A Mutual Information Characterization for Sparse Signal Processing

George Atia[1] and Venkatesh Saligrama[2]

[1] Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2] Information Systems and Sciences Lab, Boston University, Boston, MA 02215, USA

Abstract. Suppose among a set of N covariates $X_1; X_2; \ldots; X_N$ there is a subset S of covariates that are salient for predicting outcomes Y. Specifically, we suppose that when Y is conditioned on $\{X_k\}_{k \in S}$ it is independent of the other covariates. Our goal is to identify the subset S from data samples of the covariates and associated outcomes. We present a precise mutual information expression that characterizes the sample complexity for accurately identifying the subset S for many interesting scenarios and derive precise sample complexity bounds for these cases.

1 Introduction

In this paper we derive information theoretic sufficiency and necessity bounds on sparse signal processing models. In many cases, the output depends on a small relevant set of features (covariates) where relevance is expressed through conditional independence. Such models have wide applicability. Sparse models encountered in the literature include, but by no means limited to:

1. **Linear models** which naturally arise in array processing wherein the output Y is obtained as a linear (possibly noisy) transformation of some input X.
2. **Compressive sensing (CS)** with its quantized and unquantized versions.
3. **Group testing** which is essentially a form of compressive sensing with Boolean arithmetic.

While the degradation of CS with noise has been characterized (see [4,6,9,10,13,16,17] and references therein), this paper provides a universal information theoretic characterization for a wide class of problems in sparse signal processing. In the context of group testing, a significant part of the existing research is focused on combinatorial pool design (i.e. construction of measurement matrices with nice properties) to guarantee the detection of the items of interest using a small number of tests. These approaches have generally characterized fundamental tradeoffs for noiseless group testing [3, 7, 8, 12, 14, 15]. We studied the noisy counterpart of group testing in [2] wherein we derived necessary and sufficient conditions on the number of tests to identify a set of defective items in the presence of additive and dilution noise. In this paper we extend the main result and the methods that we developed in [2] to more general models of sparse signal processing. Consequently, some of the information theoretic Compressed Sensing results [1] can also be recovered using the mutual information expressions derived herein. In this paper we present a novel information theoretic approach to such problems. We formulate the salient covariates problem as a detection problem and establish its connection to Shannon coding theory [5] which, to the best of our knowledge, is explored in this paper and [2] for the first time. The new perspective allows us to easily obtain results for a wide range of models including noisy versions of group testing. Our approach, which is fairly general, is to map the problem to a corresponding channel model which allows the computation of simple mutual information expressions to derive achievable bounds on the required number of measurements. To summarize, our approach offers several advantages, including:

– **Mutual information characterization:** Our main result is a simple single letter characterization providing order-wise tight necessary and sufficient conditions on the total number of measurements.
– **Characterization for new problems:** This result allows us to verify existing bounds for some of the known scenarios and extend the analysis to many new interesting setups including noisy versions of group testing. The methods that we develop in this paper are generally applicable to sparse signal processing thus extending our results in [2].

The rest of the paper is organized as follows. In Section 2 we provide a general description of the problem. In Section 3 we derive necessary and sufficient conditions on the number of measurement required for recovery. Applications are considered in Section 4.

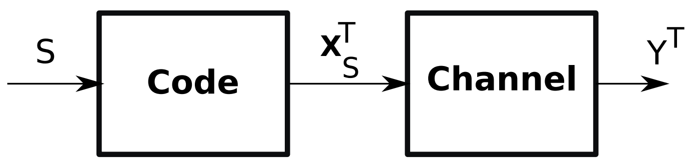

Fig. 1. channel Model

2 Problem Setup

In this section we describe the general problem. We introduce our notational convention that will be used throughout the paper. Let $X = (X_1, X_2, X_3, \ldots, X_N)$ denote a set of N independent and identically distributed random covariates with a joint probability distribution $Q(X)$. We let Y denote an observation or outcome which is generated according to an observation model $P(Y|X)$. We further assume that the outcome Y only depends on a small subset of covariates $S \subset \{1, 2, \ldots, N\}$ of known cardinality $|S| = K$ where $K << N$. In other words, Y is conditionally independent of the covariates given the subset of covariates indexed by the index set S. As such, by conditional independence

$$P(Y|X) = P(Y|X_S) \tag{1}$$

where $X_S = \{X_k\}_{k\in S}$ is the subset of covariates indexed by the set S.

We observe $(\boldsymbol{X}^T, Y^T)$ which are T i.i.d. data samples and associated outcomes $(X^{(i)}, Y^{(i)}), i = 1, 2, \ldots, T$. Our goal is to identify the subset S from the data samples and the associated outcomes $(\boldsymbol{X}^T, Y^T)$ with an arbitrarily small average error probability.

For notational convenience we use bold-face to denote matrices, while regular font is used to denote vectors, scalars and components of matrices. The vector X_j^T is a row vector of length T, corresponding to the j-th covariate, with the t-th entry denoted $X_j(t)$. Y^T is a vector of length T. Similarly, $Y(t)$ denotes the t-th component of the vector Y^T. Given a subset $S \subset \{1, 2 \ldots N\}$ with cardinality $|S|$, the matrix $\boldsymbol{X}_S$ is an $|S| \times T$ matrix formed from the rows indexed by S. Similarly, X_S denotes a vector, whose components are restricted to the set of components indexed by S. Thus, X_S is a column of the matrix $\boldsymbol{X}_S$. When time indexing is needed, $X_S(t)$ is used to specifically denote the t-th column of the matrix $\boldsymbol{X}_S$, and $X_j(t)$ is the t-th component of the vector X_j^T.

We let $\hat{S}(\boldsymbol{X}^T, Y^T)$ denote the estimate of the subset S which is random due to the randomness in X and Y. Conditioned on a particular subset S, we define the conditional error probability $P_e(S)$ as an average error probability over all possible realizations of data samples X and outcomes Y, i.e.,

$$P_e(S) = \Pr[\hat{S}(\boldsymbol{X}^T, Y^T) \neq S|S] \tag{2}$$

where the randomness is over the data X and the outcome Y. Also let $\lambda_X(S)$ denote the average error probability conditioned on a particular S and a given realization of the data samples $N \times T$ matrix $\boldsymbol{X}$. Hence,

$$\lambda_X(S) = \Pr[\hat{S}(\boldsymbol{X}^T, Y^T) \neq S|S, \boldsymbol{X}] \tag{3}$$

where the randomness is over the outcome Y. Given (2) and (3) we have

$$P_e(S) = \sum_X \lambda_X(S) Q(X) \tag{4}$$

We further let P_e denote the average probability of error, averaged over all subsets S of size K, all possible data samples X and outcomes Y, i.e.,

$$P_e = \Pr[\hat{S}(\boldsymbol{X}^T, Y^T) \neq S] \tag{5}$$

We index the different subsets of size K as S_ω with index ω. Since there are N covariates in total, there are $\binom{N}{K}$ such sets, hence

$$\omega \in \mathcal{I} = \left\{1, 2, \ldots \binom{N}{K}\right\} \tag{6}$$

Note that S_ω is a set of K indices corresponding to the ω-th set of covariates.

By symmetry, it is easy to see that $P_e = P_e(S)$, that is, the average error probability does not depend on the subset S due to averaging over the randomness of the data and we can assume without loss of generality that $\omega = 1$, i.e., S_1 is the true set.

To simplify the exposition we introduce some further notation. First, recall that the matrix $\boldsymbol{X}_S$ is an $|S| \times T$ matrix formed from rows indexed by the set S. For any 2 sets S_i and S_j, we define $S_{i,j}$, $S_{i^c,j}$, and S_{i,j^c} as the overlap set, the set of indices in S_j but not in S_i, and the set of indices in S_i but not in S_j, respectively. Namely,

$$\begin{aligned} S_{i,j} &= S_i \cap S_j \quad \text{overlap} \\ S_{i^c,j} &= S_i^c \cap S_j \quad \text{in } j \text{ but not in } i \\ S_{i,j^c} &= S_i \cap S_j^c \quad \text{in } i \text{ but not in } j \end{aligned}$$

3 Recovery: Necessary and Sufficient Conditions

3.1 Probability of error analysis

In this section, we analyze the error probability of a Maximum Likelihood (ML) decoder [11]. The decoder goes through all $\binom{N}{K}$ possible sets of size K, and chooses the set that is most likely. The decoding rule is thus defined by: given the outcomes Y^T, choose ω^* for which

$$p(Y^T|\boldsymbol{X}_{S_{\omega^*}}) > p(Y^T|\boldsymbol{X}_{S_\omega}); \quad \forall \omega \neq \omega^* \tag{7}$$

i.e., choose the set for which the given Y^T is most likely given ω. An error occurs if any set other than the true set is more likely. This ML decoder is a minimum probability of error decoder assuming uniform prior on the input messages (sets of covariates). Next, we derive an upper bound on the average error probability P_e of the ML decoder, where the average is taken over all sets, data realizations and observations.

Define the error event E_i as the event of mistaking the true set for a set which differs from the true set S_1 in exactly i covariates. The probability of such an event is denoted $P(E_i)$. The event E_i implies that there exists some set which differs from the true set in i covariates and is more likely to the decoder. Hence,

$$\begin{aligned} P(E_i) \leq \Pr\Big[&\exists j \neq 1 : p(Y^T|\boldsymbol{X}_{S_j}) \geq P(Y^T|\boldsymbol{X}_{S_1}) \\ &\text{where } |S_{1^c,j}| = |S_{1,j^c}| = i, \text{ and } |S_1| = |S_j| = K\Big] \end{aligned} \tag{8}$$

The probability $P(E_i)$ can be written as a summation over all inputs $\boldsymbol{X}_{S_1}$ and all outcomes Y^T

$$P(E_i) = \sum_{\boldsymbol{X}_{S_1}} \sum_{Y^T} Q(\boldsymbol{X}_{S_1}) p(Y^T|\boldsymbol{X}_{S_1}) \Pr[\text{error}_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] \tag{9}$$

where $\Pr[\text{error}_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T]$ is the probability of decoding error in exactly i covariates, conditioned on the true index $\omega_0 = 1$, the realization $\boldsymbol{X}_{S_1}$ for the set S_1, and on the sequence Y^T. This can be viewed as the error probability for a communication system with a transmitted message $\omega_0 = 1$, encoded message $\boldsymbol{X}_{S_1}$ and received sequence Y^T. Using the union bound, the conditional error probability averaged over data realizations is upper bounded by

$$P_e(S_1) \leq \sum_{i=1}^{K} P(E_i) = \sum_{i=1}^{K} \sum_{\boldsymbol{X}_{S_1}} \sum_{Y^T} Q(\boldsymbol{X}_{S_1} p(Y^T|\boldsymbol{X}_{S_1}) \Pr[\text{error}_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] \tag{10}$$

Next we state our main result. The following theorem provides a sufficient condition on the number of measurements T for an arbitrarily small average error probability.

Theorem 1. *(Sufficiency). Define $\Xi_S^{\{i\}}$ as the set of tuples $(\mathcal{S}^1, \mathcal{S}^2)$ partitioning the true set S into disjoint sets $\mathcal{S}^1$ and $\mathcal{S}^2$ with cardinalities i and $K-i$, respectively, i.e.,*

$$\Xi_S^{\{i\}} = \left\{ (\mathcal{S}^1, \mathcal{S}^2) : \mathcal{S}^1 \bigcap \mathcal{S}^2 = \emptyset, \mathcal{S}^1 \bigcup \mathcal{S}^2 = S, |\mathcal{S}^1| = i, |\mathcal{S}^2| = K - i \right\} \tag{11}$$

Let N be the total number of covariates and the true set S of cardinality K. If the number tests T is such that

$$T > (1+\epsilon) \cdot \max_{i:(\mathcal{S}^1,\mathcal{S}^2)\in \Xi_S^{\{i\}}} \frac{\log \binom{N-K}{i}\binom{K}{i}}{I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)}, \quad i = 1, 2, \ldots K \tag{12}$$

then asymptotically the average error probability approaches zero, namely,

$$\lim_{K\to\infty} \lim_{N\to\infty} P_e \to 0 \tag{13}$$

where $\epsilon > 0$ is a constant independent of N and K. $I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)$ is the mutual information [5] between $X_{\mathcal{S}^1}$ and $(X_{\mathcal{S}^2}, Y)$.

Theorem 1 follows from a tight bound—based on characterization of error exponents as in [11]—on the error probability $P(E_i)$. We will show that the error exponent, $E_o(\rho)$, is described by:

$$E_o(\rho) = -\log \sum_{Y\in\{0,1\}} \sum_{X_{\mathcal{S}^2}} \left[\sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1})^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad 0 \leq \rho \leq 1 \tag{14}$$

where, $(\mathcal{S}^1, \mathcal{S}^2) \in \Xi_S^{\{i\}}$, defined in (11), denoting any disjoint partitions of the set of covariates S_1 with cardinalities i and $K-i$, respectively. $X_{\mathcal{S}^1}$ and $X_{\mathcal{S}^2}$ are the corresponding disjoint partitions of the $K \times 1$ input X_{S_1} of lengths $i \times 1$ and $(K-i) \times 1$, respectively. We then have the following result:

Lemma 1. *The probability of the error event E_i defined in Eq. 9 that a set which differs from the set S_1 in exactly i covariates is selected by the ML decoder (averaged over all data realizations and outcomes) is bounded from above by*

$$P(E_i) \leq 2^{-T\left(E_o(\rho) - \rho\frac{\log\binom{N-K}{i}\binom{K}{i}}{T}\right)} \tag{15}$$

We are now ready to prove Theorem 1.

3.2 Proof of Theorem 1

Now we can readily prove our main result. First, we need to derive a sufficient condition for the error exponent of the error probability $P(E_i)$ in (15) to be positive and to drive the error probability to zero as $N \to \infty$. Specifically,

$$Tf(\rho) = TE_o(\rho) - \rho \log \binom{N-K}{i}\binom{K}{i} \to \infty \tag{16}$$

where

$$f(\rho) = E_o(\rho) - \rho \frac{\log \binom{N-K}{i}\binom{K}{i}}{T}$$

and where $E_o(\rho)$ is defined in (14).

To establish Eq. 12 we follow the argument in [11]. Note that $f(0) = 0$. Since the function $f(\rho)$ is differentiable and has a power series expansion, for a sufficiently small δ, we get by Taylor series expansion in the neighborhood of $\rho \in [0, \delta]$ that,

$$f(\rho) = f(0) + \rho \frac{df}{d\rho}\Big|_{\rho=0} + O(\rho^2)$$

But we can show that

$$\begin{aligned}\frac{\partial E_o}{\partial \rho}\Big|_{\rho=0} = \sum_Y \sum_{X_{\mathcal{S}^2}} \Bigg[&\sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1}) \log p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1}) \\ &- \sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1}) \log \sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1}) \Bigg]\end{aligned} \tag{17}$$

which simplifies to

$$\begin{aligned}\frac{\partial E_o}{\partial \rho}\Big|_{\rho=0} &= \sum_Y \sum_{X_{\mathcal{S}^2}} \sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1}) \log \frac{p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1})}{\sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2}|X_{\mathcal{S}^1})} \\ &= I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)\end{aligned} \tag{18}$$

Now it is easy to see that with $(1+\epsilon)\frac{\log \binom{N-K}{i}\binom{K}{i}}{T} < I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)$ for some constant $\epsilon > 0$, the condition in (16) is satisfied, i.e., $Tf(\rho) \to \infty$ as $N \to \infty$. To argue this we first note that from the Lagrange form of the Taylor Series expansion (an application of the mean value theorem) we can write $E_o(\rho)$ in terms of its first derivative evaluated at zero and a remainder term, i.e.,

$$E_o(\rho) = E_o(0) + \rho E_o'(0) + \frac{\rho^2}{2} E_o''(\psi) \tag{19}$$

for some $\psi \in [0, \rho]$. Hence, for the choice of T in (12) we have

$$Tf(\rho) \geq T\left(\rho \frac{\epsilon}{1+\epsilon} I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y) - \rho^2 C I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)\right)$$

where $C = -\frac{|E_o''(\psi)|}{I(X_{\mathcal{S}^1};X_{\mathcal{S}^2},Y)}$ which might depend on K. Now if we choose $\rho \leq \frac{\epsilon'}{C}$, where $\epsilon' = \frac{\epsilon}{1+\epsilon}$, then $f(\rho) = \delta$ for some $\delta > 0$ which does not depend on N or T. It follows that $Tf(\rho) \to \infty$ as $N \to \infty$.

We have just shown that for fixed K, $T > (1+\epsilon) \cdot \frac{\log\binom{N-K}{i}\binom{K}{i}}{I(X_{\mathcal{S}^1};X_{\mathcal{S}^2},Y)}$ is sufficient to ensure an arbitrarily small $P(E_i)$. Since the average error probability $P_e \leq \sum_{i=1}^{K} P(E_i)$, it follows that for any fixed K, $\lim_{N\to\infty}\sum_{i=1}^{K} P(E_i) = 0$. Consequently, since this is true for any K, $\lim_{K\to\infty}\lim_{N\to\infty}\sum_{i=1}^{K} P(E_i) = 0$. Theorem 1 now follows.

3.3 Proof of Lemma 1

We show the following weaker bound:

$$P(E_i) \leq 2^{-T\left(E_o(\rho) - \frac{\log\binom{N-K}{i}\binom{K}{i}}{T}\right)} \tag{20}$$

Note that the main difference between the above equation and Lemma 1 is the missing ρ term multiplying the binomial expression. The main result follows along the same lines and we refer the reader to [2] for further details.

To prove this weaker result we denote by $\mathcal{A}$

$$\mathcal{A} = \{\omega \in \mathcal{I} : |S_{1^c,\omega}| = i, |S_\omega| = K\} \tag{21}$$

the set of indices corresponding to sets of K covariates that differ from the true set S_1 in exactly i covariates. We can establish that,

$$\begin{aligned}\Pr[E_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] &\leq \sum_{\omega\in\mathcal{A}}\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}})\frac{p_\omega(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s} \qquad (22)\\ &= \sum_{S_{1,\omega}}\sum_{S_{1^c,\omega}}\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}})\frac{p_\omega(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s}\end{aligned}$$

Inequality (22) is established in the Appendix. It follows that,

$$\begin{aligned}\Pr[E_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] &\overset{(a)}{\leq} \left(\sum_{S_{1,\omega}}\sum_{S_{1^c,\omega}}\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}})\frac{p_\omega(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s}\right)^\rho\\ &\overset{(b)}{\leq} \left(\sum_{S_{1,\omega}}\binom{N-K}{i}\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}})\frac{p_\omega(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s}\right)^\rho\\ &\overset{(c)}{\leq} \binom{N-K}{i}\sum_{S_{1,\omega}}\left(\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}})\frac{p_\omega(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T,\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s}\right)^\rho\\ &\qquad \forall s > 0,\ 0 \leq \rho \leq 1\end{aligned}$$

Inequality (a) follows from the fact that $\Pr[E_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] \leq 1$. Consequently, if U is an upperbound of this probability then it follows that, $\Pr[E_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] \leq U^\rho$ for $\rho \in [0, 1]$. Inequality (b) follows from symmetry, namely, the inner summation is only dependent on the values of $\boldsymbol{X}_{S_{1^c,\omega}}$ and not on the items in the set $S_{1^c,\omega}$. There are exactly $\binom{N-K}{i}$ possible sets $S_{1^c,\omega}$ hence the binomial expression. Note that the sum over $S_{1,\omega}$ cannot be further simplified. This is due to the fact that $\boldsymbol{X}_{S_{1,\omega}}$ is already specified since we have conditioned on $\boldsymbol{X}_{S_1}$. Since $\boldsymbol{X}_{S_1}$ is fixed, the inner sum need not be equal for all sets $S_{1,\omega}, \omega \in \mathcal{A}$. Finally, (c) follows from standard observation that sum of positive numbers raised to ρ-th power for $\rho < 1$ is smaller than the sum of the ρ-th power of each number.

We now substitute for the conditional error probability derived above and follow the steps below:

$$P(E_i) = \sum_{\boldsymbol{X}_{S_1}} \sum_{Y^T} p_1(\boldsymbol{X}_{S_1}, Y^T) \Pr[E_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T]$$
$$\leq \binom{N-K}{i} \sum_{S_{1,\omega}} \sum_{Y^T} \sum_{\boldsymbol{X}_{S_1}} p_1(\boldsymbol{X}_{S_1}, Y^T) \left(\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) \frac{p_\omega(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s} \right)^\rho$$

Due to symmetry the summation over sets $S_{1,\omega}$ does not depend on ω. Since there are $\binom{K}{K-i}$ sets $S_{1,\omega}$ we get,

$$P(E_i) \leq \binom{N-K}{i} \binom{K}{i} \sum_{Y^T} \sum_{\boldsymbol{X}_{S_1}} p_1(\boldsymbol{X}_{S_1}, Y^T) \left(\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) \frac{p_\omega(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s} \right)^\rho$$
$$\leq \binom{N-K}{i} \binom{K}{i} \sum_{Y^T} \sum_{\boldsymbol{X}_{S_{1,\omega^c}}} \sum_{\boldsymbol{X}_{S_{1,\omega}}} Q(\boldsymbol{X}_{S_{1,\omega^c}}) p_1(\boldsymbol{X}_{S_{1,\omega}}, Y^T \mid \boldsymbol{X}_{S_{1,\omega^c}})$$
$$\left(\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) \frac{p_\omega(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p_1(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s} \right)^\rho$$
$$= \binom{N-K}{i} \binom{K}{i} \sum_{Y^T} \sum_{\boldsymbol{X}_{S_{1,\omega^c}}} \sum_{\boldsymbol{X}_{S_{1,\omega}}} Q(\boldsymbol{X}_{S_{1,\omega^c}}) p_1^{1-s\rho}(\boldsymbol{X}_{S_{1,\omega}}, Y^T \mid \boldsymbol{X}_{S_{1,\omega^c}})$$
$$\left(\sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) p_\omega(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s \right)^\rho$$
$$= \binom{N-K}{i} \binom{K}{i} \sum_{Y^T} \sum_{\boldsymbol{X}_{S_{1,\omega}}} \left(\sum_{\boldsymbol{X}_{S_{1,\omega^c}}} Q(\boldsymbol{X}_{S_{1,\omega^c}}) p_1^{1/(1+\rho)}(\boldsymbol{X}_{S_{1,\omega}}, Y^T \mid \boldsymbol{X}_{S_{1,\omega^c}}) \right)^{1+\rho}$$
$$= \binom{N-K}{i} \binom{K}{i} \sum_{Y} \sum_{X_{S_{1,\omega}}} \left(\sum_{X_{S_{1,\omega^c}}} Q(X_{S_{1,\omega^c}}) p_1^{1/(1+\rho)}(X_{S_{1,\omega}}, Y^T \mid X_{S_{1,\omega^c}}) \right)^{T(1+\rho)}$$

where the last step follows due to time independence. The step before the last follows by noting that from symmetry $\boldsymbol{X}_{S_{1^c,\omega}}$ is just a dummy variable and can be replaced by $\boldsymbol{X}_{S_{1,\omega^c}}$. This establishes the weaker bound in (20). Further details about the proof of Lemma 1 can be found in [2].

3.4 Lower Bound: Fano's inequality

In this section we also derive lower bounds on the required number of measurements using Fano's inequality [5]. We state the following theorem

Theorem 2. *For N covariates and a set S_ω of K salient covariates, a lower bound on the total number of measurement required to recover the positive set is given by*

$$T \geq \max_{i:(\mathcal{S}^1,\mathcal{S}^2)\in\Xi^{\{i\}}_{S_\omega}} \frac{\log\binom{N-K+i}{i}}{I(X_{\mathcal{S}^1};X_{\mathcal{S}^2},Y)}, \quad i=1,2,\dots K \tag{23}$$

where the set $\Xi^{\{i\}}_S$ is the set of tuples $(\mathcal{S}^1,\mathcal{S}^2)$ partitioning the set S into disjoint sets $\mathcal{S}^1$ and $\mathcal{S}^2$ with cardinalities i and $K-i$, respectively as defined in (11).

Proof. The vector of outcomes Y^T is probabilistically related to the index $\omega \in \mathcal{I} = \{1,2,\dots,\binom{N}{K}\}$. Suppose $K-i$ covariates are revealed to us. In particular, we know the $K-i\times T$ sub-matrix $\boldsymbol{X}_{\mathcal{S}^2}$ of the matrix $\boldsymbol{X}_{\mathcal{S}_\omega}$. From Y^T we estimate the set index ω. Let the estimate be $\hat{\omega}=g(Y^T)$. Define the probability of error

$$P_e = \Pr[\hat{\omega}\neq\omega]$$

If E is a binary random variable that takes the value 1 in case of an error i.e., if $\hat{\omega}\neq\omega$, and 0 otherwise. Then using the chain rule of entropies [5] [11] we have

$$\begin{aligned} H(E,\omega|Y^T,\boldsymbol{X}_{\mathcal{S}^2}) &= H(\omega|Y^T,\boldsymbol{X}_{\mathcal{S}^2}) + H(E|\omega,Y^T,\boldsymbol{X}_{\mathcal{S}^2}) \\ &= H(E|Y^T,\boldsymbol{X}_{\mathcal{S}^2}) + H(\omega|E,Y^T,\boldsymbol{X}_{\mathcal{S}^2}) \end{aligned} \tag{24}$$

The random variable E is fully determined given Y^T and ω. It follows that $H(E|\omega,Y^T,\boldsymbol{X}_{\mathcal{S}^2}) = 0$. Since E is a binary random variable $H(E|Y^T,\boldsymbol{X}_{\mathcal{S}^2}) \leq 1$. Consequently, we can bound $H(\omega|E,Y^T,\boldsymbol{X}_{\mathcal{S}^2})$ as follows

$$\begin{aligned} H(\omega|E,Y^T,\boldsymbol{X}_{\mathcal{S}^2}) &= P(E=0)H(\omega|E=0,Y^T,\boldsymbol{X}_{\mathcal{S}^2}) + P(E=1)H(\omega|E=1,Y^T,\boldsymbol{X}_{\mathcal{S}^2}) \\ &\leq (1-P_e)0 + P_e\log\left(\binom{N-K+i}{i}-1\right) \\ &\leq P_e\log\binom{N-K+i}{i} \end{aligned} \tag{25}$$

The second inequality follows from the fact that revealing $K-i$ entries, and given that $E=1$, the conditional entropy can be upper bounded by the logarithm of the number of outcomes. From (24), we obtain the genie aided Fano's inequality

$$H(\omega|Y^T,\boldsymbol{X}_{\mathcal{S}^2}) \leq 1 + P_e\log\binom{N-K+i}{i} \tag{26}$$

Since for a fixed code $\boldsymbol{X}_{S_\omega}$ is a function of ω, then

$$H(\boldsymbol{X}_{S_\omega}|Y^T, \boldsymbol{X}_{\mathcal{S}^2}) \le H(\omega|Y^T, \boldsymbol{X}_{\mathcal{S}^2})$$

and hence

$$H(\boldsymbol{X}_{S_\omega}|Y^T, \boldsymbol{X}_{\mathcal{S}^2}) \le 1 + P_e \log \binom{N-K+i}{i} \tag{27}$$

Since the set $\mathcal{S}^2$ of $K-i$ covariates is revealed, ω is uniformly distributed over the set of indices that correspond to subsets of size K containing $\mathcal{S}^2$. It follows that

$$\begin{aligned}\log \binom{N-K+i}{i} &= H(\omega|\boldsymbol{X}_{\mathcal{S}^2}) \\ &= H(\omega|Y^T, \boldsymbol{X}_{\mathcal{S}^2}) + I(\omega; Y^T|\boldsymbol{X}_{\mathcal{S}^2}) \\ &\le 1 + P_e \log \binom{N-K+i}{i} + I(\boldsymbol{X}_{S_\omega}; Y^T|\boldsymbol{X}_{\mathcal{S}^2}) \end{aligned} \tag{28}$$

Since $S_\omega = \mathcal{S}^1 \cup \mathcal{S}^2$, where $(\mathcal{S}^1, \mathcal{S}^2) \in \Xi^{\{i\}}_{S_\omega}$, we have

$$P_e \ge 1 - \frac{I(\boldsymbol{X}_{\mathcal{S}^1}; Y^T|\boldsymbol{X}_{\mathcal{S}^2}) + 1}{\log \binom{N-K+i}{i}} \tag{29}$$

Thus, for the probability of error to be asymptotically bounded away from zero, it is necessary that

$$\log \binom{N-K+i}{i} \le I(\boldsymbol{X}_{\mathcal{S}^1}; Y^T|\boldsymbol{X}_{\mathcal{S}^2}) \tag{30}$$

Following a standard set of inequalities we have

$$\begin{aligned}\log \binom{N-K+i}{i} &\le I(\boldsymbol{X}_{\mathcal{S}^1}; Y^T|\boldsymbol{X}_{\mathcal{S}^2}) \\ &= H(Y^T|\boldsymbol{X}_{\mathcal{S}^2}) - H(Y^T|\boldsymbol{X}_{S_\omega}) \\ &\overset{\text{(a)}}{=} \sum_{t=1}^{T} H(Y(t)|Y^{t-1}, \boldsymbol{X}_{\mathcal{S}^2}) - H(Y(t)|X_{S_\omega}(t)) \\ &\overset{\text{(b)}}{\le} \sum_{t=1}^{T} H(Y(t)|X_{\mathcal{S}^2}(t)) - H(Y(t)|X_{S_\omega}(t)) \\ &= \sum_{t=1}^{T} I(X_{S_\omega}(t); Y(t)|X_{\mathcal{S}^2}(t)) \\ &= \sum_{t=1}^{T} I(X_{\mathcal{S}^1}(t); Y(t)|X_{\mathcal{S}^2}(t)) \\ &\overset{\text{(c)}}{=} TI(X_{\mathcal{S}^1}; Y|X_{\mathcal{S}^2}) \\ &\overset{\text{(d)}}{=} TI(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y) \end{aligned} \tag{31}$$

In (a) we made use of the chain rule for entropy and the memoryless property of the channel and (b) is true since conditioning reduces entropy. (c) is due to the i.i.d. assumption. Finally, (d) follows from the chain rule for mutual information and the independence across covariates, i.e.,

$$I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y) = I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}) + I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}) \tag{32}$$

and $I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}) = 0$.

Since (31) has to be true for all i, a necessary condition on the total number of measurements is given by

$$T \geq \max_{i:(\mathcal{S}^1,\mathcal{S}^2)\in \Xi^{\{i\}}_{S_\omega}} \frac{\log \binom{N-K+i}{i}}{I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)} \tag{33}$$

proving theorem 2.

4 Applications

The applicability of our necessity and sufficiency results spans a wide range of applications including compressive sensing and quantized compressive sensing. In this section, we consider the Boolean model (group testing) as an example.

4.1 Boolean Model

The fundamental problem of group testing can be summarized as follows. Among a population of N items, K unknown items are of interest. The collection of these K items represents the defective set. The goal is to construct a pooling design, i.e., a collection of tests, to recover the defective set while reducing the number of required tests. In this case $\boldsymbol{X}$ is a binary measurement matrix defining the assignment of items to tests. For the noise-free case, the outcome of the tests Y^T is deterministic. It is the Boolean sum of the codewords corresponding to the defective set $\mathcal{G}$. In other words

$$Y^T = \bigvee_{i \in \mathcal{G}} X_i^T. \tag{34}$$

Alternatively, if $R_i \in \{0, 1\}$ is an indicator function for the i-th item determining whether it belongs to the defective set, i.e., $R_i = 1$ if $i \in \mathcal{G}$ and $R_i = 0$ otherwise. Then, the outcome $Y(t)$ of the t-th test in the noise-free case can be written as

$$Y(t) = \bigvee_{i=1}^{N} X_i(t) R_i \tag{35}$$

where $X_i(t)$ is the t-th entry of the vector X_i^T, or equivalently, the binary entry at cell (i, t) of the measurement matrix $\boldsymbol{X}$.

First, consider the noise free (deterministic case). The test outcome Y is 1 if and only if a defective item is pooled in that test. Hence Y is given by (34). We refer the reader to [2] for proofs of the following results which we state here for completeness. For brevity we state our results for the average error case and refer the reader to [2] for further results on worst-case error analysis.

Theorem 3. *For N items and K defectives, the number of tests $T = O(K \log N)$ is sufficient to satisfy an average error criterion, i.e., achieve an arbitrarily small average error probability. In other words, there is a constant c independent of N and K such that if $T = cK \log N$ then the probability of errors goes to zero.*

Our result also establishes upper and lower bounds on the number of tests needed for noisy versions of group testing. In particular, we consider testing with additive noise (leading to false alarms) and testing with dilution effects (leading to potential misses). We refer the reader to [2] for further details.

4.2 Compressive Sensing

The model is given by

$$Y = \mathbf{X}\beta + W \tag{36}$$

Where the output Y is a $T \times 1$ vector, the $T \times N$ codebook or the compression matrix $\mathbf{X}$ is such that $X_{i,j} \sim \mathcal{N}(0, \frac{SNR}{T})$. The unknown sparse signal β is $N \times 1$ and the noise W is a $T \times 1$ vector. The support of β denoted by the index set S, with $|S| = K$, is partitioned into two sets S_1 and S_2. We further assume that the elements in the support of β are $\mathcal{N}(\mu, \sigma^2)$.

For the achievability result we need to lower bound the mutual information expression

$$I(\mathbf{X}_{S_1}; Y|X_{S_2}) = h(Y|\mathbf{X}_{S_2}) - h(Y|\mathbf{X}_S) \tag{37}$$

Hence if $\hat{\beta}$ denotes the MMSE of β and β_S is the value of β on the support, then,

$$\begin{aligned} h(Y|\mathbf{X}_S) &\leq h(Y - \mathbf{X}_S\hat{\beta}_S|\mathbf{X}_S) \leq h(Y - \mathbf{X}_S\hat{\beta}_S) \\ &\leq \frac{1}{2}\log\left((2\pi e)^T|\mathbf{K}_V|\right) \end{aligned}$$

where $|\mathbf{K}_V|$ denotes the determinant of the covariance matrix K_V of the random vector V where

$$V = \mathbf{X}_S(\beta_S - \hat{\beta}_S) + W = \mathbf{X}_S e + W$$

The last inequality follows from the fact that Gaussian maximizes entropy. e is the error. But,

$$\mathbf{K}_V = \mathbf{X}_S\mathbf{\Lambda}_e\mathbf{X}_S^{\dagger} + \mathbf{I}$$

Alternatively, one can also upper bound the entropy using the chain rule and the fact that conditioning reduces entropy,

$$h(Y|\mathbf{X}_S) \leq \sum_{i=1}^{T} h(V_i) \leq \sum_i \frac{1}{2}\log\left(2\pi e\sigma_{V_i}^2\right)$$

V_i is i-th entry of the vector V with variance

$$\sigma_{V_i}^2 = \mathbf{X_S}(i,.)\mathbf{\Lambda}_e\mathbf{X_S}(i,.)^{\dagger} + 1$$

The error covariance Λ_e is

$$\Lambda_e = \sigma^2 \mathbf{I} - \sigma^2 \mathbf{X}_S^\dagger (\sigma^2 \mathbf{X}_S \mathbf{X}_S^\dagger + \mathbf{I})^{-1} \sigma^2 \mathbf{X}_S$$

Since we had conditioned on a particular realization of the matrix $\mathbf{X}$ we now take the expectation

$$E[h(Y|\mathbf{X}_S)] = E_{\mathbf{X}} \left[\frac{1}{2} \log \left(2\pi e \sigma_{V_i}^2 \right) \right]$$

Replacing we get that,

$$\sigma_{V_i}^2 = 1 + \mathbf{X_S}(i,.) \left[\sigma^2 \mathbf{I} - \sigma^2 \mathbf{X}_S^\dagger (\sigma^2 \mathbf{X}_S \mathbf{X}_S^\dagger + \mathbf{I})^{-1} \sigma^2 \mathbf{X}_S \right] \mathbf{X_S}(i,.)^\dagger \tag{38}$$

Now note that for large enough problem size $\mathbf{X}_S \mathbf{X}_S^\dagger \rightarrow \frac{KSNR}{T} \mathbf{I}_T$ and $\mathbf{X}_S^\dagger \mathbf{X}_S \rightarrow SNR \, \mathbf{I}_K$. Hence,

$$\begin{aligned} \sigma_{V_i}^2 &= 1 + \mathbf{X_S}(i,.) \left[\sigma^2 \mathbf{I} - \frac{\sigma^4 SNR}{1 + \frac{\sigma^2 KSNR}{T}} \mathbf{I} \right] \mathbf{X_S}(i,.)^\dagger \\ &= 1 + C(SNR, \sigma^2, K, T) ||X_S(i,.)||^2 \end{aligned} \tag{39}$$

where $C(SNR, \sigma^2, K, T) = \sigma^2 - \frac{\sigma^4 SNR}{1 + \frac{\sigma^2 KSNR}{T}}$. Now using Jensen's inequality and the concavity of the $\log$

$$\begin{aligned} E[h(Y|\mathbf{X_S})] &\leq \frac{T}{2} \log \left(2\pi e (1 + C(SNR, \sigma^2, K, T) E[||\mathbf{X}_S(i,.)]||^2) \right) \\ &= \frac{T}{2} \log \left(2\pi e \left((1 + C(SNR, \sigma^2, K, T) \frac{KSNR}{T} \right) \right) \end{aligned} \tag{40}$$

We also need to bound the first conditional entropy term from below, i.e., $H(Y|\mathbf{X}_{S_2})$. We proceed as follows

$$\begin{aligned} h(Y|\mathbf{X}_{S_2}) &\geq h(Y|\mathbf{X}_{S_2}, \beta_{S_2}) = h(\mathbf{X}_{S_1}\beta_{S_1} + \mathbf{X}_{S_2}\beta_{S_2} - \mathbf{X}_{S_2}\beta_{S_2} + W) \\ &= h(\mathbf{X}_{S_1}\beta_{S_1} + W) \\ &\geq \frac{T}{2} \log \left(2\pi e \det(\mathbf{I}_T + \sigma^2 X_{S_1} X_{S_1}^\dagger)^{\frac{1}{T}} \right) \end{aligned}$$

Note that for large enough problem sizes,

$$\det(\mathbf{I}_T + \sigma^2 X_{S_1} X_{S_1}^\dagger) = \det \left(\mathbf{I}_T + \frac{i\sigma^2 SNR}{T} \mathbf{I}_T \right) = \left(1 + \frac{i\sigma^2 SNR}{T} \right)^T$$

Hence,

$$E[h(Y|\mathbf{X}_{S_2})] \geq \frac{T}{2} \log \left(2\pi e \left(1 + \frac{i\sigma^2 SNR}{T} \right) \right) \tag{41}$$

Combining (41) and (40) with the condition

$$\log \binom{N-K}{i}\binom{K}{i} \leq I(\mathbf{X}_{S_1};Y|\mathbf{X}_{S_2})$$

we see that $T = \Omega\left(K \log N\right)$ for $SNR = \Omega(\log N)$. The condition on the SNR can be easily established from the converse bound as in [1].

Appendix A: Proof of Equation 22

Let ζ_ω, $\omega \in \mathcal{A}$ denote the event where ω is more likely than 1. Then, from the definition of $\mathcal{A}$, the 2 encoded messages differ in i covariates. Hence

$$\Pr[E_i|\omega_0 = 1, \boldsymbol{X}_{S_1}, Y^T] \leq P(\bigcup_{\omega\in\mathcal{A}} \zeta_\omega) \leq \sum_{\omega\in\mathcal{A}} P(\zeta_\omega)$$

Now note that $\boldsymbol{X}_{S_1}$ shares $(K-i)$ covariates with $\boldsymbol{X}_{S_\omega}$. Following the introduced notation, the common partition is denoted $\boldsymbol{X}_{S_{1,\omega}}$, which is a $(K-i)\times T$ submatrix. The remaining i rows which are in $\boldsymbol{X}_{S_1}$ but not in $\boldsymbol{X}_{S_\omega}$ are $\boldsymbol{X}_{S_{1,\omega^c}}$. Similarly, $\boldsymbol{X}_{S_{1^c,\omega}}$ corresponds to covariates in $\boldsymbol{X}_{S_\omega}$ but not in $\boldsymbol{X}_{S_1}$. In other words $\boldsymbol{X}_{S_1} = (\boldsymbol{X}_{S_{1,\omega}}; \boldsymbol{X}_{S_{1,\omega^c}})$ and $\boldsymbol{X}_{S_\omega} = (\boldsymbol{X}_{S_{1,\omega}}; \boldsymbol{X}_{S_{1^c,\omega}})$, where the notation $(\boldsymbol{F}^{n_1\times T}; \boldsymbol{G}^{n_2\times T})$ denotes an $(n_1+n_2)\times T$ matrix with a submatrix $\boldsymbol{F}$ in the first n_1 rows and $\boldsymbol{G}$ in the remaining n_2 rows. Thus,

$$\begin{aligned} P(\zeta_\omega) &= \sum_{\boldsymbol{X}_{S_\omega}: p(Y^T|\boldsymbol{X}_{S_\omega}) \geq p(Y^T|\boldsymbol{X}_{S_1})} Q(\boldsymbol{X}_{S_\omega}|\boldsymbol{X}_{S_1}) \\ &\leq \sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) \frac{p(Y^T|\boldsymbol{X}_{S_\omega})^s}{p(Y^T|\boldsymbol{X}_{S_1})^s} \quad \forall s>0,\ \forall \omega\in\mathcal{A} \qquad \text{(A.1)} \end{aligned}$$

By independence $Q(\boldsymbol{X}_{S_1}) = Q(\boldsymbol{X}_{S_{1,\omega}})Q(\boldsymbol{X}_{S_{1,\omega^c}})$. Similarly, $Q(\boldsymbol{X}_{S_\omega}) = Q(\boldsymbol{X}_{S_{1,\omega}})Q(\boldsymbol{X}_{S_{1^c,\omega}})$. Since we are conditioning on a particular $\boldsymbol{X}_{S_1}$, the partition $\boldsymbol{X}_{S_{1,\omega}}$ is fixed in the summation in (A.1) and

$$\begin{aligned} P(\zeta_\omega) &\leq \sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) \frac{p(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{Q(\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s} \frac{Q(\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s}{p(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s} \\ &\leq \sum_{\boldsymbol{X}_{S_{1^c,\omega}}} Q(\boldsymbol{X}_{S_{1^c,\omega}}) \frac{p(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}})^s}{p(Y^T, \boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}})^s} \quad \forall s>0 \qquad \text{(A.2)} \end{aligned}$$

where the second inequality follows from the independence across covariates, i.e. $Q(\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1,\omega^c}}) = Q(\boldsymbol{X}_{S_{1,\omega}}|\boldsymbol{X}_{S_{1^c,\omega}}) = Q(\boldsymbol{X}_{S_{1,\omega}})$.

References

1. Aeron, S., Saligrama, V., Zhao, M.: Information theoretic bounds for compressed sensing. IEEE Transactions on Information Theory 56(10), 5111–5130 (oct 2010)

2. Atia, G., Saligrama, V.: Boolean compressed sensing and noisy group testing. Accepted to IEEE Transactions on Information Theory, Available online: http://arxiv.org/abs/0907.1061 (2009)
3. Berger, T., Levenshtein, V.: Asymptotic efficiency of two-stage disjunctive testing. Information Theory, IEEE Transactions on 48(7), 1741–1749 (Jul 2002)
4. Candes, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59, 1207–1223 (2005)
5. Cover, T.M., Thomas, J.A.: Elements of Information Theory. New York: John Wiley and Sons, Inc. (1991)
6. Donoho, D.: Compressed sensing. IEEE Transactions on Information Theory 52(4), 1289 –1306 (apr 2006)
7. Du, D.Z., Hwang, F.K.: Pooling designs and non-adaptive group testing: Important tools for DNA sequencing. World Scientific Publishing Company (2006)
8. Dyachkov, A.G., Rykov, V.V., Rashad, A.M.: Bounds of the length of disjunct codes. Problems Control Information Theory 11, 7–13 (1982)
9. Fletcher, A.K., Rangan, S., Goyal, V.K.: Necessary and sufficient conditions for sparsity pattern recovery. IEEE Trans. Inf. Theor. 55(12), 5758–5772 (2009)
10. Fletcher, A.K., Rangan, S., Goyal, V.K., Ramchandran, K.: Denoising by sparse approximation: Error bounds based on rate-distortion theory. EURASIP Journal on Applied Signal Processing pp. 1–19 (2006)
11. Gallager, R.G.: Information Theory and Reliable Communication. John Wiley & Sons, Inc., New York, NY, USA (1968)
12. Gilbert, A.C., Iwen, M.A., Strauss, M.J.: Group testing and sparse signal recovery. In: 42nd Asilomar Conference on Signals, Systems and Computers (2008)
13. Reeves, G., Gastpar, M.C.: A note on optimal support recovery in compressed sensing. In: 43rd Annual IEEE Asilomar Conference on Signals, Systems, and Computers. Monterey, CA (Nov 2009)
14. Ruszinkó, M.: Family of finite sets in which no set is covered by the union of n others. Isr. J. Math. 51, 79–89 (1985)
15. Sebo, A.: On two random search problems. Journal of Statistical Planning and Inference 11, 23–31 (1985)
16. Wainwright, M.: Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting. In: International Symposium on Information Theory. Nice, France (June 2007)
17. Wainwright, M.J.: Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programs. In: Allerton Conference on Communication, Control and Computing. Monticello, IL (2006)