Lecture 34: Tackling Probability Distributions and XOR Lemma

Overview

- Until now, we have treated a distribution X over $\{0,1\}^n$ as the function $X\colon \{0,1\}^n \to \mathbb{R}$ 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 $p_0 - p_1$

Bias: Definition

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

$$\text{bias}_X(S) := N\widehat{X}(S)$$

Collision Probability

- Let X and Y be two probability distributions over $\{0,1\}^n$
- $\mathrm{col}(X, Y)$ refers to the probability that two samples drawn according to X and Y turn out to be identical. We know that

$$\mathrm{col}(X, Y) = N\langle X, Y\rangle = N \sum_{S \in \{0,1\}^n} \widehat{X}(S) \cdot \widehat{Y}(S)$$

- Equivalently, we have

$$\mathrm{col}(X, Y) = \frac{1}{N} \sum_{S \in \{0,1\}^n} \mathrm{bias}_X(S) \cdot \mathrm{bias}_Y(S)$$

Bias of XOR of two Distributions

- Recall that we had defined the distribution $(X \oplus Y)$ as a distribution over $\{0,1\}^n$ that is identical to the function $N(X * Y)$.
- We had also proven that

$$(\widehat{X * Y})(S) = \widehat{X}(S) \cdot \widehat{Y}(S)$$

- So, we can conclude that

$$\text{bias}_{X \oplus Y}(S) = \text{bias}_X(S) \cdot \text{bias}_Y(S)$$

Statistical Distance of Two Distributions I

- For two function $f, g\colon \{0,1\}^n \to \mathbb{R}$, let us define $L_1(f-g)$ as follows

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

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

$$\begin{aligned} L_1(f-g) &= \frac{1}{N} \sum_{x \in \{0,1\}^n} \left| f(x) - g(x) \right| \\ &\leqslant \frac{1}{N}\sqrt{N} \cdot \left(\sum_{x \in \{0,1\}^n} \left(f(x) - g(x)\right)^2 \right)^{1/2}, \text{ by Cauchy-Schw} \\ &= \left(\frac{1}{N} \sum_{x \in \{0,1\}^n} \left(f(x) - g(x)\right)^2 \right)^{1/2} \end{aligned}$$

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

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

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

$$=: \ell_2(\widehat{f} - \widehat{g})$$

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

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

- So, we have

$$2\mathrm{SD}(X,Y) = NL_1(X-Y) \leqslant N\ell_2(\widehat{X}-\widehat{Y}) = \ell_2(\mathrm{bias}_X - \mathrm{bias}_Y)$$

That is,

$$2\mathrm{SD}(X,Y) \leqslant \sum_{S\in\{0,1\}^n} \left(\mathrm{bias}_X(S) - \mathrm{bias}_Y(S)\right)^2$$

Summary

Functions	Probability
$\widehat{X}(S)$	$\text{bias}_X(S) := N\widehat{X}(S)$
$\langle X, Y\rangle = \sum_{S\in\{0,1\}^n} \widehat{X}(S)\widehat{Y}(S)$	$\text{col}(X, Y) = \frac{1}{N}\sum_{S\in\{0,1\}^n} \text{bias}_X(S)\text{bias}_Y(S)$
$(\widehat{X * Y})(S) = \widehat{X}(S)\widehat{Y}(S)$	$\text{bias}_{X\oplus Y}(S) = \text{bias}_X(S)\text{bias}_Y(S)$
$L_1(X - Y) \leqslant \ell_2(\widehat{X} - \widehat{Y})$	$2\mathrm{SD}\,(X, Y) \leqslant \ell_2(\text{bias}_X - \text{bias}_Y)$

XOR Lemma

- Let $\mathbb{X}$ be a distribution over $\{0,1\}$ such that $\mathbb{P}[\mathbb{X}=0]=\frac{1+\varepsilon}{2}$ and $\mathbb{P}[X=1]=\frac{1-\varepsilon}{2}$
- Note that $n=1$ and $\mathrm{bias}_X(0)=1$ and $\mathrm{bias}_X(1)=\varepsilon$
- Let $\mathbb{S}_n=\mathbb{X}^{(1)}\oplus\mathbb{X}^{(2)}\oplus\cdots\oplus\mathbb{X}^{(n)}$
- Note that

$$\mathrm{bias}_S(0)=\mathrm{bias}_{\mathbb{X}^{(1)}}(0)\cdot\mathrm{bias}_{\mathbb{X}^{(2)}}(0)\cdots\mathrm{bias}_{\mathbb{X}^{(n)}}(0)=1$$

- Note that

$$\mathrm{bias}_S(1)=\mathrm{bias}_{\mathbb{X}^{(1)}}(1)\cdot\mathrm{bias}_{\mathbb{X}^{(2)}}(1)\cdots\mathrm{bias}_{\mathbb{X}^{(n)}}(1)=\varepsilon^n$$

- From the biases, we can conclude that $\mathbb{P}[\mathbb{S}_n=0]=\frac{1+\varepsilon^n}{2}$ and $\mathbb{P}[\mathbb{S}_n=1]=\frac{1-\varepsilon^n}{2}$

- Further, we can conclude that $\mathbb{S}_n$ is very close to the uniform distribution over $\{0,1\}$, namely $\mathbb{U}_{\{0,1\}}$. Note that $\mathsf{bias}_{\mathbb{U}_{\{0,1\}}}(0) = 1$ and $\mathsf{bias}_{\mathbb{U}_{\{0,1\}}}(1) = 0$. So, the statistical distance between $\mathbb{S}_n$ and $\mathbb{U}_{\{0,1\}}$ is upper-bounded as follows.

$$2\mathrm{SD}\left(\mathbb{S}_n, \mathbb{U}_{\{0,1\}}\right) \leqslant \ell_2(\mathsf{bias}_{\mathbb{S}_n} - \mathsf{bias}_{\mathbb{U}_{\{0,1\}}}) = \ell_2\big((1,\varepsilon^n) - (1,0)\big) = \varepsilon^n$$

  That is, $\mathbb{S}_n$ is getting close to the uniform distribution exponentially fast!
- In general, we can consider the sum $\mathbb{S}_n = \mathbb{X}_1 \oplus \cdots \oplus \mathbb{X}_n$, where $\mathbb{X}_1, \dots, \mathbb{X}_n$ are independent distributions over $\{0,1\}$ with bias $\varepsilon_1, \dots, \varepsilon_n$, respectively. Then, we shall have $\mathsf{bias}_{\mathbb{S}_n}(1) = \varepsilon_1\varepsilon_2\cdots\varepsilon_n$.

- It is extremely crucial that the distributions $\mathbb{X}_1, \dots, \mathbb{X}_n$ are independent. Otherwise, we cannot multiply the biases to obtain the bias of the sum $\mathbb{S}_n$. For example, let $(\mathbb{X}_1, \dots, \mathbb{X}_n)$ 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 $\text{bias}_{\mathbb{X}_i}(1) = 0$. However, the random variable $\mathbb{S}_n$ has $\text{bias}_{\mathbb{S}_n}(1) = 1$.

A Combinatorial Proof.

- To compute the bias $\mathrm{bias}_{\mathbb{S}_n}(1)$, we need to estimate

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{S}_n = 0\right] - \mathbb{P}\left[\mathbb{S}_n = 1\right] \\
&= \sum_{i \text{ is even}} \binom{n}{i} \left(\frac{1-\varepsilon}{2}\right)^i \left(\frac{1+\varepsilon}{2}\right)^{n-i} - \sum_{i:\ \text{odd}} \binom{n}{i} \left(\frac{1-\varepsilon}{2}\right)^i \left(\frac{1+\varepsilon}{2}\right)^{n-i} \\
&= \sum_{i=1}^{n} \binom{n}{i} (-1)^i \left(\frac{1-\varepsilon}{2}\right)^i \left(\frac{1+\varepsilon}{2}\right)^{n-i} \\
&= \left(\frac{1+\varepsilon}{2} - \frac{1-\varepsilon}{2}\right)^n = \varepsilon^n
\end{aligned}$$

- Note that this conclusion followed so easily using Fourier analysis