The complexity of distributions

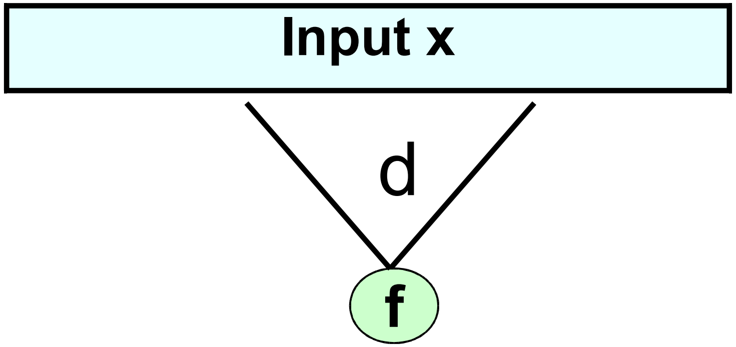
Emanuele Viola

Northeastern University

March 2011

Local functions (a.k.a. Junta, NC^0)

- $f : \{0,1\}^n \rightarrow \{0,1\}$ d-local : output depends on d input bits



- Fact: Parity(x) = 1 $\Leftrightarrow \sum x_i = 1 \bmod 2$ is not n-1 local

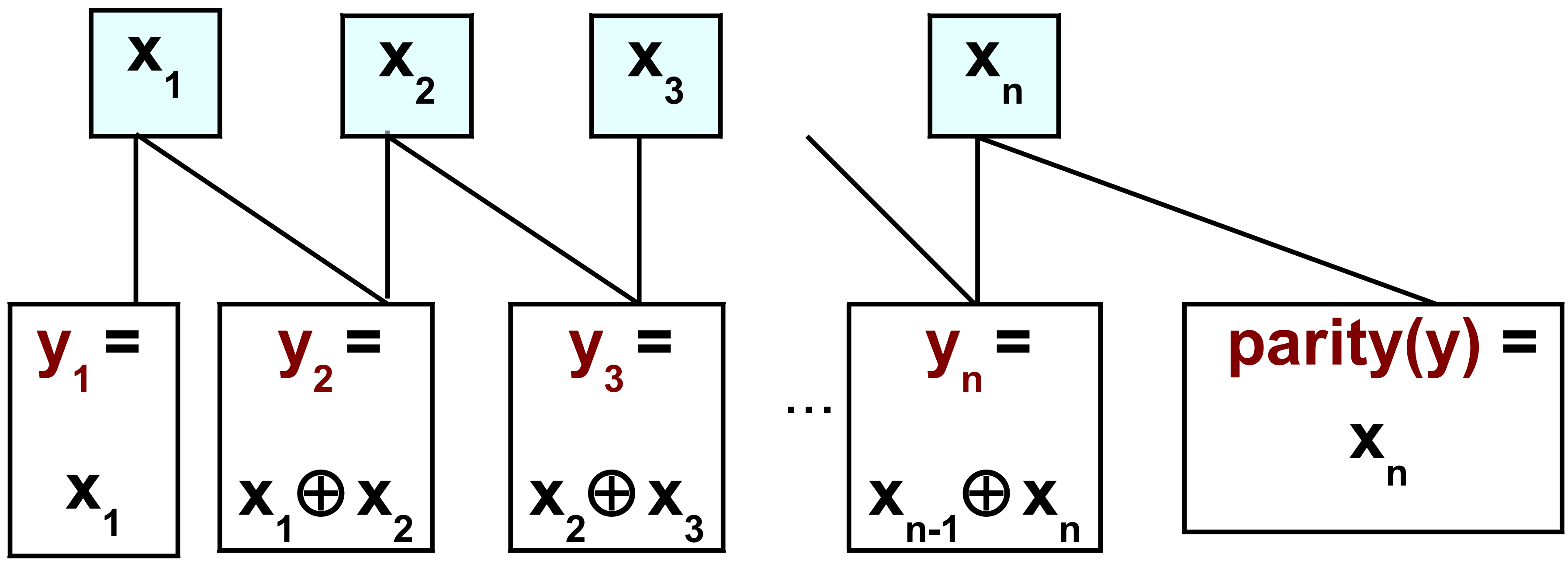
- Proof: Flip any input bit $\Rightarrow$ output flips ♦

Local generation of (Y, parity(Y))

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

There is $f : \{0,1\}^n \to \{0,1\}^{n+1}$, each bit 2-local

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



Our message

- Complexity theory of distributions (as opposed to functions)

How hard is it to generate (a.k.a. sample)
distribution D given random bits ?

E.g., D = (Y, parity(Y)), D = W_k := uniform n-bit with k 1's

Is message new?

- In addition to previous example:
- Generate Random Factored Numbers [Bach '85, Kalai]
- On the Implementation of Huge Random Objects
  [Goldreich Goldwasser Nussboim '03]
- The Equivalence of Sampling and Searching [Aaronson '10]
  (Given x, sample D_x)
- This work: first negative results (a.k.a. lower bounds)
  new connections

Outline of talk

- Generating W_k := uniform n-bit with k 1's
  - Local
  - Decision tree
- Results for (Y, b(Y))
- Bounded-depth circuit model

Our results: local

- Theorem

$f : \{0,1\}^n \rightarrow \{0,1\}^n$ 0.1 log n - local

$\Downarrow$

f(X) at Statistical Distance $> 1 - n^{-\Omega(1)}$

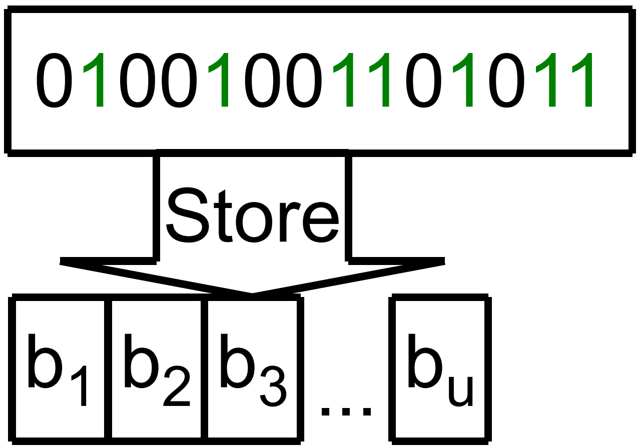
from $W_{n/2}$ = uniform w/ weight n/2

- Tight up to $\Omega()$: $f(x) = x$

- Extends to W_k, $k \neq n/2$, tight?

Our results: succinct data structures

- Problem:
  Store $S \subseteq \{1, 2, \ldots, n\}$, $|S|$ fixed
  in u = optimal + r bits,
  answer "$i \in S$?" probing d bits.



- Connection:
  Solution $\Rightarrow$ generate $W_{|S|}$ d-local, Stat. Distance < $1 - 2^{-r}$

- Corollary: Need $r > \Omega(\log n)$ if $d = 0.1 \log n$
  First lower bound for $|S| = n/2, n/4, \ldots$

Proof

- Theorem: Let $f : \{0,1\}^n \to \{0,1\}^n$: d= 0.1 log n-local.

  There is $T \subseteq \{0,1\}^n$: $\left| \Pr[f(x) \in T] - \Pr[W_{n/2} \in T] \right| > 1 - n^{-\Omega(1)}$

- Warm-up scenarios:

- $f(x) = 000111$ Low-entropy $T := \{ 000111 \}$

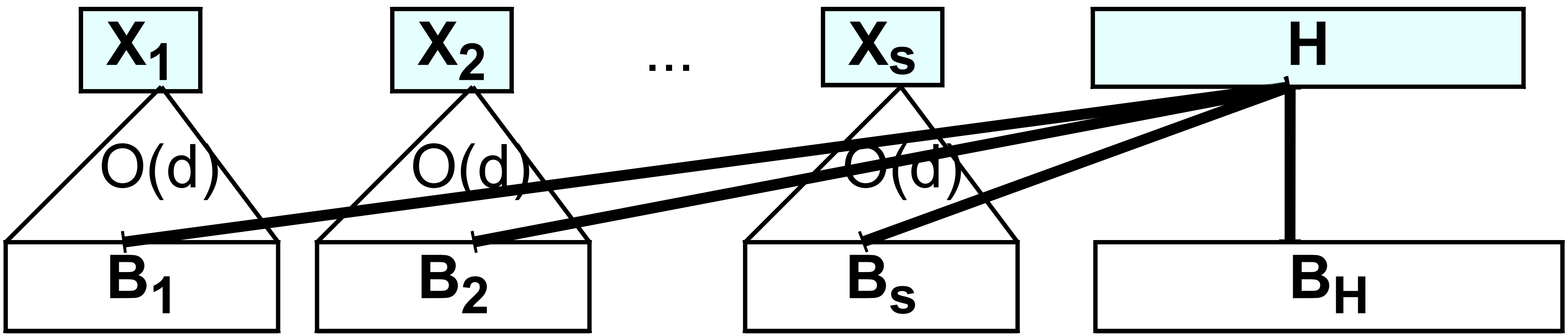
  $\left| \Pr[f(x) \in T] - \Pr[W_{n/2} \in T] \right| = \left| 1 - |T| / (n \text{ choose } n/2) \right|$

- $f(x) = x$ "Anti-concentration" $T := \{ z : \sum_i z_i = n/2 \}$

  $\left| \Pr[f(x) \in T] - \Pr[W_{n/2} \in T] \right| = \left| \Theta(1)/\sqrt{n} - 1 \right|$

Proof

- Partition input bits $X = (X_1 , X_2 , \dots , X_s , H)$

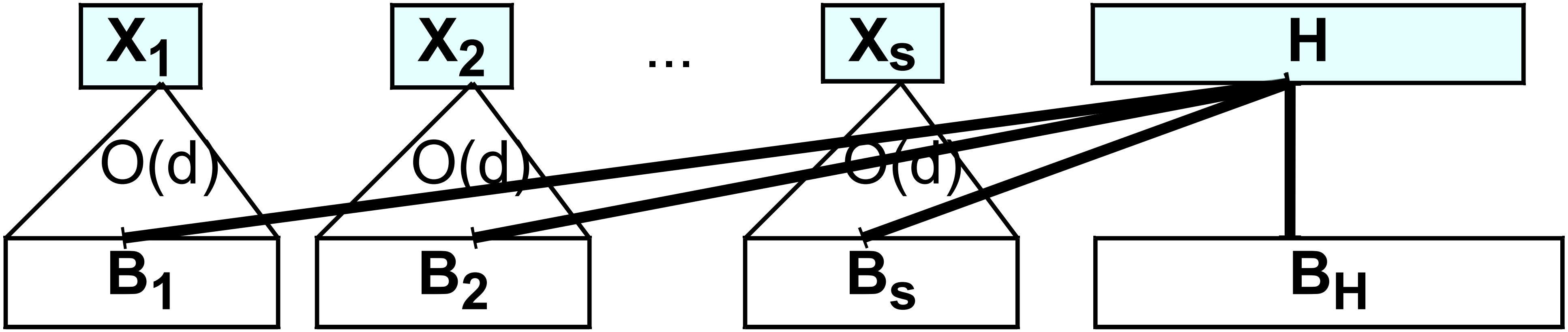


- Fix H. Output block B_i depends only on bit X_i

- Many B_i constant ($B_i(0,H) = B_i(1,H)$) $\Rightarrow$ low-entropy

- Many B_i depend on X_i ($B_i(0,H) \neq B_i(1,H)$)

  Idea: Independent $\Rightarrow$ anti-concentration: can't sum to n/2



- If many $B_i(0,H)$, $B_i(1,H)$ have different sum of bits, use

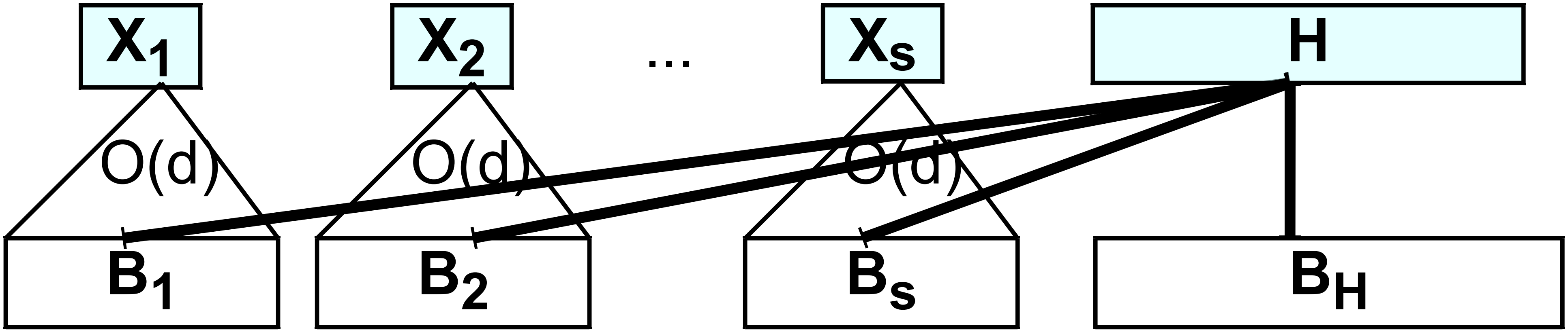
Anti-concentration Lemma [Littlewood Offord]

For $a_1, a_2, \ldots, a_s \neq 0$, any c, $\Pr_{X \in \{0,1\}^s} [\sum_i a_i X_i = c] < 1/\sqrt{n}$

- Problem: $B_i(0,H) = 100$, $B_i(1,H) = 010$

  high entropy but no anti-concentration

- Fix: want many blocks 000, so high entropy $\Rightarrow$ different sum



- Test $T \subseteq \{0,1\}^n$: $\Pr[f(X_1,\ldots,X_s,H) \in T] \approx 1$; $\Pr[W_{n/2} \in T] \approx 0$

$z \in T \Leftrightarrow$

$\exists$ H : $\exists$ $X_1,\ldots,X_s$ w/ many blocks B_i fixed : $f(X_1,\ldots,X_s,H) = z$

OR

Few blocks $z|_{B_i}$ are 000

OR

$\sum_i z_i \neq n/2$

Outline of talk

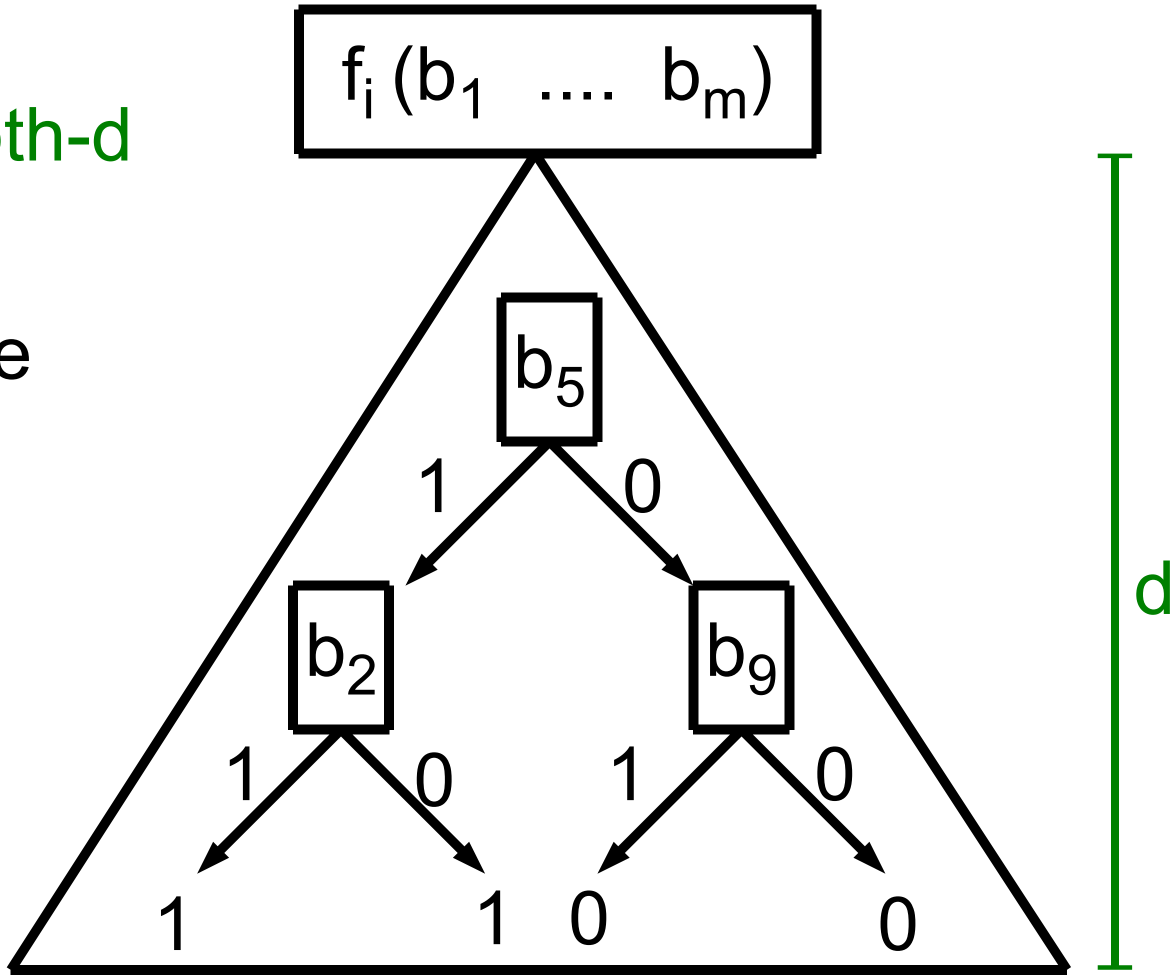
- Generating W_k := uniform n-bit with k 1's
  - Local
  - Decision tree
- Results for (Y, b(Y))
- Bounded-depth circuit model

Decision tree model

- $f : \{0,1\}^m \rightarrow \{0,1\}^n$ depth-d
  each output bit f_i
  is depth-d decision tree

- d adaptive bit-probes

- Depth d $\subseteq$ 2^d local



Our results: decision trees

- Theorem $f : \{0,1\}^* \to \{0,1\}^n$: depth $< 0.1 \log n$

  Distance($f(X)$, $W_{n/2}$) $> n^{-\Omega(1)}$

- Worse than $1 - n^{-\Omega(1)}$ lower bound for local

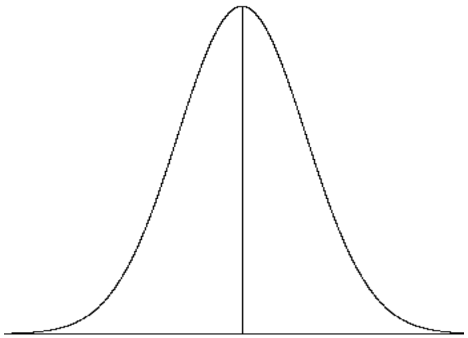
- Theorem building on [Czumaj Kanarek Lorys Kutyłowski]

  $\exists$ f : depth $O(\log n)$ and Distance($f(X)$, $W_{n/2}$) $< 1/n$

Tool for lower bound proof

- Central limit theorem:

  $x_1, x_2, \ldots, x_n$ independent $\Rightarrow \sum x_i \approx$ normal



- Bounded-independence central limit theorem [Diakonikolas Gopalan Jaiswal Servedio V.]

  $x_1, x_2, \ldots, x_n$ k-wise independent $\Rightarrow \sum x_i \approx$ normal

$$\forall\, t \quad |\Pr[\textstyle\sum x_i < t] - \Pr[\text{normal} < t]| < 1/\sqrt{k}$$

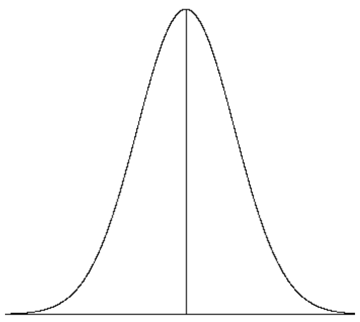
Proof

- Theorem $f : \{0,1\}^* \to \{0,1\}^n$: each bit depth $< 0.1 \log n$
  Distance(f(X), $W_{n/2}$) $> n^{-\Omega(1)}$

- Proof: Is output distribution f(X) (k = 10)-wise independent?

  NO : $W_{n/2} \approx$ k-wise independent
  Distance(those k bits, uniform on $\{0,1\}^k$) $> 2^{-k(0.1 \log n)}$
  (granularity of decision tree probability)

  YES : by prev. theorem $\sum f(X)_i \approx$ normal
  so often $\sum f(X)_i \neq n/2$ ♦



Outline of talk

- Generating W_k := uniform n-bit with k 1's
  - Local
  - Decision tree
- Results for (Y, b(Y))
- Bounded-depth circuit model

Our results for (Y, b(Y))

- Results so far: Distribution = $W_{n/2}$

  below: Distribution = (Y, b(Y)), b boolean

- Theorem: $f : \{0,1\}^n \to \{0,1\}^{n+1}$

  o(log n)-local $\Rightarrow$ Distance$\big($ f(X) , (Y, (Y mod p)>p/2) $\big)$ > 0.49

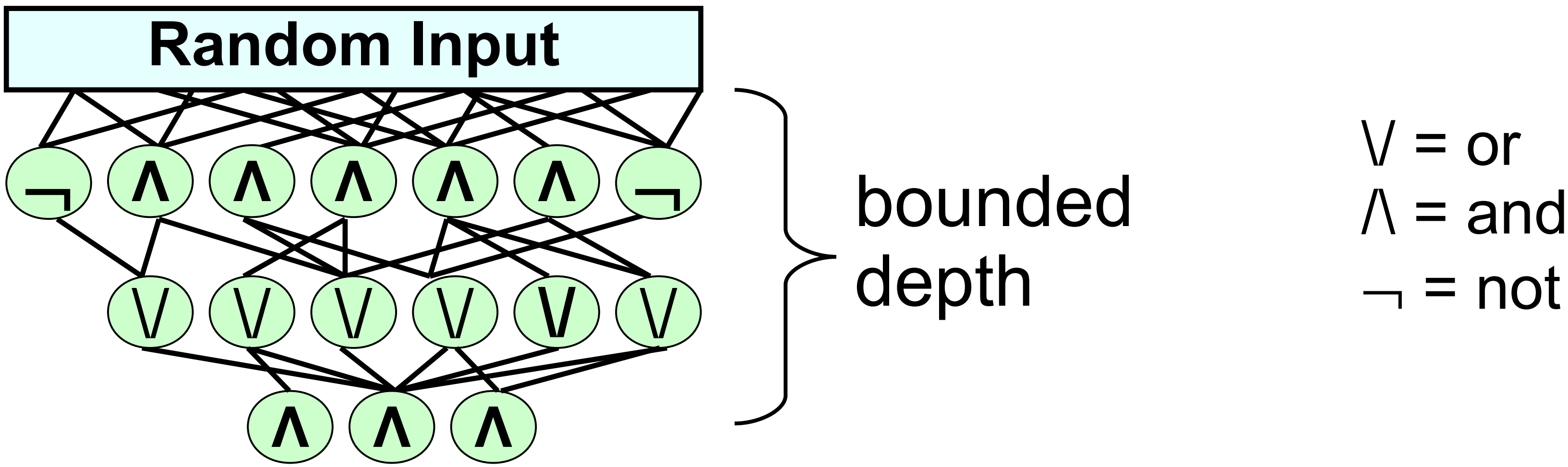
  o(log n)-depth $\Rightarrow$ Distance$\big($ f(X) , (Y, majority Y)$\big)$ > $n^{-\Omega(1)}$

Outline of talk

- Generating W_k := uniform n-bit with k 1's
  - Local
  - Decision tree
- Results for (Y, b(Y))
- Bounded-depth circuit model

Bounded-depth circuits

- More general model: small bounded-depth circuits (AC^0)



- Theorem building on [Matias Vishkin, Hagerup; '91]

  Can generate (Y, majority(Y)), error $2^{-|Y|}$

- Challenge: error 0?

Our lower bound for codes

- Theorem[Lovett V.] Cannot generate error-correcting code

  - Code $C \subseteq \{0,1\}^n$ of size $|C| = 2^{k = \Omega(n)}$

    $x \neq y \in C \Rightarrow x, y$ far : hamming distance $\Omega(n)$

  - $f : \{0,1\}^* \to \{0,1\}^n$, $f \in AC^0$

    Distance(f(X), uniform over C) $> 1 - n^{-\Omega(1)}$

Warm-up

- Fact: $f : \{0,1\}^k \to \{0,1\}^n$, $f \in AC^0$
  f cannot compute encoding function of C,
  mapping message $m \in \{0,1\}^k$ to codeword

- Proof:

- [Linial Mansour Nisan '93, Boppana] low sensitivity of AC^0:
  m, m' random at hamming distance 1
  $\Rightarrow$ f(m), f(m') close in hamming distance.

- But $f(m) \neq f(m') \in C \Rightarrow$ far in hamming distance ♦

Lower bound for codes

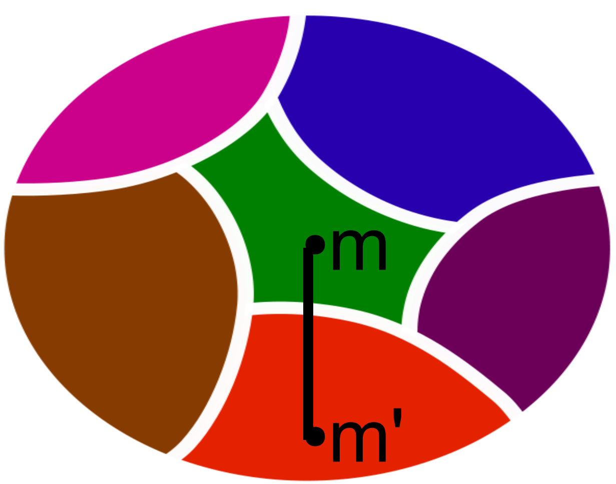
- Theorem [Lovett V.] $f : \{0,1\}^{L >> k} \rightarrow \{0,1\}^n$, $f \in AC^0$

  Distance(f(X), uniform over C) $> 1 - n^{-\Omega(1)}$

  Problem: f needs not compute encoding function.

  Input length >> message length

- Idea: Input $\{0,1\}^L$ to f partitioned in |C| sets



- Isoperimetric inequality [Harper, Hart]:

  Random m, m' at distance 1 often in ≠ sets ⇒ low sensitivity

Lower bound for codes

- Theorem [Lovett V.] $f : \{0,1\}^{L >> k} \to \{0,1\}^n$, $f \in AC^0$

  Distance(f(X), uniform over C) $> 1 - n^{-\Omega(1)}$

- Note: to get

  Need isoperimetric inequality for m, m' at distance >> 1

  Fact[thanks to Samorodnitsky] $\forall A \subseteq \{0,1\}^L$ of density α

  random m, m' obtained flipping bits w/ probability p :

$$\alpha^2 \leq \Pr[\text{both } m \in A \text{ and } m' \in A] \leq \alpha^{1/(1-p)}$$

Summary

- Complexity of distributions = uncharted territory

- Lower bounds for generating W_k locally

  $\Rightarrow$ lower bound for storing sets of size n/2, n/4, …

- More lower bounds:

  decision trees, generating (Y, b(Y)), AC^0

- Tools: Anti-concentration,

  bounded-independence central limit theorem,

  isoperimetric inequalities, ...

Two open problems

- Note $\exists$ 2-local $f : \{0,1\}^{2n} \to \{0,1\}^n$

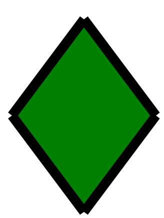
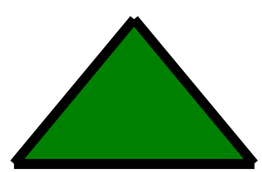
  Distance(f(X), $W_{n/4}$ = uniform w/ weight n/4) = $1 - \Theta(1)/\sqrt{n}$

- Challenge: Distance $1 - 2^{-\Omega(n)}$ input length = $H(1/4)n+o(n)$

- Recall: AC^0 can generate (Y, majority(Y)), error $2^{-|Y|}$

  Challenge: error 0?

  - Related [Lovett V.] Any bijection

    $\{0,1\}^n$ =  $\to$  $= \{x \in \{0,1\}^{n+1} : \sum x_i \geq n/2 \}$

    has large expected hamming distortion? (n even)

- ∑∏√∩∉ ∪⊃⊇⊄⊂⊆∈⇓⇒⇑⇐⇔∨∧≥≤∀∃Ωαβεγδ→
- ≠≈TΑΘ

-
- Recall: edit style changes ALL settings.
- Click on “line” for just the one you highlight

More connections

- More uses of generating W_k := uniform n-bit string with k 1's
- McEliece cryptosystem
- Switching networks, …

Previous results

- Store $S \subseteq \{1, 2, \dots, n\}$, $|S| = k$, in bits, answer "$i \in S$?"

- [Minsky Papert '69] Average-case study

- [Buhrman Miltersen Radhakrishnan Venkatesh; Pagh '00]

  Space O(optimal), probe O(1) when $k = \Theta(n)$

  Lower bounds for $k < n^{1-\varepsilon}$

- [..., Pagh, Pătraşcu] space = optimal + o(n), probe O(log n)

- [V. '09] lower bounds for $k = \Omega(n)$, except $k = n / 2^a$