or-Nothing Behavior for Bernoulli Designs

To establish the an all-or-nothing threshold for weak recovery under Bernoulli designs, it is useful to first study the weak detection problem as a stepping stone.

A. Weak Detection

Recall from Section II-C that the goal of weak detection is to distinguish $P(\mathbf{X}, \mathbf{Y})$ (defined according to the group testing model) from $Q(\mathbf{X}, \mathbf{Y}) := P(\mathbf{X})P(\mathbf{Y})$, with $P(\mathbf{X})$ being the i.i.d. Bernoulli distribution using the choice of ν in (1), and $P(\mathbf{Y}) = 2^{-n}$.

For concreteness, we consider the Bayesian setting, where the observed pair $(\mathbf{X}, \mathbf{Y})$ is drawn from P or Q with probability $\frac{1}{2}$ each. The resulting error probability of the hypothesis test is denoted by P_{e}. Trivially, choosing the hypothesis via a random guess gives $\mathsf{P}_{\mathrm{e}} = \frac{1}{2}$. It is a standard result in binary hypothesis testing that if $d_{\mathrm{TV}}(P, Q) \to 0$ as $p \to \infty$, then one cannot do better than random guessing asymptotically, i.e., it is impossible do better than $\mathsf{P}_{\mathrm{e}} = \frac{1}{2} + o(1)$ (e.g., see [35, Sec. 2.3.1]).

Let $D(P\|Q)$ denote the KL divergence between P and Q. By Pinsker's inequality, $D(P\|Q) \to 0$ implies $d_{\mathrm{TV}}(P, Q) \to 0$, and in addition, $D(P\|Q) \leq \chi^2(P\|Q)$ [36], where we define the χ^2 divergence

$$\begin{aligned}\chi^2(P\|Q) &= \mathbf{E}_Q\left[\left(\frac{P(\mathbf{X},\mathbf{Y})}{Q(\mathbf{X},\mathbf{Y})} - 1\right)^2\right] \\ &= \mathbf{E}_Q\left[\left(\frac{P(\mathbf{X},\mathbf{Y})}{Q(\mathbf{X},\mathbf{Y})}\right)^2\right] - 1 = \mathbf{E}_P\left[\frac{P(\mathbf{X},\mathbf{Y})}{Q(\mathbf{X},\mathbf{Y})}\right] - 1. \end{aligned} \tag{13}$$

[3] We assume for simplicity that $\frac{p}{k}$ is an integer; the general case follows similarly by rounding.

Hence, to prove a hardness result for distinguishing P from Q, it suffices to show that $\chi^2(P\|Q) \to 0$ as $p \to \infty$. The following theorem gives conditions under which this is the case.

Theorem 4. (Negative Result for Weak Detection) *Consider the probabilistic group testing problem with Bernoulli random testing using the choice of ν in (1), and suppose that $n \le (1-\eta)\log_2\binom{p}{k}$ for some $\eta \in (0,1)$. Then, when $k = o(p^{\frac{\eta}{1+\eta}})$ as $p \to \infty$, we have*

$$\chi^2(P\|Q) = \frac{1}{\binom{p}{k}} \sum_{l=0}^{k} \binom{k}{l}\binom{p-k}{l} 2^{n(1-\frac{l}{k})} - 1 \to 0 \quad (14)$$

as $p \to \infty$. Hence, the smallest possible error probability for the binary hypothesis test between $P(\mathbf{X},\mathbf{Y})$ and $Q(\mathbf{X},\mathbf{Y}) = P(\mathbf{X})(\mathbf{Y})$ behaves as $\mathsf{P}_{\mathrm{e}} = \frac{1}{2} + o(1)$.

Proof. The proof is based on substituting the distributions P and Q into (13) and performing asymptotic simplifications. The details are given in Appendix B. □

The condition $k = o(p^{\frac{\eta}{1+\eta}})$ holds when k grows sufficiently slowly with respect to p, e.g., it holds for arbitrarily small η when $k = \mathrm{poly}(\log p)$. On the other hand, it remains open as to whether a similar hardness result can be proved when k grows faster than $\Theta(p^{\frac{\eta}{1+\eta}})$. To address this question, we note the following:

- Theorem 2 above shows that P and Q can be reliably distinguished when $n \ge (1-\eta)\log_2\binom{p}{k}$ and $k = p^{\frac{1+\eta}{3+\eta}+\Omega(1)}$, which essentially reduces to $k \gg p^{1/3}$ for small η. Thus, $\log_2\binom{p}{k}$ no longer serves as an all-or-nothing threshold in this denser regime. This provides an interesting point of contrast with the analogous sparse linear regression problem [26], where the analogous hardness result to Theorem 4 holds for all $k = O(\sqrt{p})$. In addition, [26, App. C] provides a positive result showing that this threshold is tight.
- Theorem 5 below shows that the above χ^2-divergence does not approach zero when $n \ge (1-\eta)\log_2\binom{p}{k}$ and $k = \Omega\big(p^{\frac{\eta}{1+\eta}}\big)$. Note that χ^2-divergence approaching zero is sufficient, but not necessary, for establishing the hardness of distinguishing P from Q. Hence, this result does not establish such hardness, but it does show that any proof establishing hardness must move beyond the approach of bounding $\chi^2(P\|Q)$.
  In the sparse linear regression problem, a similar limitation regarding the χ^2 divergence is overcome by conditioning out certain "catastrophic" low-probability events [26] that blow up the divergence. Unfortunately, it appears to be difficult to identify an analogous event in the group testing problem.

Formally, the second of these is stated as follows.

Theorem 5. (A Condition for Non-Vanishing χ^2-Divergence) *Consider the probabilistic group testing problem with Bernoulli random testing using the choice of ν in (1), and suppose that $n = (1-\eta)\log_2\binom{p}{k}$ for some $\eta \in (0,1)$. Then, when $k = \Omega(p^{\frac{\eta}{1+\eta}})$ as $p \to \infty$, we have*

$$\liminf_{p\to\infty} \chi^2(P\|Q) \ge c \quad (15)$$

for some constant $c > 0$.

Proof. See Appendix B. □

B. Weak Recovery

We are now ready to present our main negative result concerning the weak recovery criterion under Bernoulli testing, and establishing an all-or-nothing threshold at $n \sim k\log_2\frac{p}{k}$.

Theorem 6. (Negative Result for Weak Recovery) *Consider the probabilistic group testing problem with Bernoulli random testing using the choice of ν in (1), and suppose that $n \le (1-\eta)\log_2\binom{p}{k}$ for some $\eta \in (0,1)$. Then, when $k \to \infty$ with $k = o(p^{\frac{\eta}{1+\eta}})$ as $p \to \infty$, for any fixed $\alpha \in (0,1)$, any decoder that outputs some estimate $\widehat{S}$ of S must yield the following as $p \to \infty$:*

$$\mathbb{P}\Big[\max\big\{|\widehat{S}\setminus S|, |S\setminus\widehat{S}|\big\} \le \alpha k\Big] \to 0 \quad (16)$$

Proof. The proof is outlined as follows. Following an idea used for sparse linear regression in [26], we study the mutual information $I(S;\mathbf{Y},Y'|\mathbf{X},X')$, where (X',Y') is an additional test independent from $(\mathbf{X},\mathbf{Y})$. Combining some manipulations of information terms with the weak detection result of Theorem 4, we show that $H(Y'|\mathbf{X},X',\mathbf{Y}) \to \log 2$ when $n \le (1-\eta)\log_2\binom{p}{k}$. On the other hand, we show that if weak recovery were possible, we would be able to predict Y' given $(\mathbf{X},X',\mathbf{Y})$ better than random guessing, meaning that $H(Y'|\mathbf{X},X',\mathbf{Y})$ would be bounded away from $\log 2$. Combining these observations, we deduce that weak recovery must be impossible. The details are given in Appendix B. □

Hence, when k is sufficiently sparse so that $k = o(p^{\frac{\eta}{1+\eta}})$ holds for any $\eta > 0$ (e.g., $k = \mathrm{poly}(\log p)$), the threshold $n^* = k\log_2\frac{p}{k}$ serves as an exact threshold between *complete success* and *complete failure*. We note that in previous works, phase transitions were proved in [13], [19], [20] regarding the error probability of *recovering the exact defective set*, whereas Theorem 6 gives the much stronger statement that one cannot even identify a small fraction of the defective set.

C. Identifying a Definite Defective

The following negative result for identifying a definite defective (see Section II-D) follows in a straightforward manner from our negative result for weak detection (Theorem 4).

Theorem 7. (Negative Result for Identifying a Definite Defective) *Consider the probabilistic group testing problem with Bernoulli random testing using the choice of ν in (1), and suppose that $n \le (1-\eta)\log_2\binom{p}{k}$ for some $\eta \in (0,1)$. Then, when $k = o(p^{\frac{\eta}{1+\eta}})$ as $p \to \infty$, for any decoder that outputs an estimate $\mathcal{I} \in \{0,1,\dots,p\}$ of a definite defective (with 0 representing "I don't know"), we have*

$$\mathbb{P}[\mathcal{I} > 0 \cap \mathcal{I} \notin S] = 0 \implies \mathbb{P}[\mathcal{I} = 0] \to 1 \quad (17)$$

as $p \to \infty$.

Proof. The idea is to note that since $\mathbb{P}[\mathcal{I} > 0 \cap \mathcal{I} \notin S] = 0$, the decoder must output $\mathcal{I} = 0$ whenever there exists some S' of cardinality k that is *disjoint from* S and consistent with the test outcomes. Using de Caen's bound [37], we show that this holds with probability at least $\frac{1}{1+\chi^2(P\|Q)}$, and the theorem follows since $\chi^2(P\|Q) \to 0$ due to Theorem 4. The details are given in Appendix B. □

D. Discussion: Bernoulli vs. General Designs

We wrap up this section and Section II by highlighting that for all three of the recovery criteria considered, Bernoulli designs can be highly suboptimal compared to general designs:

- Weak recovery is possible (for general test designs) when $n = \Omega\big(k \log \frac{n}{k}\big)$ with an arbitrarily small implied constant, whereas Bernoulli testing requires $n = (1 + o(1))k \log \frac{n}{k}$ in sufficiently sparse regimes;
- A trivial strategy exists for weak detection with only $O(1)$ tests (for constant error probability), whereas Bernoulli testing requires $n = (1 + o(1))k \log \frac{n}{k}$ in sufficiently sparse regimes;
- Identifying a definite defective is possible with $O(\log n)$ tests, whereas Bernoulli testing requires $n = (1 + o(1))k \log \frac{n}{k}$ in sufficiently sparse regimes.

Since Bernoulli designs are a prototypical example of an unstructured random design, these results indicate that despite their strong theoretical guarantees (e.g., as established in [13]), significant care should be taken in adopting them when too few tests are available, or when relaxed recovery criteria are considered to be acceptable. Essentially, in such cases, it is much better to "cut one's losses" early on via a hand-crafted design that targets the less stringent recovery criterion under consideration.

IV. Conclusion

We have established the fundamental limits of noiseless non-adaptive group testing under the weak detection criterion, and more generally approximate recovery, and shown Bernoulli designs to be highly suboptimal (in sufficiently sparse regimes) in the sense of exhibiting an all-or-nothing threshold at $n \sim k \log_2 \frac{p}{k}$. In addition, we gave similar negative results for Bernoulli testing under the criteria of weak detection and identifying a definite defective, while establishing that non-Bernoulli designs can attain these goals with very few tests.

Possible directions for future work include (i) closing the remaining gaps in between the positive and negative results in the regime $k = \Theta(p^\theta)$ with $\theta \in \big(0, \frac{1}{3}\big)$; (ii) handling Bernoulli$\big(\frac{\nu}{k}\big)$ designs with more general choices of ν; (iii) providing analogous results for the near-constant tests-per-item design [19], [23]; (iv) finding a design that attains exact or near-exact recovery with $n = (1+o(1))\big(k \log_2 \frac{p}{k}\big)$ tests while degrading gracefully below this threshold; and (v) developing analogous results under random noise models [14], [18], [38].

Appendix A
Proof of Theorem 2 (Positive Result for Weak Detection with Bernoulli Designs)

By the assumption $k = o(p)$, we have $\log_2 \binom{p}{k} = \big(k \log_2 \frac{p}{k}\big)(1 + o(1))$. Hence, it suffices to prove the theorem for $n = (1-\eta)k \log_2 \frac{p}{k}$, since we can incorporate the $1+o(1)$ term into δ.

In the following, we use the terminology that the j-th column $\mathbf{X}_j$ of $\mathbf{X}$ is *covered* by $\mathbf{Y}$ if the support of $\mathbf{X}_j$ is a subset of the support of $\mathbf{Y}$ (i.e., whenever the i-th entry of $\mathbf{X}_j$ is 1, the outcome y_i is also 1). We consider distinguishing models P and Q by counting the number of columns of $\mathbf{X}$ that are covered by $\mathbf{Y}$.

Fix a constant $\zeta \in (0,1)$, and consider the "typical" set $\mathcal{T}$ containing all the sequences $\mathbf{y} \in \{0,1\}^n$ such that the number of positive tests is between $\frac{n(1-\zeta)}{2}$ and $\frac{(1+\zeta)n}{2}$. By the law of large numbers, we have $\mathbb{P}[\mathbf{Y} \notin \mathcal{T}] \to 0$ as $p \to \infty$. Given any sequence $\mathbf{y} \in \mathcal{T}$, when an independent random column $\mathbf{X}_j$ is generated, the (conditional) probability q_0 of it being covered satisfies

$$\left(1 - \frac{\nu}{k}\right)^{\frac{(1+\zeta)n}{2}} \le q_0 \le \left(1 - \frac{\nu}{k}\right)^{\frac{(1-\zeta)n}{2}}. \tag{18}$$

For $n = (1-\eta)k \log_2 \frac{p}{k}$, recalling the choice of ν in (1) (which yields $P(\mathbf{Y}) = 2^{-n}$), we have

$$\left(1 - \frac{\nu}{k}\right)^{\frac{(1-\zeta)n}{2}} = \left(\frac{1}{2}\right)^{\frac{(1-\zeta)n}{2k}} = \left(\frac{k}{p}\right)^{\frac{(1-\zeta)(1-\eta)}{2}}, \tag{19}$$

and similarly

$$\left(1 - \frac{\nu}{k}\right)^{\frac{(1+\zeta)n}{2}} = \left(\frac{k}{p}\right)^{\frac{(1+\zeta)(1-\eta)}{2}}. \tag{20}$$

Hence, we have

$$\left(\frac{k}{p}\right)^{\frac{(1+\zeta)(1-\eta)}{2}} \le q_0 \le \left(\frac{k}{p}\right)^{\frac{(1-\zeta)(1-\eta)}{2}}. \tag{21}$$

Then, the distribution of the number $\tilde{N}(\mathbf{X}, \mathbf{y})$ of covered columns under the two hypotheses is given as follows:

- Under P: $\tilde{N}(\mathbf{X}, \mathbf{y}) \sim \big(k + \text{Binomial}(p-k, q_0)\big)$, where the addition of k is due to the fact that the defective items' columns are almost surely covered due to the definition of the group testing model.
- Under Q: $\tilde{N}(\mathbf{X}, \mathbf{y}) \sim \text{Binomial}(p, q_0)$.

We consider the following procedure for distinguishing these two hypotheses:

$$\tilde{N}(\mathbf{X}, \mathbf{Y}) \underset{Q}{\overset{P}{\gtrless}} pq_0 + \frac{k}{2}. \tag{22}$$

Then, given $\mathbf{y}$, the error probability $\mathsf{P}_{\mathrm{e}}(\mathbf{y})$ with a uniform prior satisfies

$$\mathsf{P}_{\mathrm{e}}(\mathbf{y}) = \frac{1}{2}\mathbb{P}_Q\left[\tilde{N}(\mathbf{X}, \mathbf{y}) > pq_0 + \frac{k}{2}\right] + \frac{1}{2}\mathbb{P}_P\left[\tilde{N}(\mathbf{X}, \mathbf{y}) < pq_0 + \frac{k}{2}\right]. \tag{23}$$

For the first term in (23), observe that by the Berry-Esseen Theorem [39] (see Corollary 2 in Appendix C), we have

$$\mathbb{P}_Q\left[\tilde{N}(\mathbf{X},\mathbf{y}) \geq pq_0 + \frac{k}{2}\right]$$
$$= \mathbb{P}_Q\left[\frac{\tilde{N}(\mathbf{X},\mathbf{Y}) - pq_0}{\sqrt{p(1-q_0)q_0}} \geq \frac{k}{2\sqrt{p(1-q_0)q_0}}\right] \tag{24}$$
$$\leq \mathcal{Q}\left(\frac{k}{2\sqrt{p(1-q_0)q_0}}\right) + \frac{6\rho}{\sigma^3\sqrt{p}}, \tag{25}$$

where $\mathcal{Q}(t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-u^2/2}\,\mathrm{d}u$ denotes the standard Gaussian upper tail probability function, and as also shown in Appendix C, the relevant moments are

$$\rho = (1-q_0)^3 q_0 + q_0^3(1-q_0), \tag{26}$$
$$\sigma = \sqrt{(1-q_0)q_0}. \tag{27}$$

The ratio appearing in (25) can be simplified as follows:

$$\frac{6\rho}{\sigma^3\sqrt{p}} = \frac{6(1-q_0)^3 q_0 + 6q_0^3(1-q_0)}{(\sqrt{(1-q_0)q_0})^3\sqrt{p}} \tag{28}$$
$$= \frac{6(1-q_0)^2 q_0 + 6q_0^3}{q_0^{3/2}\sqrt{(1-q_0)p}} \tag{29}$$
$$\leq \frac{12q_0}{q_0^{3/2}\sqrt{(1-q_0)p}} \tag{30}$$
$$= \frac{12}{\sqrt{pq_0(1-q_0)}}. \tag{31}$$

Similarly, again using the Berry-Essen theorem, and writing $\tilde{N}$ in place of $\tilde{N}(\mathbf{X},\mathbf{Y})$ for brevity, we have

$$\mathbb{P}_P\left[\tilde{N} \leq pq_0 + \frac{k}{2}\right]$$
$$= \mathbb{P}_P\left[\tilde{N} - k - (p-k)q_0 \leq pq_0 + \frac{k}{2} - k - (p-k)q_0\right] \tag{32}$$
$$= \mathbb{P}_P\left[\frac{\tilde{N} - k - (p-k)q_0}{\sqrt{(p-k)(1-q_0)q_0}} \leq \frac{q_0 k - \frac{k}{2}}{\sqrt{(p-k)(1-q_0)q_0}}\right] \tag{33}$$
$$\leq 1 - \mathcal{Q}\left(\frac{q_0 k - \frac{k}{2}}{\sqrt{(p-k)(1-q_0)q_0}}\right) + \frac{6\rho}{\sigma^3\sqrt{p-k}}, \tag{34}$$

and since $k = o(p)$, we have from (31) that $\frac{6\rho}{\sigma^3\sqrt{p-k}} \leq \frac{12}{\sqrt{pq_0(1-q_0)}}(1+o(1))$.

We know from (21) that $q_0 \to 0$, and combining this with $k = o(p)$, we see from (31) and (34) that $\mathsf{P}_\mathrm{e} \to 0$ as long as $pq_0 \to \infty$ and $k = \omega(\sqrt{pq_0})$. The condition $pq_0 \to \infty$ follows as an immediate consequence of (21) (with $k = o(p)$ and $\delta < 1$). In addition, again using (21), we find that the condition $k = \omega\big(\sqrt{pq_0}\big)$ is satisfied if

$$k = \omega\left(\sqrt{p}\left(\frac{k}{p}\right)^{\frac{(1-\eta)(1-\zeta)}{4}}\right). \tag{35}$$

Letting $a = (1-\eta)(1-\zeta)$, we find that this condition holds if $k^{1-\frac{a}{4}} = \omega\big(p^{\frac{1}{2}-\frac{a}{4}}\big)$, which simplifies to $k = \omega\big(p^{\frac{2-a}{4-a}}\big)$. Substituting $a = (1-\eta)(1-\zeta)$, and recalling that η is constant and ζ is arbitrarily small, we find that the preceding condition holds as long as

$$k = \Omega\big(p^{\frac{1+\eta}{3+\eta}+\delta}\big) \tag{36}$$

for arbitrarily small $\delta \in (0,1)$. This completes the proof of Theorem 2.

Appendix B
Proofs of Negative Results for Bernoulli Designs

A. Preliminary Calculations

Since the group testing model P and the null model Q have the same $\mathbf{X}$ distribution, and the null model assigns probability 2^{-n} to each $\mathbf{Y}$ sequence (due to the choice of ν in (1)), we have

$$\frac{P(\mathbf{X},\mathbf{Y})}{Q(\mathbf{X},\mathbf{Y})} = 2^n P(\mathbf{Y}|\mathbf{X}) = 2^n \sum_S \frac{1}{\binom{p}{k}} P(\mathbf{Y}|\mathbf{X},S), \tag{37}$$

where here and subsequently, the summation over S is implicitly over all $\binom{p}{k}$ subsets of $\{1,\dots,n\}$ of cardinality k. Since the observation model defining P is deterministic, $P(\mathbf{Y}|\mathbf{X},S)$ is simply 1 if $\mathbf{Y}$ is consistent with $(\mathbf{X},S)$ (according to $Y^{(i)} = \bigvee_{j\in S} X_j^{(i)}$), and 0 otherwise. Letting $\mathcal{I}_S(\mathbf{X},\mathbf{Y})$ be the corresponding indicator function, we rewrite (37) as

$$\frac{P(\mathbf{X},\mathbf{Y})}{Q(\mathbf{X},\mathbf{Y})} = \frac{2^n}{\binom{p}{k}} \sum_S \mathcal{I}_S(\mathbf{X},\mathbf{Y}), \tag{38}$$

and take the square to obtain

$$\left(\frac{P(\mathbf{X},\mathbf{Y})}{Q(\mathbf{X},\mathbf{Y})}\right)^2 = \frac{4^n}{\binom{p}{k}^2} \sum_{S,S'} \mathcal{I}_S(\mathbf{X},\mathbf{Y})\mathcal{I}_{S'}(\mathbf{X},\mathbf{Y}). \tag{39}$$

Taking the average over $(\mathbf{X},\mathbf{Y}) \sim Q$ and using the middle form in (13), we obtain

$$\chi^2(P\|Q) = \frac{4^n}{\binom{p}{k}^2} \sum_{S,S'} \mathbb{E}_Q\big[\mathcal{I}_S(\mathbf{X},\mathbf{Y})\mathcal{I}_{S'}(\mathbf{X},\mathbf{Y})\big] - 1. \tag{40}$$

The average here is the probability that a randomly generated $(\mathbf{X},\mathbf{Y})$ (independent of each other) is consistent with both S and S'. By the symmetry of $(\mathbf{X},\mathbf{Y})$ with respect to re-labeling items, we can assume without loss of generality that S equals the set

$$S_0 = \{1,2,\dots,k\}, \tag{41}$$

and average over S' alone; by splitting into S' with ℓ entries in $\{k+1,\dots,p\}$ (non-overlapping with S_0) and $k-\ell$ entries in $\{1,\dots,k\}$ (overlapping with S_0), we obtain

$$\chi^2(P\|Q)$$
$$= \frac{4^n}{\binom{p}{k}} \sum_{\ell=0}^{k} \binom{k}{\ell}\binom{p-k}{\ell} \mathbb{E}_Q\big[\mathcal{I}_{S_0}(\mathbf{X},\mathbf{Y})\mathcal{I}_{S'}(\mathbf{X},\mathbf{Y})\big] - 1, \tag{42}$$

where S' implicitly satisfies $|S_0 \cap S'| = k - \ell$ in the expectation.

Now observe that $\mathbb{E}_Q\big[\mathcal{I}_{S_0}(\mathbf{X},\mathbf{Y})\mathcal{I}_{S'}(\mathbf{X},\mathbf{Y})\big]$ is the probability (with respect to Q) that every one of the n tests satisfies *any one* of the following:

- The test outcome is negative, and all $k+\ell$ items from $S_0 \cup S'$ are excluded;
- The test outcome is positive, and at least one item from $S_0 \cap S'$ is included;
- The test outcome is positive, and no items from $S_0 \cap S'$ are included, but at least one item from each of $S_0 \setminus S'$ and $S' \setminus S_0$ are included.

For a single test, we characterize the probabilities of these three events under Q as follows follows (recalling (1)):

- The first event has probability $\frac{1}{2}\big(1-\frac{\nu}{k}\big)^{k+\ell} = \frac{1}{2}\cdot\big(\frac{1}{2}\big)^{\frac{k+\ell}{k}} = \big(\frac{1}{2}\big)^{2+\frac{\ell}{k}}$;
- The union of the second and third events above can be reformulated as the event the test outcome is positive and *none* of the following events occur: (i) All items from $S_0 \cup S'$ are excluded; (ii) All items from S_0 are excluded, but at least one from $S' \setminus S_0$ is included; (iii) All items from S' are excluded, but at least one from $S_0 \setminus S'$ is included. Using this formulation, the union of the second and third events above has probability $\frac{1}{2}\big[1-\big(1-\frac{\nu}{k}\big)^{k+\ell} - 2\cdot\frac{1}{2}\cdot\big(1-\big(1-\frac{\nu}{k}\big)^{\ell}\big)\big] = \frac{1}{2}\big[1-\big(\frac{1}{2}\big)^{1+\frac{\ell}{k}} - \big(1-\big(\frac{1}{2}\big)^{\frac{\ell}{k}}\big)\big] = \big(\frac{1}{2}\big)^{1+\frac{\ell}{k}} - \big(\frac{1}{2}\big)^{2+\frac{\ell}{k}} = \big(\frac{1}{2}\big)^{2+\frac{\ell}{k}}$.

Summing these two probabilities together gives an overall probability of $2\cdot\big(\frac{1}{2}\big)^{2+\frac{\ell}{k}} = \big(\frac{1}{2}\big)^{1+\frac{\ell}{k}}$ associated with a single test. Since the tests are independent, taking the intersection of the corresponding n events gives

$$\mathbb{E}_Q\big[\mathcal{I}_{S_0}(\mathbf{X},\mathbf{Y})\mathcal{I}_{S'}(\mathbf{X},\mathbf{Y})\big] = 2^{-n(1+\frac{\ell}{k})}, \tag{43}$$

and substitution into (42) yields

$$\chi^2(P\|Q) = \frac{4^n}{\binom{p}{k}}\sum_{\ell=0}^{k}\binom{k}{\ell}\binom{p-k}{\ell}2^{-n(1+\frac{\ell}{k})} - 1. \tag{44}$$

B. Proof of Theorem 4 (Impossibility of Weak Detection for $n \le (1-\eta)\log_2\binom{p}{k}$)

Recall the setup of weak detection described in Section II-B, and that we consider i.i.d. Bernoulli designs with ν chosen via (1). We first prove the following lemma, which provides an upper bound on the χ^2-divergence.

Lemma 2. *Assume that $n \le (1-\eta)\log_2\binom{p}{k}$ for some $\eta \in (0,1)$. Then, it holds that*

$$\chi^2(P\|Q) \le \exp\left[e^{1-\eta}k\left(\frac{k}{p}\right)^{\eta}\frac{p}{p-k+1}\right] - 1. \tag{45}$$

Proof. Using the assumption $n \le (1-\eta)\log_2\binom{p}{k}$, we bound (44) as follows:

$$\begin{aligned}
&\chi^2(P\|Q)\\
&= \frac{1}{\binom{p}{k}}\sum_{l=0}^{k}\binom{k}{l}\binom{p-k}{l}2^{n(1-\frac{l}{k})} - 1 && (46)\\
&\le \frac{1}{\binom{p}{k}}\sum_{l=0}^{k}\binom{k}{l}\binom{p-k}{l}2^{(1-\frac{l}{k})(1-\eta)\log_2\binom{p}{k}} - 1 && (47)\\
&= \frac{1}{\binom{p}{k}}\sum_{l=0}^{k}\binom{k}{l}\binom{p-k}{l}\binom{p}{k}^{(1-\eta)(1-\frac{l}{k})} - 1 && (48)\\
&= \binom{p}{k}^{-\eta}\sum_{l=0}^{k}\binom{k}{l}\binom{p-k}{l}\binom{p}{k}^{-(1-\eta)\frac{l}{k}} - 1. && (49)
\end{aligned}$$

In addition, for all $l < k$, we have

$$\begin{aligned}
\frac{\binom{p-k}{l}}{\binom{p}{k}} &\le \frac{\binom{p}{l}}{\binom{p}{k}} && (50)\\
&= \frac{k!(p-k)!}{l!(p-l)!} && (51)\\
&= \frac{k(k-1)\cdots(l+1)}{(p-l)(p-l-1)\cdots(p-k+1)} && (52)\\
&\le \left(\frac{k}{p-k+1}\right)^{k-l}. && (53)
\end{aligned}$$

Since (53) also holds for $l = k$, it follows that

$$\frac{\binom{p-k}{l}}{\binom{p}{k}} \le \left(\frac{k}{p-k+1}\right)^{k-l} \tag{54}$$

for all $0 \le l \le k$.

From (49) and (54), we obtain

$$\begin{aligned}
&\chi^2(P\|Q)\\
&\le \binom{p}{k}^{1-\eta}\sum_{l=0}^{k}\binom{k}{l}\left(\frac{k}{p-k+1}\right)^{k-l}\binom{p}{k}^{-(1-\eta)\frac{l}{k}} - 1 && (55)\\
&= \binom{p}{k}^{1-\eta}\left(\frac{k}{p-k+1}\right)^{k}\\
&\qquad\times\sum_{l=0}^{k}\binom{k}{l}\left(\frac{k}{p-k+1}\right)^{-l}\binom{p}{k}^{-(1-\eta)\frac{l}{k}} - 1 && (56)\\
&= \binom{p}{k}^{1-\eta}\left(\frac{k}{p-k+1}\right)^{k}\left[1+\frac{p-k+1}{k}\binom{p}{k}^{-\frac{1-\eta}{k}}\right]^{k} - 1 && (57)\\
&= \left[\binom{p}{k}^{\frac{1-\eta}{k}}\left(\frac{k}{p-k+1}\right)+1\right]^{k} - 1 && (58)\\
&\le \left[\left(\frac{pe}{k}\right)^{1-\eta}\left(\frac{k}{p-k+1}\right)+1\right]^{k} - 1 && (59)\\
&\le \exp\left[k\left(\frac{pe}{k}\right)^{1-\eta}\left(\frac{k}{p-k+1}\right)\right] - 1 && (60)\\
&= \exp\left[e^{1-\eta}k\left(\frac{k}{p}\right)^{\eta}\frac{p}{p-k+1}\right] - 1 && (61)
\end{aligned}$$

where (57) follows from the fact that $\sum_{j=0}^{k}\binom{k}{j}x^j = (1+x)^k$, (59) follows from $\binom{p}{k} \le \left(\frac{pe}{k}\right)^k$, and (60) uses $1+x \le e^x$. This proves (45). □

To prove Theorem 4, it suffices to show that the right-hand side of (45) tends to zero as $p \to \infty$. To see this, observe that the condition $k = o(p^{\frac{\eta}{1+\eta}})$ can equivalently be written as $k\left(\frac{k}{p}\right)^\eta = o(1)$, and this condition implies that

$$0 \le \chi^2(P\|Q) \le \exp\left[e^{1-\eta}k\left(\frac{k}{p}\right)^\eta \frac{p}{p-k+1}\right] - 1 \to 0 \tag{62}$$

as $p \to \infty$. This means that $\chi^2(P\|Q) \to 0$ when $k = o(p^{\frac{\eta}{1+\eta}})$, which proves Theorem 4.

C. Proof of Theorem 5 (A Condition for Non-Vanishing χ^2-Divergence)

Here we prove that $\liminf_{p\to\infty}\chi^2(P\|Q) \ge c$ for some $c > 0$ when $n \ge (1-\eta)\log_2\binom{p}{k}$ under the assumptions $k = o(p)$ and $k = \Omega(p^{\frac{\eta}{1+\eta}})$.[4]

As mentioned in Section III-A, $\chi^2(P\|Q) \to 0$ implies that $d_{\mathrm{TV}}(P,Q) \to 0$, which in turn implies that weak detection is impossible for any algorithm. However, Theorem 2 shows that weak detection is possible when $k = \Omega(p^{\frac{1+\eta}{3+\eta}+\delta})$ for arbitrarily small $\delta \in (0,1)$. Note that that $\frac{1+\eta}{3+\eta} < \frac{1}{2}$ for any $\eta \in [0,1)$. Thus, it must be the case that $\liminf_{p\to\infty}\chi^2(P\|Q) \ge c$ whenever $k = \Omega(\sqrt{p})$, since otherwise $\chi^2(P\|Q) = o(1)$ would be a contradiction. In the following, we therefore only consider $k = o(\sqrt{p})$.

We need to prove that $\liminf_{p\to\infty}\chi^2(P\|Q) \ge c$ for some $c > 0$ when $n \ge (1-\eta)\log_2\binom{p}{k}$ under the assumptions $k = o(\sqrt{p})$ and $k = \Omega(p^{\frac{\eta}{1+\eta}})$. Following (46)–(49) with the inequality in (47) reversed, we have

$$\chi^2(P\|Q) \ge \binom{p}{k}^{-\eta}\sum_{l=0}^{k}\binom{k}{l}\binom{p-k}{l}\binom{p}{k}^{-(1-\eta)\frac{l}{k}} - 1. \tag{63}$$

Further lower bounding the right-hand side by taking only the terms $l \in \{k-1, k\}$, we obtain

$$\begin{aligned}
\chi^2(P\|Q) &\ge -1 + \frac{\binom{p-k}{k}}{\binom{p}{k}} + k\frac{\binom{p-k}{k-1}}{\binom{p}{k}}\binom{p}{k}^{\frac{1-\eta}{k}} && (64)\\
&= -1 + \frac{\binom{p-k}{k}}{\binom{p}{k}} + \frac{k^2}{p-2k+1}\frac{\binom{p-k}{k}}{\binom{p}{k}}\binom{p}{k}^{\frac{1-\eta}{k}} && (65)\\
&= -1 + \frac{\binom{p-k}{k}}{\binom{p}{k}}\left[1 + \frac{k^2}{p-2k+1}\binom{p}{k}^{\frac{1-\eta}{k}}\right] && (66)\\
&\ge -1 + \frac{\binom{p-k}{k}}{\binom{p}{k}}\left[1 + \frac{k^2}{p-2k+1}\left(\frac{p}{k}\right)^{1-\eta}\right] && (67)\\
&= -1 + \frac{\binom{p-k}{k}}{\binom{p}{k}}\left[1 + \frac{k^{1+\eta}}{p^\eta}\left(\frac{p}{p-2k+1}\right)\right], && (68)
\end{aligned}$$

[4]We are grateful to an anonymous reviewer for helping us to significantly simplify this proof.

where (65) uses $\binom{p-k}{k-1} = \binom{p-k}{k}\frac{k}{p-2k+1}$, and (67) follows since $\binom{p}{k} \ge \left(\frac{p}{k}\right)^k$. To bound the ratio of binomial coefficients in (68), observe that

$$\begin{aligned}
1 &\ge \frac{\binom{p-k}{k}}{\binom{p}{k}} && (69)\\
&= \frac{(p-k)(p-k-1)\cdots(p-2k+1)}{p(p-1)\cdots(p-k+1)} && (70)\\
&> \left(1 - \frac{2k}{p}\right)^k && (71)\\
&\ge 1 - \frac{2k^2}{p} && (72)
\end{aligned}$$

where (72) follows since $(1+x)^k \ge 1+kx$ for all $x > -1$. Hence, since we are considering $k = o(\sqrt{p})$, it follows that $\frac{\binom{p-k}{k}}{\binom{p}{k}} = 1+o(1)$. Combining this with (68) and the assumption $k = \Omega(p^{\frac{\eta}{1+\eta}})$, we obtain

$$\liminf_{p\to\infty}\chi^2(P\|Q) \ge c \tag{73}$$

for some constant $c > 0$, completing the proof of Theorem 5.

D. Proof of Theorem 6 (Negative Result for Weak Recovery)

As mentioned in Section III-A, $\chi^2(P\|Q) \to 0$ implies that $D(P\|Q) \to 0$. Consider $(\mathbf{X},\mathbf{Y}) \sim P$, along with an additional pair $(X',Y') \in \{0,1\}^p \times \{0,1\}$ drawn from the same joint distribution as a single test in $(\mathbf{X},\mathbf{Y})$, independently from $(\mathbf{X},\mathbf{Y})$. Following the steps of [26] for sparse linear regression, we consider the following conditional mutual information term:

$$\begin{aligned}
&I(S;\mathbf{Y},Y'|\mathbf{X},X')\\
&\quad= \mathbf{E}_P\left[\log\frac{P(\mathbf{Y},Y'|\mathbf{X},X',S)}{P(\mathbf{Y},Y'|\mathbf{X},X')}\right] && (74)\\
&\quad= \mathbf{E}_P\left[\log\frac{P(\mathbf{Y},Y'|\mathbf{X},X',S)}{Q(\mathbf{Y},Y')}\right]\\
&\qquad\quad - \mathbf{E}_P\left[\log\frac{P(\mathbf{Y},Y'|\mathbf{X},X')}{Q(\mathbf{Y},Y')}\right] && (75)\\
&\quad= \mathbf{E}_P\left[\log\frac{P(\mathbf{Y},Y'|\mathbf{X},X',S)}{Q(\mathbf{Y},Y')}\right] - D(P\|Q), && (76)
\end{aligned}$$

where $D(P\|Q)$ is now defined according to P and Q containing $n+1$ tests instead of n (one extra for X', Y'). Under P, we have $P(\mathbf{Y},Y'|\mathbf{X},X',S) = 1$ almost surely, and combining this with $Q(\mathbf{Y},Y') = 2^{-(n+1)}$, it follows that

$$I(S;\mathbf{Y},Y'|\mathbf{X},X') = (n+1)\log 2 - D(P\|Q). \tag{77}$$

Moreover, by the chain rule for mutual information, we have

$$\begin{aligned}
I(S;\mathbf{Y},Y'|\mathbf{X},X') &= I(S;\mathbf{Y}|\mathbf{X},X') + I(S;Y'|\mathbf{X},X',\mathbf{Y}) && (78)\\
&\le n\log 2 + H(Y'|\mathbf{X},X',\mathbf{Y}), && (79)
\end{aligned}$$

where the two terms are attained as follows by expanding the conditional mutual information as as a difference of conditional entropies:

- For the first term, write $I(S;\mathbf{Y}|\mathbf{X},X') = H(\mathbf{Y}|\mathbf{X},X') - H(\mathbf{Y}|\mathbf{X},X',S) \le H(\mathbf{Y}|\mathbf{X},X')$, and note that $H(\mathbf{Y}|\mathbf{X},X') \le n\log 2$ since $\mathbf{Y} \in \{0,1\}^n$ and entropy is upper bounded by the logarithm of the number of outcomes;
- For the second term, write $I(S;Y'|\mathbf{X},X',\mathbf{Y}) = H(Y'|\mathbf{X},X',\mathbf{Y}) - H(Y'|\mathbf{X},X',\mathbf{Y},S)$, and note that we have $H(Y'|\mathbf{X},X',\mathbf{Y},S) = 0$ since Y' is deterministic given (X',S).

Combining (77) and (79) gives

$$H(Y'|\mathbf{X},X',\mathbf{Y}) \ge \log 2 - D(P\|Q) = \log 2 - o(1), \quad (80)$$

since $D(P\|Q) \to 0$ for $n \le (1-\eta)\log_2\binom{p}{k}$ by Theorem 4 (the replacement of of n by $n+1$ only amounts to a negligible multiplicative $1+o(1)$ change in η). Since the entropy functional is continuous, and the entropy of a binary random variable is at most $\log 2$ with equality if and only if the random variable is equiprobable on its two values, we deduce from (80) that the following holds: With probability $1-o(1)$ with respect to $(\mathbf{X},X',\mathbf{Y})$, the conditional distribution of Y' places probability $\frac{1}{2}+o(1)$ on each of $Y'=0$ and $Y'=1$.

To complete the proof of Theorem 6, we show that the preceding claim precludes the possibility of weak recovery, i.e., (16) holds. Suppose by contradiction to (16) that it were possible to use $(\mathbf{X},\mathbf{Y})$ to attain $|S\cap\hat{S}| \ge \delta k$ with probability at least δ, for some $\delta > 0$. In the following, we assume the extreme case $|S\cap\hat{S}| = \delta k$; the case of strict inequality follows similarly. Consider a procedure that uses this $\hat{S}$ to construct an estimator that takes the test vector X' as input and returns an estimate $\hat{Y}'$ of Y' as follows: Set $\hat{Y}'=1$ if the test includes any item from $\hat{S}$, and $\hat{Y}'=0$ otherwise. There are two scenarios in which the estimate is incorrect:

- The test may include no items from S (and hence $Y'=0$), but an item from $S'\setminus S$ (and hence $\hat{Y}'=1$). By the choice of ν in (1), the probability (with respect to P) of this occurring is $\frac{1}{2}\big(1-\big(1-\frac{\nu}{k}\big)^{(1-\delta)k}\big) = \frac{1}{2}\big(1-\big(\frac{1}{2}\big)^{1-\delta}\big)$.
- The test may include no items from S' (and hence $\hat{Y}'=0$), but an item from $S\setminus S'$ (and hence $Y'=1$). By the same argument as above, the probability of this occurring is $\frac{1}{2}\big(1-\big(\frac{1}{2}\big)^{1-\delta}\big)$.

Hence, when $|S\cap\hat{S}| = \delta k$, the estimator produces $\hat{Y}'=Y'$ with probability $\big(\frac{1}{2}\big)^{1-\delta}$. As a result, for any fixed $\delta>0$, the success probability behaves as $\frac{1}{2}+\Omega(1)$. This is in contradiction with the conditional distribution of Y' stated following (80) (which only permits a $\frac{1}{2}+o(1)$ probability of correctness), and this completes the proof by contradiction establishing (16).

E. Proof of Theorem 7 (Negative Result for Identifying a Definite Defective)

Consider any algorithm that, with probability one, only outputs $\mathcal{I}\neq 0$ when $\mathcal{I}$ is the index of a defective item. If S is the true defective set, then it is easy to see that an error occurs (i.e., $\mathcal{I}=0$) if some S' disjoint from S is still consistent with $(\mathbf{X},\mathbf{Y})$. Denoting this event by $\mathcal{E}_{S'}$, it follows that

$$\mathbb{P}[\mathcal{I}=0\,|\,S] \ge \mathbb{P}\Big[\bigcup_{S':\,S\cap S'=\emptyset}\{\mathcal{E}_{S'}\}\Big] \quad (81)$$

$$\ge \sum_{S':\,S\cap S'=\emptyset}\frac{\mathbb{P}[\mathcal{E}_{S'}]^2}{\sum_{S^\natural:\,S\cap S^\natural=\emptyset}\mathbb{P}[\mathcal{E}_{S'}\cap\mathcal{E}_{S^\natural}]}, \quad (82)$$

where (82) follows from de Caen's lower bound on the probability of a union [37]. However, for S' and $S^\natural$ both disjoint from S (but possibly overlapping with each other), $\mathbb{P}[\mathcal{E}_{S'}\cap\mathcal{E}_{S^\natural}]$ is exactly the same quantity as $\mathbb{E}_Q\big[\mathcal{I}_{S'}(\mathbf{X},\mathbf{Y})\mathcal{I}_{S^\natural}(\mathbf{X},\mathbf{Y})\big]$ appearing in (43). In particular, $\mathbb{P}[\mathcal{E}_{S'}]$ corresponds to the case that $S'=S^\natural$. Substituting the expression in (43) gives $\mathbb{P}[\mathcal{E}_{S'}\cap\mathcal{E}_{S^\natural}] = 2^{-n(1+\frac{\ell}{k})}$ when $|S'\cap S^\natural| = k-\ell$, and substitution into (82) gives

$$\mathbb{P}[\mathcal{I}=0\,|\,S] \ge \binom{p}{k}\frac{4^{-n}}{\sum_{\ell=0}^{k}\binom{k}{\ell}\binom{p-k}{\ell}2^{-n(1+\frac{\ell}{k})}} \quad (83)$$

$$= \frac{1}{1+\chi^2(P\|Q)}, \quad (84)$$

where (84) follows by equating with (44). From Theorem 4, we know that $\chi^2(P\|Q)\to 0$ under the conditions of Theorem 7, and we conclude that $\mathbb{P}[\mathcal{I}=0\,|\,S]\to 1$. Since this holds regardless of which S is conditioned on, we obtain $\mathbb{P}[\mathcal{I}=0]\to 1$ as desired.

Appendix C
Berry-Esseen Theorem

Our analysis makes use of the following Berry-Esseen theorem, a non-asymptotic form of the central limit theorem.

Theorem 8. (Berry-Esseen Theorem [39, Theorem 2]) *For $j=1,\dots,p$, let X_j be independent random variables with*

$$\mu_j = \mathbb{E}[X_j], \quad \sigma_j^2 = \mathsf{Var}[X_j], \quad \text{and} \quad \rho_j = \mathbb{E}[|X_j-\mu_j|^3].$$

Denote $V=\sum_{j=1}^p\sigma_j^2$ and $T=\sum_{j=1}^p\rho_j$. Then, for any $\lambda\in\mathbb{R}$, we have

$$\Big|\mathbb{P}\Big[\frac{\sum_{j=1}^p(X_j-\mu_j)}{\sqrt{V}}\ge\lambda\Big]-\mathcal{Q}(\lambda)\Big| \le \frac{6T}{V^{3/2}}, \quad (85)$$

where $\mathcal{Q}(t)=\int_t^\infty\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,\mathrm{d}u$.

More precisely, we use the following simple corollary.

Corollary 2. *Let $Z\sim$ Binomial(p,q_0). Then, for any $\lambda\in\mathbb{R}$, the following holds:*

$$\Big|\mathbb{P}\Big[\frac{Z-pq_0}{\sigma\sqrt{p}}\ge\lambda\Big]-\mathcal{Q}(\lambda)\Big| \le \frac{6\rho}{\sigma^3\sqrt{p}}, \quad (86)$$

where

$$\rho = (1-q_0)^3q_0+q_0^3(1-q_0), \quad (87)$$

$$\sigma = \sqrt{(1-q_0)q_0}. \quad (88)$$

Proof. Since $Z\sim$ Binomial(p,q_0), we can write $Z=\sum_{j=1}^pZ_j$, where the Z_j are i.i.d. with distribution

Bernoulli(q_0). We shift to a zero-mean summation by writing $Z - pq_0 = \sum_{j=1}^{p}(Z_j - q_0)$, and observe that

$$\rho_1 := \mathbb{E}[|Z_1 - q_0|^3] = (1-q_0)^3 q_0 + |0-q_0|^3(1-q_0) = \rho, \tag{89}$$

and

$$\sigma_1^2 = \mathbb{E}[(Z_1 - q_0)^2] = (1-q_0)^2 q_0 + (0-q_0)^2(1-q_0) = \sigma^2 \tag{90}$$

for ρ and σ defined in (87)–(88). Hence, (86) follows directly from Theorem 8 with $T = p\rho_1^3 = p\rho^3$ and $V = p\sigma_1^2 = p\sigma^2$. □

References

[1] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: An information theory perspective," *Found. Trend. Comms. Inf. Theory*, vol. 15, no. 3–4, pp. 196–392, 2019.

[2] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 363–377, 1964.

[3] D. Du and F. K. Hwang, *Combinatorial group testing and its applications*. World Scientific, 2000, vol. 12.

[4] M. Malyutov, "The separating property of random matrices," *Math. Notes Acad. Sci. USSR*, vol. 23, no. 1, pp. 84–91, 1978.

[5] A. de Bonis, L. Gasieniec, and U. Vaccaro, "Optimal two-stage algorithms for group testing problems," *SIAM Journal on Computing*, vol. 34, no. 5, pp. 1253–1270, 2005.

[6] A. G. D'yachkov and V. V. Rykov, "A survey of superimposed code theory," *Prob. Contr. Inf.*, vol. 12, no. 4, pp. 1–13, 1983.

[7] A. Rashad, "Random coding bounds on the rate for list-decoding superimposed codes," *Prob. Ctrl. Inf. Theory*, vol. 19, no. 2, pp. 141–149, 1990.

[8] M. Cheraghchi, "Noise-resilient group testing: Limitations and constructions," in *Int. Symp. Found. Comp. Theory*, 2009, pp. 62–73.

[9] P. Indyk, H. Q. Ngo, and A. Rudra, "Efficiently decodable non-adaptive group testing," in *ACM-SIAM Symp. Disc. Alg. (SODA)*, 2010.

[10] H. Q. Ngo, E. Porat, and A. Rudra, "Efficiently decodable error-correcting list disjunct matrices and applications," in *Int. Colloq. Automata, Lang., and Prog.*, 2011.

[11] J. Scarlett and V. Cevher, "How little does non-exact recovery help in group testing?" in *IEEE Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP)*, 2017.

[12] K. Lee, K. Chandrasekher, R. Pedarsani, and K. Ramchandran, "SAFFRON: A fast, efficient, and robust framework for group testing based on sparse-graph codes," *IEEE Trans. Sig. Proc.*, vol. 67, no. 17, pp. 4649–4664, Sept. 2019.

[13] J. Scarlett and V. Cevher, "Phase transitions in group testing," in *Proc. ACM-SIAM Symp. Disc. Alg. (SODA)*, 2016.

[14] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, "Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms," in *Allerton Conf. Comm., Ctrl., Comp.*, Sep. 2011, pp. 1832–1839.

[15] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli, "Group testing with probabilistic tests: Theory, design and application," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 7057–7067, Oct 2011.

[16] M. B. Malyutov, "Search for sparse active inputs: A review," in *Inf. Theory, Comb. and Search Theory*, 2013, pp. 609–647.

[17] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, June 2014.

[18] J. Scarlett and V. Cevher, "Near-optimal noisy group testing via separate decoding of items," *IEEE Trans. Sel. Topics Sig. Proc.*, vol. 2, no. 4, pp. 625–638, 2018.

[19] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, "Information-theoretic and algorithmic thresholds for group testing," in *Int. Colloq. Aut., Lang. and Prog. (ICALP)*, 2019.

[20] ——, "Optimal group testing," in *Conf. Learn. Theory (COLT)*, 2020.

[21] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3019–3035, May 2014.

[22] L. Baldassini, O. Johnson, and M. Aldridge, "The capacity of adaptive group testing," in *IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2676–2680.

[23] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with near-constant tests-per-item," *IEEE Trans. on Inform. Th.*, vol. 65, no. 2, pp. 707–723, 2019.

[24] M. Hahn-Klimroth and P. Loick, "Optimal adaptive group testing," 2019, https://arxiv.org/abs/1911.06647.

[25] D. Gamarnik and I. Zadik, "High dimensional regression with binary coefficients. estimating squared error and a phase transtition," in *Conf. Learn. Theory (COLT)*, 2017, pp. 948–953.

[26] G. Reeves, J. Xu, and I. Zadik, "The all-or-nothing phenomenon in sparse linear regression," in *Conf. Learn. Theory (COLT)*, 2019.

[27] ——, "All-or-nothing phenomena: From single-letter to high dimensions," 2019, https://arxiv.org/abs/1912.13027.

[28] A. Gilbert, M. Iwen, and M. Strauss, "Group testing and sparse signal recovery," in *Asilomar Conf. Sig., Sys. and Comp.*, Oct. 2008, pp. 1059–1063.

[29] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, March 2012.

[30] J. Scarlett, "Noisy adaptive group testing: Bounds and algorithms," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3646–3661, June 2019.

[31] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[32] F. Hwang, "A method for detecting all defective members in a population by group testing," *J. Amer. Stats. Assoc.*, vol. 67, no. 339, pp. 605–608, 1972.

[33] G. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[34] S. Y. T. Soon, "Binomial approximation for dependent indicators," *Statistica Sinica*, vol. 6, no. 3, pp. 703–714, 1996.

[35] J. Duchi, "Lecture notes for statistics 311/electrical engineering 377," http://stanford.edu/class/stats311/.

[36] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

[37] D. de Caen, "A lower bound on the probability of a union," *Discrete mathematics*, vol. 169, no. 1, pp. 217–220, 1997.

[38] M. B. Malyutov and P. S. Mateev, "Screening designs for non-symmetric response function," *Mat. Zametki*, vol. 29, pp. 109–127, 1980.

[39] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. John Wiley and Sons, 1971.