Lecture 20: Extractors (Small-bias Masking)

Randomness Extraction: Motivation

- Randomized algorithms that we use assume that we are provided access to uniform independent random bits
- But this is a very stringent requirement
- Can we use sources of randomness that provide only a weak guarantee?

Min-Entropy Source

Definition (Min-Entropy Source)

A distribution X over the sample space $\{0,1\}^n$ is said to have min-entropy k, if for all $x \in \{0,1\}^n$, we have $\mathbb{P}[X = x] \leqslant 2^{-k}$. The distribution X is also referred to as a source with min-entropy k or (n,k)-source.

- Intuitively, a source with min-entropy k has the probability of sampling every element $\leqslant 2^{-k}$
- A source with high min-entropy has (exponentially) low probability of sampling each element in the sample space
- Intuitively, it suffices to think of a k-source to be a uniform distribution over a subset $S \subseteq \{0,1\}^n$ such that $|S| = 2^k$

Extractor

- An extractor is a function that takes as input (1) a sample from a weak randomness source, and (2) a small seed (uniform random bits)
- The extractor outputs a (large) number of uniform random bits

Definition (Randomness Extractor)

A function $\mathrm{Ext}\colon \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^n$ is an (n,ℓ,k,ε)-extractor if it takes as input (1) a sample of from an (n,k)-source, and (2) a uniform random seed from $\{0,1\}^\ell$, and it outputs m bits that are ε-close to the uniform distribution over $\{0,1\}^n$.

Impossibility of Deterministic Extraction

- A deterministic extractor is a function $\mathrm{Ext}\colon \{0,1\}^n \to \{0,1\}^m$ such that, for all sources X with min-entropy k, the distribution $f(X)$ is close to the uniform distribution $\mathbb{U}_{\{0,1\}}$ (note, we have $\ell = 0$)
- Note that we want to design one function Ext that works for all min-entropy sources
- We will show that deterministic extraction is impossible even when we want to extract $m = 1$ bit from $k = (n-1)$
  - Let $S \subseteq \{0,1\}^n$ be the larger of the two sets $\mathrm{Ext}^{-1}(0)$ and $\mathrm{Ext}^{-1}(1)$
  - Note that $|S| \geqslant 2^{n-1}$
  - Let X be the distribution $\mathbb{U}_S$
  - Note that X has min-entropy $(n-1)$ and $\mathrm{Ext}(X)$ is constant
- So, we construct randomized algorithms Ext that take a small uniformly random bit-string seed as input and output a long random bit-string, i.e. m is larger than ℓ

Bias of a Distribution

Definition (Bias of a Distribution)

We say that a distribution f has ε-bias, if $N\widehat{f}(S) \leqslant \varepsilon$, for all $\emptyset \neq S \in \{0,1\}^n$.

Theorem

Let X be a k-source and M be an ε-bias distribution. Then, we have

$$2\mathrm{SD}\left(X \oplus M, \mathbb{U}_{\{0,1\}^n}\right) \leqslant \varepsilon\sqrt{\frac{N}{K}}$$

- Recall (k-source): If X is a k-source, then

$$\sum_{S\in\{0,1\}^n} \widehat{X}(S)^2 \leqslant \frac{1}{NK}$$

- By definition (ε-Bias): If M is an ε-bias distribution, then

$$\widehat{M}(S) \leqslant \frac{\varepsilon}{N},$$

for all $\emptyset \neq S \in \{0,1\}^n$

Extraction via Small-bias Masking II

- Recall (Bound on SD): We have proven that

$$2\mathrm{SD}\left(f, \mathbb{U}_{\{0,1\}^n}\right) \leqslant N \left(\sum_{S \neq \emptyset} \widehat{f}(S)^2\right)^{1/2}$$

- Recall (Def. of Conv.): By definition of convolution, we have

$$(\widehat{X \oplus M})(S) = N\widehat{X}(S)\widehat{M}(S)$$

- We will use all these properties below

$$\begin{aligned}
2\mathrm{SD}\left(X \oplus M, \mathbb{U}_{\{0,1\}^n}\right) &\leqslant N\left(\sum_{S\neq\emptyset} (\widehat{X \oplus M})(S)^2\right)^{1/2} && \text{Bound on SD}\\
&= N\left(\sum_{S\neq\emptyset} N^2 \widehat{X}(S)^2 \widehat{M}(S)^2\right)^{1/2} && \text{Def. of Conv.}\\
&\leqslant N\left(\sum_{S\neq\emptyset} \widehat{X}(S)^2 \varepsilon^2\right)^{1/2} && \varepsilon\text{-Bias}\\
&\leqslant N\varepsilon\left(\frac{1}{NK}\right)^{1/2} && k\text{-Source}
\end{aligned}$$

This completes the proof.