STRONG IMPOSSIBILITY RESULTS FOR NOISY GROUP TESTING

Vincent Y. F. Tan⋆† *George K. Atia*‡

⋆ Department of Electrical and Computer Engineering, National University of Singapore
† Data Analytics Department, Institute for Infocomm Research (I^2R), A*STAR, Singapore
‡ Department of Electrical Engineering and Computer Science, University of Central Florida

ABSTRACT

Strong impossibility results for noisy group testing are derived. It is shown that regardless of the allowed error probability in identifying the defective set, the required of number of measurements is almost the same as that required for the error probability to be arbitrarily small. Our proof technique involves the use of the blowing-up lemma.

***Index Terms*— Noisy group testing, Defective set, Strong converse, Necessary conditions, Blowing-up lemma**

1. INTRODUCTION

Group testing [1, 2] can be regarded as a non-linear Boolean version of the well-known compressive sensing model wherein a binary measurement matrix is applied to a sparse vector, with the goal of reconstructing the support, or equivalently identifying a set of interest in a large population of items.

Related Work: A large body of research on the topic has focused on combinatorial pool design and matrix constructions with favorable separability and covering properties to guarantee the detection of the items of interest using a small number of tests. The covering property ensures that a test pattern obtained by taking any K columns of the measurement matrix does not cover any other boolean sum of K or smaller number of columns. Matrices that satisfy this property are often referred to as superimposed codes and combinatorial constructions were extensively developed by [3–5].

A different approach to group testing based on a probabilistic method has also been advocated by several researchers [6–9], and upper and lower bounds on the number of rows T for a matrix to be K-disjunct (bounds on lengths of superimposed codes) were developed. Random designs were used to compute upper bounds on the lengths of superimposed codes by investigating when randomly generated matrices have the desired covering/separability properties. Sebo [7] investigated average error probabilities and showed that for an arbitrarily small error probability, a randomly generated matrix will be K-disjunct if $T = O(K \log N)$ as $N \to \infty$, where N denotes the total number of items.

This work was supported by a startup grant from the National University of Singapore, A*STAR and NSF Grant CCF-1320547.

The problem of group testing was further investigated from an information-theoretic perspective in [10–12]. In particular, in [12] the problem of group testing (with its noisy versions) is mapped to a channel coding problem. Sufficient conditions on the number of measurements needed to identify the defective set are derived based on the analysis of a maximum-likelihood (ML) decoder, and were further generalized to other sparse signal processing models in [13–15].

Main Contributions: Although achievability results (sufficient conditions) are abound, systematic studies of converses (necessary conditions) for group testing are largely lacking, with the exception of combinatorial bounds in [1] (and references therein), and information-theoretic converses in [12, 14, 16]. Nevertheless therein, the necessary conditions involved Fano's inequality [17, Th. 2.10.1], and hence are weak converses, since Fano's inequality only establishes a condition for the error probability to be bounded away from zero. In this paper, we improve on the weak converse results by Atia-Saligrama [12, Th. IV.1] and Aldridge [18, Th. 1]. We establish new strong converse bounds based on the blowing-up lemma [19, 20]. To the best of our knowledge, this is the first time this approach is used to establish converse results for noisy group testing and its generalizations.

Paper Outline: The rest of the paper is organized as follows. In Section 2, the basic group testing problem and the blowing-up lemma are introduced. In Section 3, we present strong converse bounds for the maximum and average error probabilities. The proof is presented in Section 4 and we point out generalizations in Section 5. We conclude in Section 6.

2. PROBLEM SETUP AND BLOWING-UP LEMMA

2.1. The Basic Group Testing Problem

We have N items of a population and they are indexed by the integers as $[N] := \{1, \ldots, N\}$. There are no more than K defective (or salient) items of interest and the set of defective items is denoted as $\mathcal{K} \subset [N]$. We wish to detect the set $\mathcal{K}$ using a (small) number tests $T \in \mathbb{N}$. A *pooling strategy* to detect $\mathcal{K}$ is defined by a *testing matrix* $\mathbf{X} = \{x_{nt}\} \in \{0, 1\}^{N \times T}$, which is a *random* binary matrix, where $x_{nt} = 1$ means that item n is in the pool for test t and $x_{nt} = 0$ means that it is not. Test t produces an output $Y_t \in \{0,1\}$. Let $k_t := |\{n \in \mathcal{K} : x_{nt} = 1\}|$ denote the number of defective items in test t. In the noiseless group testing problem, the output $Y_t = 1$ iff $k_t \geq 1$, i.e., there exist a defective item in test t. The vector of all Y_t is denoted as $Y^T = (Y_1, \ldots, Y_T)$.

We will consider group testing with noise as in [12]. In this case, $Y_t = 1$ with high probability (but not necessarily probability 1) if $k_t \geq 1$. Conversely, $Y_t = 0$ with high probability if $k_t = 0$. We will not be concerned with the exact noise models as our results are general and can be specialized to the various noise models considered in [12] using the same techniques. Instead, we make *strong impossibility statements*.

2.2. Definitions of Error Probability

Let g be a function, called an *estimator*, that maps an element in $\{0,1\}^T$ to a subset of $[N]$ of size no more than K. This will be our decoder that decides, based on Y^T, what the set of defectives $\mathcal{K}$ is. For any estimator g, the *probability of error* for the set of defectives $\mathcal{K}$ is defined as

$$\lambda_{\mathcal{K}} := \Pr\left[g(Y^T) \neq \mathcal{K} \,\middle|\, \mathcal{K} \text{ is the defective set}\right]. \quad (1)$$

Note that the probability is over the noisy observations Y^T that are generated based on the testing matrix $\mathbf{X}$ and the true defective set $\mathcal{K}$. There are several such noise models of practical interest that are discussed in [12, Sec. II-C] such as the *additive* and *dilution* models. These noise models can be regarded as noisy channels from $\mathbf{X}$ to Y^T. Additionally, since $\mathbf{X}$ is also random, the probability is also over $\mathbf{X}$. The *maximum probability of error* is defined as

$$\lambda_{\max} := \max_{\mathcal{K} \in \mathcal{S}_{N,K}} \lambda_{\mathcal{K}}, \quad (2)$$

where $\mathcal{S}_{N,K} := \{\mathcal{K} \subset [N] : |\mathcal{K}| \leq K\}$. The *average probability of error* is defined as

$$\lambda_{\mathrm{ave}} := \frac{1}{\sum_{m=0}^{K} \binom{N}{m}} \sum_{\mathcal{K} \in \mathcal{S}_{N,K}} \lambda_{\mathcal{K}}. \quad (3)$$

Let $(T,N,K,\varepsilon)_{\max}$ (resp. $(T,N,K,\varepsilon)_{\mathrm{ave}}$) be called an *achievable quadruple for the noisy group testing problem* if there exists an estimator g that can detect a defective set of size no more than K among N items with T tests and maximum (resp. average) probability of error at most $\varepsilon \in (0,1)$. Note that unlike most works on noisy group testing [12, 15], we do not require ε to be arbitrarily small. Let $h(\alpha) := -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$ denote the binary entropy.

2.3. The Blowing-Up Lemma

We now state a simple version of the blowing-up lemma [19] proved by Marton [20]. Let $\mathcal{Y}$ be a finite set $\mathcal{Y}^T$ its T-fold Cartesian product. For two strings $y^T, z^T \in \mathcal{Y}^n$, we let $d_{\mathrm{H}}(y^T, z^T)$ be the *Hamming distance*, or the number of locations at which y^T and z^T differ. For any set $\mathcal{A} \subset \mathcal{Y}^T$ and arbitrary number $0 \leq l \leq T$, we denote the *l-blowup of $\mathcal{A}$* as $\Gamma^l(\mathcal{A}) := \left\{z^T : \min_{y^T \in \mathcal{A}} d_{\mathrm{H}}(y^T, z^T) \leq l\right\}$. That is, $\Gamma^l(\mathcal{A})$ is the set of T-tuples at distance no larger than l from $\mathcal{A}$.

Lemma 1 (Blowing-Up)**.** *Let $X_1, \ldots, X_T$ be independent random variables with distribution P^T. Let $\gamma_T = o(1)$ be a sequence. There exists sequences $\delta_T, \zeta_T = o(1)$ such that if*

$$P^T\left(\mathcal{A}_T\right) \geq \exp(-T\gamma_T), \quad \textit{then} \quad (4)$$

$$P^T\left(\Gamma^{T\delta_T}(\mathcal{A}_T)\right) \geq 1 - \zeta_T. \quad (5)$$

The intuition behind the blowing-up lemma is as follows: Suppose $\mathcal{A}_T$ has P^T-probability that is "not too small" in the sense of (4), then by enlarging $\mathcal{A}_T$ slightly by including those T-tuples at a sublinear Hamming distance from $\mathcal{A}_T$, the resultant set $\Gamma^{T\delta_T}(\mathcal{A}_T)$ has P^T-probability arbitrarily close to one. This concentration of measure phenomenon has been studied extensively in many domains. See [21].

In fact, Lemma 1 can written in a more quantitative (and non-asymptotic) manner. In our application of the blowing-up lemma,

$$\gamma_T := \frac{1}{T} \log \frac{1}{1-\varepsilon} \quad (6)$$

for some $0 < \varepsilon < 1$ and thus γ_T tends to zero. Marton's remark in [20] then says that δ_T and ζ_T can be chosen as

$$\delta_T := T^{-1/4}, \quad \text{and} \quad \zeta_T := T^{-1/4}\sqrt{\frac{1}{1-\varepsilon}}. \quad (7)$$

3. MAIN RESULTS

Theorem 2 (Strong Converse for Maximum Error Probability)**.** *Let $\varepsilon \in (0,1)$. Let the components of $\mathbf{X}$ be independent and identically distributed (i.i.d.). The following is a necessary condition on all achievable quadruples for the noisy group testing problem $(T,N,K,\varepsilon)_{\max}$:*

$$T \geq \max_{\mathcal{L} \subset \mathcal{K}} \frac{(1-\zeta_T)\log\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}}; \mathbf{X}_{\mathcal{L}}, Y) + \eta_T} \quad (8)$$

where $\eta_T = O(T^{-1/4}\log T)$ is a sequence defined as

$$\eta_T := T^{-1} + h(\delta_T). \quad (9)$$

A strong converse statement can also be made for the average probability of error by relating the two quandruples $(T,N,K,\varepsilon)_{\max}$ and $(T,N,K,\varepsilon)_{\mathrm{ave}}$. In order to make such a statement, we require that $\lambda_{\mathcal{K}'} \leq \lambda_{\mathcal{K}}$ for all pairs of sets satisfying $|\mathcal{K}'| \leq |\mathcal{K}|$. Intuitively, this means that larger subsets have error probabilities no smaller than smaller subsets.

Corollary 3 (Strong Converse for Average Error Probability)**.** *Let $\varepsilon \in (0,1)$ and let for each $1 < \tau < \frac{1}{\varepsilon}$, let*

$$K'(\tau) := \max\left\{K' \in [K] : \binom{N}{K'} \leq \left(1 - \frac{1}{\tau}\right)\binom{N}{K}\right\}. \quad (10)$$

Let the components of $\mathbf{X}$ *be i.i.d. The following is a necessary condition on all achievable quadruples for the noisy group testing problem* $(T, N, K, \varepsilon)_{\text{ave}}$*:*

$$T \geq \sup_{1<\tau<\frac{1}{\varepsilon}} \max_{\mathcal{L}\subset\mathcal{K}} \frac{(1-\zeta_T)\log\binom{N-|\mathcal{L}|}{K'(\tau)-|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}};\mathbf{X}_{\mathcal{L}},Y)+\eta_T}. \tag{11}$$

In the asymptotic setting, τ will be chosen to approach 1 (as the problem size increases) so $K'(\tau)$ approaches K, implying that (8) and (11) are almost identical.

We remark that Theorem 2 and Corollary 3 are *significantly stronger* than the converse result stated in [12]. Recall that the converse statement in [12] is as follows:

Theorem AS (Th. IV.1 in [12])**.** *Let the components of* $\mathbf{X}$ *be i.i.d. In order to ensure that either* $\lambda_{\max} \to 0$ *or* $\lambda_{\text{ave}} \to 0$*, the number of tests must satisfy the following inequality as the parameters* (N, K, T) *tend to infinity:*

$$T \geq \max_{\mathcal{L}\subset\mathcal{K}} \frac{\log\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}};\mathbf{X}_{\mathcal{L}},Y)} \tag{12}$$

Remark: Note that Theorem 2 does not require $\lambda_{\max}$ to vanish. It can be fixed at say some arbitrarily large value $\varepsilon = 1-10^{-20}$. Thus, no matter how large the error probability is allowed to be (as long as it is strictly smaller than 1), a converse bound arbitrarily close to the asymptotic converse bound in (12) must be satisfied. This is the essence of the strong converse theorems. In other words, if (12) is violated, $\lambda_{\max} \to 1$ as the parameters of the problem size grow. In the weak converse statement in [12], the statement is that if (12) is violated then the only conclusion we can make is that the error probability is asymptotically bounded away from 0.

Another salient observation is that our bounds in Theorem 2 and Corollary 3 are non-asymptotic. They are in fact, very close to the asymptotic bound in (12) because the terms η_T and ζ_T in (8) and (11) vanish as T tends to infinity.

4. PROOF OF MAIN RESULT

Proof of Theorem 2. Since this is an impossibility result, it suffices to consider those sets $\mathcal{K} \in \mathcal{S}_{N,K}$ whose sizes are *exactly* K. Since $\lambda_{\mathcal{K}} \leq \varepsilon$, we have

$$\Pr\left[g(Y^T) = \mathcal{K} \,\middle|\, \mathcal{K} \text{ is the defective set}\right] \geq 1-\varepsilon \tag{13}$$

for all $\mathcal{K} \subset [N]$ of size exactly K. Note that $\mathcal{K}$ is a random set. Let $\mathcal{D}_{\mathcal{K}} := \{y^T : g(y^T) = \mathcal{K}\}$ denote the *decoding region* for set $\mathcal{K}$. Then (13) can be rewritten as

$$W_{\mathcal{K}}^T(\mathcal{D}_{\mathcal{K}}) \geq 1-\varepsilon \tag{14}$$

for all $\mathcal{K}$ of size K where $W_{\mathcal{K}}^T(y^T)$ is the probability of observing y^T conditioned on $\mathcal{K}$ being the true defective set. The blowing-up lemma (Lemma 1) applied to (14) with γ_T in (6) states that there exists positive sequences $\delta_T, \zeta_T \in o(1)$ (both not depending on $\mathcal{K}$) satisfying

$$W_{\mathcal{K}}^T\left(\Gamma^{\delta_T T}(\mathcal{D}_{\mathcal{K}})\right) \geq 1-\zeta_T. \tag{15}$$

Now suppose that a genie reveals $|\mathcal{L}|$ defective items $\mathcal{L} \subset \mathcal{K}$ items to us. Then, we need to estimate the remaining $\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}$ equally likely subsets of defective items. Thus,

$$H(\mathcal{K}|\mathcal{L}) = \log\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}. \tag{16}$$

Now, by the chain rule for entropy,

$$H(\mathcal{K}|\mathcal{L}) = I(\mathcal{K};Y^T|\mathcal{L}) + H(\mathcal{K}|Y^T,\mathcal{L}). \tag{17}$$

By using the argument in the Appendix of [18] which applies for the more general adaptive setting (cf. Section 27), we know that $I(\mathcal{K};Y^T|\mathcal{L})$ can be single-letterized to be

$$I(\mathcal{K};Y^T|\mathcal{L}) \leq T I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}};\mathbf{X}_{\mathcal{L}},Y) \tag{18}$$

so it remains to bound the conditional entropy $H(\mathcal{K}|Y^T,\mathcal{L})$. Define the "error" random variable

$$E := \mathbf{1}\{Y^T \notin \Gamma^{\delta_T T}(\mathcal{D}_{\mathcal{K}})\}. \tag{19}$$

From the chain rule of entropy, we have

$$H(E,\mathcal{K}|Y^T,\mathcal{L}) = H(\mathcal{K}|E,Y^T,\mathcal{L}) + H(E|Y^T,\mathcal{L}) \tag{20}$$
$$= H(E|\mathcal{K},Y^T,\mathcal{L}) + H(\mathcal{K}|Y^T,\mathcal{L}). \tag{21}$$

Clearly, $H(E|\mathcal{K},Y^T,\mathcal{L}) = 0$ and $H(E|Y^T,\mathcal{L}) \leq 1$. Thus,

$$H(\mathcal{K}|Y^T,\mathcal{L}) \leq H(\mathcal{K}|E,Y^T,\mathcal{L}) + 1. \tag{22}$$

Further expanding the conditional entropy of the RHS yields

$$H(\mathcal{K}|E,Y^T,\mathcal{L}) = \Pr(E=0)H(\mathcal{K}|E=0,Y^T,\mathcal{L}) + \Pr(E=1)H(\mathcal{K}|E=1,Y^T,\mathcal{L}). \tag{23}$$

We now record the following simple facts:

$$\Pr(E=1) \leq \zeta_T, \tag{24}$$

$$H(\mathcal{K}|E=1,Y^T,\mathcal{L}) \leq \log\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}, \tag{25}$$

where (24) follows from (15) and (25) follows from the fact that there are $\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}$ choices for $\mathcal{K}$ given that $\mathcal{L}$ is known. The quantity $H(\mathcal{K}|E=0,Y^T,\mathcal{L})$ is the most interesting. To bound this, we define the set

$$\mathcal{N}_K(y^T) := \left\{\mathcal{K} \subset [N] : |\mathcal{K}| = K, y^T \in \Gamma^{\delta_T T}(\mathcal{D}_{\mathcal{K}})\right\}. \tag{26}$$

Now we have the following important estimate for $|\mathcal{N}_K(y^T)|$:

$$|\mathcal{N}_K(y^T)| \overset{(a)}{\leq} |\Gamma^{\delta_T T}(y^T)| \overset{(b)}{=} \sum_{k=0}^{\lfloor \delta_T T\rfloor}\binom{T}{k}(|\mathcal{Y}|-1)^k \overset{(c)}{=} \sum_{k=0}^{\lfloor \delta_T T\rfloor}\binom{T}{k} \overset{(d)}{\leq} 2^{Th(\delta_T)}. \tag{27}$$

In inequality (a), we note that $|\mathcal{N}_K(y^T)| = \sum_{\mathcal{K}} \mathbf{1}\{y^T \in \Gamma^{\delta_T T}(\mathcal{D}_{\mathcal{K}})\}$ where the sum is over those subsets of $[N]$ of size K. To each $\mathcal{K}$ for which the indicator returns 1, there exists at least one vector $\tilde{y}_{\mathcal{K}}^T \in \mathcal{D}_{\mathcal{K}}$ for which $d_{\mathrm{H}}(\tilde{y}_{\mathcal{K}}^T, y^T) \le \delta_T T$. Because the decoding sets are disjoint (i.e., $\mathcal{D}_{\mathcal{K}} \cap \mathcal{D}_{\mathcal{K}'} = \emptyset$ for $\mathcal{K} \neq \mathcal{K}'$), the vectors $\{\tilde{y}_{\mathcal{K}}^T\}_{\mathcal{K}}$ are also distinct. Thus, we can conservatively upper bound $|\mathcal{N}_K(y^T)|$ by the size of the entire $(\delta_T T)$-blowup of $\{y^T\}$, i.e., $|\Gamma^{\delta_T T}(y^T)|$. For equality (b), $\tilde{y}^T \in \Gamma^{\delta_T T}(\mathcal{D}_{\mathcal{K}})$ if and only if there exists a $y^T \in \mathcal{D}_{\mathcal{K}}$ differing from $\tilde{y}^T$ in no more than $\delta_T T$ locations. Thus, we run over all possible such locations from $k=0$ to $k = \lfloor \delta_T T \rfloor$ and count the number of of sequences differing from $\tilde{y}^T$ in exactly k locations. Equality (c) follows because $|\mathcal{Y}| = 2$ and inequality (d) from an elementary result [22, Ex. 2.8(a)] involving sum of the first $\lfloor \delta_T T \rfloor$ binomial coefficients. Here we assumed that T is so large that $\delta_T < \frac{1}{2}$. Note that the bounds in (27) do not depend on K or y^T. Thus, we have

$$H(\mathcal{K}|E=0, Y^T, \mathcal{L}) \le T h(\delta_T), \tag{28}$$

which is sublinear. Combining (22), (24), (25) and (28) yields

$$H(\mathcal{K}|Y^T, \mathcal{L}) \le 1 + T h(\delta_T) + \zeta_T \log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|}. \tag{29}$$

Uniting this with (16), (17) and (18), we obtain

$$\log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} \le 1 + T h(\delta_T) + \zeta_T \log \binom{N - |\mathcal{L}|}{K - |\mathcal{L}|} + T I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}}; \mathbf{X}_{\mathcal{L}}, Y). \tag{30}$$

Hence, by rearranging (30) and maximizing over all $\mathcal{L} \subset \mathcal{K}$, the proof of (8) is complete. □

Proof of Corollary 3. Fix $1 < \tau < \frac{1}{\varepsilon}$. For every achievable quadruple $(T, N, K'(\tau), \tau\varepsilon)_{\max}$, there exists a $(T, N, K, \varepsilon)_{\mathrm{ave}}$ achievable quadruple where $K'(\tau)$ satisfies (10). This is true because we can simply disregard those large subsets (i.e., subsets whose sizes are larger than $K'(\tau)$) which have error probability larger than some $\tau\varepsilon$. This is where the monotonicity of $\lambda_{\mathcal{K}}$ comes into play. In this way, since total number of subsets is $\binom{N}{K'(\tau)}$ and every error probability is no larger than $\tau\varepsilon$, the average error probability is no larger than ε. □

5. GENERALIZATIONS

In this section, we discuss generalizations of our results. For brevity, the results are stated here without proof.

5.1. Adaptive Group Testing

Aldridge [18] considered a variant of the noisy group testing problem where the makeup of a testing pool can depend on the outcomes of earlier tests, i.e. $x_{it} = x_{it}(Y_1, \ldots, Y_{t-1})$ for each $t \in [T]$. The author showed using weak converse techniques such as Fano's inequality that in the adaptive scenario, the required number of tests is no more than the non-adaptive case. Since the only part of our proof that involves adaptivity is the single-letterization step in (18) which holds for both the adaptive and non-adaptive setting, our strong impossibility results also hold for the adaptive setting.

5.2. Sparse Signal Processing Models

5.2.1. Models with Latent Variables

In many sparse signal processing problems [15] such as 1-bit quantized compressive sensing, the model contains latent variables. The observations may be related to the data as

$$Y^T = q(X^T \beta + W^T), \tag{31}$$

where $q(\cdot)$ is a quantizer, X^T are the covariates, $\boldsymbol{\beta}$ is the vector of latent variables and W^T is a noise vector. By sparse, we mean that there are no more than $K \ll N$ salient features or, more precisely, $\Pr(Y=y|X=x) = \Pr(Y=y|X_{\mathcal{K}} = x_{\mathcal{K}})$. That is, the observations Y depend only on a subset of covariates indexed by $\mathcal{K} \in \mathcal{S}_{N,K}$ and latent variables $\boldsymbol{\beta}_{\mathcal{K}}$. By modifying the single-letterization step in (18), we can show that the strong converse bound in (8) holds with the mutual information in the denominator replaced by $I(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}}; \mathbf{X}_{\mathcal{L}}, Y | \boldsymbol{\beta}_{\mathcal{K}})$. The same technique goes through if the testing matrix $\mathbf{X}$ is observed partially, e.g., entries of $\mathbf{X}$ are deleted at random.

5.2.2. Models with Exchangeable Covariates

An exchangeable sequence of random variables (covariates) has the property that the joint distribution is invariant to any permutation of the covariates, i.e., for any permutation π

$$P_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = P_{X_{\pi(1)},\ldots,X_{\pi(N)}}(x_1, \ldots, x_N). \tag{32}$$

Now, as in [15, 16] and Section 5.2.1, assume that the (discrete) output Y of a model is independent of all covariates conditioned on a salient set of covariates $\mathbf{X}_{\mathcal{K}}$. The goal is to identify the set $\mathcal{K}$ from T independent realizations of covariates/output pairs $(\mathbf{X}^T, Y^T)$. It can be easily shown that the same strong converse result derived herein using the blowing-up lemma also extends to this case, thereby, strengthening the weak converse of [15], which is based on Fano's inequality.

6. CONCLUSION

The application of the blowing-up technique can also be used to strengthen the converse statements for learning the structure of discrete graphical models [23,24] and estimating low-rank matrices over finite fields [25]. However, the discreteness of the observations is crucial in applying the blowing-up technique (cf. (27)). Future work involves extending the blowing-up technique to continuous observations. Alternative techniques such as those proposed by Das *et al.* [26] may prove to be useful in strengthening impossibility results in models with continuous observations.

7. REFERENCES

[1] D.-Z. Du and F. K. Hwang, *Combinatorial Group Testing and Its Applications*, World Scientific Publishing Company, second edition, 2000.

[2] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Stat.*, 1943.

[3] W. H. Kautz and R.C. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. on Inf. Th.*, vol. 10, no. 4, pp. 363–377, Oct. 1964.

[4] A. G. Dyachkov and V. V. Rykov, "A survey of superimposed code theory," *Problems of Control and Information Theory*, vol. 12, no. 4, pp. 1–13, 1983.

[5] P. Erdös, P. Frankl, and Z. Füredi, "Family of finite sets in which no set is covered by the union of n others," *Isr. J. Math.*, vol. 51, pp. 79–89, 1985.

[6] A. G. Dyachkov, V. V. Rykov, and A. M. Rashad, "Bounds of the length of disjunct codes," *Problems Control Information Theory*, vol. 11, pp. 7–13, 1982.

[7] Andras Sebo, "On two random search problems," *Journal of Statistical Planning and Inference*, vol. 11, pp. 23–31, 1985.

[8] A. C. Gilbert, M. A. Iwen, and M. J. Strauss, "Group testing and sparse signal recovery," in *42nd Asilomar Conference on Signals, Systems and Computers*, 2008.

[9] M. Ruszinkó, "Family of finite sets in which no set is covered by the union of n others," *Isr. J. Math.*, vol. 51, pp. 79–89, 1985.

[10] M. B. Malyutov, P. S. Mateev, and M. Zhao, "Screening design for non-symmetric reponse function," *Mat. Zemetki*, vol. 27, pp. 109–127, 1980.

[11] A. Dyachkov, "Lectures on designing screening experiments," *Moscow State Univ.*, 2003.

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

[13] G. Atia and V. Saligrama, "A mutual information characterization for sparse signal processing," in *The 38th Int. Colloq. on Autom., Lang. and Prog. (ICALP)*, Switzerland, Jul 2011.

[14] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and non-linear observations: A unified Shannon-theoretic approach," *CoRR*, vol. abs/1304.0682, 2013.

[15] C. Aksoylar, G. Atia, and V. Saligrama, "Sample complexity of salient feature identification for sparse signal processing," in *IEEE Statistical Signal Processing Workshop*, 2012.

[16] V. Y. F. Tan, M. J. Johnson, and A. S. Willsky, "Necessary and sufficient conditions for high-dimensional salient subset recovery," in *Int. Symp. Inf. Th.*, 2010.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2nd edition, 2006.

[18] M. Aldridge, "Adaptive group testing as channel coding with feedback," in *Int. Symp. Inf. Th.*, 2012.

[19] R. Ahlswede and P. Gács and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, no. 3, pp. 157–177, 1976.

[20] K. Marton, "A simple proof of the blowing-up lemma (corresp.)," *IEEE Trans. on Inf. Th.*, vol. 32, pp. 445–446, Mar 1986.

[21] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013.

[22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2011.

[23] N. P. Santhanam and M. J. Wainwright, "Information-theoretic limits of selecting binary graphical models in high dimensions," *IEEE Trans. on Inf. Th.*, vol. 58, no. 7, pp. 4117–4134, 2012.

[24] A. Anandkumar, V. Y. F. Tan, F. Huang, and A. S. Willsky, "High-dimensional structure estimation of Ising models: Local separation criterion," *Annals of Statistics*, vol. 40, no. 3, pp. 1346–1375, 2012.

[25] V. Y. F. Tan, L. Balzano, and S. C. Draper, "Rank minimization over finite fields: Fundamental limits and coding-theoretic interpretations," *IEEE Trans. on Inf. Th.*, vol. 58, no. 4, pp. 2018–2039, Apr 2012.

[26] A. K. Das, P. Netrapalli, S. Sanghavi, and S. Vishwanath, "Learning Markov graphs up to edit distance," in *Int. Symp. Inf. Th.*, 2012.