1 Administrative Notes

- Project Presentations on Wed 5/1, emphasize one interesting point if do not have enough time
- Writeup (~ 5 pages) due Wed 5/8
- Polished Scribe Notes (including worked out exercises) due Wed 5/8

2 Today’s Agenda

- Amplification/Polarization of $SD$
- $SD \leq SD$

3 Amplification/Polarization of SD

Recall that to define sampleable distributions, we define a circuit $C$ with $m$ inputs and $n$ outputs: $c : \{0,1\}^m \rightarrow \{0,1\}^n$ where $m, |c| = poly(n)$. Then for $X \sim Bern(\frac{1}{2})^m$, $C(X)$ defines a distribution on $\{0,1\}^n$, and we use $C$ to represent this distribution for simplicity of notations.

Given sampleable distributions and two parameters $c, f$ ($0 \leq c \leq f \leq 1$), we can define two sets:

$$\text{CLOSE}^c = \{(c_1, c_2) | \delta(c_1, c_2) \leq c\}$$
$$\text{FAR}^f = \{(c_1, c_2) | \delta(c_1, c_2) \geq f\}$$

where $\delta(P, Q)$ is the statistical difference between distributions $P$ and $Q$ (see last lecture for the definition of statistical difference).

**Definition 1** (Statistical Difference Problem). Given $(c_1, c_2) \in \text{CLOSE}^c \cup \text{FAR}^f$, the statistical difference problem $SD_{c,f}$ is to decide whether $(c_1, c_2) \in \text{CLOSE}^c$ (return YES) or $(c_1, c_2) \in \text{FAR}^f$ (return NO).

For completeness of this definition, we can consider its complement problem.

**Definition 2** (Complement of the Statistical Difference Problem). The complement of $SD_{c,f}$ is $SD_{c,f}$, which returns NO if $(c_1, c_2) \in \text{CLOSE}^c$ and YES if $(c_1, c_2) \in \text{FAR}^f$ for $(c_1, c_2) \in \text{CLOSE}^c \cup \text{FAR}^f$.

As mentioned in the last lecture, we are interested in $SD^{\frac{1}{2}, \frac{1}{2}}$ due to its relation to the problem of Graph Isomorphism. For simplicity we omit the superscripts and call it $SD$. We want to ask if we can amplify this problem into $SD^{2^{-n}, 1-2^{-n}}$ which is more polarized since $c$ comes closer to 0 and $f$ comes closer to 1.

**Theorem 1.** $SD^{\frac{1}{2}, \frac{1}{2}} \leq SD^{2^{-n}, 1-2^{-n}}$, where we can make $\varepsilon$ arbitrarily close to 1. More generally, the proof goes through as long as $c < f^2$.

To prove Theorem 1 we need to use two kinds of reductions.

**Lemma 1** (The Direct Product reduction). $SD_{c,f} \leq SD_{c,1-2^{exp(-tf^2/2)}}$.
We begin with the direct product reduction because it’s simpler to prove (during class this was called Ingredient 2). To prove Lemma 1, we simply map any \((c_1, c_2)\) to \((c_1', c_2')\), where
\[
(c_1, c_2) \rightarrow (c_1', c_2') = (c(X_1), c(X_2), \ldots c(X_t))
\]
To prove that the reduction finds a solution to \(SD^{c_1, t} \leq 1 - 2\exp(-tf^2/2)\), we need to prove that

(i) \(\delta(c'_1, c'_2) \leq t\delta(c_1, c_2)\)

(ii) \(\delta(c'_1, c'_2) \geq f, \delta(c'_1, c'_2) \geq 1 - 2\exp(-tf^2/2)\)

To prove (i), notice that \(\delta\) is a distance metric and we can apply triangular inequality multiple times, each time replacing one element in the sequence beginning with \(c_1(X_1), \ldots c_t(X_t)\), until we arrive at \(c_2(X_1), \ldots c_2(X_t)\).

To prove (ii), from the definition of statistical difference, because \(c_1\) and \(c_2\) are apart from each other by at least \(f\), \(\exists\) test \(T\) and value \(\alpha\) such that
\[
\begin{align*}
\mathbb{E}_{z \sim c_1}[T(z)] &\geq \alpha + f \\
\mathbb{E}_{z \sim c_2}[T(z)] &\leq \alpha
\end{align*}
\]
Then our new test for \((c'_1, c'_2)\) returns 1 if \(T(z_1) + T(z_2) + \cdots T(z_t) \geq (\alpha + \frac{f}{2})t\) else 0.

Exercise 1. Prove that \(\delta(c'_1, c'_2) \geq 1 - 2\exp(-tf^2/2)\) for \(\delta(c'_1, c'_2) \geq f\) in a rigorous way.

Proof: For the test defined above, we get 1 when \(T(z_1) + T(z_2) + \cdots T(z_t) \geq (\alpha + \frac{f}{2})t\), or equivalently
\[
\frac{1}{t} \sum_{i=1}^{t} T(z_i) \geq \alpha + \frac{f}{2}
\]

If we sample from \(c_1\), the mean of \(T(z)\) for \(z \sim c_1\) is \(\alpha + f\), so applying the Chernoff Bound, we have w.p. at least \(1 - \exp(-tf^2/2)\) that
\[
\frac{1}{t} \sum_{i=1}^{t} T(z_i) \geq \alpha + f - \frac{f}{2} = \alpha + \frac{f}{2}
\]
On the other hand, if we sample from \(c_2\), the mean of \(T(z)\) for \(z \sim c_1\) is \(\alpha\), so applying the Chernoff Bound, we have w.p. at least \(1 - \exp(-t\delta^2/2)\)
\[
\frac{1}{t} \sum_{i=1}^{t} T(z_i) \leq \alpha + \frac{f}{2}
\]
Using union bound, we’ve found a test such that w.p. at least \(1 - 2\exp(-tf^2/2)\) it takes 1 under \(c_1\) and 0 under \(c_2\). Therefore the statistical difference between \(c_1\) and \(c_2\) is at least \(1 - 2\exp(-tf^2/2)\).

\(\square\)

Lemma 2 (The XOR reduction). \(SD^{c_1, t} \leq SD^{c'_1, t'}\)

We construct the new distributions by mapping \((c_0, c_1)\) to
\[
(D_0, D_1) = ((c_0, c_1)_{c_1} \oplus t, (c_0, c_1)_{c_0, c_1} \oplus t)
\]
We will define \(((c_0, c_1)_{c_0} \oplus t, (c_0, c_1)_{c_1} \oplus t)\) recursively below.

For pairs of random variables \((X_0, X_1)\) and \((y_0, y_1)\), we construct a new pair \((Z_0, Z_1)\) as follows:

\[
Z_0 = \begin{cases} 
(X_0, y_0) & \text{wp } \frac{1}{2} \\
(X_1, y_1) & \text{wp } \frac{1}{2}
\end{cases}
\]

CS 229r Information Theory in Computer Science-2
and

\[ Z_1 = \begin{cases} (X_0, y_1) \text{ wp } \frac{1}{2} \\ (X_1, y_0) \text{ wp } \frac{1}{2} \end{cases} \]

Then \( \forall \alpha, \beta \), we have

\[
P[Z_0 = (\alpha, \beta)] - P[Z_1 = (\alpha, \beta)] = \frac{1}{2}(P[X_0 = \alpha]P[y_0 = \beta] + P[X_1 = \alpha]P[y_1 = \beta] - (P[X_0 = \alpha]P[y_1 = \beta] + P[X_1 = \alpha]P[y_0 = \beta])
\]

Therefore, \( \delta(Z_0, Z_1) = \delta(X_0, X_1)\delta(y_0, y_1) \).

We define \((c_0 \oplus t, c_1 \oplus t)\) by recursively apply the above operations. \( \delta(D_0, D_1) = \delta(C_0, C_1)^t \) follows by induction on \( t \). Lemma \[2\] then follows.

To prove Theorem \[1\] we apply the direct product reduction and the XOR reduction multiple times.

(i) \( SD^{\frac{1}{2}, \frac{3}{4}} \rightarrow SD^{\frac{1}{3}, \frac{1}{3}} \). We achieve this by using the XOR reduction with \( t = O(\log n) \)

(ii) \( SD^{\frac{1}{4}, \frac{1}{4}} \rightarrow SD^{\frac{1}{4}, \exp(-n^{0.4})} \). We achieve this by using the direct product reduction with \( t = \frac{n^2}{t} \).

(iii) \( SD^{\frac{1}{4}, \exp(-n^{0.4})} \rightarrow SD^{\frac{1}{4}, \exp(-n^{0.3})} \). We achieve this by using the XOR reduction with \( t = n^{1.4} \).

\[ \text{C_1} \quad \text{C_2} \]

\[ y \]

## 4 Reduction of SD to its complement

In the last lecture, we mentioned that \( SD^{c,f} \equiv SD^{\overline{c}, \overline{f}} \), which means they are computationally equivalent, up to poly-time reductions. We only need to prove \( SD \leq SD \) since we can then apply this to \( SD \). The below proofs were originally proposed in \[2\] and was then presented in \[1\] and \[3\].

**Theorem 2.** \( SD \leq SD \), which means we can find a polytime reduction \((c_1, c_2) \rightarrow (D_1, D_2)\) such that

\[
(D_1, D_2) \text{ are } \begin{cases} \text{far if } (c_1, c_2) \text{ are close} \\ \text{close if } (c_1, c_2) \text{ are far} \end{cases}
\]

### 4.1 Entropy Difference

We consider the computational problem of entropy difference.

**Definition 3.** For distributions \((c_1, c_2)\), the entropy difference problem \( ED^k \) is to decide whether \( H(c_1) \geq H(c_2) + k \) (YES) or \( H(c_2) \geq H(c_1) + k \) (NO).
We need the following properties to finish the proof of Theorem 2.

(i) \( \text{ED} \leq \sqrt{n} \) (like before we repeat: \((c_0, c_1) \rightarrow (c_0', c_1')\))

(ii) SD \( \leq \) ED

(iii) ED \( \leq \) E\(D\) (the proof is trivial, we just map \((c_1, c_2) \rightarrow (c_2, c_1)\))

(iv) ED \( \leq \) SD (we can convert it to \((\text{ED} \leq \text{SD})\))

To prove (ii) SD \( \leq \) ED, given \((c_0, c_1)\), we map it to \((D_0, D_1)\) such that

- \( D_0 = (b, c_b(X)) \) where \( b \) and \( X \) are picked at random
- \( D_1 = (b', c_b(X)) \) where \( b', b \) and \( X \) are picked at random

We can see that \( H(D_1) = H(c_b) + 1 \). If \( c_0 \) and \( c_1 \) are very far apart, then we can infer \( b \) from \( c_b(X) \), so \( H(D_0) \) would be close to \( H(c_b) + 1 \) and be significantly smaller than \( H(D_1) \). Otherwise, if \( c_0 \) and \( c_1 \) are really close, then \( b \) cannot be inferred from \( c_b(X) \) and \( H(D_1) \approx H(D_0) \).

To prove (iii), we use “extractors” to manipulate random variables. If there is already sufficient entropy, we can transform random variable with entropy \( k \) to uniform distribution on \( k - o(1) \) bits. However, we cannot do this with insufficient entropy. To do this, there are three problems:

1. Need Extractor + analysis (easy, pairwise independence)
2. This works for min entropy, but we have entropy. (“entropy flattening”, “AEP”, \( C_1 \rightarrow C_1' \))
3. Even if we have flat source, we don’t know entropy of \( c_1 \) or \( c_2 \), so we have no idea of which hash function to use to extract entropy.

The key solution is proposed in [2]. Assume \( C_2 \) has more entropy than \( C_1 \). Consider a random output of \( C_1 \) and a random hash function \( h \). We map \((C_1, C_2)\) to \(((C_1(X), h, h(X), C_2(y)))\). Then we can prove that

\[
H(X|C_1(X)) = m - H(C_1(X))
\]

So \( H(X, C_2(y)|C_1(X)) = m - H(C_1(X)) + H(C_2(y)) \). Because the entropy of \( X \) and \((X, C_2(y))\) differ (conditioning on \( C_1(X) \)), when we feed them into \( h \), we will get distinguishable distributions (one should be much “more uniform” than the other). Notice we don’t need to know how much entropy in \( C_1 \).

With the above properties, we can readily prove Theorem 2.

**Exercise 2.** Prove Theorem 2 using the above proved properties.

**Proof.**

\[
\text{SD} \leq \text{ED} \text{ (property (ii))} \\
\leq \text{E}\text{D} \text{ (property (iii))} \\
\leq \text{S}\text{D} \text{ (converted property (iv))}
\]

CS 229r Information Theory in Computer Science-4
References

