Lecture 24: Tackling Probability Distributions and XOR Lemma
Until now, we have treated a distribution $X$ over $\{0, 1\}^n$ as the function $X : \{0, 1\}^n \rightarrow \mathbb{R}$ such that $X(\omega) := \mathbb{P}[X = \omega]$

However, for intuition purposes, we want to develop concepts that are unique to distributions that are analogous to the concepts in Fourier analysis of functions.
Bias of a Distribution: Intuition

- Let $X$ be a distribution over $\{0, 1\}^n$
- Consider the following algorithm for a fixed $S \in \{0, 1\}^n$
  
  1. Sample $x \sim X$
  2. Output $S \cdot x$

  The output distribution is over the sample space $\{0, 1\}$. Let $p_0$ represent the probability that the output of this algorithm is 0; and, $p_1$ represent the probability of the output being 1.

  We want to say that the output is “unbiased” (or, “has bias 0”) if $p_0 = p_1 = 1/2$. Similarly, we want to say that the output “has bias 1” if $p_0 = 1$ and $p_1 = 0$. Finally, we want to say that the output “has bias $-1$” if $p_0 = 0$ and $p_1 = -1$.

  Interpolating this intuition, we want to say that the bias of the output distribution of the algorithm above is $p_0 - p_1$.
Definition

Let $X$ be a distribution over the sample space $\{0, 1\}^n$. For any $S \in \{0, 1\}^n$, we define the bias of $X$ with respect to (the linear test) $S$ as

$$\text{Bias}_X(S) := \mathcal{N}(X)(S)$$
Collision Probability

Let $X$ and $Y$ be two probability distributions over $\{0, 1\}^n$.

$\text{Col}(X, Y)$ refers to the probability that two samples drawn according to $X$ and $Y$ turn out to be identical. We know that

$$\text{Col}(X, Y) = N \langle X, Y \rangle = N \sum_{S \in \{0, 1\}^n} \hat{X}(S) \cdot \hat{Y}(S)$$

Equivalently, we have

$$\text{Col}(X, Y) = \frac{1}{N} \sum_{S \in \{0, 1\}^n} \text{Bias}_X(S) \cdot \text{Bias}_Y(S)$$
Recall that we had defined the distribution \((X \oplus Y)\) as a distribution over \(\{0, 1\}^n\) that is identical to the function \(N(X \ast Y)\).

We had also proven that

\[
(X \ast Y)(S) = \hat{X}(S) \cdot \hat{Y}(S)
\]

So, we can conclude that

\[
\text{Bias}_{X \oplus Y}(S) = \text{Bias}_X(S) \cdot \text{Bias}_Y(S)
\]
For two functions $f, g : \{0, 1\}^n \rightarrow \mathbb{R}$, let us define $L_1(f - g)$ as follows:

$$L_1(f - g) := \frac{1}{N} \sum_{x \in \{0, 1\}^n} |f(x) - g(x)|$$

We can upper-bound $L_1(f - g)$ using $\hat{f}$ and $\hat{g}$ as follows:

$$L_1(f - g) = \frac{1}{N} \sum_{x \in \{0, 1\}^n} |f(x) - g(x)|$$

$$\leq \frac{1}{N} \sqrt{N} \cdot \left( \sum_{x \in \{0, 1\}^n} (f(x) - g(x))^2 \right)^{1/2}, \text{ by Cauchy-Schwarz}$$

$$= \left( \frac{1}{N} \sum_{x \in \{0, 1\}^n} (f(x) - g(x))^2 \right)^{1/2}$$
Statistical Distance of Two Distributions

\[= \left( \frac{1}{N} \sum_{x \in \{0,1\}^n} (f - g)(x)^2 \right)^{1/2} \]

= \left( \sum_{S \in \{0,1\}^n} (\hat{f} - \hat{g})(S)^2 \right)^{1/2}, \text{ by Parseval's}

= \left( \sum_{S \in \{0,1\}^n} (\hat{f}(S) - \hat{g}(S))^2 \right)^{1/2}

=: l_2(\hat{f} - \hat{g})

XOR Lemma
We can obtain a similar result for statistical distance, which is the analogue of $L_1(\cdot)$ for functions

$$2SD(X, Y) := \sum_{x \in \{0, 1\}^n} |X(x) - Y(x)|$$

So, we have

$$2SD(X, Y) = NL_1(X - Y) \leq N \ell_2(\hat{X} - \hat{Y}) = \ell_2(Bias_X - Bias_Y)$$

That is,

$$2SD(X, Y) \leq \sum_{S \in \{0, 1\}^n} (Bias_X(S) - Bias_Y(S))^2$$
<table>
<thead>
<tr>
<th>Functions</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{X}(S)$</td>
<td>$\text{Bias}_X(S) := N\hat{X}(S)$</td>
</tr>
<tr>
<td>$\langle X, Y \rangle = \sum_{S \in {0,1}^n} \hat{X}(S)\hat{Y}(S)$</td>
<td>$\text{Col}(X, Y) = \frac{1}{N} \sum_{S \in {0,1}^n} \text{Bias}_X(S)\text{Bias}_Y(S)$</td>
</tr>
<tr>
<td>$(X \star Y)(S) = \hat{X}(S)\hat{Y}(S)$</td>
<td>$\text{Bias}_{X \oplus Y}(S) = \text{Bias}_X(S)\text{Bias}_Y(S)$</td>
</tr>
<tr>
<td>$L_1(X - Y) \leq \ell_2(\hat{X} - \hat{Y})$</td>
<td>$2\text{SD}(X, Y) \leq \ell_2(\text{Bias}_X - \text{Bias}_Y)$</td>
</tr>
</tbody>
</table>
Let $X$ be a distribution over $\{0, 1\}$ such that $\Pr[X = 0] = \frac{1+\varepsilon}{2}$ and $\Pr[X = 1] = \frac{1-\varepsilon}{2}$.

Note that $n = 1$ and $\text{Bias}_X(0) = 1$ and $\text{Bias}_X(1) = \varepsilon$.

Let $S_n = X^{(1)} \oplus X^{(2)} \oplus \ldots \oplus X^{(n)}$.

Note that

$$\text{Bias}_S(0) = \text{Bias}_{X^{(1)}}(0) \cdot \text{Bias}_{X^{(2)}}(0) \cdots \text{Bias}_{X^{(n)}}(0) = 1$$

Note that

$$\text{Bias}_S(1) = \text{Bias}_{X^{(1)}}(1) \cdot \text{Bias}_{X^{(2)}}(1) \cdots \text{Bias}_{X^{(n)}}(1) = \varepsilon^n$$

From the biases, we can conclude that $\Pr[S_n = 0] = \frac{1+\varepsilon^n}{2}$ and $\Pr[S_n = 1] = \frac{1-\varepsilon^n}{2}$. 
Further, we can conclude that $S_n$ is very close to the uniform distribution over $\{0, 1\}$, namely $U_{\{0,1\}}$. Note that $\text{Bias}_{U_{\{0,1\}}} (0) = 1$ and $\text{Bias}_{U_{\{0,1\}}} (1) = 0$. So, the statistical distance between $S_n$ and $U_{\{0,1\}}$ is upper-bounded as follows.

$$2 \text{SD} \left( S_n, U_{\{0,1\}} \right) \leq \ell_2 \left( \text{Bias}_{S_n} - \text{Bias}_{U_{\{0,1\}}} \right) = \ell_2 \left( (1, \varepsilon^n) - (1, 0) \right) = \varepsilon^n$$

That is, $S_n$ is getting close to the uniform distribution exponentially fast!

In general, we can consider the sum $S_n = X_1 \oplus \cdots \oplus X_n$, where $X_1, \ldots, X_n$ are independent distributions over $\{0, 1\}$ with bias $\varepsilon_1, \ldots, \varepsilon_n$, respectively. Then, we shall have $\text{Bias}_{S_n} (1) = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_n$. 
It is extremely crucial that the distributions $X_1, \ldots, X_n$ are independent. Otherwise, we cannot multiply the biases to obtain the bias of the sum $S_n$. For example, let $(X_1, \ldots, X_n)$ be uniform random variables over $\{0, 1\}^n$ such that their parity is 0 (that is, they have even number of 1s). Each random variable has $\text{Bias}_{X_i}(1) = 0$. However, the random variable $S_n$ has $\text{Bias}_{S_n}(1) = 1$. 
A Combinatorial Proof.

To compute the bias $\text{Bias}_{S_n}(1)$, we need to estimate

$$P[S_n = 0] - P[S_n = 1]$$

$$= \sum_{i \text{ is even}} \binom{n}{i} \left( \frac{1 - \varepsilon}{2} \right)^i \left( \frac{1 + \varepsilon}{2} \right)^{n-i} - \sum_{i: \text{ odd}} \binom{n}{i} \left( \frac{1 - \varepsilon}{2} \right)^i \left( \frac{1 + \varepsilon}{2} \right)^{n-i}$$

$$= \sum_{i=1}^{n} \binom{n}{i} (-1)^i \left( \frac{1 - \varepsilon}{2} \right)^i \left( \frac{1 + \varepsilon}{2} \right)^{n-i}$$

$$= \left( \frac{1 + \varepsilon}{2} - \frac{1 - \varepsilon}{2} \right)^n = \varepsilon^n$$

Note that this conclusion followed so easily using Fourier analysis.