Fooling Near-Maximal Decision Trees

@ UChicago CS Theory Lunch

January 29, 2025

William M. Hoza

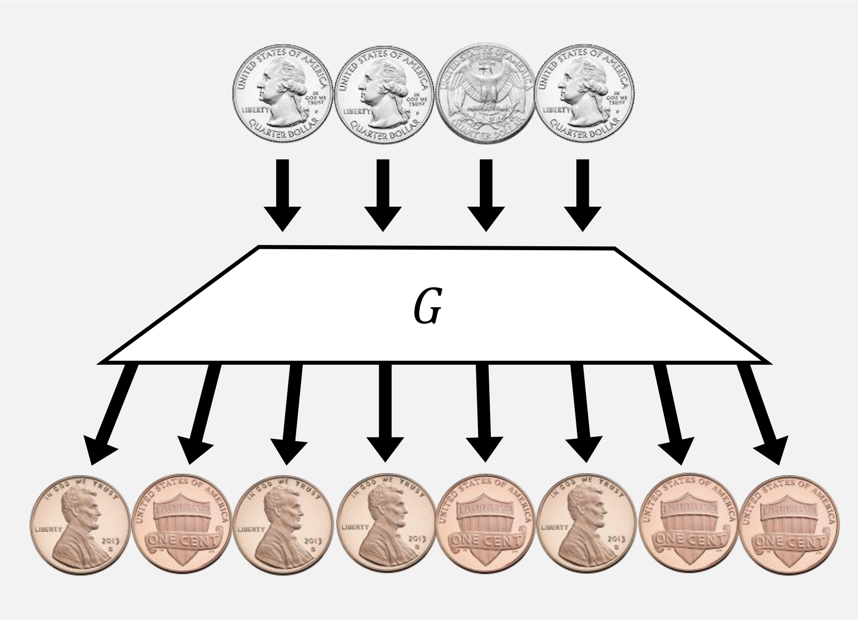
The University of Chicago

Pseudorandom generators

- A pseudorandom generator (PRG) is a function $G: \{0,1\}^s \to \{0,1\}^n$
- The PRG fools $T: \{0,1\}^n \to \{0,1\}$ with error ε if

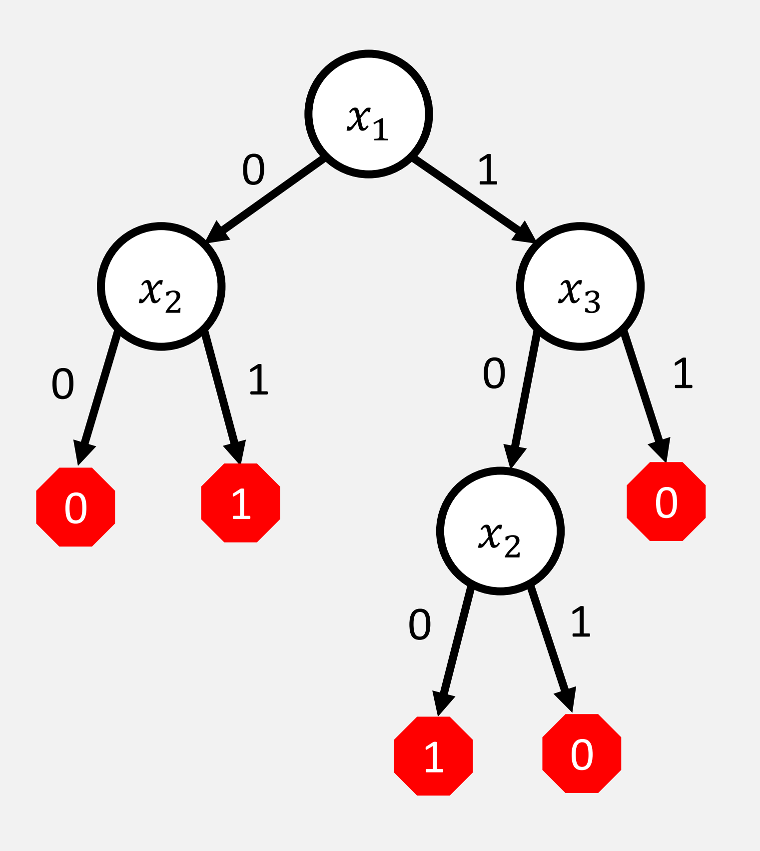
$$\left|\Pr[T(U_n) = 1] - \Pr\left[T\left(G(U_s)\right) = 1\right]\right| \leq \varepsilon$$

- Notation: U_n = uniform distribution over $\{0,1\}^n$
- Today's talk: PRGs that fool decision trees

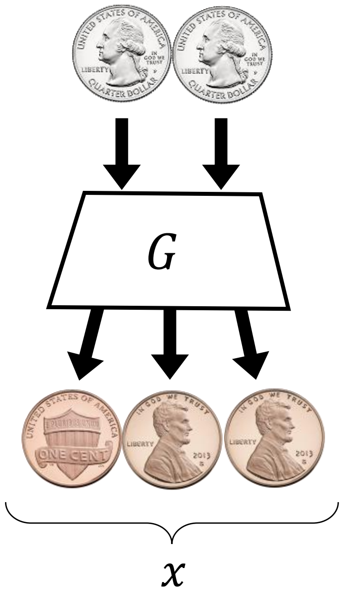


Decision trees

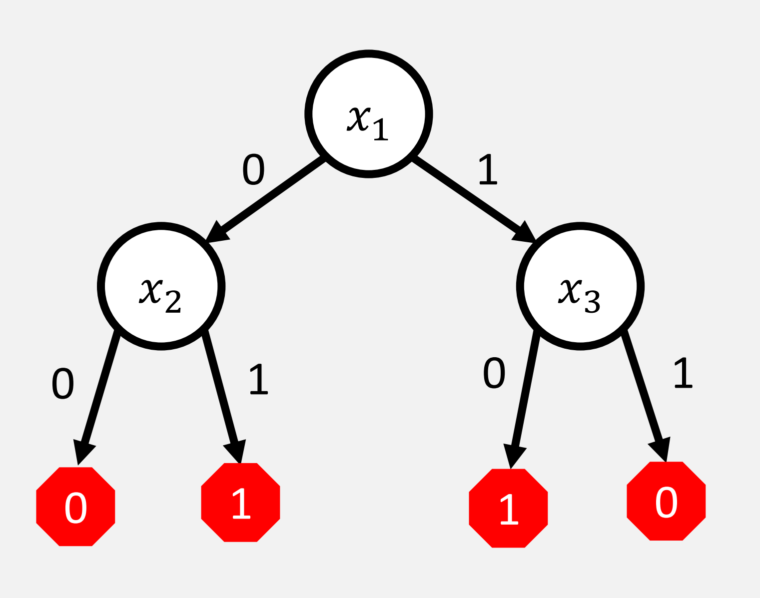
- In each step, the tree may observe any one bit of the input $x \in \{0,1\}^n$
- Eventually, the tree must halt and outputs a value $T(x) \in \{0,1\}$
- Depth = maximum length of path from root to leaf
- Size = number of leaves



Example: Fooling depth-2 decision trees



- **Claim:** There exists a PRG $G: \{0,1\}^2 \to \{0,1\}^3$ that fools depth-2 decision trees with error 0
- **Proof sketch:** Let $G(a, b) = (a, b, a \oplus b)$
- Those three bits are pairwise independent



Fooling depth-k decision trees

Theorem [Naor, Naor 1993] [Alon, Goldreich, Håstad, Peralta 1992] [Kushilevitz, Mansour 1993]:

$\forall n, k, \varepsilon, \exists$ explicit PRG $G: \{0,1\}^s \to \{0,1\}^n$ that fools depth-k decision trees with error ε and seed length $s = 2k + O(\log(k/\varepsilon) + \log\log n)$.

Theorem (this work): Let $\alpha > 0$ be an arbitrarily small constant.

$\forall n, k, \varepsilon, \exists$ explicit PRG $G: \{0,1\}^s \to \{0,1\}^n$ that fools depth-k decision trees with error ε and seed length $s = (1+\alpha) \cdot k + O(\log(1/\varepsilon) + \log\log n)$.

Fooling size-2^k decision trees

Theorem [Naor, Naor 1993] [Alon, Goldreich, Håstad, Peralta 1992] [Kushilevitz, Mansour 1993]:

$\forall n, k, \varepsilon$, $\exists$ explicit PRG $G: \{0,1\}^s \to \{0,1\}^n$ that fools size-2^k decision trees

with error ε and seed length $s = 2k + O(\log(k/\varepsilon) + \log\log n)$.

Theorem (this work): Let $\alpha > 0$ be an arbitrarily small constant.

$\forall n, k, \varepsilon$, $\exists$ explicit PRG $G: \{0,1\}^s \to \{0,1\}^n$ that fools size-2^k decision trees

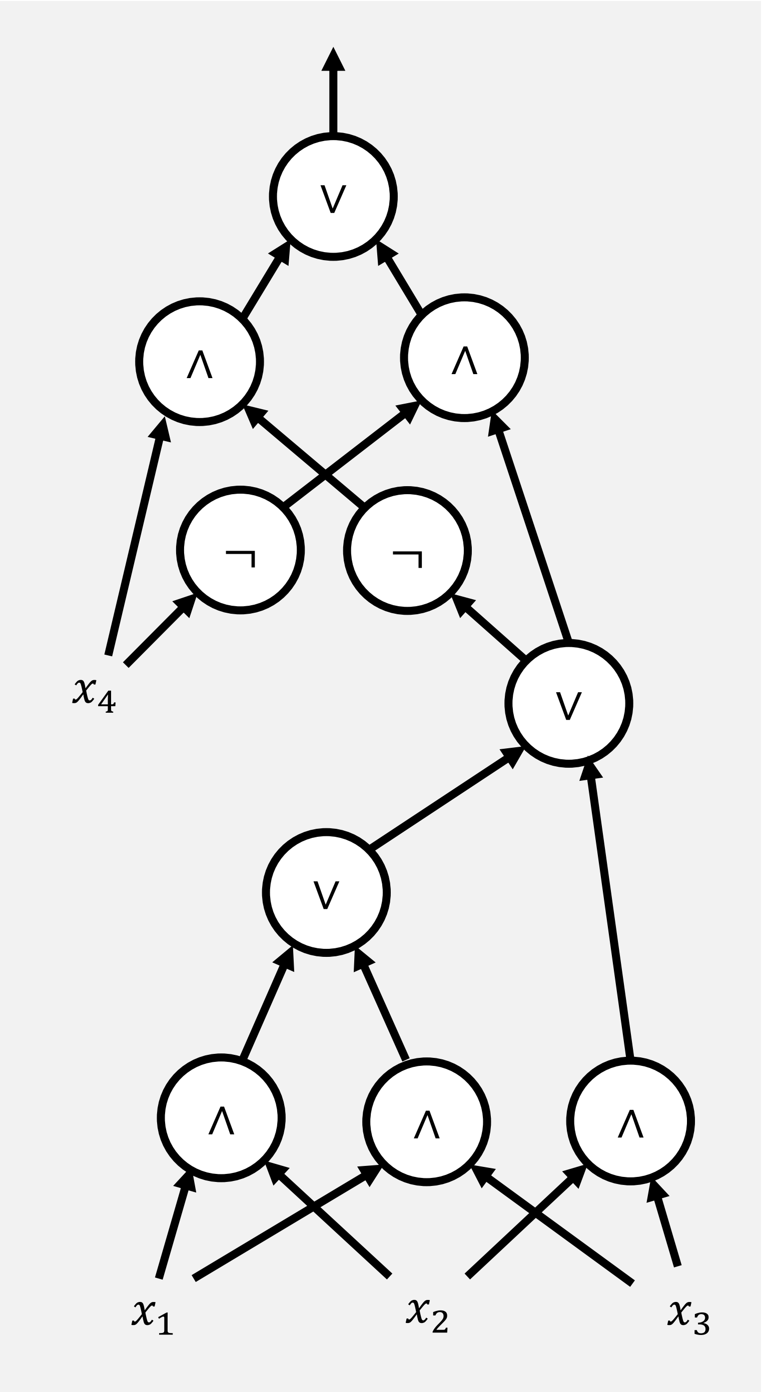
with error ε and seed length $s = (1+\alpha) \cdot k + O(\log(1/\varepsilon) + \log\log n)$.

Why care about this factor of two?

- Answer 1: It’s a fundamental problem
- Answer 2: One can prove a lower bound of $1 \cdot k$
- Answer 3 (main): There is a connection with circuit complexity!

Circuits over the U_2 basis

- **Definition:** A U_2-circuit is a network of AND/OR/NOT gates applied to Boolean variables
- Each AND/OR gate has only two incoming wires
- The size of the circuit is the total number of AND/OR gates
  - NOT gates are not counted



Circuits are poorly understood

- **Theorem** [Shannon 1949]: There exists a function $h: \{0,1\}^n \to \{0,1\}$ such that every U_2-circuit computing h has size $\Omega(2^n/n)$
- What if we want an explicit hard function $h \in \mathrm{NP}$?
  - **Theorem** [Schnorr 1974]: $\exists h \in \mathrm{NP}$ such that every U_2-circuit computing h has size $3n - O(1)$
  - **Theorem** [Zwick 1991]: $\exists h \in \mathrm{NP}$ such that every U_2-circuit computing h has size $4n - O(1)$
  - **Theorem** [Lachish and Raz 2001]: $\exists h \in \mathrm{NP}$ such that every U_2-circuit computing h has size $4.5n - o(n)$
  - **Theorem** [Iwama and Morizumi 2002]: $\exists h \in \mathrm{NP}$ such that every U_2-circuit computing h has size $5n - o(n)$

Our contribution: A PRG that fools U_2-circuits

Theorem (this work): $\forall n \in \mathbb{N}$, $\exists$ explicit PRG $G: \{0,1\}^s \rightarrow \{0,1\}^n$ that fools U_2-circuits of size $2.99 \cdot n$ with error $2^{-\Omega(n)}$ and seed length $\left(1 - \Omega(1)\right) \cdot n$.

Theorem [Chen, Kabanets 2016]: If a function $f: \{0,1\}^n \rightarrow \{0,1\}$ can be computed by a U_2-circuit of size $(3 - \alpha) \cdot n$, then it can also be computed by a decision tree of size 2^k where $k = \left(1 - \Omega(\alpha^2)\right) \cdot n$.

"It would be interesting to get pseudorandom generators for general boolean circuits" [Chen, Kabanets 2016]

How we construct our new PRG

- Our approach for fooling decision trees is based on a new kind of "almost k-wise independence"
- Let's start by reviewing exact k-wise independence

k-wise uniform bits

- Let X be a distribution over $\{0,1\}^n$
- **Definition:** X is k-wise uniform if, for every set $S \subseteq [n]$ with $|S| = k$, the subsequence X_S is distributed uniformly over $\{0,1\}^k$
- A k-wise uniform generator is a function $G: \{0,1\}^s \to \{0,1\}^n$ such that $G(U_s)$ is k-wise uniform

A classic k-wise uniform generator

Theorem [Lancaster 1965, Joffe 1971, Joffe 1974, …]: $\forall n, k$, $\exists$ explicit k-wise uniform generator $G: \{0,1\}^s \to \{0,1\}^n$ with seed length $s = O(k \cdot \log n)$.

- Proof sketch: (Assume WLOG that $n = 2^r \cdot r$ for some $r \in \mathbb{N}$)
  - Use the seed to pick a random polynomial $p: \mathbb{F} \to \mathbb{F}$ of degree less than k, where $\mathbb{F} = \mathrm{GF}(2^r)$
  - Output $p(x)$ for every $x \in \mathbb{F}$
  - This works because of polynomial interpolation

The regime $k = \Theta(n)$

- Any k-wise uniform generator fools depth-k decision trees with error 0
- **Theorem (good news)** [Cheng, Li 2021]: $\forall n, k, \exists$ explicit k-wise uniform generator $G: \{0,1\}^s \to \{0,1\}^n$ with seed length $s = O(k \cdot \log(n/k))$.
- **Theorem (bad news)** [Karloff, Mansour 1997]: If $k \geq \big(1/2 + \Omega(1)\big) \cdot n$, then every k-wise uniform generator has seed length at least $n - O(1)$.

Almost k-wise uniformity

- Let X be a distribution over $\{0,1\}^n$
- **Definition**: X is ε-almost k-wise uniform if, for every function $f:\{0,1\}^n \to \{0,1\}$ that depends on at most k variables, we have

$$|\Pr[f(X) = 1] - \Pr[f(U_n) = 1]| \leq \varepsilon$$

- **Equivalent:** For every set $S \subseteq [n]$ with $|S| = k$, the subsequence X_S is uniform to within total variation distance ε

Almost k-wise uniformity

- **The good news:** There are constructions of ε-almost k-wise uniform generators with seed length $k + O(\log(k/\varepsilon) + \log\log n)$

  [Alon, Goldreich, Håstad, Peralta 1992]

- **The bad news:** The condition of being ε-almost k-wise uniform is weaker than that of fooling depth-k decision trees, because depth-k decision trees can be adaptive

Key new concept: k-wise probable uniformity

- Let X be a distribution over $\{0,1\}^n$
- **Definition (new)**: X is k-wise ε-probably uniform if, for every function $f: \{0,1\}^n \to \{0,1\}$ that depends on at most k variables, we have

$$\Pr[f(X) = 1] \geq (1 - \varepsilon) \cdot \Pr[f(U_n) = 1]$$

- **Equivalent**: For every set $S \subseteq [n]$ with $|S| = k$, the subsequence X_S has a mixture distribution: sample from U_k with probability $1 - \varepsilon$, and sample from some other distribution with probability ε

"Probably uniform"

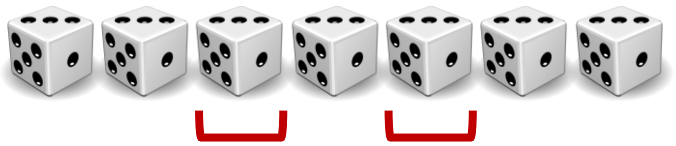
Main technical contribution

Theorem (this work): $\forall n, k, \varepsilon$, $\exists$ explicit k-wise ε-probably uniform generator $G: \{0,1\}^s \to \{0,1\}^n$ with seed length

$$s = k + O\left(k^{2/3} \cdot \log^{1/3}(k/\varepsilon) + \log(1/\varepsilon) + \log\log n\right).$$

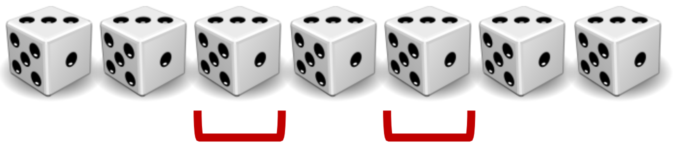
- Today's talk: A couple of elements of the construction

Pairwise uniform hash functions



- Let $\mathcal{H}$ be a family of hash functions $h: \{0,1\}^s \to \{0,1\}^n$
- **Definition:** $\mathcal{H}$ is pairwise uniform (aka "strongly universal") if, for every pair of distinct $x_1, x_2 \in \{0,1\}^s$, when we sample $h \sim \mathcal{H}$, the pair $\left(h(x_1), h(x_2)\right)$ is distributed uniformly over $\{0,1\}^{2n}$
- **Fact:** $\forall s, n, \exists$ explicit pairwise uniform family $\mathcal{H}$ such that sampling $h \sim \mathcal{H}$ costs $O(s+n)$ truly random bits

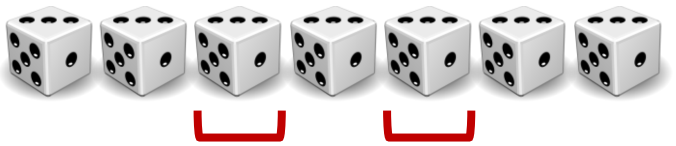
A sampling lemma



- Let $\mathcal{H}$ be a pairwise uniform family of hash functions $h: \{0,1\}^s \to \{0,1\}^n$
- Let $f: \{0,1\}^n \to \{0,1\}$ and let $\mu = \mathbb{E}[f(U_n)]$
- Think of a single h in $\mathcal{H}$ as a PRG that we can use to try to fool f
- **Lemma** (standard): If we sample $h \sim \mathcal{H}$, then for any $\varepsilon \in (0,1)$,

$$\Pr_h[h \text{ fools } f \text{ with error } \varepsilon \cdot \mu] \geq 1 - \frac{1}{2^s \cdot \varepsilon^2 \cdot \mu}.$$

Proof of sampling lemma



- For each fixed $x \in \{0,1\}^s$, define a random variable $Z_x = f\big(h(x)\big)$
- Then $\mathbb{E}[Z_x] = \mu$ and $\mathrm{Var}[Z_x] = \mu \cdot (1-\mu) \le \mu$
- Let $Z = \sum_x Z_x$
- Then $\mathbb{E}[Z] = \mu \cdot 2^s$ and $\mathrm{Var}[Z] \le \mu \cdot 2^s$ by pairwise independence
- Now apply Chebyshev's inequality:

$$\Pr\left[\left|\frac{Z}{2^s} - \mu\right| > \varepsilon \cdot \mu\right] \le \frac{\mathrm{Var}[Z]}{(2^s \cdot \varepsilon \cdot \mu)^2} \le \frac{1}{2^s \cdot \varepsilon^2 \cdot \mu}.$$

We can tolerate "bad events"

- Our k-wise probably uniform generator involves sampling a hash function $h \sim \mathcal{H}$ and then using it several times (see paper for more)
- There is some "bad event" B where $\Pr[B] \approx \varepsilon$
- This is okay: $\mathbb{E}[f(X)] \geq \Pr[\neg B] \cdot \mathbb{E}[\, f(X) \mid \neg B\,] \geq (1-\varepsilon) \cdot \mathbb{E}[f]$
- Crucially, we do not claim $\mathbb{E}[f(X)] \leq (1+\varepsilon) \cdot \mathbb{E}[f]$!

Fooling decision trees

- **Claim:** If X is k-wise ε-probably uniform, then X fools depth-k decision trees with error ε
- **Proof:** Let A be the set of accepting leaves in a depth-k decision tree T
- For each leaf $u \in A$, let $T_u(x)$ indicate whether $T(x)$ reaches u

$$\mathbb{E}[T(X)] = \sum_{u \in A} \mathbb{E}[T_u(X)] \geq \sum_{u \in A} (1-\varepsilon) \cdot \mathbb{E}[T_u(U_n)] \geq \mathbb{E}[T(U_n)] - \varepsilon$$

- $\mathbb{E}[T(X)] \leq \mathbb{E}[T(U_n)] + \varepsilon$, because $1 - T$ is another depth-k decision tree

Conclusions

- We construct PRGs fooling near-maximal decision trees and U_2-circuits of size $2.99 \cdot n$
- The construction is based on a new kind of almost k-wise independence, called k-wise probable uniformity
- Open problem: Find more applications of k-wise probable uniformity
- **Thank you!**