Bounded indistinguishability of simple sources

Andrej Bogdanov
K. Dinesh

Yuval Filmus
Avi Kaplan
Yuval Ishai

Akshay Srinivasan

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 $a_1 + \cdots + a_m = 0$

Fooling AC^0

[BIVW] 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 $\vec{r}$
- Local sources:
  Y_i depends on few bits of $\vec{r}$

Crypto motivations.

Results at a glance

If X, Y are 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 OR

... Y has constant locality then X, Y fool OR

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

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

$\exists$ $\sqrt{n}$-indistinguishable sources of degree $O(\log n)$ distinguished by OR

$\sqrt{n}$-indistinguishable log degree sources not fooling OR

- Since $\widetilde{\deg}(OR) = \Omega(\sqrt{n})$, by LP duality OR distinguishes some pair X, Y of $\sqrt{n}$-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 $O(\log n)$.

$$\sum_{\text{leaves}} \ell_1 \wedge \cdots \wedge \ell_w \Rightarrow \sum_{\text{leaves}} \prod_{k} \left(1 + \sum_{j} (1+\ell_j) r_{kj}\right)$$

Predictability

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

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

— If S ε-predicts Y and X, Y are $(|S|+1)$-indist. then S $(n\varepsilon)$-predicts X

— If S δ-predicts X, Y then X, Y δ-fool OR and $(s\delta)$-fool decision trees of size s

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

Predicting linear sources

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

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

Y is **linear** if each Y_i is linear function of $\vec{r}$

Case 1: there exist $\log_2(\frac{1}{\varepsilon})$ linearly independent coordinates $\Rightarrow \Pr[Y|_S = 0] = \varepsilon$

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 ε-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 r_j's

Choose maximal set T of indices depending on disjoint coordinates

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

Choose $S \subseteq T$ 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 ε-predicts Y for f if for $1-\varepsilon$ fraction of leaves (wrt Y), value of f is determined.

If a depth d DT ε-predicts Y for f and X, Y are d-indist. then X, Y ε-fool f.

Any linear source is ε-predicted for any width w DNF by a DT of depth $O(w 2^w \log \frac{1}{\varepsilon})$.

Proof: combination of arguments for linear and local sources.

Connection with linear IPPP

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

$\Downarrow$

$= r_1, \dots, r_n$

if an AC^0 circuit can predict $\ell(\vec{r})$ from $\ell_1(\vec{r}), \dots, \ell_m(\vec{r})$
then ℓ spanned by polylog(n) many ℓ_i

$\Downarrow$

if $m = \mathrm{poly}(n)$ then for any $\ell_1, \dots, \ell_m$ there exists ℓ s.t.
no AC^0 circuit can predict $\ell(\vec{r})$ given $\ell_1(\vec{r}), \dots, \ell_m(\vec{r})$

$\Downarrow$

no AC^0 circuit can predict $\langle r, s \rangle$ given $\ell_1(\vec{r}), \dots, \ell_m(\vec{r}), \varphi_1(\vec{s}), \dots, \varphi_m(\vec{s})$

$\Downarrow$

no $AC^0 \circ XOR$ 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 polylog(n)-indist, $Y \in \mathcal{Y} \Rightarrow X, Y$ 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 ...