CS 6815: Lecture 12

Instructor: Eshan Chattopadhyay Scribe: Juan C. Martínez Mori

October 4, 2018

1 Randomness Extractors

Definition 1.1 (Min-entropy)**.** *The min-entropy of a random variable X is defined as*

$$H_\infty(X) = \min_{x \in \sup(X)} \left\{ \log \left(\frac{1}{\Pr[X = x]} \right) \right\}.$$

Definition 1.2 ((n, k)-sources)**.** *A random variable X is a (n, k)-source if X is distributed on $\{0, 1\}^n$ and $H_\infty(X) \geq k$.*

2 Convex Combinations of Distributions

Let $\mathcal{X}$ be a family of distributions, each on $\{0, 1\}^n$.

Definition 2.1 (Mixture distributions)**.** *Let D be a distribution on $\{0, 1\}^n$. Then, D can be expressed as a convex combination of distributions in $\mathcal{X}$ if there exists an integer $t > 0$, $\lambda_1, \cdots, \lambda_t \in \mathbb{R}^{\geq 0}$, and $X_1, \cdots, X_t \in \mathcal{X}$ satisfying $\sum_{i=1}^t \lambda_i = 1$ and $D = \sum_{i=1}^t \lambda_i X_i$. In turns, this means that for all $y \in \{0, 1\}^n$, $\Pr[D = y] = \sum_{i=1}^t \lambda_i \cdot \Pr[X_i = y]$.*

Definition 2.2 (Flat distributions)**.** *D is a flat distribution if there exists $S \subseteq \{0, 1\}^n$ such that D is uniform on S.*

Fact 2.3. *Any (n, k)-source X is a convex combination of flat sources, each with support size 2^k. That is, each with min-entropy k, since each probability is upper bounded by 2^{-k}.*

Note that in the case of (n, k)-sources, the flat sources form a convex polytope with $\binom{2^n}{2^k}$ vertices.

3 Seeded Extractors

The intuition is as follows. We take a (n, k)-source, which by definition is a distribution on $\{0, 1\}^n$ with min-entropy k, together a uniformly distributed seed in $\{0, 1\}^d$, to obtain a uniform distribution on $\{0, 1\}^m$. We want m to be as close to $d + k$ as possible. In other words, an extractor Ext gets $x \in X$, which is a (n, k)-source, and $y \in U_d$, which is an uniformly distributed seed, to produce $\text{Ext}(x, y) = z \in \{0, 1\}^m$.

Fact 3.1. *Let D_1, D_2 be distributions on $\{0, 1\}^n$. Then,*

$$|D_1 - D_2| = \frac{1}{2} \|D_1 - D_2\| = \max_{S \subseteq \{0,1\}^n} |\Pr[D_1 \in S] - \Pr[D_2 \in S]|.$$

Definition 3.2 (Seeded Extractors)**.** *A function Ext* : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a seeded extractor if for all* (n,k)*-sources* X*, we have*

$$|Ext(X, U_d) - U_m| \leq \epsilon.$$

Definition 3.3 (Short Seeded Extractors)**.** *A function Ext*: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a short seeded extractor if for all* (n,k)*-sources* X*, wre have*

$$|(Ext(X, U_d), U_d) - (U_m, U_d)| \leq \epsilon.$$

Lemma 3.4 (Proposition 6.14, Vadhan S., Pseudorandomness)**.** *Seeded extractors exist.*

Proof. This proof uses the Probabilistic Method on a randomly chosen extractor. Recall Ext : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$. We will use the notation $N = 2^n$, $D = 2^d$, $M = 2^m$, and $K = 2^k$. Let X be a flat (n,k)-source, that is, with support size $K = 2^k$. Let $T \subseteq \{0,1\}^m$ be arbitrary. We want to show that for all T we have

$$\left| \Pr[\text{Ext}(X, U_d) \in T] - \frac{|T|}{M} \right| \not> \epsilon,$$

where $\frac{|T|}{M} = \Pr[U_m \in T]$. Note that there are $K \cdot D$ random strings in $\{0,1\}^m$. Let

$$\mathbf{1}_{x,y} = \begin{cases} 1, & \text{if } \text{Ext}(x,y) \in T \\ 0, & \text{o.w.} \end{cases}$$

For each of the K points $x \in \sup(X)$ and each of the D strings $y \in \{0,1\}^d$, we have $\Pr[\text{Ext}(x,y) \in T] = \frac{|T|}{M}$, and these events are independent. Then, for a fixed T and a fixed flat source X,

$$\begin{aligned} \Pr[\text{Ext}(X, U_d) \in T] &= \frac{1}{K \cdot D} \sum_{\substack{x \in \sup(X) \\ y \in \{0,1\}^d}} \mathbf{1}_{x,y} \\ &\leq \exp\left(-\frac{-\epsilon^2}{4} K \cdot D \right), \end{aligned}$$

where the inequality follows from Chernoff's Bound. Now, note that there are $\binom{N}{K}$ possible flat sources, and that there are 2^M possible tests. Then, the probability that the condition is violated for at least one T for at least one flat source is

$$\leq 2^M \binom{N}{K} \exp\left(-\frac{-\epsilon^2}{4} K \cdot D \right).$$

One can verify that this bound on the probability of the extractor failing is less than one for $m = k + d - 2\log(1/\epsilon) - O(1)$ and $d \geq \log(n-k) + 2\log(1/\epsilon) + O(1)$. □

4 Extractors for Hash Functions

Lemma 4.1 (Leftover Hash Lemma)**.** *Let* $\mathcal{H}$ *be a cardinality* N *family of hash functions* h : $\{0,1\}^n \to \{0,1\}^m$ *satisfying*

$$\Pr_{h \sim \mathcal{H}}[h(x_1) = h(x_2)] \leq \frac{1}{M}.$$

for all $x_1 \neq x_2 \in \{0,1\}^n$*,* $M = 2^m$*. Then, for any* $0 \leq l \leq n/2$*,* $Ext(x,h) = h(x)$ *is a strong-seeded extractor for min-entropy at least* $n - l$ *with output length* $m = n - 2l$ *and error* $2^{-l/2}$*.*

Proof. Let X be a (n,k)-source and H be chosen uniformly at random from $\mathcal{H}$. Let $\text{Ext}(X,H) = H(X)$, with seed length $n = \log N$, $m = n - 2l$, and $\epsilon = 2^{-l/2}$.

We are interested in $|(H, H(X)) - (H, U_m)| \leq \epsilon$. Note that we can bound the collision probability, which we denote by C_P, by

$$\begin{aligned} C_P(H, H(X)) &= \frac{1}{N} \Pr_{\substack{h \sim \mathcal{H} \\ x_1, x_2 \sim X}} [h(x_1) = h(x_2)] \\ &\leq \frac{1}{N}\left(\frac{1}{K} + \frac{1}{M}\right) \\ &= \frac{1 + (M/K)}{NM}. \end{aligned}$$

Claim 4.2. *Let D be a distribution on a set T. Suppose $C_P(D) \leq \frac{1+4\epsilon^3}{|T|}$. Then, $|D - U_t| \leq \epsilon$.*

Sketch. Take $[n] = T$. Then, $C_P(D) = \sum_i D_i^2 = \|D\|_2^2$. We have

$$\begin{aligned} |D - U_{[n]}| &= \frac{1}{2}\left\|D - U_{[n]}\right\| \\ &\leq \frac{1}{2}\sqrt{n}\left\|D - U_{[n]}\right\| \\ &= \frac{1}{2}\sqrt{n}\left(\|D\|_2 - \frac{1}{2}\right)^{1/2}, \end{aligned}$$

and so on. □

Lastly, given the claim above we find

$$\begin{aligned} |(H, H(x) - (H, U_m)| &\leq \sqrt{\frac{M}{4K}} \\ &= 2^{-(k-m)/2}. \end{aligned}$$

For all l, take $m = n - 2l$, $k = n - l$ so long as $k > n/2$. Ultimately we have $\epsilon = 2^{-l/2}$. □

5 Extractors from Codes

Let $\mathcal{C} : [\bar{n}, n, (1-\delta)\bar{n}]_q$, with encoder $C : \{0,1\}^n \to \{0,1\}^{\bar{n}}$. Define $\text{Ext}(x,y) = C(x)_{|y}$, that is $\text{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where $d = \log(\bar{n})$ and $m = \log q$. Denote the collision probability by C_P. We are interested in $Y, C(x)_{|y} \approx Y, U_m$.

$$\begin{aligned} C_P\left(Y, C(X)_{|Y}\right) &= \frac{1}{\bar{n}} \Pr_{\substack{y \sim U_d \\ x_1, x_2 \sim X}} \left[C(x_1)_{|y} = C(x_2)_{|y}\right] \\ &= \frac{1}{\bar{n}}\left(\frac{1}{k} + \Pr_{\substack{y \sim U_d \\ x_1 \neq x_2 \sim X}} \left[C(x_1)_{|y} = C(x_2)_{|y}\right]\right) \\ &\leq \frac{1}{\bar{n}}\left(\frac{1}{k} + \delta\right) \\ &= \frac{1}{\bar{n}q}\left(\frac{q}{K} + \delta q\right). \end{aligned}$$

This is to be continued in the **next lecture**.