Exact Thresholds in Noisy Non-Adaptive Group Testing

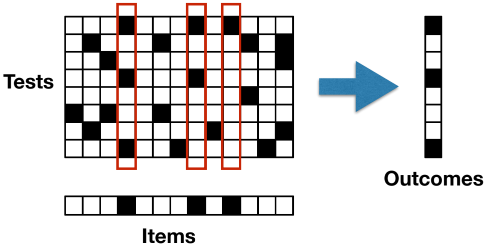
Junren Chen
The University of Hong Kong

January 15, 2025
SODA25, New Orleans

Joint work with Jonathan Scarlett (NUS)

I. Problem Setup

(Noisy) Group Testing



In this talk, we consider probabilistic group testing:

- Defective set $S \sim \mathrm{Uniform}\binom{p}{k}$ (i.e., k out of p items with a uniform prior)
- Non-adaptive: the test design $\mathbf{X} = (X_{ij}) \in \{0,1\}^{n \times p}$ is fixed before observing any outcome
- Noiseless:

$$Y_i = \bigvee_{j \in S} X_{ij} \tag{1}$$

- Noisy:

$$Y_i = \Big(\bigvee_{j \in S} X_{ij}\Big) \oplus Z_i \tag{2}$$

with $Z \sim \mathrm{Bernoulli}(\rho)$ for some noise level $\rho \in \left(0, \frac{1}{2}\right)$

Recovery Criteria

- We consider two popular random designs:
  - Bernoulli design: $X_{ij} \overset{iid}{\sim} \mathrm{Bernoulli}(\frac{\nu}{k})$; each item is independently placed in each test with probability $\frac{\nu}{k}$ for some $\nu > 0$
  - Near-constant weight design: each item is independently placed in $\Delta = \frac{\nu n}{k}$ tests chosen uniformly at random with replacement for some $\nu > 0$

- Given a decoder $\widehat{S}$, we define error probability as

$$P_{\mathrm{e}} := \mathbb{P}[\widehat{S} \neq S]$$

  taken w.r.t. randomness of $(S, \mathbf{X}, \mathbf{Z})$

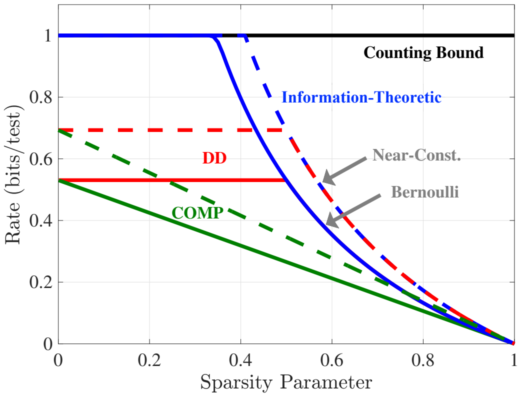
- **Goal**: Conditions on n for $P_{\mathrm{e}} \to 0$ in the large-system limit
- Sublinear sparsity: $k = \Theta(p^{\theta})$ for $\theta \in (0, 1)$
- Our work establishes the exact thresholds $n^* = Ck \log \frac{p}{k}$ with precise constants C for both designs, such that:
  - (Exact achievability)

    When $n > (1 + o(1))n^*$, some decoder gives $P_e \to 0$;

  - (Exact converse)

    When $n < (1 - o(1))n^*$, any decoder suffers from $P_e \to 1$

Milestones in Noiseless GT ($rate = \lim_{p\to\infty} \frac{\log_2 \binom{p}{k}}{n}$)



- Exact thresholds for bernoulli design [1] (ensemble tightness) [2]
- Exact thresholds for NCC design and ensemble tightness [3]
- The blue dashed curve is near optimal for arbitrary design [4]

---

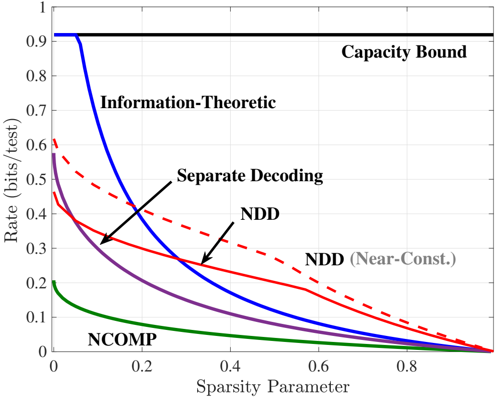
[1] Phase transitions in group testing, J. Scarlett and V. Cevher, 16 SODA

[2] The capacity of Bernoulli nonadaptive group testing, M. Aldridge, 17 T-IT

[3] Information-theoretic and algorithmic thresholds for group testing, A. Coja-Oghlan et al., 20 T-IT

[4] Optimal group testing, Coja-Oghlan et al., 20 COLT

Noisy GT Bounds Before Our Work ($\rho = 0.01$)

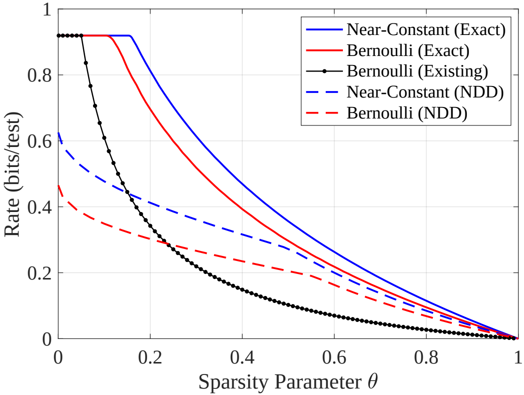


- Information-theoretic upper bounds that are tight for very small values of θ [SC16]
- The information-theoretic upper bound is an attempt to exact thresholds under Bernoulli design [5]
- Compared to the noiseless case, the prior work is much less complete!

[5] Noisy non-adaptive group testing: A (near-)definite defectives approach, J. Scarlett and O. Johnson, 20 T-IT

II. Exact Thresholds

Graphical Illustration of Our Exact Thresholds ($\rho = 0.01$)



- The best exisiting efficient algorithms fall short of information theoretic thresholds

Preliminaries

Notation:

$$a \star b = ab + (1-a)(1-b) \tag{3}$$

$$D(a\|b) = a\log\left(\frac{a}{b}\right) + (1-a)\log\left(\frac{1-a}{1-b}\right) \tag{4}$$

$$H_2(a) = a\log\left(\frac{1}{a}\right) + (1-a)\log\left(\frac{1}{1-a}\right) \tag{5}$$

Technical Lemma:
Consider $X \sim \mathrm{Bin}(N, q)$, then we have

- Chernoff bound:

$$\mathbb{P}\big(X \le k\big) \le \exp\Big(-N \cdot D\big(\frac{k}{N}\|q\big)\Big), \quad \text{if } k \le Nq \tag{6}$$

$$\mathbb{P}\big(X \ge k\big) \le \exp\Big(-N \cdot D\big(\frac{k}{N}\|q\big)\Big), \quad \text{if } k \ge Nq \tag{7}$$

- Anti-concentration:

$$\mathbb{P}(X = k) \ge \underbrace{\frac{1}{2\sqrt{2k(1-\frac{k}{N})}}\exp\Big(-N \cdot D\Big(\frac{k}{N}\|q\Big)\Big)}_{\text{often } = \exp\left(-N\cdot\left[D\left(\frac{k}{N}\|q\right)+o(1)\right]\right)}, \quad k = 1, 2, ..., N-1 \tag{8}$$

Thresholds for Bernoulli Designs

Thresholds for Bernoulli design with i.i.d. $\mathrm{Bernoulli}(\frac{\nu}{k})$ entries:

$$n^*_{\mathrm{Bern}} = \max\Bigg\{ \frac{k \log \frac{p}{k}}{H_2(e^{-\nu} \star \rho) - H_2(\rho)}, \quad \text{(first branch)}$$

$$\frac{k \log \frac{p}{k}}{(1-\theta)\nu e^{-\nu} \min_{\substack{C>0 \\ \zeta \in (0,1)}} \max\{\frac{1}{\theta} f_1^{\mathrm{Bern}}(C,\zeta,\rho), f_2^{\mathrm{Bern}}(C,\zeta,\rho)\}} \quad \text{(second branch)} \Bigg\},$$

$$f_1^{\mathrm{Bern}}(C,\zeta,\rho) = C \log C - C + C \cdot D(\zeta \| \rho) + 1,$$

$$f_2^{\mathrm{Bern}}(C,\zeta,\rho) = \min_{d \geq \max\{0, C(1-2\zeta)/\rho\}} g^{\mathrm{Bern}}(C,\zeta,d,\rho),$$

$$g^{\mathrm{Bern}}(C,\zeta,d,\rho) = \rho d \log d + \big(\rho d - C(1-2\zeta)\big) \log\left(\frac{\rho d - C(1-2\zeta)}{1-\rho}\right) + 1 - 2\rho d + C(1-2\zeta)$$

recall some notation:

- p items, k defectives, $k \sim p^\theta$
- ν: design parameter
- ρ: noise level
- C, ζ: optimization parameters

Thresholds for Near-Constant Weight Designs

Thresholds for near-constant weight design with $\Delta = \frac{\nu n}{k}$ placements per item:

$$n^*_{\mathrm{NC}} = \max\Bigg\{ \frac{k \log \frac{p}{k}}{H_2(e^{-\nu} \star \rho) - H_2(\rho)}, \quad \text{(first branch)}$$

$$\frac{k \log \frac{p}{k}}{(1-\theta)\nu e^{-\nu} \min_{C \in (0, e^{\nu}), \zeta \in (0,1)} \max\{\frac{1}{\theta} f_1^{\mathrm{NC}}(C, \zeta, \rho, \nu), f_2^{\mathrm{NC}}(C, \zeta, \rho, \nu)\}} \quad \text{(second branch)} \Bigg\},$$

$$f_1^{\mathrm{NC}}(C, \zeta, \rho, \nu) = e^{\nu} D(Ce^{-\nu} \| e^{-\nu}) + C \cdot D(\zeta \| \rho),$$

$$f_2^{\mathrm{NC}}(C, \zeta, \rho, \nu) = \min_{d \,:\, |C(1-2\zeta)| \le d \le e^{\nu}} g^{\mathrm{NC}}(C, \zeta, d, \rho, \nu),$$

$$g^{\mathrm{NC}}(C, \zeta, d, \rho, \nu) = e^{\nu} \cdot D\big(de^{-\nu} \| e^{-\nu}\big) + d \cdot D\Big(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \| \rho\Big).$$

High-level Intuitions

Two branches appear in the final thresholds:

1. The common first branch $\frac{k \log(p/k)}{H_2(e^{-\nu} \star \rho) - H_2(\rho)}$ is related to the Shannon capacity of the binary symmetric channel.
   - Established by analyzing $\ell = |\widehat{S} \setminus S| = k$ (high ℓ, low overlap)
2. The more complicated second branches involve f_1 and f_2:
   - Established by analyzing $\ell = |\widehat{S} \setminus S| = 1$ (low ℓ, high overlap)
   - The optimization constants (C, ζ, d) that are introduced to characterize certain quantities in the error events.

III. Proofs for Converse

The First Branch $\frac{k \log(p/k)}{H_2(e^{-\nu} \star \rho) - H_2(\rho)}$

Intuition:

- The test has probability about $e^{-\nu}$ of containing no defectives;
- (Roughly) $e^{-\nu} \star \rho$ of being positive;
- Thus, each test can only reveal $H_2(e^{-\nu} \star \rho) - H_2(\rho)$ bits of information;
- With $\binom{p}{k}$ possible defective sets, we need (roughly) $\log \binom{p}{k} \sim k \log \frac{p}{k}$ bits; comparing them gives the capacity branch.

Sketch of technical argument:

- For any $\delta_1 > 0$, we have [SC16]

$$P_e \geq \mathbb{P}\left(\imath^n(\mathbf{X}_s, \mathbf{Y}) \leq \log \left(\delta_1 \binom{p}{k} \right) \right) - \delta_1 \tag{9}$$

$$\approx \mathbb{P}\left(\imath^n(\mathbf{X}_s, \mathbf{Y}) \leq k \log \binom{p}{k} \right) \tag{10}$$

  where $\imath^n(\mathbf{X}_s, \mathbf{Y}) = \log \frac{\mathbb{P}(\mathbf{Y}|\mathbf{X}_s)}{\mathbb{P}(\mathbf{Y})} = \log \mathbb{P}(\mathbf{Y}|\mathbf{X}_s) - \log \mathbb{P}(\mathbf{Y})$.

- Establish *upper concentration bound* for $\imath^n(\mathbf{X}_s, \mathbf{Y})$ by separately analyzing $\log \mathbb{P}(\mathbf{Y}|\mathbf{X}_s)$ and $\log \mathbb{P}(\mathbf{Y})$.

The Second Branch – Failure of MLE

Challenge:

- This kind of terms appeared in thresholds for noiseless case, based on such a central idea:
  if a defective item is *masked* (i.e., every test it is in also contains at least one other defective), then even an optimal decoder will be unable to identify it.
- However, this is no longer the dominant error event in the noisy case, thus cannot be used to derive the exact/tight converse bounds

Ideas:

- MLE is the optimal decoding strategy, and we only need to show MLE fails when n is below n^*;
- Given $(\mathbf{X}, \mathbf{Y})$, the likelihood of an estimate s is

$$\mathcal{L}_{\mathbf{X},\mathbf{Y}}(s) = \rho^{N_{\mathbf{X},\mathbf{Y}}(s)}(1-\rho)^{n-N_{\mathbf{X},\mathbf{Y}}(s)} \tag{11}$$

  where $N_{\mathbf{X},\mathbf{Y}}(s)$ denotes the number of "correct tests"
- Error event: for some defective j and nondefective j' it holds that have

$$N_{\mathbf{X},\mathbf{Y}}\big(\underbrace{(S \setminus \{j\}) \cup \{j'\}}_{:=\widehat{S}}\big) > N_{\mathbf{X},\mathbf{Y}}(S) \iff N_{\mathbf{X},\mathbf{Y}}(\widehat{S}) - N_{\mathbf{X},\mathbf{Y}}(S) > 0 \tag{12}$$

The Second Branch – Failure of MLE

$$\widehat{S} = (S \setminus \{j\}) \cup \{j'\}$$

Counting:

- Only two types of tests contribute to $N_{\mathbf{X},\mathbf{Y}}(\widehat{S})$ and $N_{\mathbf{X},\mathbf{Y}}(S)$ differently:

$$\text{contain } j \text{ as the only defective but not contain } j' \begin{cases} \text{positive } (I_1) : N_{\mathbf{X},\mathbf{Y}}(S) \leftarrow N_{\mathbf{X},\mathbf{Y}}(S) + 1 \\ \text{negative } (I_2) : N_{\mathbf{X},\mathbf{Y}}(\widehat{S}) \leftarrow N_{\mathbf{X},\mathbf{Y}}(\widehat{S}) + 1 \end{cases} \tag{13}$$

$$\text{contain no defective but contain } j' \begin{cases} \text{positive } (I_3) : N_{\mathbf{X},\mathbf{Y}}(\widehat{S}) \leftarrow N_{\mathbf{X},\mathbf{Y}}(\widehat{S}) + 1 \\ \text{negative } (I_4) : N_{\mathbf{X},\mathbf{Y}}(S) \leftarrow N_{\mathbf{X},\mathbf{Y}}(S) + 1 \end{cases} \tag{14}$$

- Failure condition:

$$N_{\mathbf{X},\mathbf{Y}}(\widehat{S}) - N_{\mathbf{X},\mathbf{Y}}(S) > 0 \Longrightarrow I_2 + I_3 > I_1 + I_4 \tag{15}$$

Analytical formulation:

- Notation:
  - $\mathcal{M}_j$: tests in which j is the only defective
  - $\mathcal{N}_0$: tests containing no defective
  - $\mathcal{M}_{j1}$ ($\mathcal{M}_{j0}$): the positive (negative) tests in $\mathcal{M}_j$
  - $\mathcal{N}_{01}$ ($\mathcal{N}_{00}$): the positive (negative) tests in $\mathcal{N}_0$
  - $G_{j,j',1}$: number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{j1}$ that contain j'
  - $G_{j,j',2}$: number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{j0}$ that contain j'
- For some $(C, \zeta) \in (0, \infty) \times (0, 1)$ such that $\frac{Cn\nu e^{-\nu}}{k}, \frac{\zeta \cdot Cn\nu e^{-\nu}}{k} \in \mathbb{Z}$ we have
  - **(C1)** There exists some $j \in S$ such that

$$M_j = |\mathcal{M}_j| = \frac{Cn\nu e^{-\nu}}{k} \tag{16}$$

$$M_{j0} = |\mathcal{M}_{j0}| = \zeta \cdot M_j \tag{17}$$

  - **(C2)** Failure condition: For some $j' \in [p] \setminus S$,

$$I_2 + I_3 > I_1 + I_4 \implies G_{j,j',1} - G_{j,j',2} > (1 - 2\zeta)\frac{Cn\nu e^{-\nu}}{k} \tag{18}$$

The Second Branch – Failure of MLE

The second branch takes the form $\frac{k\log(p/k)}{(1-\theta)\nu e^{-\nu}\min_{C,\zeta}\max\{\frac{f_1}{\theta},f_2\}}$

Step I. Ensuring (C1) leads to f_1:

- **(C1)** for some $j \in S$, $M_j = \frac{Cn\nu e^{-\nu}}{k}$ and $M_{j0} = \zeta \cdot \frac{Cn\nu e^{-\nu}}{k}$
- Challenge lies on M_j
- (Bernoulli) Poisson approximation for the multinomial distribution

$$\left(M_1, M_2, \cdots, M_{k\xi}\right) \tag{19}$$

- (Near-Constant)
  - Work with a surrogate of M_j—
    M_j': the number of tests in which j is the only defective and *is placed exactly once*
  - Interpret the placements of items into tests as edges in a bipartite graph [CGHL20], and use symmetry to show $(M_1', M_2', \cdots, M_{k\xi}')$ obeys

$$\begin{aligned} M_1' &\sim \mathrm{Hg}(k\Delta, e^{-\nu}k\Delta, \Delta) \\ M_2'|M_1' &\sim \mathrm{Hg}(k\Delta, e^{-\nu}k\Delta, \Delta) \\ &\cdots \\ M_{k\xi}'\,|\left(M_1', M_2', \cdots, M_{k\xi-1}'\right) &\sim \mathrm{Hg}(k\Delta, e^{-\nu}k\Delta, \Delta) \end{aligned} \tag{20}$$

The Second Branch – Failure of MLE

The second branch takes the form $\frac{k \log(p/k)}{(1-\theta)\nu e^{-\nu} \min_{C,\zeta} \max\{\frac{f_1}{\theta}, f_2\}}$

Step II. Ensuring (C2) leads to f_2:

- **(C2)**: $G_{j,j',1} - G_{j,j',2} > (1-2\zeta)\frac{Cn\nu e^{-\nu}}{k}$
- This comes down to the analysis of the difference of *two independent binomial random variables*, and can be handled by anti-concentration
- Many technical challenges/details omitted ...

Step III. Optimizing (C, ζ)

- **(C1) and (C2)**
- Optimizing (C, ζ) to establish the strongest converse bound

IV. Proofs for Achievability

Information density Decoder in [SC16]

Existing Information density decoder (Scarlett & Cevher, 16 SODA):

- We assume $S = s$ is the defective set
- We consider partitioning s into $(s_{\text{dif}}, s_{\text{eq}})$ with $s_{\text{dif}} \neq \emptyset$, and define for each s_{dif} the information density as

$$\imath^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) := \log \frac{\mathbb{P}(\mathbf{Y} | \mathbf{X}_{s_{\text{dif}}}, \mathbf{X}_{s_{\text{eq}}})}{\mathbb{P}(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}})}. \tag{21}$$

- Its expectation depends only on $\ell := |s_{\text{dif}}|$ and is defined as

$$\mathbb{E}\left[\imath^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}})\right] := I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) := I_\ell^n \tag{22}$$

- **Information density decoder**:
  - Fix the constants $\{\gamma_\ell\}_{\ell=1}^k$, and search for a set s of cardinality k such that

$$\imath^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) \geq \gamma_{|s_{\text{dif}}|}, \quad \forall (s_{\text{dif}}, s_{\text{eq}}) \text{ such that } |s_{\text{dif}}| \neq 0. \tag{23}$$

- **Intuition:** $\imath^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}})$ tends to be high for the actual defective set s;
- **Limitation:** Analyzing this decoder under small ℓ leads to sub-optimal threshold for noisy GT.

Our hybrid decoding rule

- We resort to MLE for low-ℓ case
- **Hybrid Decoder:** We search for a set $\hat{s}$ of cardinality k that satisfies
  - (Low ℓ: MLE) It holds that

$$\mathcal{L}_{\mathbf{X},\mathbf{Y}}(\hat{s}) > \mathcal{L}_{\mathbf{X},\mathbf{Y}}(s'), \quad \forall s' \text{ such that } 1 \leq |\hat{s} \setminus s'| \leq \frac{k}{\log k}, \tag{24}$$

    where we implicitly also constrain s' to have cardinality k.

  - (High-ℓ: information density) For suitably chosen $\{\gamma_\ell\}_{\frac{k}{\log k} < \ell \leq k}$, it holds that

$$\imath^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \geq \gamma_{|s_{\mathrm{dif}}|}, \quad \forall (s_{\mathrm{dif}}, s_{\mathrm{eq}}) \text{ such that } |s_{\mathrm{dif}}| > \frac{k}{\log k}, \tag{25}$$

    where $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ is a disjoint partition of $\hat{s}$.

- **Success conditions**:
  - Success condition for Low-ℓ:

$$(24) \text{ holds for } \hat{s} = s \tag{26}$$

  - Success condition for high-ℓ:

$$(25) \text{ holds for } \hat{s} = s \tag{27}$$

$$\forall \tilde{s} \text{ with } |\tilde{s}| = k,\ |s \setminus \tilde{s}| > \frac{k}{\log k}, \text{ it holds that } \imath^n(\mathbf{X}_{\tilde{s}\setminus s}; \mathbf{Y}|\mathbf{X}_{\tilde{s}\cap s}) < \gamma_{|s\setminus\hat{s}|} \tag{28}$$

The First Branch $\frac{k\log(p/k)}{H_2(e^{-\nu}\star\rho)-H_2(\rho)}$

- **Starting point:** [SC16] for any $\delta_1 > 0$, $\mathbb{P}\big((27)\&(28)\text{ fail}\big) \le$

$$\mathbb{P}\Bigg[\bigcup_{(s_{\text{dif}},s_{\text{eq}}):|s_{\text{dif}}|\ge\ell_{\min}}\left\{\imath^n(\mathbf{X}_{s_{\text{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\text{eq}}}) \le \log\binom{p-k}{|s_{\text{dif}}|} + \log\left(\frac{k}{\delta_1}\binom{k}{|s_{\text{dif}}|}\right)\right\}\Bigg] + \delta_1$$

$$\overset{\delta_1\to 0}{\approx} \mathbb{P}\Bigg[\bigcup_{(s_{\text{dif}},s_{\text{eq}}):|s_{\text{dif}}|\ge\ell_{\min}}\left\{\imath^n(\mathbf{X}_{s_{\text{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\text{eq}}}) \le (1+o(1))\ell\log\left(\frac{p}{k}\right)\right\}\Bigg] \quad (29)$$

- **Concentration bound:** for any $\delta_2 \in (0,1)$,

$$\mathbb{P}\big[\imath^n(\mathbf{X}_{s_{\text{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\text{eq}}}) \le (1-\delta_2)I_\ell^n\big] \le \psi_\ell(n,\delta_2), \quad (30)$$

- Therefore, if it holds that

$$\max_{\ell > \frac{k}{\log k}} \frac{(1+o(1))\ell\log(\frac{p}{k})}{I_\ell^n(1-\delta_2)} \le 1, \quad (31)$$

  we are able to enforce

$$\mathbb{P}\big((27)\&(28)\text{ fail}\big) \le \sum_{\ell=\ell_{\min}}^{k}\binom{k}{\ell}\psi_\ell(n,\delta_2) \to 0 \quad (32)$$

- Combining with the asymptotic of scaling of I_ℓ^n, (31) yields $n \ge$ the first branch

The Second Branch – Success of MLE

- The second branch, $\frac{k\log(p/k)}{(1-\theta)\nu e^{-\nu}\min_{C,\zeta}\max\{\frac{f_1}{\theta},f_2\}}$, is used to ensure

$$\mathcal{L}_{\mathbf{X},\mathbf{Y}}(s) > \mathcal{L}_{\mathbf{X},\mathbf{Y}}(s'), \quad \forall s' \text{ such that } 1 \le |s \setminus s'| \le \frac{k}{\log k} \tag{33}$$

  so that the MLE part succeeds.
- Similar arguments with nontrivial generalizations

Preparations:

- $\mathcal{L}_{\mathbf{X},\mathbf{Y}}(s) > \mathcal{L}_{\mathbf{X},\mathbf{Y}}(s') \iff N_{\mathbf{X},\mathbf{Y}}(s) > N_{\mathbf{X},\mathbf{Y}}(s')$
- $\mathcal{J} = s \setminus s'$ (defective) and $\mathcal{J}' = s' \setminus s$ (non-defective)
- Only two types of tests matter

$$\text{only contain defectives in } \mathcal{J} \text{ but not contain items in } \mathcal{J}' \begin{cases} \text{positive } (I_1) : N_{\mathbf{X},\mathbf{Y}}(s)++ \\ \text{negative } (I_2) : N_{\mathbf{X},\mathbf{Y}}(s')++ \end{cases} \tag{34}$$

$$\text{contain no defective but contain items from } \mathcal{J} \begin{cases} \text{positive } (I_3) : N_{\mathbf{X},\mathbf{Y}}(s')++ \\ \text{negative } (I_4) : N_{\mathbf{X},\mathbf{Y}}(s)++ \end{cases} \tag{35}$$

- Success condition: $N_{\mathbf{X},\mathbf{Y}}(s) > N_{\mathbf{X},\mathbf{Y}}(s') \Longrightarrow I_1 + I_4 > I_2 + I_3$

Analytical formulation:

- Notation:
  - $\mathcal{M}_{\mathcal{J}}$: tests in which items in $\mathcal{J}$ are the only defectives
  - $\mathcal{M}_{\mathcal{J}1}(\mathcal{M}_{\mathcal{J}0})$: the positive (negative) tests in $\mathcal{M}_{\mathcal{J}}$
  - $\mathcal{N}_0, \mathcal{N}_{00}, \mathcal{N}_{01}$: as before
  - $G_{\mathcal{J},\mathcal{J}',1}$: number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}1}$ that contain some item from $\mathcal{J}'$
  - $G_{\mathcal{J},\mathcal{J}',2}$: number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}0}$ that contain some item from $\mathcal{J}'$
- For any $\ell \leq \frac{k}{\log k}$ and any pairs of $(C, \zeta) \in [0, \infty) \times [0, 1]$ such that $\frac{Cn\nu e^{-\nu}\ell}{k}, \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \in \mathbb{Z}$, one of the following conditions hold:
  - **(C1)** $\mathcal{K}_{\ell,C,\zeta} = \varnothing$ where

$$\mathcal{K}_{\ell,C,\zeta} = \left\{ \mathcal{J} \subset s \,:\, |\mathcal{J}| = \ell, M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k},\ M_{\mathcal{J}0} = \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \right\} \tag{36}$$

  - **(C2)** If $\mathcal{K}_{\ell,C,\zeta} \neq \varnothing$, then for any $\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta}$ and $\mathcal{J}' \subset [p] \setminus s$ and $|\mathcal{J}'| = \ell$,

$$I_1 + I_4 > I_2 + I_3 \iff G_{\mathcal{J},\mathcal{J}',1} - G_{\mathcal{J},\mathcal{J}',2} < (1 - 2\zeta)\frac{Cn\nu e^{-\nu}\ell}{k} \tag{37}$$

The Second Branch – Success of MLE

The second branch takes the form $\frac{k \log(p/k)}{(1-\theta)\nu e^{-\nu} \min_{C,\zeta} \max\{\frac{f_1}{\theta}, f_2\}}$

Overview:

- **Step 1.** Ensuring **(C1)** yields the f_1 part of the second branch
  - Unlike in the converse, we utilize a first-order method which first bounds

$$\mathbb{E}|\mathcal{K}_{\ell,C,\zeta}| = \binom{k}{\ell}\mathbb{P}\left(\text{for fixed } \mathcal{J} \subset s \text{ with } |\mathcal{J}| = \ell,\ M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k},\ M_{\mathcal{J}0} = \zeta M_{\mathcal{J}}\right) \tag{38}$$

  and then utilize Markov's inequality to enforce $|\mathcal{K}_{\ell,C,\zeta}| = 0$

- **Step 2.** Ensuring **(C2)** yields the f_2 part of the second branch. More complicated than the converse as we need to handle all (ℓ, C, ζ)
  - Working with the sum/difference of two independent binomial variables
- **Step 3.** Optimizing over (C, ζ)

Summary & Future Directions

Summary:

- We established exact thresholds for noisy group testing with Bernoulli design and near-constant weight design
- For converse analysis, the main innovation is to identify a novel set of dominant error events
- For achievablity analysis, we introduce a hybrid decoder that combines the exsiting information density approach and MLE

Future Directions:

1. **Efficient and Optimal Algorithm:** Devise an *efficient* algorithm to achieve the exact thresholds for near-constant weight design.
   - A concurrent work solved this problem via spatial coupling designs.[6]
2. **Converse for Arbitrary Design:** Investigate whether the n^*_{NC} is the general converse for arbitrary design.
   - This is true in the noiseless case [CGHL20]

Thank You

[6] Noisy group testing via spatial coupling, Coja-Oghlan et al., Comb. Probab. Comput., 2024