Lecture 7 : Improved δ-almost k-wise independence and K-universal set

Topics in Pseudo-randomness and Complexity Theory (Spring 2018)
Rutgers University
Swastik Kopparty
Scribe: Vishwas Bhargava, Vibha Sridhar

1 Review:

In last lecture we saw an explicit construction of ϵ-biased sets and used them generate δ-almost k-wise independence. Specifically, we proved the following lemma. Specifically, we proved the following lemma.

Lemma 1. *We can obtain a δ-almost k-wise independent distribution which we can sample with $2\log n + \log\frac{1}{\delta} + k$ random bits.*

This followed simply by taking μ to be an ϵ-biased distribution on $\mathbb{F}_2^n$ and sampling elements from it. A natural question arises from this: Can this be done using less randomness?

2 Improved δ-almost k-wise independence

We have δ-almost k-wise independence using $\mathcal{O}(\log n + \log\frac{1}{\delta} + k)$ random bits. We want to generate a distribution δ-close to uniform on $\{0,1\}^n$ How many random-bits?

Recall, if X,Y are random variables on $\{0,1\}^n$ such that X,Y are δ close in statistical distance, that is $\Delta(X,Y) \leq \delta$. Then for *any* $f : \{0,1\}^n \longrightarrow \{0,1\}$, $\Delta(f(X), f(Y)) \leq \delta$.

Since, in proof of *Lem.* 1 we didn't use the full power of ϵ-biased sets, perhaps we can have δ-almost k-wise independence with much less randomness. Let's follow the most natural plan:

Find a distribution Y that is (0.01) close to U_n =uniform on $\{0,1\}^n$. then, $\Delta(f(U_n), f(Y)) \leq 0.01$. Hopefully, Y needs fewer bits to sample.

Recall, any distribution with support size $\geq 2^{n-1}$ has $\Delta(Y, U_n) \leq \frac{1}{2}$. So, if $\delta < \frac{1}{2}$, we need n random-bits to be δ-close to uniform.

So, we cant improve k. But we can improve $O(\log n)$ to $O(\log\log n)$.

Plan : To produce a $(X_1, X_2 \ldots X_n)$ distribution s.t. $\forall T \subseteq [N]$ $1 \leq |T| \leq k$

$$Pr\big[\oplus_{i\in T} X_i = 1\big] \in \Big(\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\Big) \tag{1}$$

Property 1 inherits the notion of being ϵ-biased for small sets(of size k).

In order to construct such a distribution we will need a matrix $\mathbf{M} \in \mathbb{F}_2^{k \log n \times n}$, such that every k column of $\mathbf{M}$ are linearly independent. That is,

$$\mathbf{M} = \begin{bmatrix} | & \cdots & | \\ v_1 & \ddots & v_n \\ | & \cdots & | \end{bmatrix}$$

where $v_i \in \{0,1\}^{k \log n}$ and every k column are linearly independent. We will defer explicit construction of $\mathbf{M}$ to next subsection.

Now, consider an ϵ-biased distribution $\mathbf{Z}$ on $\{0,1\}^{k \log n}$

Output: $\mathbf{ZM} \in \{0,1\}^n$.

Claim 2. *The distribution* $\boldsymbol{ZM}$ *is δ-almost k-wise independence.*

Proof. Let $(X_1, X_2 \dots X_n) = \mathbf{ZM}$. Take $T \subseteq [n]$, $1 \leq |T| \leq k$. We need to study,

$$\bigoplus_{i \in T} X_i = \bigoplus_{i \in T} \langle v_i, z \rangle,$$

where $v_T \neq 0$. Notice that, $Pr\big[\bigoplus_{i \in T} \langle v_i, z \rangle = 1\big] \in (\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2})$ by ϵ-biasedness of $\mathbf{z}$.

Property 1 $\implies$ $(X_1 \dots X_n)$ is $(\epsilon 2^{-k/2})$-almost k wise independent so, set $\epsilon = \delta 2^{-k/2}$)

$\square$

We need to sample a $\mathbf{z} \in \{0,1\}^{k \log n}$ which is ϵ-biased. This needs $O(\log k \log n + \log \frac{1}{\epsilon})$ bits. Therefore, the total amount of randomness used is $O(\log \log n + \log \frac{1}{\epsilon} + k)$.

Construction of M

Let $m = \log n$. Recall that, $\phi : \mathbb{F}_{2^m} \longrightarrow \mathbb{F}_2^m$ is a linear bijection, as discussed in previous lecture.

$$\mathbf{M} = \overbrace{\begin{bmatrix} \phi(1) & \cdots & \cdots & \phi(1) \\ \phi(\alpha_1) & \cdots & \cdots & \phi(\alpha_n) \\ \vdots & \vdots & \vdots & \vdots \\ \phi(\alpha_1^{k-1}) & \cdots & \cdots & \phi(\alpha_n^{k-1}) \end{bmatrix}}^{\text{elements of } \mathbb{F}_{2^m}}$$

Here $\alpha_1 \dots \alpha_n$ are elements of $\mathbb{F}_{2^m}$. Since, $\phi(\alpha_i) \in \mathbb{F}_2^m$, $\phi(\alpha_i)$ represents a block of size $\log n$. Thus $\mathbf{M}$ is a $\{0,1\}$ matrix.

3 K-Universal Set

Consider a subset of $\{0,1\}^n$ such that for any $S \subseteq [n]$, $|S| \leq k$ and any $y \in \{0,1\}^S$, there is some $x \in X$ such that $x|_S = y$.

Observation: Support of a $\frac{1}{2^{2k}}$ - almost k-wise independent distribution is k-universal.

There are explicit such distributions that can be sampled with $O(k + log(logn))$ random bits. Size of support is now $2^{O(k)}$ poly-$\log n$ versus $O(2^k \log n)$ for non explicit distributions.

3.1 Designing Tests

F is a collection of functions from $\{0,1\}^n \rightarrow \{0,1\}$.

Definition 3. *A distribution $(x_1, x_2, ...x_n)$ ϵ - Fools F, if $f \in F$ such that*

$$|Pr[f(x_1, ...x_n) = 1] - Pr[f(U_n) = 1]| \leq \epsilon.$$

Examples:

1. F = all functions such that $(x_1, ...x_n)$ is any distribution ϵ-close to U_n.
2. F = linear function ($f(x) = \langle, x\rangle$) such that $(x_1, ...x_n)$ is ϵ-biased.
3. F = k-juntas, $(x_1, ...x_n)$ is ϵ-almost k-wise independent.

   Definition 4. *A k-junta is a $f : \{0,1\}^n \rightarrow \{0,1\}$ such that f is dependent on only k variables.*

   Even though a k-junta takes in 'n' input values, the function depends only on some k variables.

Consider the Probability masses p, q where D $\rightarrow$ [0,1] and $f : D \rightarrow R$. For Δ(p,q) :

$$\begin{aligned}
\Delta(f(p), f(q)) &= \sum_{r \in R} \big|\text{Pr[r under p] - Pr[r under q]}\big| \\
&= \sum_{r} \big| \sum_{d \in f^{-1}(r)} \big|p(d) - q(d)\big| \\
&\leq \sum_{r, \sum d \in f^{-1}(r)} \big|p(d) - q(d)\big| \qquad \text{(by triangle inequality)} \\
&= \Delta(p, q)
\end{aligned}$$

Here, f^{-1} corresponds to the inverse of f.

So, $\Delta(p, q) = \max_{s \subseteq D}$[p(s) - q(s)].

Theorem 5. *Let F be a collection of functions. Then $\exists$ a multiset $S \subseteq \{0,1\}^n$,*

$$| \ S \ | \leq \frac{log(F)}{\epsilon}, \textit{ such that}$$

the uniform distribution over S, ϵ-fools F.

Proof. Take $x_1, \ldots, x_t \in \{0,1\}^m$ uniformly at random. Fix $f \in F$. The probability that $\underset{x_1,\ldots x_t}{\mathrm{Pr}}[| \ \frac{1}{t} \sum_{i=1}^{t} f(x_i)$ - $\underset{y \in \{0,1\}^n}{\mathbb{E}}[[] \ \mathrm{f(y)}] \ | \geq \epsilon \ ] \leq \mathrm{e}^{-\omega(\epsilon^2 t)}$

Here, $f(x_i)$ measures the probability of seeing 1 when you feed in a uniform string from x_1,...x_t, if the distribution was uniform distribution of $x_1, \ldots, x_t$.

For a fixed f the probability that $x_1, \ldots, x_t$ doesn't work is really small.

Taking Union Bound over all F $\in$ F.

$$Pr_{x_1,\ldots,x_t}[\exists F \in F \text{ such that } x_1, \ldots, x_t \text{ is bad for } F] \leq |F| e^{-\omega(\epsilon^2 t)}$$

So, $\exists$ $x_1, \ldots, x_t$ such that for all $f \leftarrow F$, $x_1, \ldots, x_t$ is not bad for f.

□

Example: F = all functions computable by a Boolean circuit of size n^{10}.

Here, n is the number of input, so there will be n^{10} gates, with each gate choosing which source it comes from.

$F \leq (\mathrm{n}^{10^{2^{n^{10}}}}) \leq 2^{n^{11}}$

So, $\exists$ S of size $O(\mathrm{n}^{11})$ such that uniform distribution over S, (0.1)-fools F.

If we could deterministically enumerate S on poly(n) time, then we can derandomize any randomized algorithm that runs in time $O\mathrm{n}^{10}$ and uses n random bits.

For specific F of certain sizes, there are lower bounds.

4 Low Space Algorithms

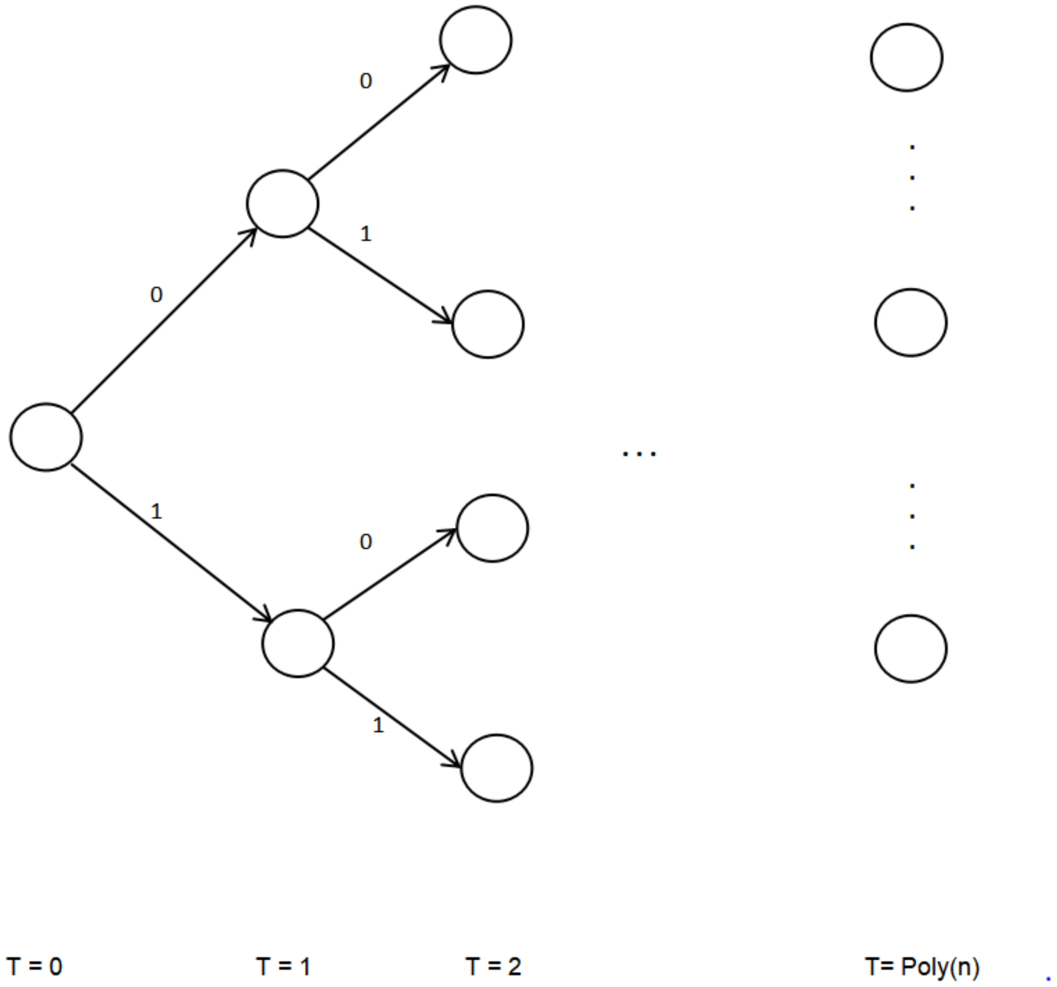
A Low Space Algorithm uses at most log n of it's own space.

1. L - denotes a function that can be computed by deterministic log space algorithm.
2. BPL - denotes a function that can be computed by a randomized algorithm using $O(\log \mathrm{n})$ space and poly(n) time.

4.1 Layered Branching Programming

Definition 6. *Layered Branching Programs are branching programs where the nodes are partitioned into a number of layers, and edges go only from nodes in one layer to nodes in the next. The start node is in the first layer and the sink nodes in the last.*

There are at most Poly(n) states per layer.



Goal: Produce distribution $(x_1, ...x_n)$ such that $(x_1, ...x_n)$ ϵ-fools read-once branching program that reads x_1, x_2,...x_n in that order.

5 Randomness Extractors

Definition 7. *A randomness extractor, is a function, which being applied to output from a weakly random entropy source, together with a short, uniformly random seed, generates a highly random output that appears independent from the source and uniformly distributed i.e, it takes in some pretty random bits and outputs very random bits.*

Example Application : To sample uniformly random n-bit prime.

The two approaches discussed in class are :

Approach 1 :

1. Pick a uniform random number
2. Check if it is prime. If so, output it.
3. Else, repeat.

Approach 2 : This approach is more efficient and reduces the amount of randomization used.

1. Sample an n-bit integer X
2. If X is prime, output X. Done.
3. Else, Extract randomness from X.
4. Repeat using this randomness and possibly more randomness.

Definition 8. *A deterministic 1-bit extractor from a k-bit source is a function* $f : \{0,1\}^n \rightarrow \{0,1\}$ *such that for any distribution* X *which is uniform over a set of size* 2^k, $f(X)$ *is (0.1)-close to uniform on* $\{0,1\}$.