Bounded indistinguishability of simple sources

Andrej Bogdanov
K. Dinesh

Yuval Filmus
Avi Kaplan
Yuval Ishai

CUHK
Technion
TIFR
K-indistinguishability

Two sources $X,Y$ of $n$ random bits are $K$-indistinguishable if

$$X_S \approx Y_S$$

for any set $S \subseteq [n]$ of size $K$.

Example: $X = a_1, b_1, a_1+b_1, \ldots, a_m, b_m, a_m+b_m$

$Y = X$ conditioned on $a_1 + \ldots + a_m = 0$
Fooling $AC^0$

[B1VW] constructed a pair $X, Y$ of $\sqrt{n}$-indistinguishable sources that can be distinguished by OR.

Braverman proved that if $X, Y$ are polylog($n$)-indistinguishable and $Y$ is the uniform distribution then $X, Y$ fool $AC^0$.

Can we close this gap?
Simple sources

We consider sources samplable from an infinite supply of iid uniformly random bits $r_i$.

- Low degree sources:
  $Y_i$ is low-degree polynomial in $r$

- Local sources:
  $Y_i$ depends on few bits of $r$

Crypto motivations.
Results at a glance

If $X, Y$ are $\text{polylog}(n)$-indistinguishable and...

... $Y$ is uniform then $X, Y$ fool $AC^0$ (Braverman)
... $Y$ is linear then $X, Y$ fool decision trees & narrow DNFs
... $Y$ is quadratic then $X, Y$ fool decision trees
... $Y$ has constant degree then $X, Y$ fool $\text{OR}$
... $Y$ has constant locality then $X, Y$ fool $\text{OR}$

polylog($n$)-indistinguishable linear sources fool $AC^0$

$\Rightarrow$ Inner-Product & $AC^0 \circ \text{XOR}$

$n^{\text{polylog}(n)}$-indistinguishable sources
of degree $O(\log n)$ distinguished by $\text{OR}$
$\tilde{\Omega}(n)$-indistinguishable log degree sources not fooling OR

- Since $\deg(\text{OR}) = \Omega(n^2)$, by LP duality
  OR distinguishes some pair $X, Y$ of $\tilde{\Omega}(n)$-indis. sources
- "Resampling": wlog, $X,Y$ are mixtures of iid
- Can sample $X,Y$ using poly size decision trees
- Use Razborov–Smolensky randomized encoding to consistently approximate $X,Y$ using polynomials of degree $O(\log n)$.

\[ \sum_{\text{leaves}} l_1 \cdots l_m \Rightarrow \sum_{\text{leaves}} \prod \left(1 + \sum_{j} (1 + \lambda_j) r_{k,j} \right) \]
Predictability

A subset $S \subseteq [n]$ $\varepsilon$-predicts $Y$ if

$$\Pr[Y|S=0 \text{ but } Y\neq 0] \leq \varepsilon.$$ 

- If $S$ $\varepsilon$-predicts $Y$ and $X,Y$ are $(|S|+1)$-indist. then $S$ $(n\varepsilon)$-predicts $X$.
- If $S$ $\delta$-predicts $X,Y$ then $X,Y$ $\delta$-fool
  and $(\delta\delta)$-fool decision trees of sizes.

$\Rightarrow$ Goal: if $Y$ is simple then $Y$ is
$\varepsilon$-predicted by set of size $\text{polylog}(\frac{1}{\varepsilon})$.
Predicting linear sources

A subset \( S \subseteq [n] \) \( \epsilon \)-predicts \( Y \) if

\[
\Pr[Y|S = 0 \text{ but } Y \neq 0] \leq \epsilon.
\]

\( Y \) is linear if each \( Y_i \) is linear function of \( X \)

Case 1: there exist \( \log_2(\frac{1}{\epsilon}) \) linearly independent coordinate \( S \) \( \Rightarrow \) \( \Pr[Y|S = 0] = \epsilon \)

Case 2: otherwise, choose a basis \( S \)

\( \Rightarrow \) \( \Pr[Y|S = 0 \text{ but } Y \neq 0] = 0 \)

Generalization to higher degree uses higher-order Fourier analysis
Predicting local sources

$S$ $\varepsilon$-predicts $Y$ if $\Pr [Y|S = 0 \text{ but } Y \neq 0] \leq \varepsilon$

$Y$ is $t$-local: every $Y_i$ depends on $t$ many $S_j$'s

Choose maximal set $T$ of indices depending on disjoint coordinates

Case 1: $|T| \geq 2^t \log(\frac{1}{\varepsilon})$

Choose set $T$ of that size $\Rightarrow \Pr [Y|S = 0] \leq \varepsilon$

Case 2: $|T| \leq 2^t \log(\frac{1}{\varepsilon})$

For each assignment to coords appearing in $Y|T$, source simplifies to $(t-1)$-local source; induction
Prediction for narrow DNFs

A decision tree $\epsilon$-predicts $Y$ for $f$ if for $1-\epsilon$ fraction of leaves (wrt $Y$), value of $f$ is determined.

If a depth $d$ DT $\epsilon$-predicts $Y$ for $f$ and $x, y$ are $d$-indist. then $x, y$ $\epsilon$-fool $f$.

Any linear source is $\epsilon$-predicted for any width $w$ DNF by a DT of depth $O(w^{2\epsilon}/\log 2\epsilon)$.

Proof: combination of arguments for linear and local sources.
Connection with linear IPPP

polylog(n)-indist. Linear sources fool AC^0

\[ \downarrow \]

if an AC^0 circuit can predict \( l(F^2) \) from \( l_1(F^2), \ldots, l_m(F^2) \)

then \( l \) spanned by polylog(n) many \( l_i \)

\[ \downarrow \]

if \( m = \text{poly}(n) \) then for any \( l_1, \ldots, l_m \) there exists \( l \) s.t.

no AC^0 circuit can predict \( l(F^2) \) given \( l_1(F^2), \ldots, l_m(F^2) \)

\[ \downarrow \]

no AC^0 circuit can predict \( \langle r, s \rangle \) given \( l_1(F^2), \ldots, l_m(F^2), y_1(s), \ldots, y_m(s) \)

\[ \downarrow \]

no AC^0\text{-XOR} circuit can predict \( \langle r, s \rangle \)
Open Questions

A class of sources $\mathcal{Y}$ is simple for a class of functions $\mathcal{F}$ if

$$X, Y \text{ polylog}(n)\text{-indist, } Y \in \mathcal{Y} \Rightarrow X, Y \text{ fool } \mathcal{F}$$

1. Maximal $d$ s.t. degree $d$ sources are simple for OR?
   Know: $d = \omega(1)$, $d = O(\log n)$

2. Maximal $t$ s.t. $t$-local sources are simple for OR?
   Know: $t = \omega(1)$, $t = \tilde{O}(n)$

3. Same for decision trees, DNFs, $AC^0$...