Improved Pseudorandom Generators for Depth-2 Circuits

Anindya De[1] Omid Etesami[1] Luca Trevisan[2] Madhur Tulsiani[3]

[1]UC Berkeley

[2]UC Berkeley and Stanford University

[3]IAS and Princeton University

Fooling DNF formulas

- DNF: Disjunction of m terms $A_1, \ldots, A_m$. Each term is a conjunction of literals.
  e.g. $(x_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_2 \wedge \bar{x}_3 \wedge x_4 \wedge \bar{x}_5) \vee (x_1 \wedge x_5)$

Fooling DNF formulas

- DNF: Disjunction of m terms $A_1, \ldots, A_m$. Each term is a conjunction of literals.
  e.g. $(x_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_2 \wedge \bar{x}_3 \wedge x_4 \wedge \bar{x}_5) \vee (x_1 \wedge x_5)$

  width-w: Each term has at most w literals.

  read-once: Each variable appears at most once.

Fooling DNF formulas

- DNF: Disjunction of m terms $A_1, \ldots, A_m$. Each term is a conjunction of literals.
  e.g. $(x_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_2 \wedge \bar{x}_3 \wedge x_4 \wedge \bar{x}_5) \vee (x_1 \wedge x_5)$

  width-w: Each term has at most w literals.

  read-once: Each variable appears at most once.

- A depth-2 circuit is a DNF or its complement.

Fooling DNF formulas

- DNF: Disjunction of m terms $A_1, \ldots, A_m$. Each term is a conjunction of literals.
  e.g. $(x_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_2 \wedge \bar{x}_3 \wedge x_4 \wedge \bar{x}_5) \vee (x_1 \wedge x_5)$

  width-w: Each term has at most w literals.

  read-once: Each variable appears at most once.

- A depth-2 circuit is a DNF or its complement.

- Distribution D δ-fools $f : \{0,1\}^n \to \mathbb{R}$ if

$$|\mathbb{E}_{x \sim D}[f(x)] - \mathbb{E}_{x \sim U_n}[f(x)]| \leq \delta .$$

  Gives PRG with seed length s if D can be sampled using s bits.

Results

Goal: δ-fool a DNF with m terms and n variables.

	DNF Family	Seed length
[Nisan 91]	all	$O(\log^{10}(mn/\delta))$
[LVW93]	all	$O(\log^{4}(mn/\delta))$
[Bazzi 07]	all	$O(\log n \cdot \log^{2}(m/\delta))$
This work	all	$O(\log n + \log^{2}(m/\delta) \cdot \log\log(m/\delta))$

Results

Goal: δ-fool a DNF with m terms and n variables.

	DNF Family	Seed length
[Nisan 91]	all	$O(\log^{10}(mn/\delta))$
[LVW93]	all	$O(\log^{4}(mn/\delta))$
[Bazzi 07]	all	$O(\log n \cdot \log^{2}(m/\delta))$
This work	all	$O(\log n + \log^{2}(m/\delta) \cdot \log\log(m/\delta))$
[LV91]	width-w	$O(\log n + w2^{w} \cdot \log(1/\delta))$
This work	width-w	$O(\log n + w \log w \cdot \log(m/\delta))$

Results

Goal: δ-fool a DNF with m terms and n variables.

	DNF Family	Seed length
[Nisan 91]	all	$O(\log^{10}(mn/\delta))$
[LVW93]	all	$O(\log^{4}(mn/\delta))$
[Bazzi 07]	all	$O(\log n \cdot \log^{2}(m/\delta))$
This work	all	$O(\log n + \log^{2}(m/\delta) \cdot \log\log(m/\delta))$
[LV91]	width-w	$O(\log n + w2^{w} \cdot \log(1/\delta))$
This work	width-w	$O(\log n + w \log w \cdot \log(m/\delta))$
[Bazzi 03]	read-once	$O(\log n \cdot \log m \cdot \log(1/\delta))$
This work	read-once	$O(\log n + \log m \cdot \log(1/\delta))$

Another look at the results

Results for general DNFs, for the special case $m = poly(n)$ and $\delta = 1/poly(n)$.

	Seed length
[Nisan 91]	$O(\log^{10}(n))$
[LVW93]	$O(\log^{4}(n))$
[Bazzi 07]	$O(\log^{3} n)$
This work	$O(\log^{2} n \cdot \log\log(n))$

Another look at the results

Results for general DNFs, for the special case $m = poly(n)$ and $\delta = 1/poly(n)$.

	Seed length
[Nisan 91]	$O(\log^{10}(n))$
[LVW93]	$O(\log^{4}(n))$
[Bazzi 07]	$O(\log^{3} n)$
This work	$O(\log^{2} n \cdot \log\log(n))$

- The $O(\log^3 n)$ seed length in [Bazzi 07] is achieved by using an $O(\log^2 n)$-wise independent distribution.
- The improvement to $O(\log^2 n \cdot \log\log n)$ is obtained by using ϵ-biased distributions with $\epsilon = (\log n)^{-O(\log^2 n)}$.

A little Fourier analysis

- Can decompose every $f : \{0,1\}^n \to \mathbb{R}$ as a combination of *characters*.

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x), \qquad \chi_S(x) \stackrel{\text{def}}{=} (-1)^{\sum_{i \in S} x_i}$$

- $\mathbb{E}_x\left[\chi_S(x)\right] = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{otherwise} \end{cases}$ $\qquad \mathbb{E} f = \widehat{f}(\emptyset)$

A little Fourier analysis

- Can decompose every $f : \{0,1\}^n \to \mathbb{R}$ as a combination of *characters*.

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x), \qquad \chi_S(x) \stackrel{\text{def}}{=} (-1)^{\sum_{i \in S} x_i}$$

- $\mathbb{E}_x[\chi_S(x)] = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{otherwise} \end{cases}$ $\qquad \mathbb{E} f = \widehat{f}(\emptyset)$

- D is ϵ-biased if it ϵ-fools all characters i.e.

$$\forall S \neq \emptyset \quad |\mathbb{E}_{x \sim D}[\chi_S(x)]| \leq \epsilon .$$

- [NN 93]: Can sample an ϵ-biased distribution using a seed of size $2\log n + 2\log(1/\epsilon)$.

Functions fooled by ϵ-biased distributions

- $\left\|\widehat{f}\right\|_1 \stackrel{\text{def}}{=} \sum_{S \neq \emptyset} \left|\widehat{f}(S)\right|$.

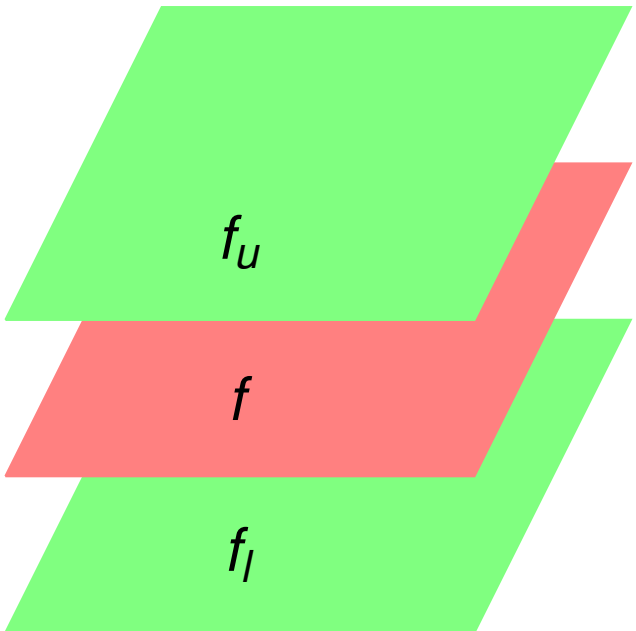
Functions fooled by ϵ-biased distributions

- $\left\|\widehat{f}\right\|_1 \stackrel{\text{def}}{=} \sum_{S \neq \emptyset} \left|\widehat{f}(S)\right|$.
- ϵ-biased distributions fool functions with small ℓ_1 norm.

$$\begin{aligned} \left|\mathbb{E}_{x \sim U_n}[f(x)] - \mathbb{E}_{x \sim D}[f(x)]\right| &= \left|\sum_{S \neq \emptyset} \widehat{f}(S) \cdot \mathbb{E}_{x \sim D}\left[\chi_S(x)\right]\right| \\ &\leq \epsilon \cdot \left\|\widehat{f}\right\|_1 . \end{aligned}$$

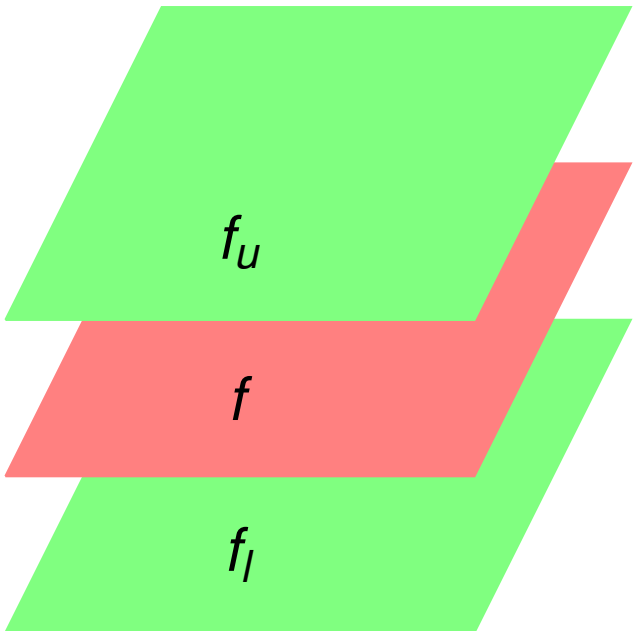
Fooling other functions by sandwiching [Bazzi]

Let $f_l \leq f \leq f_u$ and f_l, f_u be δ-fooled by D. If $\mathbb{E}_{x \sim U_n}[f_u - f_l] \leq \delta'$, then D also $(\delta + \delta')$-fools f.



Fooling other functions by sandwiching [Bazzi]

Let $f_l \leq f \leq f_u$ and f_l, f_u be δ-fooled by D. If $\mathbb{E}_{x \sim U_n}[f_u - f_l] \leq \delta'$, then D also $(\delta + \delta')$-fools f.



$$\begin{aligned}\mathbb{E}_{x \sim D}[f(x)] &\leq \mathbb{E}_{x \sim D}[f_u(x)] \\ &\leq \mathbb{E}_{x \sim U_n}[f_u(x)] + \delta \\ &\leq \mathbb{E}_{x \sim U_n}[f(x)] + \delta' + \delta\end{aligned}$$

Similarly for f_l.

Sandwiching functions for DNFs

[Bazzi]: Enough to find one function g s.t. $f = 0 \implies g = 0$.

$f = \bigvee_{i=1}^{m} A_i$

Sandwiching functions for DNFs

[Bazzi]: Enough to find one function g s.t. $f = 0 \implies g = 0$.

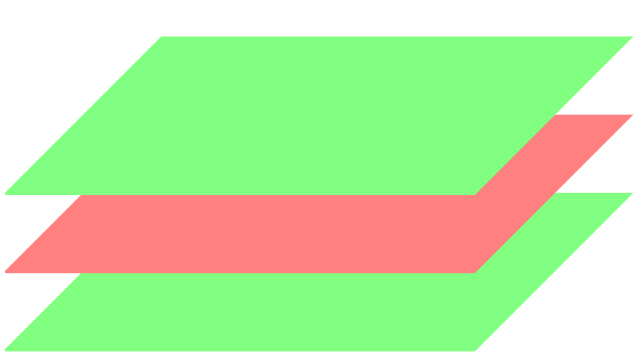
$f_u = 1 - (1 - \sum_{i=1}^{m} A_i)(1 - g)^2$

$f = \bigvee_{i=1}^{m} A_i$

$f_l = 1 - (1 - g)^2$

Sandwiching functions for DNFs

[Bazzi]: Enough to find one function g s.t. $f = 0 \implies g = 0$.



$f_u = 1 - (1 - \sum_{i=1}^{m} A_i)(1 - g)^2$

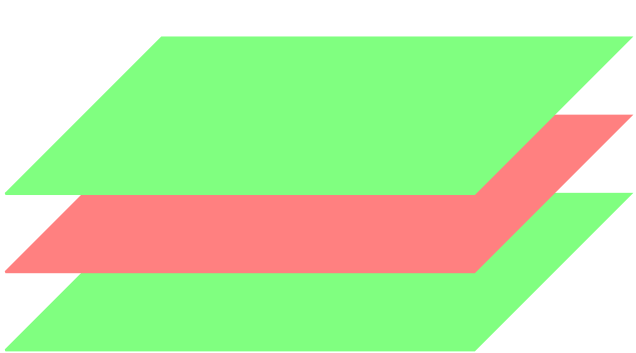
$f = \bigvee_{i=1}^{m} A_i$

$f_l = 1 - (1 - g)^2$

- $f = 0 \implies g = 0 \implies f_l = 0$. $f_l \leq 1$.
- $f = 1 \implies f_u \geq 1$. $f = 0 \implies f_u \geq 0$.

Sandwiching functions for DNFs

[Bazzi]: Enough to find one function g s.t. $f = 0 \implies g = 0$.



$f_u = 1 - (1 - \sum_{i=1}^{m} A_i)(1 - g)^2$

$f = \bigvee_{i=1}^{m} A_i$

$f_l = 1 - (1 - g)^2$

- $f = 0 \implies g = 0 \implies f_l = 0$. $f_l \leq 1$.
- $f = 1 \implies f_u \geq 1$. $f = 0 \implies f_u \geq 0$.
- $\left\|\widehat{f_l}\right\|_1, \left\|\widehat{f_u}\right\|_1 = O(m \cdot \left\|\widehat{g}\right\|_1^2)$.
- $\mathbb{E}[f_u - f_l] = O(m \cdot \|f - g\|_2^2)$.

Simplifying further

- [Razborov 09]: Suffices to focus on ℓ_1 and ℓ_2 norms.

Simplifying further

- [Razborov 09]: Suffices to focus on ℓ_1 and ℓ_2 norms.
- $f = \bigvee_{i=1}^{m} A_i = \sum_{i=1}^{m} A_i \cdot \left(1 - \bigvee_{j=1}^{i-1} A_j\right)$. Let $f_i = \bigvee_{j=1}^{i-1} A_j$.

  Approximate each f_i by g_i such that $\left\|\widehat{g_i}\right\|_1 \leq B$ and $\|f_i - g_i\| \leq \eta$.

  Take $g = \sum_{i=1}^{m} A_i \cdot (1 - g_i)$.

Simplifying further

- [Razborov 09]: Suffices to focus on ℓ_1 and ℓ_2 norms.
- $f = \bigvee_{i=1}^{m} A_i = \sum_{i=1}^{m} A_i \cdot \left(1 - \bigvee_{j=1}^{i-1} A_j\right)$. Let $f_i = \bigvee_{j=1}^{i-1} A_j$.

  Approximate each f_i by g_i such that $\left\|\widehat{g_i}\right\|_1 \leq B$ and $\|f_i - g_i\| \leq \eta$.

  Take $g = \sum_{i=1}^{m} A_i \cdot (1 - g_i)$.
- $f = 0 \implies g = 0$ (as $A_i = 0$ for all i). Also, by triangle inequality:
  - $\left\|\widehat{g}\right\|_1 \leq m \cdot B$
  - $\|f - g\|_2 \leq m \cdot \eta$

Finding good approximating functions

- [Mansour 95]: For any $\eta > 0$ and any width-w DNF formula f, there is a family of subsets $\mathcal{F} \subseteq 2^{[n]}$ satisfying:
  - $|\mathcal{F}| \leq w^{O(w \cdot \log(1/\eta))}$
  - For $g = \sum_{S \in \mathcal{F}} \widehat{f}(S) \cdot \chi_S$, $\|f - g\|_2 \leq \epsilon$.

  (Also, $\left\|\widehat{g}\right\|_1 \leq |\mathcal{F}| = w^{O(w \cdot \log(1/\eta))}$)

Finding good approximating functions

- [Mansour 95]: For any $\eta > 0$ and any width-w DNF formula f, there is a family of subsets $\mathcal{F} \subseteq 2^{[n]}$ satisfying:
  - $|\mathcal{F}| \leq w^{O(w \cdot \log(1/\eta))}$
  - For $g = \sum_{S \in \mathcal{F}} \widehat{f}(S) \cdot \chi_S$, $\|f - g\|_2 \leq \epsilon$.

  (Also, $\left\|\widehat{g}\right\|_1 \leq |\mathcal{F}| = w^{O(w \cdot \log(1/\eta))}$)

- Any term of width more than $\log(m/\delta)$ is satisfied with probability at most δ/m. Can assume $w = O(\log(m/\delta))$.

Finding good approximating functions

- [Mansour 95]: For any $\eta > 0$ and any width-w DNF formula f, there is a family of subsets $\mathcal{F} \subseteq 2^{[n]}$ satisfying:
  - $|\mathcal{F}| \leq w^{O(w \cdot \log(1/\eta))}$
  - For $g = \sum_{S \in \mathcal{F}} \widehat{f}(S) \cdot \chi_S$, $\|f - g\|_2 \leq \epsilon$.

  (Also, $\left\|\widehat{g}\right\|_1 \leq |\mathcal{F}| = w^{O(w \cdot \log(1/\eta))}$)

- Any term of width more than $\log(m/\delta)$ is satisfied with probability at most δ/m. Can assume $w = O(\log(m/\delta))$.

- An ϵ-biased set with $\epsilon = (\log(m/\delta))^{-O(\log^2(m/\delta))}$ suffices, giving PRG with seed length $O(\log n + \log^2(m/\delta) \cdot \log\log(m/\delta))$.

A lower bound

- Exhibit a read-once DNF formula f and for every integer d, a distribution D such that
  - D has bias at most $\epsilon = (2/m)^d$.
  - f distinguishes D from uniform with probability at least $\delta = \exp(-O(d \log d))$.

A lower bound

- Exhibit a read-once DNF formula f and for every integer d, a distribution D such that
  - D has bias at most $\epsilon = (2/m)^d$.
  - f distinguishes D from uniform with probability at least $\delta = \exp(-O(d \log d))$.

- Shows that one needs $\epsilon = \exp\left(-\Omega\left(\log m \cdot \frac{\log(1/\delta)}{\log\log(1/\delta)}\right)\right)$, and hence, seed length $\Omega\left(\log n + \log m \cdot \frac{\log(1/\delta)}{\log\log(1/\delta)}\right)$.

A lower bound

- Exhibit a read-once DNF formula f and for every integer d, a distribution D such that
  - D has bias at most $\epsilon = (2/m)^d$.
  - f distinguishes D from uniform with probability at least $\delta = \exp(-O(d \log d))$.

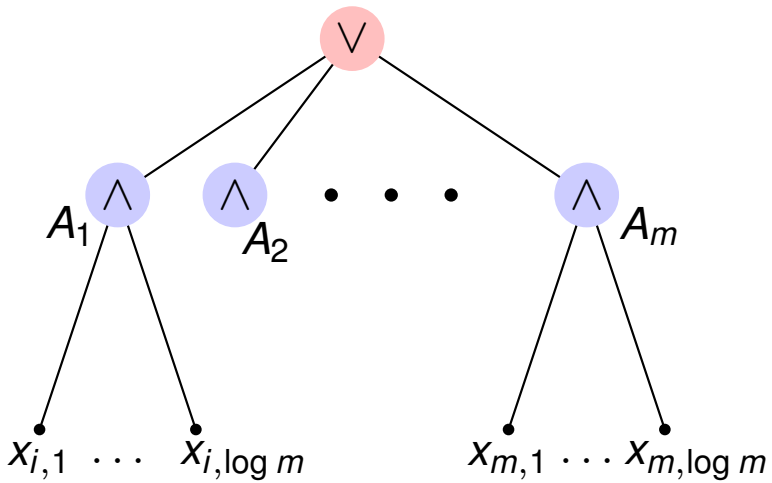
- Shows that one needs $\epsilon = \exp\left(-\Omega\left(\log m \cdot \frac{\log(1/\delta)}{\log\log(1/\delta)}\right)\right)$, and hence, seed length $\Omega\left(\log n + \log m \cdot \frac{\log(1/\delta)}{\log\log(1/\delta)}\right)$.

- Almost tight for read-once DNFs. Also provide a slightly tighter example (without the $\log\log(1/\delta)$) for general DNFs.

Constructing the lower bound

The DNF

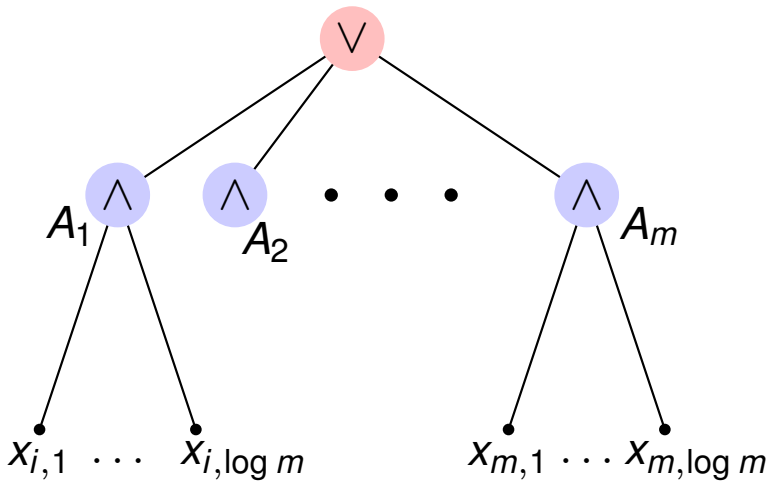
$$f \stackrel{\text{def}}{=} \bigvee_{i=1}^{m} \left(\bigwedge_{j=1}^{\log m} x_{i,j} \right) \quad \text{(Tribes)}$$



Constructing the lower bound

The DNF

$$f \stackrel{\text{def}}{=} \bigvee_{i=1}^{m} \left(\bigwedge_{j=1}^{\log m} x_{i,j} \right) \quad \text{(Tribes)}$$



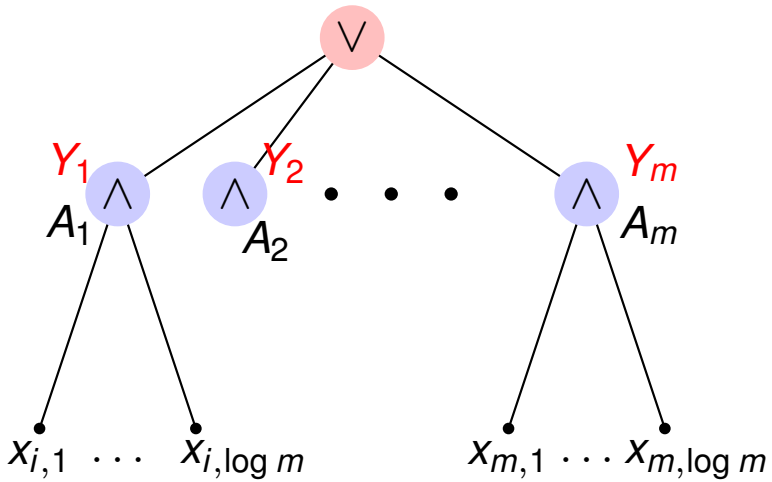
The distribution

- Choose $Y_1, \ldots, Y_m \in \{0,1\}$ satisfying
  - d-wise independent
  - $\sum_i Y_i \leq d$
  - $\mathbb{P}[Y_i = 1] = 1/m$

Constructing the lower bound

The DNF

$$f \stackrel{\text{def}}{=} \bigvee_{i=1}^{m} \left(\bigwedge_{j=1}^{\log m} x_{i,j} \right) \qquad \text{(Tribes)}$$



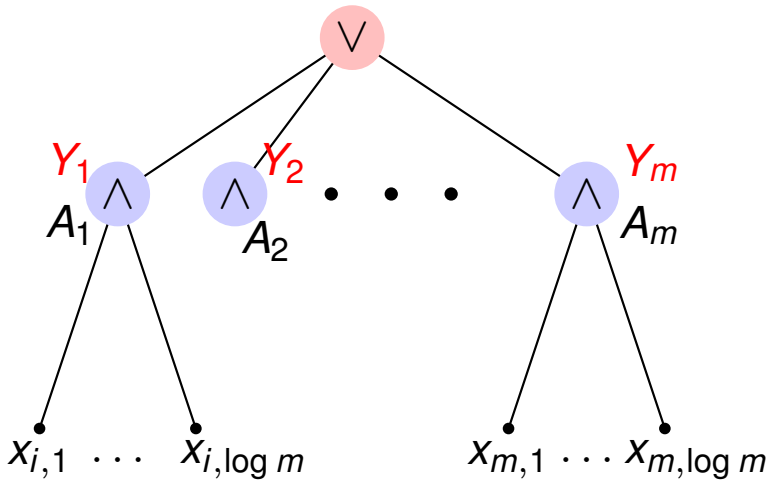
The distribution

- Choose $Y_1, \ldots, Y_m \in \{0,1\}$ satisfying
  - d-wise independent
  - $\sum_i Y_i \leq d$
  - $\mathbb{P}[Y_i = 1] = 1/m$
- Choose random $x_{i,1}, \ldots, x_{i,\log m}$ satisfying $\bigwedge_{j=1}^{\log m} x_{i,j} = Y_i$ i.e.
  - all 1 if $Y_i = 1$
  - random in $\{0,1\}^{\log m} \setminus 1^{\log m}$ if $Y_i = 0$

Constructing the lower bound

The DNF

$$f \stackrel{\text{def}}{=} \bigvee_{i=1}^{m} \left(\bigwedge_{j=1}^{\log m} x_{i,j} \right) \qquad \text{(Tribes)}$$



The distribution

- Choose $Y_1, \ldots, Y_m \in \{0,1\}$ satisfying
  - d-wise independent
  - $\sum_i Y_i \leq d$
  - $\mathbb{P}[Y_i = 1] = 1/m$
- Choose random $x_{i,1}, \ldots, x_{i,\log m}$ satisfying $\bigwedge_{j=1}^{\log m} x_{i,j} = Y_i$ i.e.
  - all 1 if $Y_i = 1$
  - random in $\{0,1\}^{\log m} \setminus 1^{\log m}$ if $Y_i = 0$
- For each i, the variables $x_{i,1}, \ldots, x_{i,\log m}$ in the i^{th} term are uniform.
- Given $Y_1, \ldots, Y_m$, the x bits are independent.

Generating $Y_1, \ldots, Y_m$

- Let P be a random degree-d polynomial in a field of size m.
- Take $Y_i = 1$ iff $P(i) = 0$.

Generating $Y_1, \ldots, Y_m$

- Let P be a random degree-d polynomial in a field of size m.
- Take $Y_i = 1$ iff $P(i) = 0$.
- $\mathbb{P}[Y_i = 1] = 1/m$, as all values of $P(i)$ are equiprobable.
- $\sum_i Y_i$ = number of roots $\leq d$.
- Random polynomial $\equiv$ upto d independent roots.

Distinguishing D from uniform

- $f = 1$ when at least one of $A_1, \ldots, A_m$ is 1.
- For any set $S \subseteq [m]$, $|S| \leq d$ the event $\{A_i = 1 \text{ for all } i \in S\}$ have equal probability $(1/m^{|S|})$ according to D and the uniform distribution.

Distinguishing D from uniform

- $f = 1$ when at least one of $A_1, \ldots, A_m$ is 1.
- For any set $S \subseteq [m]$, $|S| \leq d$ the event $\{A_i = 1 \text{ for all } i \in S\}$ have equal probability $(1/m^{|S|})$ according to D and the uniform distribution.
- (Slightly inaccurate) The distinguishing probability is roughly the probability that for independent Bernoulli variables $Z_1, \ldots, Z_m$ with expected value $1/m$ each,

$$\mathbb{P}[Z_1 + \ldots + Z_m > d] \approx \exp(-d \log d).$$

Finishing up: bounding the bias

- By d-wise independence, $\mathbb{E}_{x \sim D}[\chi_S(x)] = 0$ if S involves $\leq d$ terms.

Finishing up: bounding the bias

- By d-wise independence, $\mathbb{E}_{x\sim D}[\chi_S(x)] = 0$ if S involves $\leq d$ terms.
- Let S intersect (say) $A_1, \ldots, A_t$ with $t > d$. Let $S_i = S \cap A_i$ so that $\chi_S = \prod_i \chi_{S_i}$. Then each χ_{S_i} is almost unbiased except when $Y_i = 1$.

$$\begin{aligned}\mathbb{E}_{Y_i}\left|\mathbb{E}_x\left[\chi_{S_i}(x) \mid Y_i\right]\right| &= \mathbb{P}[Y_i = 1]\cdot 1 + \mathbb{P}[Y_i = 0]\cdot\frac{1}{m-1} \\ &= \frac{1}{m}\cdot 1 + \left(1-\frac{1}{m}\right)\cdot\frac{1}{m-1} = \frac{2}{m}\end{aligned}$$

Finishing up: bounding the bias

- By d-wise independence, $\mathbb{E}_{x\sim D}[\chi_S(x)] = 0$ if S involves $\leq d$ terms.
- Let S intersect (say) $A_1, \ldots, A_t$ with $t > d$. Let $S_i = S \cap A_i$ so that $\chi_S = \prod_i \chi_{S_i}$. Then each χ_{S_i} is almost unbiased except when $Y_i = 1$.

$$\begin{aligned}\mathbb{E}_{Y_i}\left|\mathbb{E}_x\left[\chi_{S_i}(x) \mid Y_i\right]\right| &= \mathbb{P}[Y_i = 1]\cdot 1 + \mathbb{P}[Y_i = 0]\cdot\frac{1}{m-1} \\ &= \frac{1}{m}\cdot 1 + \left(1-\frac{1}{m}\right)\cdot\frac{1}{m-1} = \frac{2}{m}\end{aligned}$$

- Hence, $\left|\mathbb{E}_Y\mathbb{E}_x\left[\chi_S(x) \mid Y\right]\right| \leq \mathbb{E}_Y\prod_{i=1}^{t}\left|\mathbb{E}\left[\chi_{S_i}(x) \mid Y_i\right]\right| \leq (2/m)^d$.

Conclusion

- Improve seed length of generators by sandwiching between functions with low ℓ_1 norm , instead of low-degree as in previous constructions.

Conclusion

- Improve seed length of generators by sandwiching between functions with low ℓ_1 norm , instead of low-degree as in previous constructions.
- Lower bounds also show that ϵ-biased distributions require seed length $O(\log^2 m)$ when $\delta = 1/poly(m, n)$. Hence they cannot provide $O(\log(mn))$ seed PRGs (e.g. for hardness amplification).

Conclusion

- Improve seed length of generators by sandwiching between functions with low ℓ_1 norm , instead of low-degree as in previous constructions.
- Lower bounds also show that ϵ-biased distributions require seed length $O(\log^2 m)$ when $\delta = 1/poly(m, n)$. Hence they cannot provide $O(\log(mn))$ seed PRGs (e.g. for hardness amplification).
- Seed length is also $O(\log^2 m)$ for constant error δ while the lower bound is only $O(\log m)$.

Thank You

Questions?