Lecture 26: Left-over Hash Lemma & Bonami-Beckner Noise Operator

Objective

- Suppose we have access to a sample from a probability distribution $\mathbb{X}$ that only has very weak randomness guarantee. For example, $\mathbb{X}$ is a probability distribution over the sample space $\{0,1\}^n$ such that $\mathrm{H}_\infty(X) \geqslant k$. That is, the output of $\mathbb{X}$ is very unpredictable and for all $x \in \{0,1\}^n$

$$\mathbb{P}[\mathbb{X} = x] \leqslant \frac{1}{2^k} = \frac{1}{K}$$

- Our objective is to generate uniform random bits from any distribution with $\mathrm{H}_\infty(\mathbb{X}) \geqslant k$

- Ideally, we will prefer to have one function $f\colon \{0,1\}^n \to \{0,1\}^m$ such that it can its output $f(\mathbb{X})$ is close to the uniform distribution $\mathbb{U}_m$ (the uniform distribution over $\{0,1\}^m$)
- However, we shall show that it is impossible that one function can extract random bits from all high min-entropy sources. This impossibility is in the strongest possible sense.
- We shall show that for every extraction function $f\colon \{0,1\}^n \to \{0,1\}$, there exists a min-entropy source $\mathbb{X}$ such that $\mathrm{H}_\infty(\mathbb{X}) \geqslant n-1$ such that $f(\mathbb{X})$ is constant. We cannot even extract one random bit from sources with $(n-1)$ min-entropy.

- The proof is as follows. Consider $S_0 = f^{-1}(0)$ and $S_1 = f^{-1}(1)$. Note that either S_0 or S_1 has at least 2^{n-1} entries. Suppose without loss of generality, $|S_0| \geqslant 2^{n-1}$. Consider $\mathbb{X}$, the uniform distribution over the set S_0. Note that $\mathbb{P}\left[\mathbb{X} = x\right] \leqslant \frac{1}{2^{n-1}}$. We have $\mathrm{H}_\infty(\mathbb{X}) \geqslant n - 1$.

Universal Hash Function Family

Definition (Universal Hash Function Family)

Let $\mathcal{H} = \{h_1, h_2, \dots, h_\alpha\}$ be a collection of hash functions such that, for each $1 \leqslant i \leqslant \alpha$, we have $h_i \colon \{0,1\}^n \to \{0,1\}^m$. Let $\mathbb{H}$ be a probability distribution over the hash functions in $\mathcal{H}$. The family $\mathcal{H}$ is a *universal hash function family* with respect to the probability distribution $\mathbb{H}$ if it satisfies the following condition. For all distinct inputs $x, x' \in \{0,1\}^n$, we have

$$\mathbb{P}\left[h(x) = h(x') \colon h \sim \mathbb{H}\right] \leqslant \frac{1}{2^m} = \frac{1}{M}$$

- Recall that we have seen that it is impossible for a deterministic function to extract even one random bit from sources with $(n-1)$ bits of min-entropy.
- We shall now show that choosing a hash function from a universal hash function family suffices

Theorem (Left-over Hash Lemma)

Let $\mathcal{H}$ be a universal hash function family $\{0,1\}^n \to \{0,1\}^m$ with respect to the probability distribution $\mathbb{H}$ over $\mathcal{H}$. Let $\mathbb{X}$ be any min-entropy source over $\{0,1\}^n$ such that $\mathrm{H}_\infty(\mathbb{X}) \geqslant k$. Then, we have

$$\mathrm{SD}\left((\mathbb{H}(\mathbb{X}), \mathbb{H}), (\mathbb{U}_m, \mathbb{H})\right) \leqslant \frac{1}{2}\sqrt{\frac{M}{K}}$$

- **Remark.** Note that we are claiming that $\mathbb{H}(\mathbb{X})$ is close to the uniform distribution $\mathbb{U}_m$ over $\{0,1\}^m$ even given the hash function $\mathbb{H}$.

- The proof proceeds in the following steps.

$$
\begin{aligned}
& 2\mathrm{SD}\left((\mathbb{H}(\mathbb{X}), \mathbb{H}), (\mathbb{U}_m, \mathbb{H})\right) \\
=& \mathbb{E}\left[2\mathrm{SD}\left((\mathbb{H}(\mathbb{X}) | \mathbb{H} = h), (\mathbb{U}_m | \mathbb{H} = h)\right) \,:\, h \sim \mathbb{H}\right] \\
=& \mathbb{E}\left[2\mathrm{SD}\left(h(\mathbb{X}), \mathbb{U}_m\right) \,:\, h \sim \mathbb{H}\right] \\
\leqslant& \mathbb{E}\left[\ell_2\left(\mathrm{Bias}_{h(\mathbb{X})} - \mathrm{Bias}_{\mathbb{U}_m}\right) \,:\, h \sim \mathbb{H}\right] \\
=& \mathbb{E}\left[\sqrt{\sum_{S \in \{0,1\}^m} \mathrm{Bias}_{h(\mathbb{X})}(S)^2 - 1} \colon h \sim \mathbb{H}\right] \\
\leqslant& \sqrt{\mathbb{E}\left[\sum_{S \in \{0,1\}^m} \mathrm{Bias}_{h(\mathbb{X})}(S)^2 - 1 \colon h \sim \mathbb{H}\right]}
\end{aligned}
$$

The last inequality is due to Jensen's inequality.

- Let us continue our simplification.

$$
\begin{aligned}
& 2\mathrm{SD}\left((\mathbb{H}(\mathbb{X}), \mathbb{H}), (\mathbb{U}_m, \mathbb{H})\right) \\
\leqslant & \sqrt{\mathbb{E}\left[\sum_{S \in \{0,1\}^m} \mathrm{Bias}_{h(\mathbb{X})}(S)^2 - 1 \colon h \sim \mathbb{H}\right]} \\
= & \sqrt{\mathbb{E}\left[\sum_{S \in \{0,1\}^m} \mathrm{Bias}_{h(\mathbb{X})}(S)^2 \colon h \sim \mathbb{H}\right] - 1} \\
= & \sqrt{\mathbb{E}\left[M \cdot \mathrm{Col}\left(h(\mathbb{X}), h(\mathbb{X})\right) \colon h \sim \mathbb{H}\right] - 1}
\end{aligned}
$$

- Note that one sample of $h(\mathbb{X})$ collides with a second sample of $h(\mathbb{X})$ due to the following cases
  1. The first sample of $\mathbb{X}$ collides with the second sample of $\mathbb{X}$. Since, $\mathrm{H}_\infty(\mathbb{X}) \geqslant k$, we have
  $$\mathrm{Col}(\mathbb{X}, \mathbb{X}) \leqslant \frac{1}{K}$$
  2. If the first and the second samples from $\mathbb{X}$ are different, then they collide with probability $\leqslant \frac{1}{M}$ when $h \sim \mathbb{H}$.

  Overall, by union bound, we get that
  $$\mathbb{E}\left[\mathrm{Col}\left(h(\mathbb{X}), h(\mathbb{X})\right) : h \sim \mathbb{H}\right] \leqslant \frac{1}{K} + \frac{1}{M}$$

- Substituting this estimation, we obtain

$$\begin{aligned}
&2\mathrm{SD}\left((\mathbb{H}(\mathbb{X}),\mathbb{H}),(\mathbb{U}_m,\mathbb{H})\right)\\
\leqslant&\sqrt{\mathbb{E}\left[M\cdot\mathrm{Col}\left(h(\mathbb{X}),h(\mathbb{X})\right)\,:\,h\sim\mathbb{H}\right]-1}\\
=&\sqrt{M\cdot\left(\frac{1}{K}+\frac{1}{M}\right)-1}=\sqrt{\frac{M}{K}}
\end{aligned}$$

- Note that this result says that we must ensure $m<k$ for the output of the extraction to be close to the uniform distribution

- Today, we shall introduce the basics of the "noise operator"
- This operator is crucial to one of the most powerful technical tools in Fourier Analysis, namely, the Hypercontractivity

Noise Operator

- Let $\mathbb{N}_\varepsilon$ be a probability distribution over the sample space $\{0,1\}^n$ such that

$$\mathbb{P}\left[\mathbb{N}_\varepsilon = x\right] = (1-\varepsilon)^{n-|x|}\varepsilon^{|x|}$$

  Here $|x|$ represents the number of 1s in x (or, equivalently, the Hamming weight of x)

- Intuitively, imagine a channel through which 0^n is fed as input. The channel converts each bit independently as follows. It converts $0 \mapsto 1$ with probability ε; and $1 \mapsto 0$ with probability $(1-\varepsilon)$. Note that the probability of the output being x is $(1-\varepsilon)^{n-|x|}\varepsilon^{|x|}$

- Our objective is to prove that

$$\mathrm{Bias}_{\mathbb{N}_\varepsilon}(S) = (1-2\varepsilon)^{|S|}$$

  We shall prove this result using a highly modular and elegant approach

- For $1 \leqslant i \leqslant n$, let $\mathbb{N}_{\varepsilon,i}$ be the probability distribution defined below

$$\mathbb{P}\left[\mathbb{N}_{\varepsilon,i} = x\right] = \begin{cases} (1-\varepsilon), & \text{if } x = 0^n \\ \varepsilon, & \text{if } x = \delta_i \\ 0, & \text{otherwise} \end{cases}$$

- Intuitively, 0^n is fed through a channel. All bits except the i-th bit are left unchanged. The i-th bit is converted as follows. It maps $0 \mapsto 1$ with probability ε; and $0 \mapsto 0$ with probability $(1-\varepsilon)$.

- Let us compute the bias of this distribution. For any $S \in \{0,1\}^n$, note that, if $S_i = 0$, we have

$$\mathrm{Bias}_{\mathbb{N}_{\varepsilon,i}}(S) = 1$$

For any $S \in \{0,1\}$, if $S_i = 1$, we have

$$\mathrm{Bias}_{\mathbb{N}_{\varepsilon,i}}(S) = (1-\varepsilon) - \varepsilon = (1-2\varepsilon)$$

- Succinctly, we can express this as

$$\mathrm{Bias}_{\mathbb{N}_{\varepsilon,i}}(S) = (1-2\varepsilon)^{S_i}$$

- So, we can conclude that

$$\mathrm{Bias}_{\bigoplus_{i=1}^{n} \mathbb{N}_{\varepsilon,i}}(S) = (1-2\varepsilon)^{\sum_{i=1}^{n} S_i} = (1-2\varepsilon)^{|S|}$$

- It is left as an exercise to prove that the distribution $\mathbb{N}_\varepsilon$ is identical to the distribution $\bigoplus_{i=1}^{n} \mathbb{N}_{\varepsilon,i}$

Noisy Version of a Function

- Let $f\colon \{0,1\}^n \to \mathbb{R}$ be any function
- Define the noisy version of f as follows

$$\widetilde{f}(x) = T_\rho(x) := \mathbb{E}\left[f(x+e)\colon e \sim \mathbb{N}_\varepsilon\right],$$

  where $\rho = 1 - 2\varepsilon$
- So, we have

$$\widetilde{f}(x) = \sum_{e\in\{0,1\}^n} \mathbb{N}_\varepsilon(e) f(x+e) = N(\mathbb{N}_\varepsilon * f)$$

  Equivalently, we have $\widetilde{f} = \mathbb{N}_\varepsilon \oplus f$ (we emphasize that f need not be a probability distribution to use the notation of $\oplus$ of two functions)
- Therefore, we get

$$\mathrm{Bias}_{\widetilde{f}}(S) = \mathrm{Bias}_{\mathbb{N}_\varepsilon}(S) \cdot \mathrm{Bias}_f(S) = \rho^{|S|}\mathrm{Bias}_f(S)$$

- That is, we conclude that

$$\widehat{T_\rho(f)}(S) = \rho^{|S|}\widehat{f}(S)$$