# New sampling lower bounds via the separator

Emanuele Viola

Northeastern University

CCC 2023

# Two goals

- Classical goal of complexity theory: Lower bounds for computing functions:

  Example: Parity not in AC0

- This work: Lower bounds for sampling distributions, given uniform bits

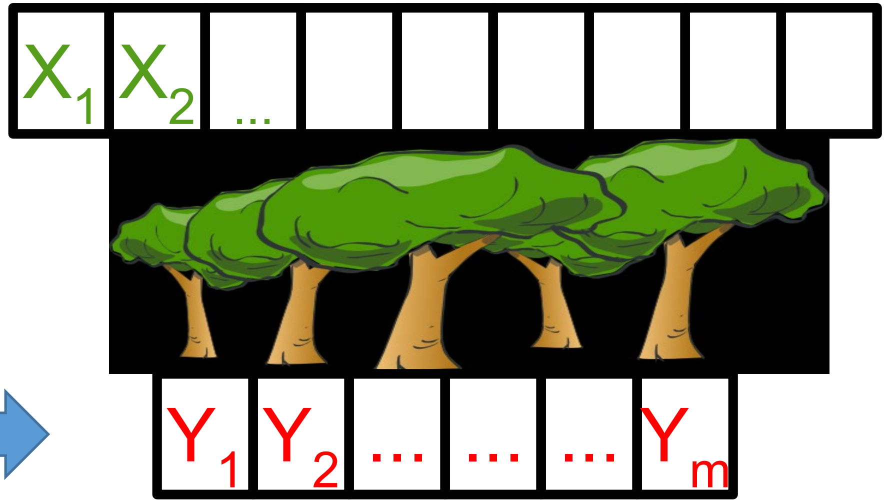
  Line of research spanning 10+ years

  Connections with 1) randomness extractors [will not see]

  2) data structures [will touch on this later]

# The model in this work: Forest

- **Input: Uniform, independent cells of w bits, no restriction on number of cells. Write input as $[W]^L$ for some $L$; $W = 2^w$ .**




- **Output: m cells of w bits**

- **Each output cell computed by a $W$-ary tree of depth q, querying input cells. A.k.a. time-q cell-probe algorithm**

- **Have m distinct trees, one per output cell**

- **Think W = m.  Generalizes boolean decision trees (w = 1)**

# Previous lower bounds for forests

- **Follow from lower bounds for AC0 [V]**

  **Apply to "pseudorandom objects" like extractors, codes**

- **Shortcomings:**

  **Cannot prove separation between AC0 and forest samplers**
  **(cf. known separations for computing)**

  **Do not apply to "simple" distributions**

# Overview of this work

- New lower bounds for sampling by forests
    - Separate AC0 and forest samplers
    - Prove a hierarchy for forest samplers: more depth, more power
    - Apply to “simple” distributions
      Reprove some data-structure lower bounds as corollary
- New tool: The separator:
  Can restrict input so that output of forest is “close to” pair-wise independent

# Outline

- Overview
- Two sampling lower bounds
- The separator

# Lower bound for Rank (a.k.a. prefix sums)

- Definition: Rank(x)=$(x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + x_2 + \cdots + x_m) \in [m]^m$

  Where $x \in \{0,1\}^m$, sum over integers (sum mod 2 easy to sample)

- Theorem: For any depth-q forest sampler f:

  Statistical-Distance( f($[W]^L$), Rank($\{0,1\}^m$) ) > 1 – $2^{-m/w^{O(q)}}$ – Tight

- Note: $[W]^L, \{0,1\}^m$ also denote uniform distribution on those sets

- Distance close to 1 ⇒ lower bounds for succinct data structures

  ⇒ reprove Patrascu-V data-structure lower bound for Rank

- Rank can be sampled by quasi-polynomial AC0. Open: Poly-size AC0

# Lower bound for Predecessor

- Definition: Pred(x) = $y \in \{0,1,\ldots,m\}^m$
  where $y_i = \max\{j \leq i : x_j = 1\}$ is predecessor of $i$, and x $\in \{0,1\}^m$

- Pred(U) easy to sample.
- Consider Pred(H) for distribution H encoding "direct product" predecessor

- Theorem: For any depth-q forest sampler f:
  Statistical-Distance( f($[W]^L$), Pred(H) ) > 1 – $2^{-m/w^{O(q)}}$ — Tight

- Pred(H) can be sampled in poly-size AC0 => separating forest & AC0 samplers

- Also gives sampling hierarchy: depth O(q) samples more than depth q

# Outline

- Overview
- Two sampling lower bounds
- The separator

# The separator

- Theorem: Let $f = (f_1, f_2, \dots, f_m)$ be depth-q forest, S a distribution
  If Statistical-Distance ( f($[W]^L$), S) < 1 - $\epsilon$ then ∃ large $D \subseteq [W]^L$:
  (1) f(D) is suitably close to S
  (2) Most pairs of output words of f(D) are almost independent

- $[W]^L, D, S$ denote sets as well as uniform distributions over them

- Suitably close := Supp($f(D)$) ⊆ Supp($S$) and $H_\infty\left(f(D)\right) \geq H_\infty(S) - c \log 1/\epsilon$

- Key: Number of pairs in (2) compares favorably to entropy loss in (1)

- Distributions suitably close to Rank/Pred do not satisfy (2) ⇒ lower bounds

# The separator

- Theorem: Let $f = (f_1, f_2, \ldots, f_m)$ be depth-q forest, S a distribution
  If Statistical-Distance ( f($[W]^L$), S) < 1 - $\epsilon$ then ∃ large $D \subseteq [W]^L$:

  (1) f(D) is suitably close to S

  (2) Most pairs of output words of f(D) are almost independent

- High-level proof idea:
  If (2) Does not hold
  ⇒ trees $f_i$ intersect queries often
  ⇒ can fix some queries, further restrict $D$, and reduce depth of forest.

  Implementation is somewhat technical. Next some proof highlights.

# The separator

- Theorem: Let $f = (f_1, f_2, \dots, f_m)$ be depth-q forest, S a distribution
  If Statistical-Distance ( f($[W]^L$), S) < 1 - $\epsilon$ then ∃ large $D \subseteq [W]^L$:
  (1) f(D) is suitably close to S
  (2) Most pairs of output words of f(D) are almost independent

- How to get started:
  Particular way in which assumption can be satisfied:
  $f\left([W]^L\right) = S$ with probability $\epsilon$, and $f\left([W]^L\right) = 0$ otherwise

- Lemma: Particular way is general way: ∃ large $D \subseteq [W]^L$ : (1) holds

- Now "forget" S; goal is to restrict D to ensure (2)

# The separator

- Theorem: Let $f = (f_1, f_2, \ldots, f_m)$ be depth-q forest, S a distribution
  If Statistical-Distance ( f($[W]^L$), S) < 1 - $\epsilon$ then ∃ large $D \subseteq [W]^L$:

  (1) f(D) is suitably close to S
  (2) Most pairs of output words of f(D) are almost independent

- How to iterate

- Fixed-Set Lemma [GSV]: ∃ large $D' \subseteq D \subseteq [W]^L$ :
  $D'$ looks like product distribution to small-depth trees

- If $(f_i, f_j)(D')$ not close to independent, by Fixed-Set Lemma
  $f_i, f_j$ intersect queries often ⇒ can restrict $D'$ to reduce forest depth

# Overview of this work

- New lower bounds for sampling by forests
  - Separate AC0 and forest samplers
  - Prove a hierarchy for forest samplers: more depth, more power
  - Apply to “simple” distributions
    Reprove some data-structure lower bounds as corollary
- New tool: The separator:
  Can restrict input so that output of forest is “close to” pair-wise independent

# Open problems

- Sampling lower bounds still uncharted area

- Open: Sample Rank by poly-size AC0

- Open: Sample a uniform permutation by a forest.
  Can you even settle depth 2?

# Thanks!