Lecture 22: Left-over Hash Lemma & Bonami-Beckner Noise Operator
Objective

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

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

- Our objective is to generate uniform random bits from any distribution with $H_\infty(\mathcal{X}) \geq k$
Ideally, we will prefer to have one function \( f : \{0, 1\}^n \rightarrow \{0, 1\}^m \) such that it can its output \( f(\mathbf{X}) \) 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 \( f : \{0, 1\}^n \rightarrow \{0, 1\} \), there exists a min-entropy source \( \mathbf{X} \) such that \( H_{\infty}(\mathbf{X}) \geq n - 1 \) such that \( f(\mathbf{X}) \) is constant. That is, 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, $|S_0| \geq 2^{n-1}$.

Consider $X$ that has uniform distribution over the set $S_0$. Note that $P[X = x] \leq \frac{1}{2^{n-1}}$. That is, we have $H_\infty(X) \geq n - 1$. 

Definition (Universal Hash Function Family)

Let \( \mathcal{H} = \{ h_1, h_2, \ldots, h_\alpha \} \) be a collection of hash functions such that, for each \( 1 \leq i \leq \alpha \), we have \( h_i : \{0, 1\}^n \rightarrow \{0, 1\}^m \). 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

\[
P \left[ h(x) = h(x') : h \sim \mathbb{H} \right] \leq \frac{1}{2^m} = \frac{1}{M}
\]
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 \(\{0, 1\}^n \rightarrow \{0, 1\}^m\) with respect to the probability distribution \(\mathbb{H}\) over \(\mathcal{H}\). Let \(X\) be any min-entropy source over \(\{0, 1\}^n\) such that \(H_\infty(X) \geq k\). Then, we have

\[
\text{SD} \left( (\mathbb{H}(X), \mathbb{H}), (U_m, \mathbb{H}) \right) \leq \frac{1}{2} \sqrt{\frac{M}{K}}
\]
Remark. Note that we are claiming that $\mathbb{H}(X)$ 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.

\[
2\text{SD}((H(X), H), (U_m, H)) = \mathbb{E}\left[2\text{SD}((H(X)|H = h), (U_m|H = h)) : h \sim H\right]
\]

\[
= \mathbb{E}\left[2\text{SD}(h(X), U_m) : h \sim H\right]
\]

\[
\leq \mathbb{E}\left[\ell_2\left(\text{Bias}_{h(X)} - \text{Bias}_{U_m}\right) : h \sim H\right]
\]

\[
= \mathbb{E}\left[\sqrt{\sum_{S \in \{0,1\}^m} \text{Bias}_{h(X)}(S)^2 - 1} : h \sim H\right]
\]

\[
\leq \sqrt{\mathbb{E}\left[\sum_{S \in \{0,1\}^m} \text{Bias}_{h(X)}(S)^2 - 1 : h \sim H\right]}
\]
The last inequality is due to Jensen's inequality.

- Let us continue our simplification.

\[
2\text{SD} \left( (\mathbb{H}(X), \mathbb{H}), (\mathbb{U}_m, \mathbb{H}) \right) \leq \sqrt{\mathbb{E} \left[ \sum_{S \in \{0,1\}^m} \text{Bias}_{h(X)}(S)^2 - 1 : h \sim \mathbb{H} \right]} \]

\[
= \sqrt{\mathbb{E} \left[ \sum_{S \in \{0,1\}^m} \text{Bias}_{h(X)}(S)^2 : h \sim \mathbb{H} \right] - 1}
\]

\[
= \sqrt{\mathbb{E} \left[ M \cdot \text{Col} \left( h(X), h(X) \right) : h \sim \mathbb{H} \right] - 1}
\]
Note that one sample of $h(\mathbf{X})$ collides with a second sample of $h(\mathbf{X})$ due to the following cases:

1. The first sample of $\mathbf{X}$ collides with the second sample of $\mathbf{X}$. Since, $H_{\infty}(\mathbf{X}) \geq k$, we have
   \[
   \text{Col}(\mathbf{X}, \mathbf{X}) \leq \frac{1}{K}
   \]

2. If the first and the second samples from $\mathbf{X}$ are different, then they collide with probability $\leq \frac{1}{M}$ when $h \sim \mathbb{H}$.

Overall, by union bound, we get that
\[
\mathbb{E} \left[ \text{Col} \left( h(\mathbf{X}), h(\mathbf{X}) \right) : h \sim \mathbb{H} \right] \leq \frac{1}{K} + \frac{1}{M}
\]
Substituting this estimation, we obtain

\[
2SD \left( (H(X), H), (U_m, H) \right) \\
\leq \sqrt{E \left[ M \cdot \text{Col} (h(X), h(X)) : h \sim H \right]} - 1 \\
= \sqrt{M \cdot \left( \frac{1}{K} + \frac{1}{M} \right) - 1} = \sqrt{\frac{M}{K}}
\]

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

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

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 being fed as input. The channel converts each bit independently as follows. It converts $0 \mapsto 1$ with probability $\varepsilon$; and $1 \mapsto 0$ with probability $(1 - \varepsilon)$. Note that the probability of the output being $x$ is $(1 - \varepsilon)^{n-|x|}\varepsilon^{|x|}$.\[20pt]\]

Our objective is to prove that

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

We shall prove this result using a highly modular and elegant approach.
For $1 \leq i \leq n$, let $\mathcal{N}_{\varepsilon,i}$ be the probability distribution defined below

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

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

$$\text{Bias}_{N_\varepsilon, i}(S) = 1$$

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

$$\text{Bias}_{N_\varepsilon, i}(S) = (1 - \varepsilon) - \varepsilon = (1 - 2\varepsilon)$$

Succinctly, we can express this as

$$\text{Bias}_{N_\varepsilon, i}(S) = (1 - 2\varepsilon)^{S_i}$$

So, we can conclude that

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

It is left as an exercise to prove that the distribution $N_\varepsilon$ is identical to the distribution $\bigoplus_{i=1}^n N_\varepsilon, i$.
Noisy Version of a Function

- Let $f : \{0, 1\}^n \rightarrow \mathbb{R}$ be any function.
- Define the noisy version of $f$ as follows:
  \[
  \tilde{f}(x) = T_\rho(x) := \mathbb{E} \left[ f(x + e) : e \sim \mathcal{N}_\varepsilon \right],
  \]
  where $\rho = 1 - 2\varepsilon$.
- So, we have
  \[
  \tilde{f}(x) = \sum_{e \in \{0, 1\}^n} \mathcal{N}_\varepsilon(e)f(x + e) = \mathcal{N}(\mathcal{N}_\varepsilon \ast f)
  \]
  Equivalently, we have $\tilde{f} = \mathcal{N}_\varepsilon \oplus f$ (we emphasize that $f$ need not be a probability distribution to use the notation of $\oplus$ of two functions).
- Therefore, we get
  \[
  \text{Bias}_{\tilde{f}}(S) = \text{Bias}_{\mathcal{N}_\varepsilon}(S) \cdot \text{Bias}_f(S) = \rho^{\|S\|} \text{Bias}_f(S)
  \]
- That is, we conclude that
  \[
  \mathcal{T}_\rho(f)(S) = \rho^{\|S\|} \hat{f}(S)
  \]