Sampling lower bounds

Emanuele Viola

Northeastern University

September 2018

The complexity of distributions

- Leading goal: lower bounds
  for computing a function on a given input

- This talk: lower bounds
  for sampling distributions, given uniform bits

- Several papers, connections,
  still uncharted

The complexity of distributions

- 2-source extractors [Chattopadhyay Zuckerman, …, Ben-Aroya Doron Ta-Shma]
- Data structure lower bounds ?

- This ta… s
  for sampling d… ons, given uniform bits

- Several papers, connections, still uncharted

Outline

- A couple of problems for decision trees

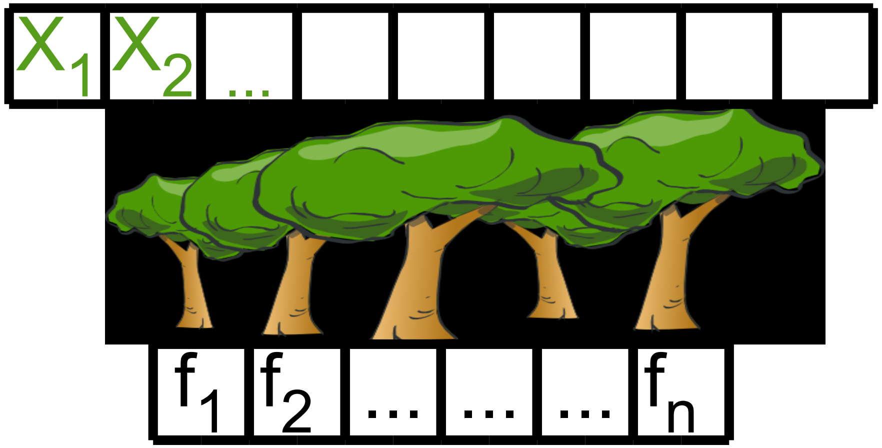
- AC^0

  - Upper bounds

  - Lower bounds

Sampling Hamming slices

- S = n uniform bits of weight n/4
- X uniform

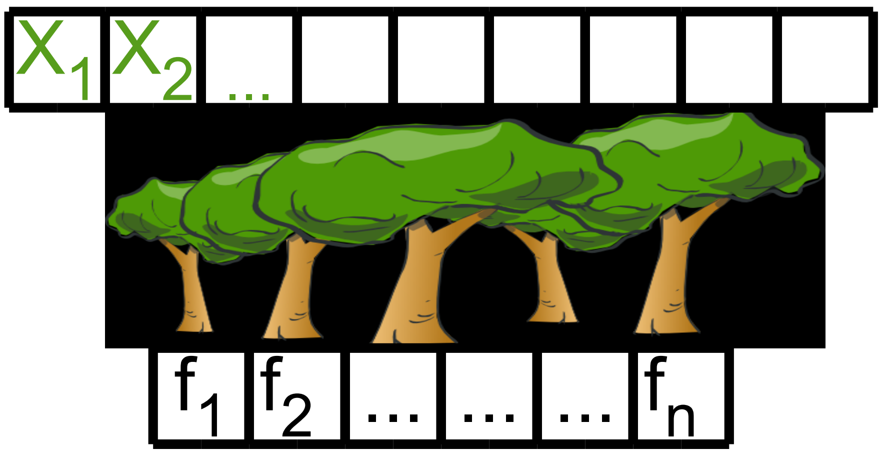


- $f : \{0,1\}^* \rightarrow \{0,1\}^n$
  depth-d forest

- Statistical distance Δ (f(X), S) $\geq$?

Sampling Hamming slices

- S = n uniform bits of weight n/4
- X uniform

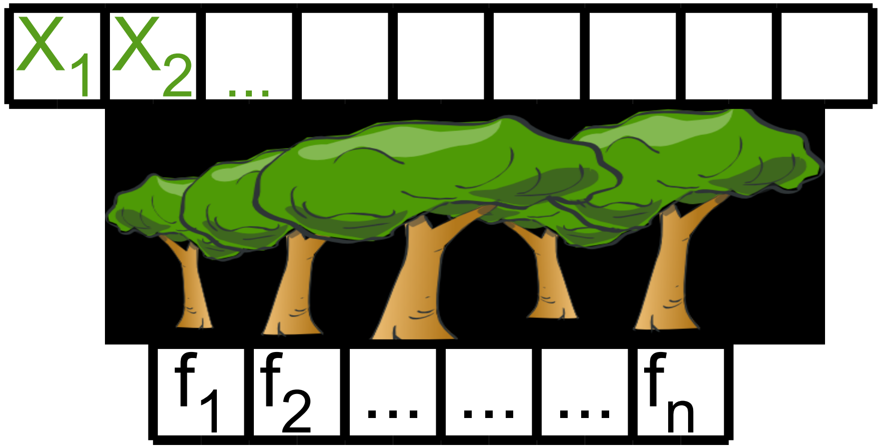


- $f : \{0,1\}^* \to \{0,1\}^n$
  depth-d forest

- Statistical distance $\Delta\,(f(X), S) \geq \Omega(1/2^d)$ [V]

Sampling Hamming slices

- S = n uniform bits of weight n/4
- X uniform

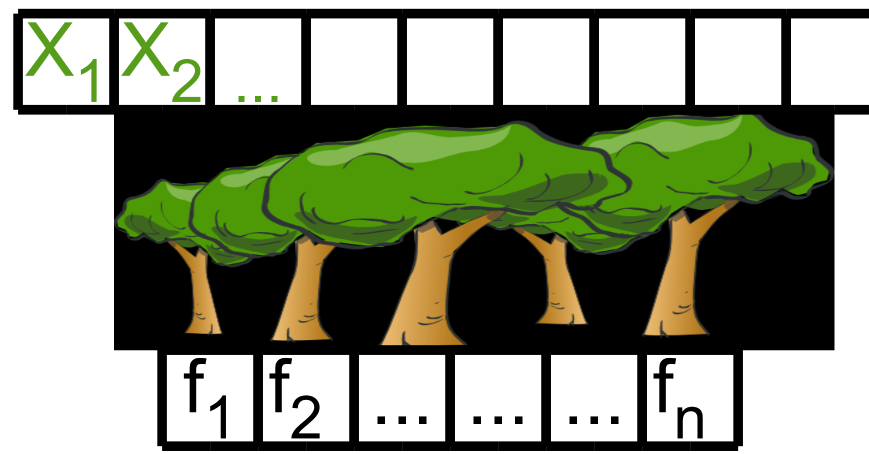


- $f : \{0,1\}^* \to \{0,1\}^n$
  depth-d forest

- Statistical distance $\Delta(f(X), S) \geq \Omega(1/2^d)$ [V]
  $\leq 1/n$ for $d = O(\log n)$ [CKKL]

Sampling Hamming slices

- S = n uniform bits of weight n/4
- X uniform



- $f : \{0,1\}^* \to \{0,1\}^n$
  depth-d forest

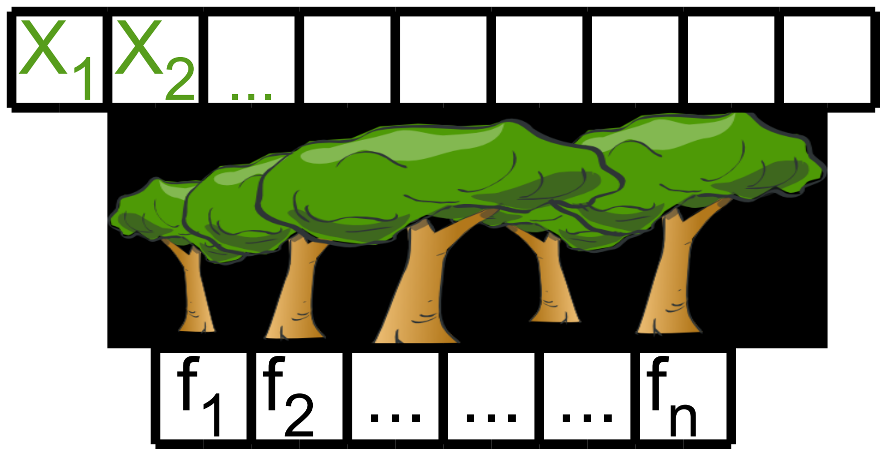
- Statistical distance Δ (f(X), S) $\geq \Omega(1/2^d)$ [V]
  $\leq 1/n$ for d = O(log n)
  [CKKL]

- Open: Δ (f(X), S) for d = O(1)?
- Note: $\Delta \geq 1 - o(1)$ ➔ data structure lower bound

Sampling permutations

- $\prod$:= uniform permutations of [n]
- $f : [n]^* \rightarrow [n]^n$

  depth-2 forest



- Statistical distance Δ (f(X), $\prod$) ≥ ?
- $\Delta \geq 1\text{-}o(1)$ ➔ data structure lower bound

Outline

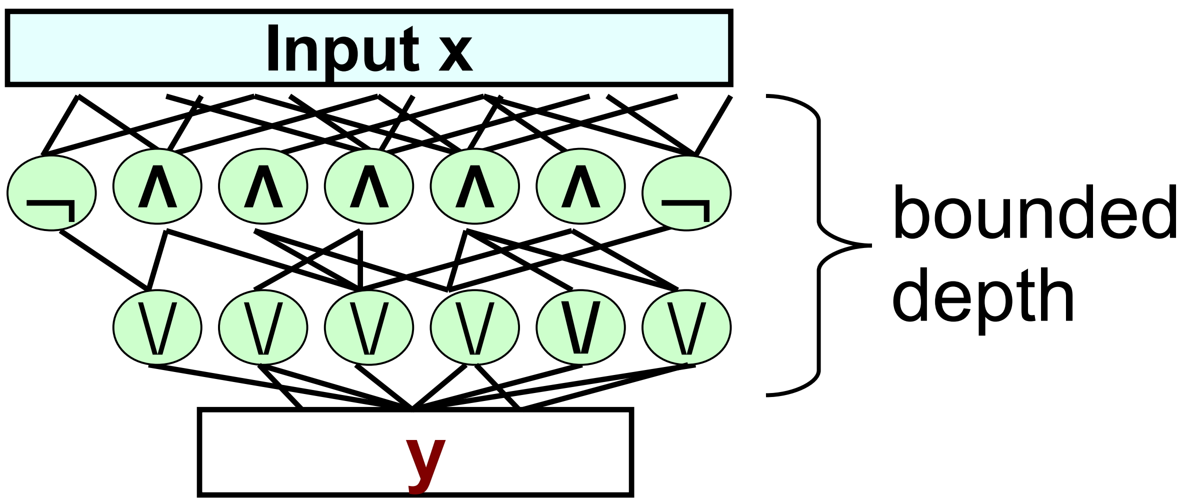
- A couple of problems for decision trees

- AC^0

  - Some upper bounds

  - Lower bounds

Bounded-depth circuits (AC^0)



$\vee$ = or
$\wedge$ = and
$\neg$ = not

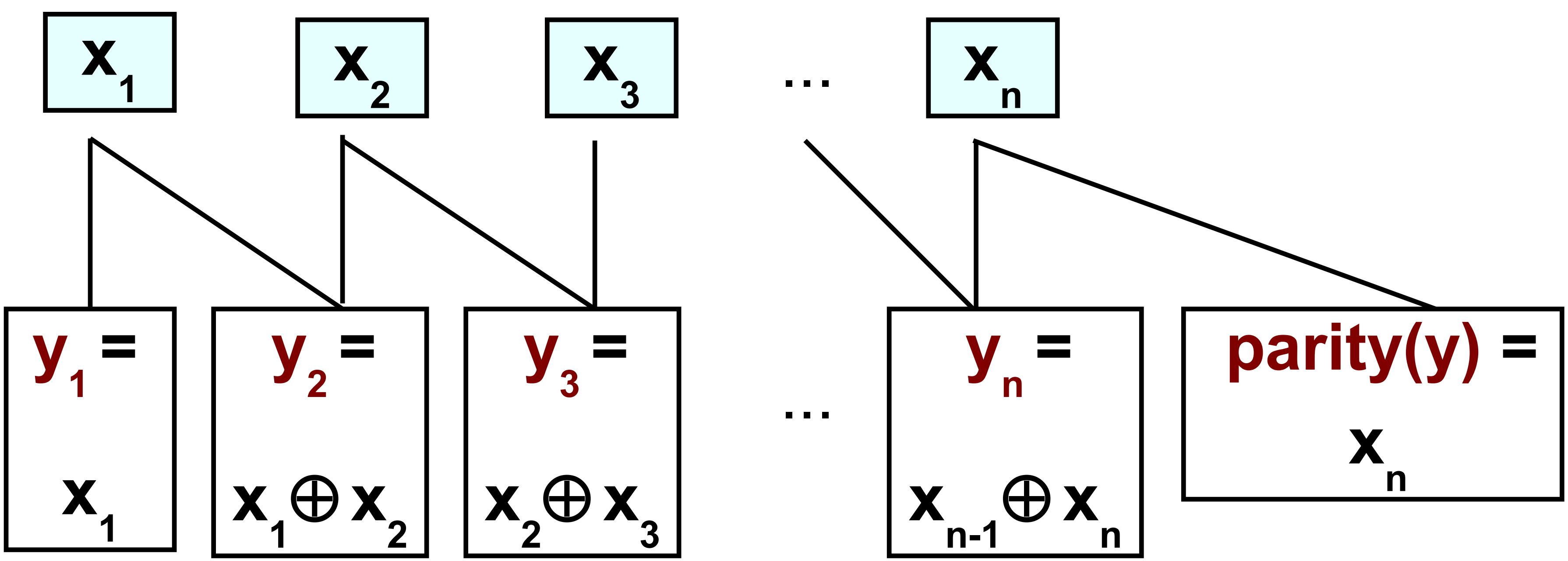
- AC^0 cannot compute parity
  [1980's: Furst Saxe Sipser, Ajtai, Yao, Hastad, ….]

Sampling (Y, parity(Y))

- Theorem [Babai '87; Boppana Lagarias '87]

There is $f : \{0,1\}^n \rightarrow \{0,1\}^{n+1}$, in AC^0

Distribution $f(X) \equiv$ (Y, parity(Y)) $\quad$ ($X, Y \in \{0,1\}^n$ uniform)



AC^0 can sample

- (Y, Inner-Product(Y)) [Impagliazzo Naor]
- Permutations (error 2^{-n}) [Matias Vishkin, Hagerup]
- (Y, f(Y)), any symmetric f (error 2^{-n}) [V]

  e.g. f=Majority
- Open: (Y, Majority(Y)) with error 0?

AC^0 can sample

Next

- (Y, Inner-Product(Y)) [Impagliazzo Naor]

- Permutations (error 2^{-n}) [Matias Vishkin, Hagerup]

- (Y, f(Y)), any symmetric f (error 2^{-n}) [V]

  e.g. f=Majority

- Open: (Y, Majority(Y)) with error 0?

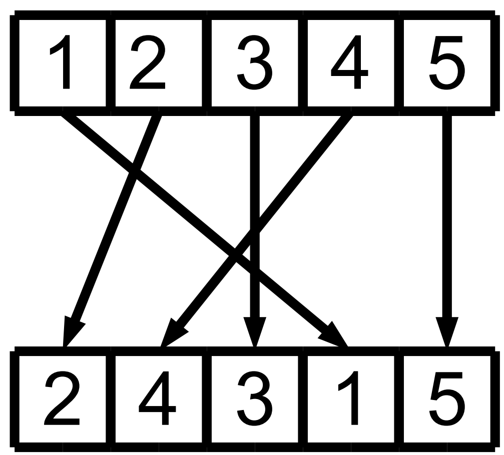
Sampling permutations in AC^0

- Dart throwing Place i = 1..n in A[1..n] uniformly

Sampling permutations in AC^0

- Dart throwing Place i = 1..n in A[1..n] uniformly

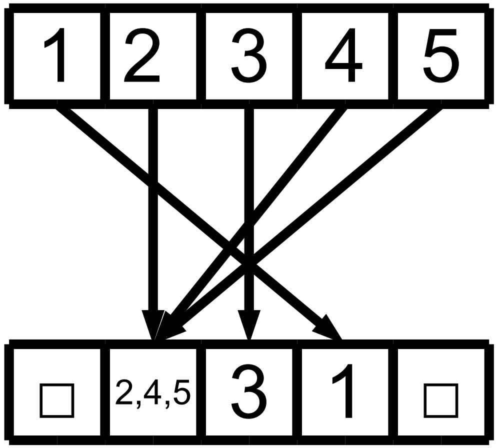
- If no collisions, done



Sampling permutations in AC^0

- Dart throwing Place i = 1..n in A[1..n] uniformly
- ~~If no collisions, done~~

  There will be collisions



Sampling permutations in AC^0

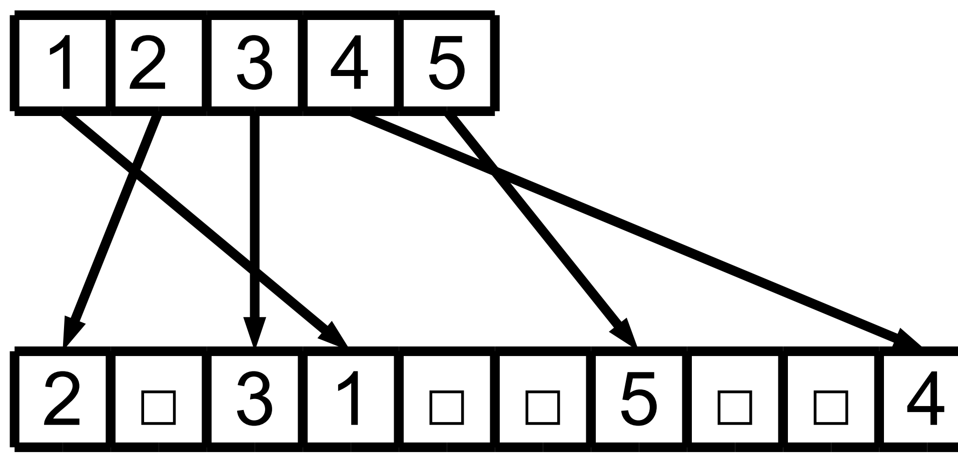
- Dart throwing Place i = 1..n in A[1..m] uniformly

- Enlarge A.

  No collisions,

  and I just need

  to remove the □



Sampling permutations in AC^0

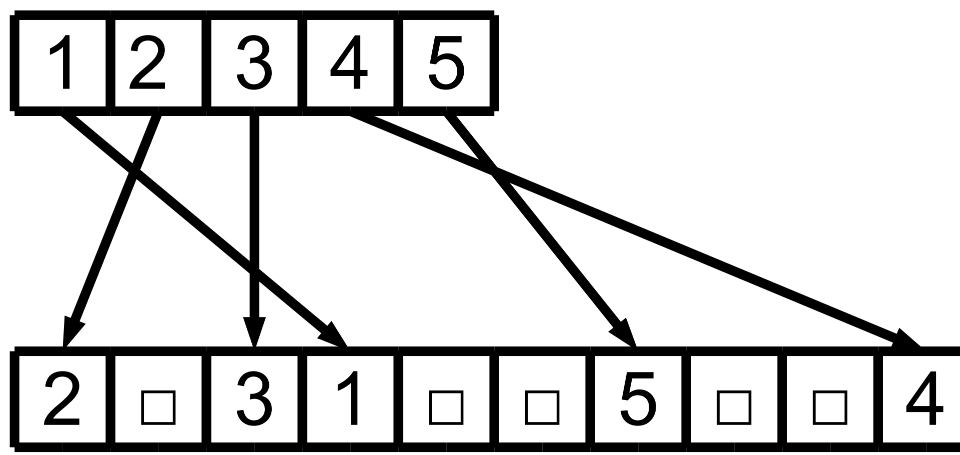
- Dart throwing Place i = 1..n in A[1..m] uniformly

- Enlarge A.

  No collisions,

  and I just need

  to remove the □

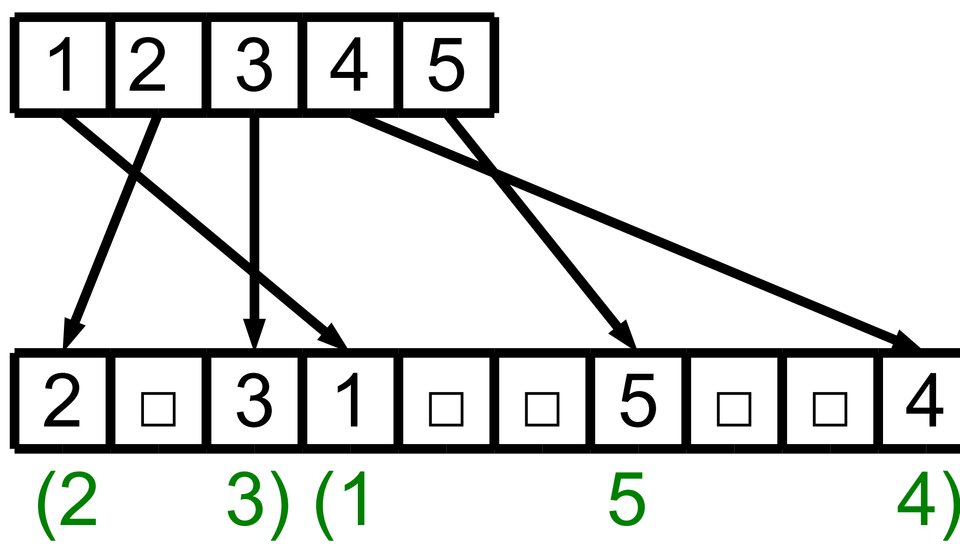


impossible

Sampling permutations in AC^0

- Dart throwing Place i = 1..n in A[1..m] uniformly

- Cycle format.
  Each cycle starts with
  least element.
  Least elements sorted.



- Next element in cycle computable in AC^0

Qed

Outline

- A couple of problems for decision trees
- AC^0
  - Some upper bounds
  - Lower bounds

AC^0 cannot sample

AC^0 cannot sample

- Error-correcting codes [Lovett V 2011, Beck Impagliazzo Lovett]

  Z = uniform on good binary code $\subseteq \{0,1\}^n$

  AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$

  ➔ Statistical-Distance(Z, $C(X)$) $\geq 1 - \exp(-n^{0.1})$

AC^0 cannot sample

- Error-correcting codes [Lovett V 2011, Beck Impagliazzo Lovett]

  Z = uniform on good binary code $\subseteq \{0,1\}^n$

  AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$

  ➔ Statistical-Distance(Z, C(X)) $\geq 1 - \exp(-n^{0.1})$

- (Y, f(Y)) for bit-block extractor $f : \{0,1\}^n \to \{0,1\}$

  Statistical-Distance((Y, f(Y) ,C(X)) > 0 [V 2011]

AC^0 cannot sample

- Error-correcting codes [Lovett V 2011, Beck Impagliazzo Lovett]

  Z = uniform on good binary code $\subseteq \{0,1\}^n$

  AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$

  ➔ Statistical-Distance(Z, $C(X)$) $\geq 1 - \exp(-n^{0.1})$

- $(Y, f(Y))$ for bit-block extractor $f : \{0,1\}^n \to \{0,1\}$

  Statistical-Distance($(Y, f(Y)$,$C(X)$) > 0 [V 2011]

  $> 1/2 - 1/n^{\omega(1)}$ [now]

"Cannot compute f better than tossing a coin,
even if you can sample the input yourself"

AC^0 cannot sample

- Error-correcting codes [Lovett V 2011, Beck Impagliazzo Lovett]

  Z = uniform on good binary code $\subseteq \{0,1\}^n$

  AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$

  ➔ Statistical-Distance(Z, C(X)) $\geq 1 - \exp(-n^{0.1})$

Next

- (Y, f(Y)) for bit-block extractor $f : \{0,1\}^n \to \{0,1\}$

  Statistical-Distance((Y, f(Y) ,C(X)) > 0 [V 2011]

  $> 1/2 - 1/n^{\omega(1)}$ [now]

"Cannot compute f better than tossing a coin, even if you can sample the input yourself"

- Theorem: AC^0 circuit C
  min-entropy $C(X) \geq k$ ($\forall$ a, $\Pr[C(X) = a] \leq 2^{-k}$)
  ➔ C(X) close to convex combination of bit-block sources
  with min-entropy $\geq k^2/n^{1.01}$

- Bit-block source: each bit is either constant or literal
  Example: $(0, 1, z_5 , 1-z_3 , z_3 , z_3 , 0 , z_2)$

- Corollary: f bit-block extractor ➔ $C(X) \neq (Y, f(Y))$

- Proof:

- Theorem: AC^0 circuit C
  min-entropy $C(X) \geq k$ ($\forall$ a, $Pr[C(X) = a] \leq 2^{-k}$)
  ➔ C(X) close to convex combination of bit-block sources
  with min-entropy $\geq k^2/n^{1.01}$

- Bit-block source: each bit is either constant or literal
  Example: $(0, 1, z_5, 1\text{-}z_3, z_3, z_3, 0, z_2)$

- Corollary: f bit-block extractor ➔ $C(X) \neq (Y, f(Y))$

- Proof: $C(X) = (Y, f(Y))$ ➔ min-entropy $C(X) \geq |Y| = n$
  ➔ convex combination high min-entropy bit-block sources
  can fix "f(Y)" bit leaving high min-entropy
  contradicts extractor property QED

- Theorem: AC[0]
min-entropy C
➔ C(X) close
with min-e

Heads up
Rules out Statistical-Distance 0, but not 0.1

Possible:
Statistical-Distance(C(X), (Y,f(Y)) ≤ 0.1,
but min-entropy C(X) = O(1)

Example later

- Bit-block sour
Example: (0,

- Corollary: f bit

- Proof: C(X) = (Y, f(Y)) ➔ min-entropy C(X) ≥ |Y| = n
➔ convex combination high min-entropy bit-block sources
can fix "f(Y)" bit leaving high min-entropy
contradicts extractor property QED

- Theorem: AC^0 circuit C
  min-entropy $C(X) \geq k$ ($\forall$ a, $\Pr[C(X) = a] \leq 2^{-k}$)
  ➔ C(X) close to convex combination of bit-block sources
  with min-entropy $\geq k^2/n^{1.01}$

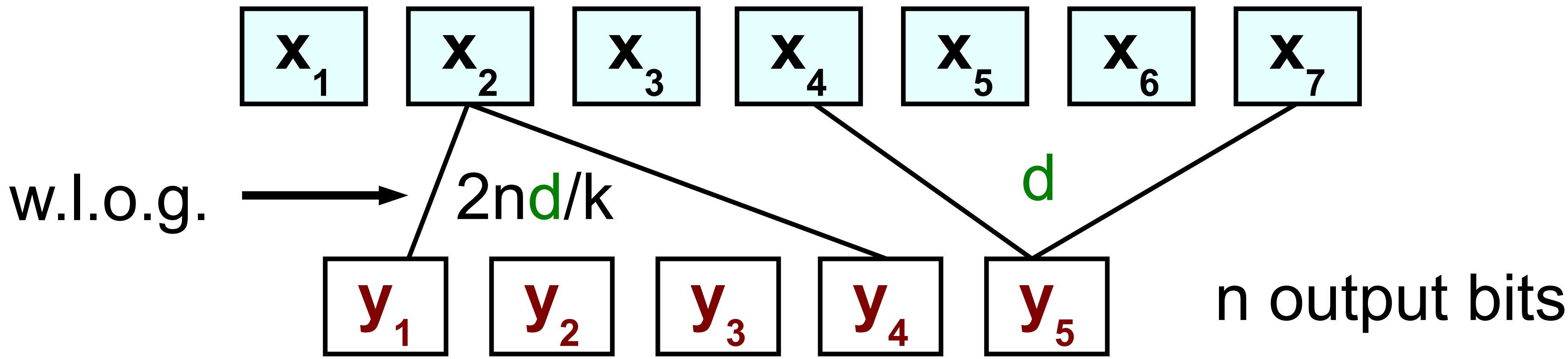
- Proof:

  (1) Prove when C is d-local (each output bit depends on d input bits)

  (2) For AC^0 use random restrictions
  - switching lemma collapses AC^0 to d-local
  - New: entropy is preserved

Proof

- d-local n-bit source min-entropy k: convex combo bit-block



- Output entropy > Ω(k) ➔ $\exists$ $\mathbf{y}_i$ with variance > Ω(k/n)
- Isoperimetry ➔ $\exists$ $\mathbf{x}_j$ with influence > Ω(k/nd)
- Set uniformly $N(N(\mathbf{x}_j)) \setminus \{\mathbf{x}_j\}$ (N(v) = neighbors of v)

  with prob. > Ω(k/nd), $N(\mathbf{x}_j)$ non-constant block of size 2nd/k
- Repeat $\Omega(k) / |N(N(\mathbf{x}_j))|$ times ➔ expect $\Omega(k^3/n^2d^3)$ blocks

♦

Proof

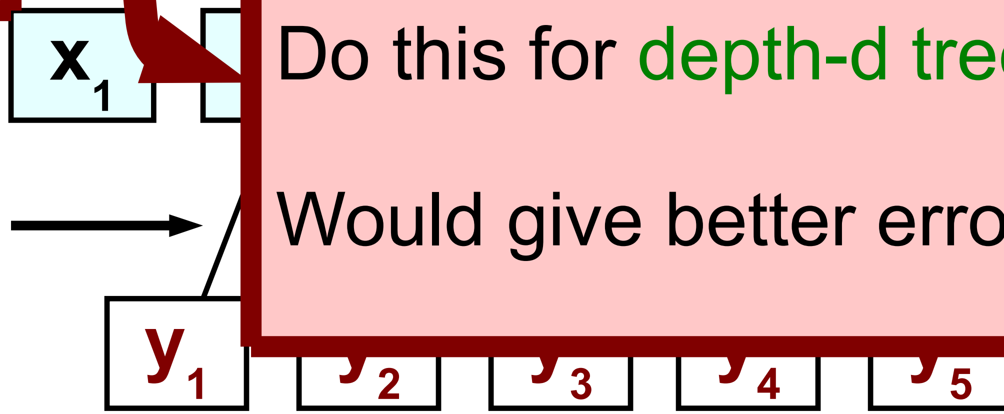
Open problem:

Do this for depth-d trees

Would give better error bounds

- d-local n-bit sc

w.l.o.g. →



- Output entropy > $\Omega(k)$ ➔ $\exists$ y_i with variance > $\Omega(k/n)$
- Isoperimetry ➔ $\exists$ x_j with influence > $\Omega(k/nd)$
- Set uniformly $N(N(x_j)) \setminus \{x_j\}$ (N(v) = neighbors of v)

  with prob. > $\Omega(k/nd)$, $N(x_j)$ non-constant block of size $2nd/k$
- Repeat $\Omega(k)$ / $|N(N(x_j))|$ times ➔ expect $\Omega(k^3/n^2d^3)$ blocks

♦

- Theorem: AC^0 circuit C
  min-entropy $C(X) \geq k$ ($\forall$ a, $Pr[C(X) = a] \leq 2^{-k}$)
  ➔ C(X) close to convex combination of bit-block sources
  with min-entropy $\geq k^2/n^{1.01}$

- Proof:

✓ (1) Prove when C is d-local (each output bit depends on d input bits)

(2) For AC^0 use random restrictions
- switching lemma collapses AC^0 to d-local
- New: entropy is preserved

The effect of restrictions on entropy

- Theorem $f : \{0,1\}^* \to \{0,1\}^n$: $f(X)$ has min-entropy k

  Let R be random restriction with $\Pr[*] = p$

  W.h.p. $f|_R(X)$ has min-entropy $\Omega(pk)$

- Proof:
- Bound collision probability $\Pr[\, f|_R(X) = f|_R(X)\,]$

- Isoperimetric inequality for noise [Lovett V]

  $\forall A \subseteq \{0,1\}^L$ of density α, uniform X, p-noise vector N :

$$\alpha^2 \leq \Pr[X \in A \wedge (X+N) \in A] \leq \alpha^{1+p}$$

Proof of isoperimetric inequality

- $\forall$ $A \subseteq \{0,1\}^L$ of density α random X, p-noise vector N :
  $\Pr[X \in A \wedge (X+N) \in A] \leq \alpha^{1+p}$

- Proof:
  $f := 1_A$
  $E_{X,N}[\ f(X) \cdot f(X+N)\ ]$
  $\quad = E_X[\ f(X) \cdot E_N[f(X+N)]\ ]$

Proof of isoperimetric inequality

- $\forall A \subseteq \{0,1\}^L$ of density α random X, p-noise vector N :
  $\Pr[X \in A \wedge (X+N) \in A] \le \alpha^{1+p}$

- Proof:
  $f := 1_A$
  $E_{X,N}[\ f(X) \cdot f(X+N)\ ]$

  $= E_X[\ f(X) \cdot E_N[f(X+N)]\ ]$

  $\le \sqrt{E_X[\ f^2(X)\ ]} \cdot \sqrt{E_X\ [E_N{}^2\ [f(X+N)]\ ]}$ Cauchy-Schwarz

Proof of isoperimetric inequality

- $\forall A \subseteq \{0,1\}^L$ of density α random X, p-noise vector N :

  $\Pr[X \in A \wedge (X+N) \in A] \leq \alpha^{1+p}$

- Proof:

  $f := 1_A$

  $E_{X,N}[\ f(X) \bullet f(X+N)\ ]$

  $= E_X[\ f(X) \bullet E_N[f(X+N)]\ ]$

  $\leq \sqrt{E_X[\ f^2(X)\ ]} \bullet \sqrt{E_X\ [E_N^2\ [f(X+N)]\ ]}$ Cauchy-Schwarz

  $\leq \sqrt{E_X[\ f^2(X)\ ]} \bullet E_X[f^{\,2-O(p)}\ (X)]^{1/(2-O(p))}$ Hypercontractivity

  $= \sqrt{\alpha} \bullet \alpha^{\,1/(2-O(p))}$ Qed

Recap

- Showed high-entropy AC^0 ➔ high-entropy bit-block sources
- Implies sampling lower bounds
- But only Statistical-Distance $\Delta > 0$, not 0.1

Possible:
Δ (C(X), (Y,f(Y)) $\leq$ 0.1, but min-entropy C(X) = O(1)

Example next

Example

- Circuit C: “On input x:

  If first 4 bits are 0 output the all-zero string

  Otherwise sample (Y, f(Y)) exactly”

- Statistical-Distance(C(X) , (Y, f(Y)) $\leq$ 0.1,

  but min-entropy C(X) = O(1)

- Observation: If you fix first 4 bits,

  min-entropy polarizes: either zero or very large

  We show this happens for every AC^0 circuit

Polarizing min-entropy

- Theorem: For every AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$
  $\exists$ set S of $\leq 2^n$ restrictions such that:

  (1) preserve output distribution

  $\Delta(C|_r (X), C(X)) \leq \varepsilon$, for uniform $r \in S$, X

  (2) polarize min-entropy

  $\forall\ r \in S$, $C|_r$ has min-entropy 0 or $n^{0.8}$

- ? ? ? ? ?

Polarizing min-entropy

- Theorem: For every AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$

  $\exists$ set S of $\leq 2^n$ restrictions such that:

  (1) preserve output distribution

  $\Delta(C|_r (X), C(X)) \leq \varepsilon$, for uniform $r \in S$, X

  (2) polarize min-entropy

  $\forall\, r \in S$, $C|_r$ has min-entropy 0 or $n^{0.8}$

- Trivial:

  $S :=$ one input for each of $\leq 2^n$ outputs, entropy always 0

Polarizing min-entropy

- Theorem: For every AC^0 circuit $C : \{0,1\}^* \to \{0,1\}^n$

  $\exists$ set S of $\leq 2^{n - n^{0.9}}$ restrictions such that:

  (1) preserve output distribution

  $\Delta(C|_r (X), C(X)) \leq \varepsilon$, for uniform $r \in S$, X

  (2) polarize min-entropy

  $\forall\, r \in S$, $C|_r$ has min-entropy 0 or $n^{0.8}$

Polarization lemma

- Lemma: For every $f : \{0,1\}^* \rightarrow \{0,1\}^n$

  $\exists$ set S of $\leq 2^{n - n^{0.9}}$ restrictions s.t.

  $\Delta(f|_r (X), f(X)) \leq \varepsilon$, for uniform $r \in S$, X

- Proof:
- Pick S randomly with Pr[*] = $n^{-0.9}$; fix $A = f^{-1}(y)$ of density α

  Show: $\Pr_S\Big[\Pr_{r,X}[X|_r \in A] < \alpha - \varepsilon 2^{-n} \Big] < 2^{-n}$

  Note: Deviation $\varepsilon 2^{-n}$ but $|S| < 2^n$

Polarization lemma

- Lemma: For every $f : \{0,1\}^* \rightarrow \{0,1\}^n$

  $\exists$ set S of $\leq 2^{n - n^{0.9}}$ restrictions s.t.

  $\Delta(f|_r (X), f(X)) \leq \varepsilon$, for uniform $r \in S$, X

- Proof:
- Pick S randomly with $\Pr[*] = n^{-0.9}$; fix $A = f^{-1}(y)$ of density α

  Show: $\Pr_S\left[\Pr_{r,X}[X|_r \in A] < \alpha - \varepsilon 2^{-n} \right] < 2^{-n}$

  Note: Deviation $\varepsilon 2^{-n}$ but $|S| < 2^n$

  Isoperimetric inequality ➔ $\Pr_{r,X}[X|_r \in A]$ “small variance”

Polarization lemma

- Lemma: For every $f : \{0,1\}^* \to \{0,1\}^n$

  $\exists$ set S of $\leq 2^{n - n^{0.9}}$ restrictions s.t.

  $\Delta(f|_r (X), f(X)) \leq \varepsilon$, for uniform $r \in S$, X

- Proof:
- Pick S randomly with $\Pr[*] = n^{-0.9}$; fix $A = f^{-1}(y)$ of density α

  Show: $\Pr_S\left[\Pr_{r,X}[X|_r \in A] < \alpha - \varepsilon 2^{-n} \right] < 2^{-n}$

  Note: Deviation $\varepsilon 2^{-n}$ but $|S| < 2^n$

  Isoperimetric inequality ➔ $\Pr_{r,X}[X|_r \in A]$ “small variance”

  Use specific lower-tail concentration bound Qed

Putting things together

- In the end, lower bound for sampling (Y, f(Y))

$$f : \{0,1\}^n \to \{0,1\} \text{ bit-block extractor}$$

- Given circuit C, statistical distance $1/2 - 1/n^{\omega(1)}$ witness:

  A U B =

  $\{ z : z$ one of those $2^{n - n^{0.9}}$ restrictions s.t. C is constant$\}$

  U $\{ (y,b) : b \neq f(y) \}$

- Proof: Think of C(X) as $C|_r$ (X) for uniform $r \in S$

  $C|_r$ constant ➔ $C|_r$ (X) $\in$ A, but (Y, f(Y)) not in A w.h.p.

  else $\Pr[C|_r (X) \in B] > 1/2 - 1/n^{\omega(1)}$, but (Y, f(Y)) never in B

More open problems and conclusion

- Open problem: Statistical distance 1/2 - $\exp(-n^{0.1})$
- Derandomize entropy polarization
- Much more to chart...