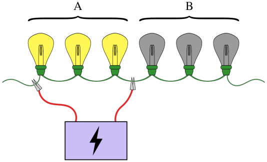
Dilution Group Testing: Novel Bounds via Practical Decoders

Gabriel Arpino
w. Nicolò Grometto
Supervision: Prof. Dr. Afonso Bandeira (D-MATH), Prof. Dr. Helmut Bölcskei (D-ITET)

Overview

- Motivation
- Problem Setup
- NDD Algorithm
- Achievability Result for NDD
- Converse Results
- Experimental Results

Motivation



Contributions:

- First achievability results for DCN
- Preliminary results on the algorithm-independent converse

Problem Setup

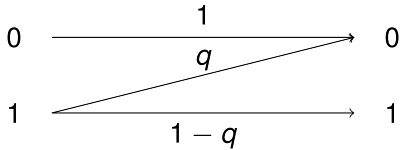
- n: total number of items;
- $d = \Theta(n^{\theta})$: number of defective items for $\theta \in (0, 1)$;
- $T = \beta d \log n$: number of tests sufficient for vanishing P_e;
- $\boldsymbol{\sigma}$: Parameter of interest (Boolean vec.) defectives (1) & non-defectives (0);
- $\hat{\boldsymbol{\sigma}}$: Estimate (Boolean vec.) defectives (1) & non-defectives (0);
- M: $n \times T$ design matrix ($m_{(i,j)} \sim \text{Bernoulli}(\frac{\alpha}{d})$);

- $\mathcal{D}$: set of indices for defective items
- Noiseless Model:

$$Y_j = \bigvee_{i \in \mathcal{D}} M_{i,j}$$

- Dilution Model: Samples can get diluted. Z-Channel with transition probability q:

$$Y_j = \bigvee_{i \in \mathcal{D}} \mathcal{Z}(M_{i,j})$$



NDD Algorithm [Scarlett & Johnson 2020]

NDD Algorithm for Dilution Noise

1. For each $i \in \{1, \ldots, n\}$, let $N_{\text{neg,i}}$ be the number of negative tests in which item i is included. In the first step, fix a constant $\xi \in (q, 1)$ and construct the following set of possibly defective items:

$$\hat{PD} = \left\{ i \in \{1, \ldots, n\} : N_{\text{neg,i}} < \frac{\xi T \alpha}{d} \right\}$$

2. For each $i \in \hat{PD}$, let $N'_{\text{pos,i}}$ be the number of positive tests that include item j and no other item from $\hat{PD}$. We estimate the defective set as follows:

$$\hat{\sigma} = \left\{ i \in \hat{PD} : N'_{\text{pos,i}} > 0 \right\}$$

Achievability Result for NDD

Theorem 2. For noisy group testing under dilution channel noise with parameter $q \in (0, 1)$, in the sparsity regime $d = \Theta(n^{\theta})$, with $\theta \in (0, 1)$, *NDD* achieving $P_e \to 0$ requires no more than

$$T > \max\{T_a, T_b, T_c\}$$

tests, where

$$T_a = \frac{d \log d \cdot (1 + o(1))}{\alpha q e^{-\alpha(1-q)} \cdot D\left(\frac{\xi}{q} e^{\alpha(1-q)}\right)}$$

$$T_b = \frac{(1-\nu) \cdot e^{\alpha(1-q)} \cdot d \log n \cdot (1 + o(1))}{\alpha \cdot D\left(\xi e^{\alpha(1-q)}\right)}$$

$$T_c = \frac{e^{\alpha} d \log d \cdot (1 + o(1))}{\alpha(1-q)}$$

for Bernoulli designs with parameter $\frac{\alpha}{d}$, $\xi \in (q, e^{-\alpha(1-q)})$, $D(x) := x\log(x) - x + 1$, and ν arbitrarily close to θ.

Converse Results

Claim 5.

The strong converse number of tests T_{sc} below which $P_e \to 1$ as $n \to \infty$ for any dilution group testing algorithm is given by:

$$T_{sc} = \frac{\log_2 \binom{n}{k}}{\max_{\alpha \in [0,d]} H_b(e^{-\alpha(1-q)}) - \mathbb{E}_X[H_b(q^X)]}(1-\eta)(1+o(1))$$

where H_b is the binary entropy function, $X \sim \text{Poisson}(\alpha)$, $\eta > 0$.

Claim 6.

For $q \in (0, 0.5)$, $T_{sc} \in [d \log_2(\frac{n}{d}), \frac{d \log_2(\frac{n}{d})}{1-H_b(q)})$.

Experimental Results

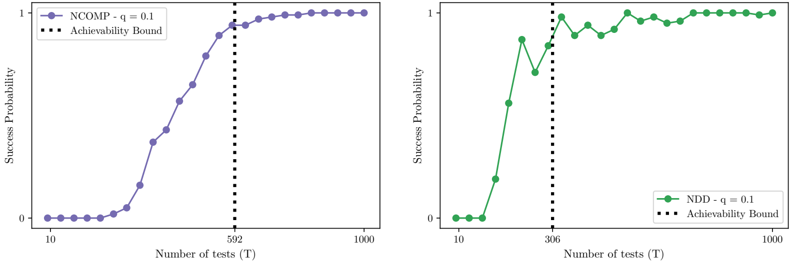


Figure: *NCOMP* (left) and *NDD* (right) with dilution noise; $q = 0.1$, $n = 10000$, $\theta = 0.15$, $d = 6 \sim O(2n^{0.15})$

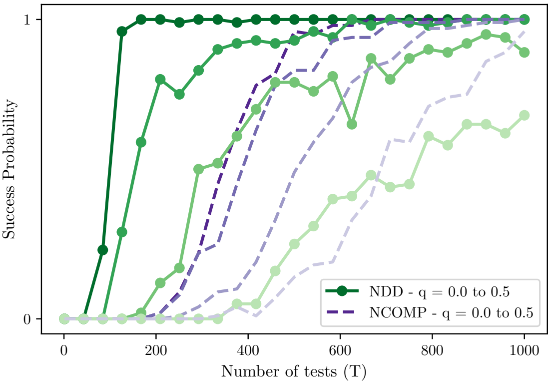


Figure: Comparison of the *NCOMP* and *NDD* algorithms; $q = \{0.00001, 0.1, 0.3, 0.5\}$, $n = 10000$, $\theta = 0.15$, $d = 6 \sim O(2n^{0.15})$

Thank you!

References

[SJ 2020]: J. Scarlett and O. Johnson, "Noisy non-adaptive group testing: A (near-)definite defectives approach," IEEE Transactions on Information Theory, 2020.

Pictures on slide 3 obtained from:

- https://upload.wikimedia.org/wikipedia/commons/thumb/f/fa/Group_testing_lightbulbs.svg/1200px-Group_testing_lightbulbs.svg.png
- https://images.agoramedia.com/EverydayHealth/gcms/Blood-Test-for-Pancreatic-Cancer-Shows-Early-Promise-1 jpg

Sketch of Proof

Step 1 (T_a): Ensure $\hat{PD}$ contains all defectives. Bound $N_{\text{neg, i}}$, obtain T_a.

$$N_{\text{neg,i}} \sim \text{Binomial}\left(T, q\frac{\alpha}{d} \cdot \left(1 - \frac{\alpha(1-q)}{d}\right)^{d-1}\right)$$

$$\mathbb{P}[\exists k \in \mathcal{K} : k \notin \hat{\text{PD}}] \leq e^{f(T)}$$

Step 1 (T_b): Ensure $\hat{PD}$ contains no more than n^{ν} non-defective items, $\nu \in (0, \theta)$. For $i \notin \mathcal{D}$:

$$\mathbb{P}[i \in \hat{\mathrm{PD}} | N_{\mathrm{neg}} = Te^{-\alpha}] \leq \exp\left(-(Te^{-\alpha})\frac{\alpha}{d}D(1-\eta) \cdot (1 + o(1))\right)$$

Step 2: (T_c): Ensure all defectives in $\hat{PD}$ are correctly classified.

$$\left(N'_{\text{pos,j}} | \hat{\text{pd}}, g, n_{\text{neg}}\right) \sim \text{Binomial}\left(T\left(1 - e^{\alpha(1-q)}\right), \frac{\alpha(1-q)e^{-\alpha}}{d(1 - e^{-\alpha(1-q)})}\right)$$