Distributed goodness-of-fit: when you can't talk much and have little in common

Jayadev Acharya (Cornell), Clément Canonne (Stanford), Yanjun Han (Stanford), Ziteng Sun (Cornell), and Himanshu Tyagi (IISc Bangalore)
Insert motivating story here.

(something about bees?)
Perform statistical inference in this distributed setting
Perform statistical inference in this distributed setting.
Perform statistical inference in this **distributed** setting.
Identity testing

Known distribution $q$ on $[k] = \{1, 2, \ldots, k\}$

Distance parameter $\varepsilon$

$H_0: \ p = q$

vs.

$H_1: \ TV(p, q) > \varepsilon$
What's known

- Centralized: $\Theta\left( \frac{\sqrt{k}}{\varepsilon^2} \right)$ [Pan’08, VV’17]

- Our setting:
  - Public-coin: $\Theta\left( \frac{k}{2^{\varepsilon} \varepsilon^2} \right)$ [ACT’19,12]
  - Private-coin: $\Theta\left( \frac{k^{3/2}}{2^{\varepsilon} \varepsilon^2} \right)$ [ACT’19,12]

$L$: # bits each player can send
Oh yes, about that...

Public-coin: \( \forall \) and all \( \forall \)'s share common random seed (hardcoded or broadcast to all by \( \forall \))

Private-coin: every \( \forall \) for itself (indep randomness)
Moreover, the optimal public-coin protocol only requires $O(\log k)$ shared random bits.
However...

What happens when we have only few shared random bits? Say, $\sqrt{\log k}$, or 15?
Theorem. For any $l \geq 1$, $s > 0$ s.t. $l + s \leq \log k$, there is a protocol for identity testing w/ 
\[ \frac{\sqrt{k}}{\varepsilon^2} \sqrt{\frac{k}{2^l}} \sqrt{\frac{k}{2^{s+l}}} \]
phones. Moreover, this is tight.
Proof. Ideas.
Generalize the minmax and minmax decoupled $\chi^2$-fluctuations notions from [ACT'19] to limited randomness: Deminmaxmin.

$$\overline{\chi}(\mathcal{W}, \varepsilon, s) = \sup_{\mathcal{W} \subseteq \mathcal{W}^n} \inf_{\mathbb{P} \in \mathbb{P}_\varepsilon} \mathbb{E}_n \left[ \chi_{(2)}^{(2)}(\mathcal{W}^n|\mathbb{P}) \right]$$
"Derandomization" of key anticoncentration lemma in [ACT'19]

Theorem. \( \forall n, \forall x \in \mathbb{R}^n, \exists m \leq n \) subsets \( S_1, \ldots, S_m \subseteq [n] \) s.t.

\[
\mathbb{P}_{i \sim [m]} \left\{ \left| \sum_{j \in S_i} x_j \right|^2 \geq \|x\|_2^2 \right\} \geq 1
\]
Apply this with \( n := 2^s \) to reduce the problem to private-coin identity testing over a domain of size \( L \leq \frac{k}{2^s} \), with distance \( \varepsilon' \geq \frac{\varepsilon}{\sqrt{2^s}} \).

Using the protocol from [ACT'19]:

\[
\frac{L^{3/2}}{2^{1/2} \varepsilon'^2} \times \frac{k^{3/2}}{2^{1/2} \sqrt{2^s} \varepsilon^2} = \frac{\sqrt{k}}{\varepsilon^2} \times \frac{k}{\sqrt{2}} \times \frac{k}{\sqrt{2^{s+1}}}.
\]

This suffices, as claimed.*
*OK, with a catch.*

This gives a test with low soundness (big Type-II error).
And we cannot amplify by repetition: our 4 bits are gone!
*Ok, with a catch.

This gives a tester cu/ low soundness (big Type-II error). And we cannot amplify by repetition: our s bits are gone!

Solution: boost using the same randomness (but 100x more phones)!

Details omitted. See board.
Open questions

What uses of this "derandomization" lemma?

Of this probability amplification technique?

Extend to other local constraints?

Is my handwriting that terrible?