The complexity of distributions

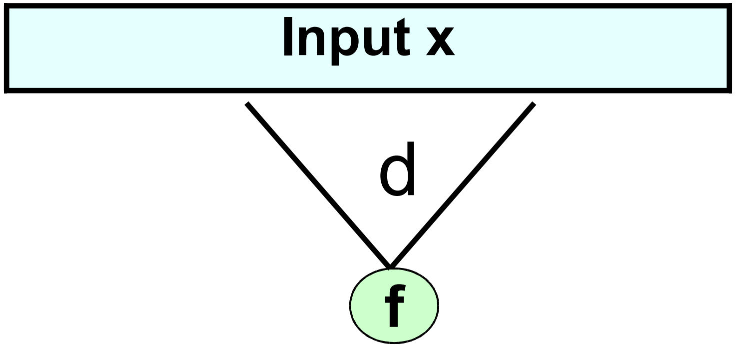
Emanuele Viola

Northeastern University

March 2012

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

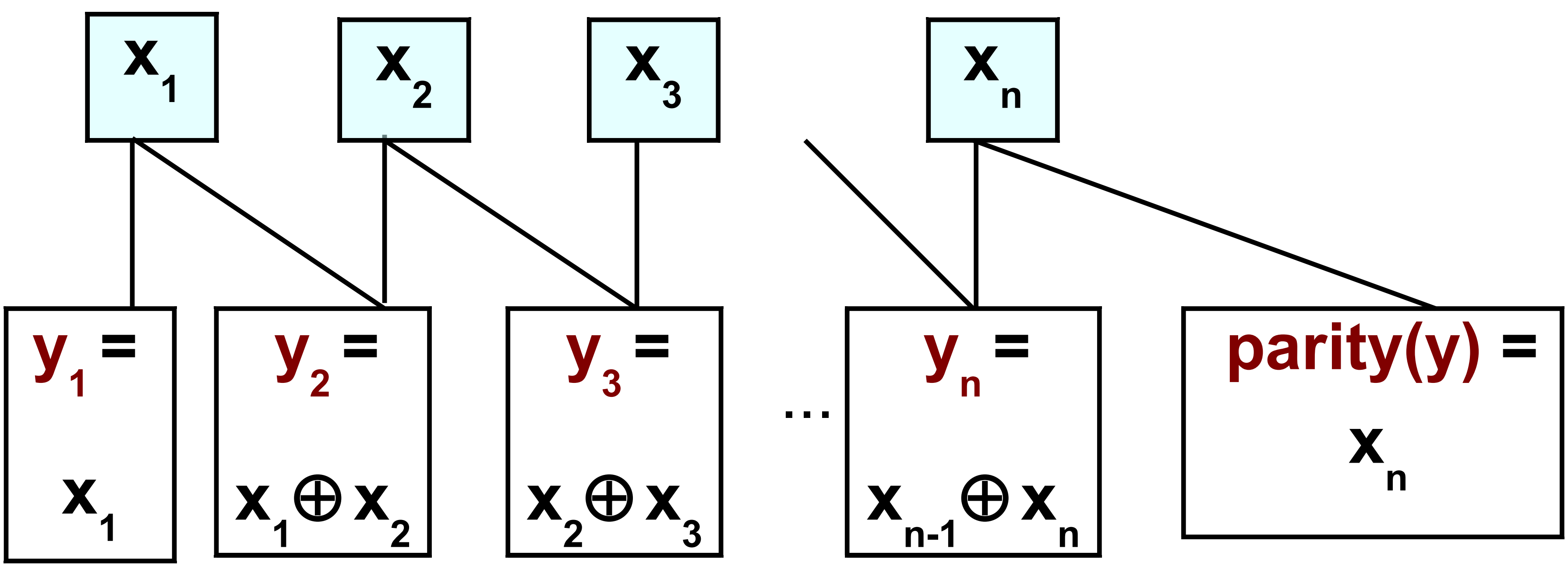
- Proof: Flip any input bit $\Rightarrow$ output flips ♦

Local generation of (Y, parity(Y))

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

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

- Generate Random Factored Numbers [Bach '85, Kalai]
- Random Generation of Combinatorial Structures from a Uniform Distribution [Jerrum Valiant Vazirani '86]
- The Quantum Communication Complexity of Sampling [Ambainis Schulman Ta-Shma Vazirani Wigderson]
- On the Implementation of Huge Random Objects [Goldreich Goldwasser Nussboim]
- Our line of work: 1) first negative results (lower bounds) for local, AC^0, Turing machines, etc.
  2) new connections

Outline of talk

- Lower bound for sampling W_k = uniform weight-k string

- Randomness extractors
  - Local sources
  - Bounded-depth circuit (AC^0)
  - Turing machine

- Theorem [V.]

$f : \{0,1\}^n \to \{0,1\}^n$ 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?
- Also open: remove bound on input length

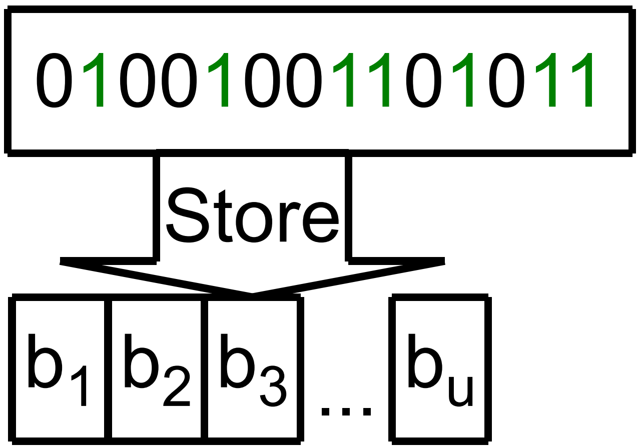
Succinct data structures

- Problem:

  Store $S \subseteq \{1, 2, \ldots, n\}$, $|S|$ fixed

  in u = optimal + r bits,

  answer "$i \in S$?" probing d bits.



- Connection [V.]:

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

- Warm-up scenarios:

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

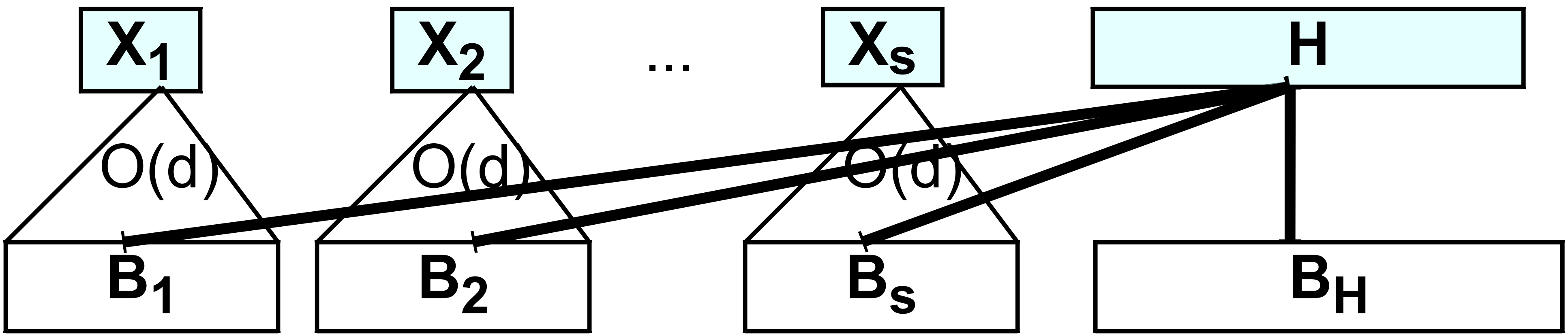
  $\big| \Pr[f(x) \in T] - \Pr[W_{n/2} \in T] \big| = \big| 1 - |T| / (n \text{ choose } n/2) \big|$

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

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

Proof

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

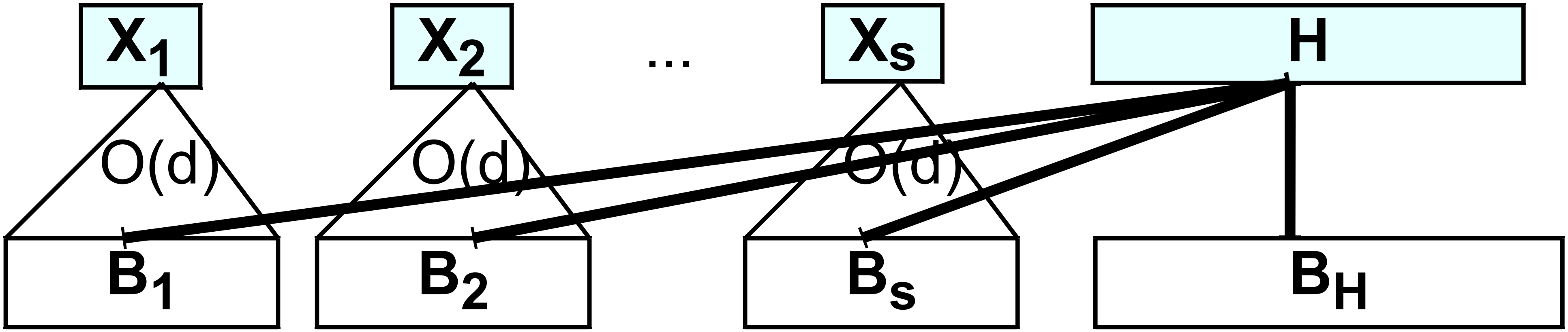


- Fix H. Output block B_i depends only on bit X_i

- 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



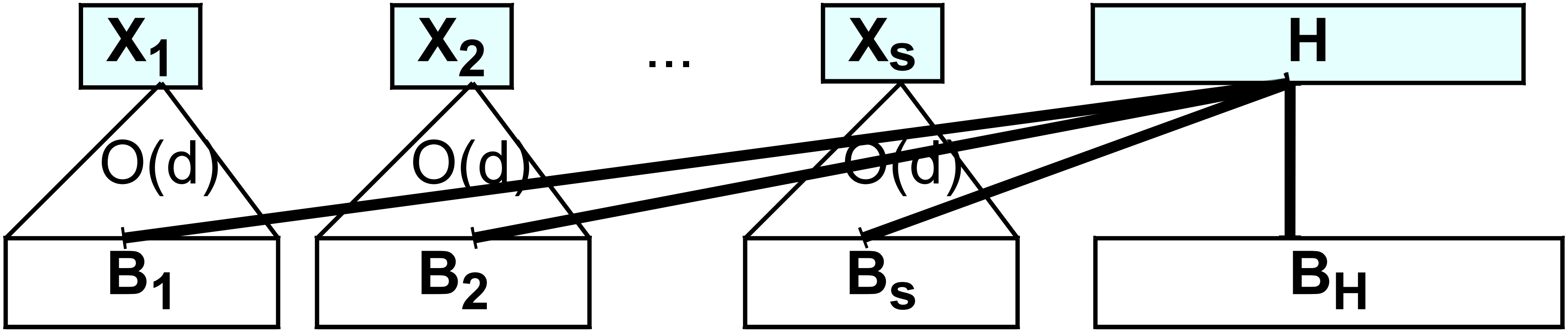
- If many $B_i(0,H)$, $B_i(1,H)$ have different sum of bits, use

> Anti-concentration Lemma [Littlewood Offord]
>
> For $a_1, a_2, \ldots, a_s \neq 0$, any c, $\Pr_{X \in \{0,1\}^s} \left[\sum_i a_i X_i = c\right] < 1/\sqrt{n}$

- 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

$\sum_i z_i \neq n/2$

Open problem

- Understand complexity of W_k = uniform weight-k string

  for all choices of: k,

  model (local, decision tree, etc.),

  statistical distance,

  randomness complexity

- Similar problems in combinatorics, coding, ergodic theory

- One example

  $\exists$ 2-local $f : \{0,1\}^{2n} \to \{0,1\}^n$ Distance$(f(X), W_{n/4}) \leq 1-\Theta(1)/\sqrt{n}$

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

Outline of talk

- Lower bound for sampling W_k = uniform weight-k string

- Randomness extractors
  - Local sources
  - Bounded-depth circuit (AC^0)
  - Turing machine

Randomness extractors

- Want: turn weak randomness (correlation, bias, ...)
  into close to uniform

- **Extractor** for sources (distributions) S on $\{0,1\}^n$

  Deterministic, efficient map : $\{0,1\}^n \to \{0,1\}^m$

  $\forall$ D $\in$ S, **Extractor**(D) ε-close to uniform

- Starting with [Von Neumann '51] major line of research

Sources

- Independent blocks [Chor Goldreich 88, Barak Bourgain Impagliazzo Kindler Rao Raz Shaltiel Sudakov Wigderson ...]
- Some bits fixed, others uniform & indep. [Chor Friedman Goldreich Hastad Rudich Smolensky '85, Cohen Wigderson, Kamp Zuckerman, …]
- One-way, space-bounded algorithm [Blum '86, Vazirani, Koenig Maurer, Kamp Rao Vadhan Zuckerman]
- Affine set [BKSSW, Bourgain, Rao, Ben-Sasson Kopparty, Shaltiel]
- Our results: first extractors for circuit sources: local, AC^0 and for Turing-machine sources

Trevisan Vadhan; 2000

- Sources D with min-entropy k ($\Pr[D = a] < 2^{-k}$ $\forall$ a)
  sampled by small circuit C: $\{0,1\}^* \to \{0,1\}^n$
  given random bits.

- Extractor $\Longrightarrow$ Lower bound for C
  (even 1 bit
  from k=n-1)

- Extractor $\Longleftarrow$ Time($2^{O(n)}$) $\not\subseteq$ $\exists\forall\exists\forall\exists$–circuit size $2^{o(n)}$

[V.]

- Extractor $\Leftrightarrow$ Sampling lower bound

(1 bit from k=n-1)

$f : \{0,1\}^n \to \{0,1\}$ (balanced) $\Leftrightarrow$ small circuits cannot sample $f^{-1}(0)$ (uniformly, given random bits)

(lower bound we just saw $\Rightarrow$

extract 1 bit, error < 1, from entropy k=n-1, O(1)-local source)

Outline of talk

- Lower bound for sampling W_k = uniform weight-k string

- Randomness extractors

  – Local sources

  – Bounded-depth circuit (AC^0)

  – Turing machine

Extractors for local functions

- $f : \{0,1\}^* \to \{0,1\}^n$ d-local : each output bit depends on d input

- Theorem[V.] From d-local n-bit source with min-entropy k:
  Let T := k poly(k/nd)
  Extract T bits, error exp(-T)

- E.g. $T = k^c$ from $k = n^{1-c}$, $d = n^c$

- Note: always need k > d

- $d = O(1) \Rightarrow NC^0$ source. Independently [De Watson]

High-level proof

- Theorem d-local n-bit min-entropy k source (T:=k poly(k/nd))

  Is convex combination of bit-block source

  block-size = dn/k, entropy T, error exp(-T)

- Bit-block source with entropy T:

  $(0, 1, X_1, 1- X_5, X_3, X_3, 1- X_2, 0, X_7, 1- X_8, 1, X_1)$

  $X_1, X_2, ..., X_T \in \{0,1\}$

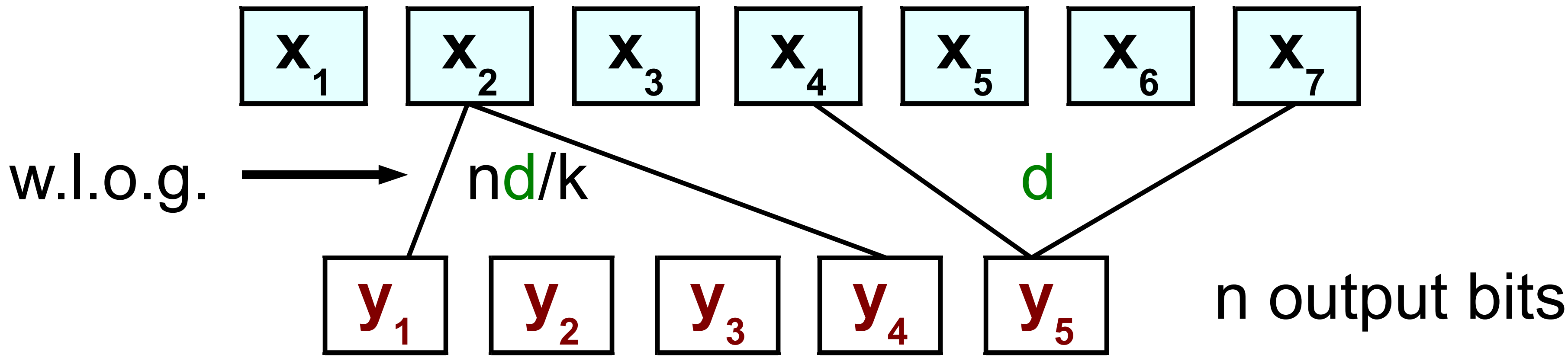
  0 < occurrences of X_i < block-size = dn/k

- Special case of low-weight affine sources

  Use [Rao 09]

Proof

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



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

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

♦

Open problem

- Does previous result hold for decision-tree sources?

- May use isoperimetric inequality for decision trees [O'Donnell Saks Schramm Servedio]

Outline of talk

- Lower bound for sampling W_k = uniform weight-k string

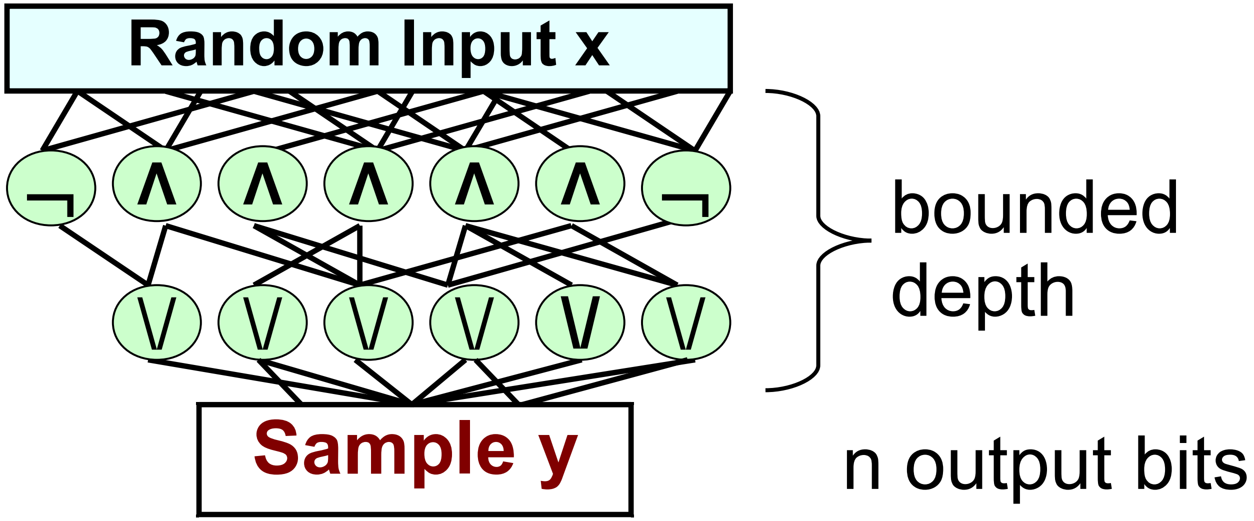
- Randomness extractors

  - Local sources

  - Bounded-depth circuit (AC^0)

  - Turing machine

Bounded-depth circuits (AC^0)



∨ = or
∧ = and
¬ = not

- Theorem [V.]

From AC^0 n-bit source with min-entropy k:

Extract $k \,\mathrm{poly}\left(k / n^{1.001}\right)$ bits, error $1/n^{\omega(1)}$

High-level proof

- Apply random restriction [Furst Saxe Sipser, Ajtai, Yao, Hastad]

- Switching lemma: Circuit collapses to $d=n^{\varepsilon}$-local
  apply previous extractor for local sources

- Problem: fix 1-o(1) input variables, entropy?

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 $\Pr[*] = p$

  With high prob., $f|_R(X)$ has min-entropy pk

- Parameters: $k = \mathrm{poly}(n)$, $p = 1/\sqrt{k}$

  After restriction both circuit collapsed

  and min-entropy $pk = \sqrt{k}$ still poly(n)

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 $\Pr[*] = p$

  With high prob., $f|_R(X)$ has min-entropy pk

- Proof: Builds on [Lovett V]
- Isoperimetric inequality for noise: $\forall A \subseteq \{0,1\}^L$ of density α random m, m' obtained flipping bits w/ probability p :

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

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

Bounded-depth circuits (AC^0)

- Corollary to AC^0 extractor

  Explicit boolean f : AC^0 cannot sample (Y, f(Y))

  f := 1-bit affine extractor for min-entropy $k = n^{0.99}$

- Note: For $k > 1/2$, Inner Product 1-bit affine extractor, and AC^0 can sample (Y, InnerProduct(Y)) [Impagliazzo Naor]

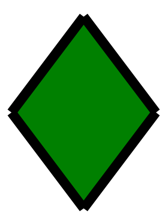
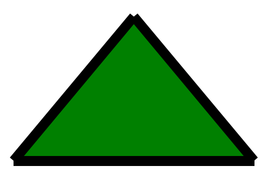
- Explains why affine extractors for $k < 1/2$ more complicated

Open problem

- Theorem[V.] AC^0 can generate (Y, majority(Y)), error $2^{-|Y|}$

- Challenge: error 0?

  - Related [Lovett V.] Does every bijection

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

    have large expected hamming distortion? (n even)

Outline of talk

- Lower bound for sampling W_k = uniform weight-k string

- Randomness extractors

  - Local sources

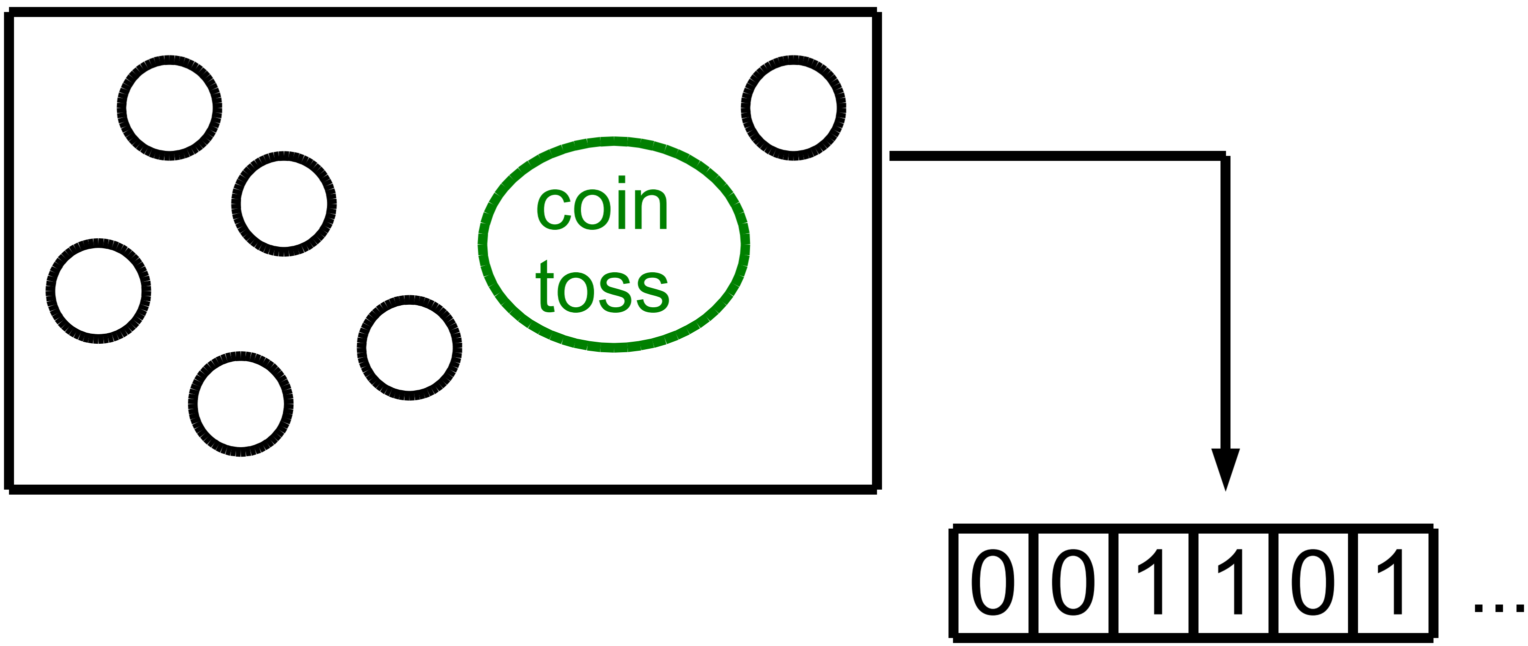
  - Bounded-depth circuit (AC^0)

  - Turing machine

Turing-machine source

- Machines start on blank (all-zero) tape

have "coin-toss" state: writes random bit on tape



- When computation is over, first n bits on tape are sample

Extractors

- Theorem [V.] From Turing-machine n-bit source running in time $\leq n^{1.9}$ and with min-entropy $k \geq n^{0.9}$: Extract $n^{\Omega(1)}$ bits, error $\exp(-n^{\Omega(1)})$

- Proof: Variant of crossing-sequence technique ⇨ TM source = convex combo of independent-block source (no error)
  Use e.g. [Kamp Rao Vadhan Zuckerman]

Extractors

- Corollary [V.] Turing-machine running in time $\leq n^{1.9}$ cannot sample (X, Y, InnerProduct(X,Y)) for $|X| = |Y| = n$

- Proof: As before, but use extractor in [Chor Goldreich]

Summary

- Complexity of distributions = uncharted research direction

- New connections to data structures,
  randomness extractors, and
  various combinatorial problems

- First sampling lower bounds and extractors for
  local,
  decision tree (not in this talk),
  AC^0
  Turing machines