

Lecture 23

Instructor: Madhu Sudan

Scribe: Conlan Olson

1 Logistics

- Project presentations will be on Wednesday May 1.
- The writeup (approximately 5 pages) will be due on Wednesday May 8.
- Polished scribe notes with worked out exercises will be due on Wednesday May 8.

Send over any suggestions or requests for topics for the last two lectures.

2 Agenda for today

- Amplification/polarization of SD .
- $SD \leq \overline{SD}$.

3 Review from last time

A **sampleable distribution** comes from a circuit $C : \{0, 1\}^m \rightarrow \{0, 1\}^n$ (with $m, |C| = \text{poly}(n)$) and is the distribution of $C(X)$ when $X \sim \text{Bern}(1/2)^m$.

This is a (small) subset of all possible distributions on $\{0, 1\}^n$. Even if we discretize probabilities, there are exponentially many choices for the probability of each string. Since there are exponentially many strings, this gives a *doubly* exponential number of distributions on $\{0, 1\}^n$. On the other hand, there are only exponentially many sampleable distributions because they are specified by circuits.

The statistical difference problem $SD^{c,f}$ for $0 \leq c < f \leq 1$ is based on the classes

$$CLOSE^c = \{(C_1, C_2) \mid \delta(C_1, C_2) \leq c\}, \quad FAR^f = \{(C_1, C_2) \mid \delta(C_1, C_2) \geq f\}.$$

Given $(C_1, C_2) \in CLOSE^c \cup FAR^f$, we want to report *YES* if $(C_1, C_2) \in CLOSE^c$ and *NO* if $(C_1, C_2) \in FAR^f$.

For \overline{SD} , we'll report *YES* if *FAR* and *NO* if *CLOSE*.

When we omit the superscript c, f , we'll take them to be $1/3, 2/3$.

The key theorem today will be that $SD < \overline{SD}$, which we'll show by showing there exists a poly time reduction $(C_1, C_2) \mapsto (D_1, D_2)$ such that if $(C_1, C_2) \in CLOSE$, then $(D_1, D_2) \in FAR$ and vice versa.

First however, we'll show $SD^{1/3, 2/3} \leq SD^{2^{-n^\varepsilon}, 1-2^{-n^\varepsilon}}$ where ε can be arbitrarily close to 1. The choice of $1/3, 2/3$ here isn't that important; we could just as well work with any c, f such that $c < f^2$.

4 Amplification/polarization of SD

i.e. $SD^{1/3,2/3} \leq SD^{2^{-n^\epsilon}, 1-2^{-n^\epsilon}}$.

The proof will use two ingredients:

Ingredient 1: $SD^{c,f} \leq SD^{c^t, f^t}$ for any t . Note that as t increases, the *gap* between the cutoffs will decrease, but the *ratio* will increase.

Ingredient 2: $SD^{c,f} \leq SD^{ct, 1-2\exp(-tf^2/2)}$. We wish we could have the upper cutoff be tf , but we can't quite do this. This ingredient will serve to "translate" the gap over but mostly preserve the ratio.

4.1 Ingredient 2

Consider the reduction $(C_1, C_2) \mapsto (C_1^t, C_2^t)$ where $C^t(X_1, \dots, X_t) = (C(X_1), \dots, C(X_t))$. This should make two slightly distinguishable distributions more distinguishable. We need to show

$$(i) \delta(C_1^t, C_2^t) \leq t\delta(C_1, C_2)$$

$$(ii) \delta(C_1^t, C_2^t) \geq 1 - 2\exp\left(-t \cdot \frac{\delta(C_1, C_2)^2}{2}\right)$$

For (i), we'll use a "hybrid argument." Consider the sequence

$$\begin{array}{ccccccc}
 C_1(X_1) & C_1(X_2) & C_1(X_3) & \cdots & C_1(X_{t-2}) & C_1(X_{t-1}) & C_1(X_t) \\
 C_2(X_1) & C_1(X_2) & C_1(X_3) & \cdots & C_1(X_{t-2}) & C_1(X_{t-1}) & C_1(X_t) \\
 C_2(X_1) & C_2(X_2) & C_1(X_3) & \cdots & C_1(X_{t-2}) & C_1(X_{t-1}) & C_1(X_t) \\
 & & & \vdots & & & \\
 C_2(X_1) & \cdots & C_2(X_i) & C_1(X_{i+1}) & C_1(X_{i+2}) & \cdots & C_1(X_t) \\
 C_2(X_1) & \cdots & C_2(X_i) & C_2(X_{i+1}) & C_1(X_{i+2}) & \cdots & C_1(X_t) \\
 & & & \vdots & & & \\
 C_2(X_1) & C_2(X_2) & C_2(X_3) & \cdots & C_2(X_{t-2}) & C_1(X_{t-1}) & C_1(X_t) \\
 C_2(X_1) & C_2(X_2) & C_2(X_3) & \cdots & C_2(X_{t-2}) & C_2(X_{t-1}) & C_1(X_t) \\
 C_2(X_1) & C_2(X_2) & C_2(X_3) & \cdots & C_2(X_{t-2}) & C_2(X_{t-1}) & C_2(X_t)
 \end{array}$$

(i.e. in each row, we apply C_2 to one more X_i).

The distance between row j and row $j+1$ is equal to $\delta(C_1, C_2)$ because everything on either side of the $j+1$ -th place is equal.

By the triangle inequality, $\delta(C_1^t, C_2^t) \leq t\delta(C_1, C_2)$.

For (ii), given Z_1, \dots, Z_t we want to tell whether it came from $C_1(X_1), \dots, C_1(X_t)$ or $C_2(X_1), \dots, C_2(X_t)$.

Define $f = \delta(C_1, C_2)$. By the definition of statistical distance, there exists a 0-1 valued test T and a value α such that

$$\begin{aligned}
 \mathbb{E}_{Z \sim C_1}[T(Z)] &\geq \alpha + f \\
 \mathbb{E}_{Z \sim C_2}[T(Z)] &\leq \alpha.
 \end{aligned}$$

Look at $T(Z_1) + \dots + T(Z_t)$. We'll use the test

$$\begin{aligned}
 &\text{If } T(Z_1) + \dots + T(Z_t) \geq (\alpha + f/2)t, \text{ report } C_1^t \\
 &\text{else, report } C_2^t.
 \end{aligned}$$

Exercise 1. Using Chernoff bounds, show that this test fails with exponentially small probability (with some repetition).

Solution. We might as well change T so that $\alpha = 0$. Then define the random variable $S = T(X_1) + \dots + T(X_t)$. For the probability that we report C_2 when the answer is C_1 , we want $\Pr[S < ft/2]$. The expected value of S in this case is ft . Using the Chernoff bound, this probability is less than $e^{-\frac{ft}{10}}$. Since $f \in [0, 1]$, this is less than $e^{-t^2/10}$.

On the other side, we can't let $\alpha = 0$ because then the Chernoff bound doesn't say anything. The probability that we report C_1 when the answer is C_2 is $\Pr[S > (\alpha + f/2)t]$ when $\mathbb{E}(S) = \alpha t$. The Chernoff bound then says the probability is less than $e^{-\frac{(f/2\alpha)^2 \alpha t}{3}} = e^{-\frac{f^2 t}{12\alpha}}$. We can shift α to be sufficiently small, then we have a total failure probability bounded by $e^{-\frac{f^2 t}{2}}$. \square

As a remark, this ingredient is where we lose tightness. For example, tc might be bigger than $1 - 2\exp(-t^2/2)$. We could avoid this problem if we used KL divergence, but here we have to use statistical distance.

Exercise 2. The “hybrid argument” method is common in cryptography. We say a poly time $G : \{0, 1\}^n \rightarrow \{0, 1\}^m$ is a (t, ε) pseudorandom number generator if for any t -time circuit A ,

$$|\Pr[A(G(S)) = 1] - \Pr[A(R) = 1]| < \varepsilon$$

where S is uniform on $\{0, 1\}^n$ and R is uniform on $\{0, 1\}^m$. Use a hybrid argument to show that if no t -time algorithm B can predict $G(S)_i$ given $G(S)_0, \dots, G(S)_{i-1}$ with better than ε/m probability, then G is a (t, ε) pseudorandom number generator.

Solution. Proof adapted from [Lyn]. Consider the distributions R_0, \dots, R_{m-1} where R_i is formed by taking the first $m - i + 1$ bits of $G(S)$ and then the last i bits of R (the truly random string).

We claim if we have a t -time circuit A such that $|\Pr[A(G(S)) = 1] - \Pr[A(R) = 1]| \geq \varepsilon$, then we can find a t -time circuit B such that $|\Pr[B(G(S)_i) = 1] - \Pr[B(R_i) = 1]| \geq \varepsilon/m$ for some i .

$$\begin{aligned} \varepsilon &\leq |\Pr[A(G(S)) = 1] - \Pr[A(R) = 1]| \\ &\leq |\Pr[A(G(S)) = 1] - \Pr[A(R_1) = 1]| \\ &\quad + \Pr[A(R_1) = 1] - \Pr[A(R_2) = 1] \\ &\quad + \dots + \Pr[A(R_{m-2}) = 1] - \Pr[A(R) = 1]| \\ &\leq \sum_{i=0}^{m-1} |\Pr[A(R_i) = 1] - \Pr[A(R_{i+1}) = 1]| \end{aligned}$$

so there must be some i such that $|\Pr[A(R_i) = 1] - \Pr[A(R_{i+1}) = 1]| \geq \varepsilon/m$. Therefore A is a circuit that can distinguish with probability at least ε between

$$G(S)_0, \dots, G(S)_i, R_{i+1}, \dots, R_{m-1}$$

and

$$G(S)_0, \dots, G(S)_{i+1}, R_{i+2}, \dots, R_{m-1}.$$

This means that, given $G(S)_0, \dots, G(S)_i$, it can guess $G(S)_{i+1}$ with probability ε better than random guessing. The bits from R after the first i bits of $G(S)$ don't matter because they are just uniform randomness. \square

4.2 Ingredient 1

To do this, we'll give another construction. If we have $(X_0, X_1), (Y_0, Y_1)$, we'll make (Z_0, Z_1) such that $\delta(Z_0, Z_1) = \delta(X_0, X_1) \cdot \delta(Y_0, Y_1)$.

Consider the "XOR" construction

$$\begin{aligned} Z_0 &= (X_0, Y_0) \text{ w.p. } \frac{1}{2} \\ &= (X_1, Y_1) \text{ w.p. } \frac{1}{2} \\ Z_1 &= (X_0, Y_1) \text{ w.p. } \frac{1}{2} \\ &= (X_1, Y_0) \text{ w.p. } \frac{1}{2}. \end{aligned}$$

We can verify that this has the desired property. The important step is that we can write

$$\begin{aligned} \Pr[Z_0 = (\alpha, \beta)] - \Pr[Z_1 = (\alpha, \beta)] &= \frac{1}{2}(\Pr[X_0 = \alpha] \Pr[Y_0 = \beta] + \Pr[X_1 = \alpha] \Pr[Y_1 = \beta] \\ &\quad - \Pr[X_0 = \alpha] \Pr[Y_1 = \beta] - \Pr[X_1 = \alpha] \Pr[Y_0 = \beta]) \\ &= \frac{1}{2}(\Pr[X_0 = \alpha] - \Pr[X_1 = \alpha])(\Pr[Y_0 = \beta] - \Pr[Y_1 = \beta]). \end{aligned}$$

Call this transformation *XOR*.

Then we can define $(C_1, C_2)^{\oplus t}$ recursively by $(C_1, C_2)^{\oplus t} = \text{XOR}((C_1, C_2)^{\oplus t-1}, (C_1, C_2))$ with a base case of $(C_1, C_2)^{\oplus 1} = (C_1, C_2)$.

Now we'll use the transformation

$$(C_1, C_2) \mapsto ((C_1, C_2)_0^{\oplus t}, (C_1, C_2)_1^{\oplus t}).$$

This will satisfy $\delta(D_1, D_2) = \delta(C_1, C_2)^t$.

4.3 Combining the ingredients

We start with an input for $SD^{1/3, 2/3}$. Then we can apply the following chain of transformations:

$$\begin{aligned} SD^{\frac{1}{3}, \frac{2}{3}} &\xrightarrow{\text{apply ingredient 1 with } t=O(\log n)} SD^{\frac{1}{n^2}, \frac{1}{n^{0.8}}} \\ SD^{\frac{1}{n^2}, \frac{1}{n^{0.8}}} &\xrightarrow{\text{apply ingredient 2 with } t=O(n^2)} SD^{\frac{1}{4}, 1-\exp(-n^{0.4})} \\ SD^{\frac{1}{4}, 1-\exp(-n^{0.4})} &\xrightarrow{\text{apply ingredient 1 with } t=O(n^{0.1})} SD^{\frac{1}{4n^{0.1}}, 1-\exp(-n^{0.3})} \end{aligned}$$

where the last step uses the approximation $1 - \exp(-x) \approx x$. Using this method (possibly with different constants), we can get $SD^{1/3, 2/3} \leq SD^{2^{-n^\epsilon}, 1-2^{-n^\epsilon}}$.

5 SD reduces to its complement

In 1996, Okamoto showed that $SD \leq \overline{SD}$ in a difficult-to-read paper. This was presented more clearly by Sahai and Vadhan. Good references for this topic are [Vad99] and [GV99].

We'll use another problem to prove this. Define the entropy difference problem ED^k that takes two distributions (C_1, C_2) . It should sort these into two categories

$$YES = \{H(C_1) \geq H(C_2) + k\}$$

$$NO = \{H(C_2) \geq H(C_1) + k\}.$$

We'll prove several properties of this problem:

0. $ED^{1/\sqrt{n}} \leq ED^{\sqrt{n}}$
1. $SD \leq ED$
2. $ED \leq \overline{ED}$
3. $ED \leq SD \Leftrightarrow \overline{ED} \leq \overline{SD}$

2. is easy, the reduction is $(C_1, C_2) \mapsto (C_2, C_1)$.

0. can be proved using the reduction $(C_1, C_2) \mapsto (C_1^t, C_2^t)$. This will amplify the entropy gap.

1. can be showed by making the transformation $(C_0, C_1) \mapsto (D_0, D_1)$ where

$$D_0 = (b, C_b(X)) \text{ where } b \text{ is selected at random}$$

$$D_1 = (b', C_b(X)) \text{ where } b, b' \text{ are selected at random.}$$

For the cases:

- i. $\delta(C_0, C_1) \leq 2^{-n^\epsilon}$
- ii. $\delta(C_0, C_1) \geq 1 - 2^{n^\epsilon}$

$H(b | C_b(X))$ is big in case i. because b doesn't give much information about $C_b(X)$. As $\epsilon \rightarrow 0$, $H(b | C_b(X))$ goes to $H(b) = 1$. Because $H(D_0) = H(b | C_b(X))$, if ϵ is very small, we can say $H(D_0)$ is close to $H(C_b(X)) + 1$.

In case ii., $H(b | C_b(X))$ is small because C_0, C_1 are far apart so we b is almost fully determined by $C_b(X)$. Therefore $H(D_0)$ in this case is close to just $H(C_b(X))$.

On the other hand $H(D_1) = H(C_b(X)) + 1$ always. Therefore we can solve SD by solving ED .

3. is the hardest property. There are things known as "extractors" that transform random variables with entropy k into the uniform distribution on $k - O(1)$ bits. On the other hand, if the entropy is small we can't get the uniform distribution on many bits using a deterministic transformation.

Some problems to solve

- We need to prove some things about expanders. This turns out to be straightforward.
- This works for min-entropy $H_\infty = \min_\omega \log \frac{1}{\Pr[X=\omega]}$ but we want regular entropy. This turns out to be ok because we can use the asymptotic equipartition principle to "flatten" the entropy so these are equivalent quantities.
- Even if we have a completely flat source, we still don't know the entropy of C_1 or C_2 .

The key idea to resolve this is to use a hash function. If $C_1, C_2 : \{0, 1\}^m \rightarrow \{0, 1\}^n$, choose a has function $h : \{0, 1\}^{m+n} \rightarrow \{0, 1\}^m$. Then we'll make the transformation

$$(C_1, C_2) \mapsto ((C_1(X), h, h(X, C_2(Y))), (C_1(X), h, Y))$$

for X, Y uniformly random over m bits.

These are only similar if there is enough entropy in $X, C_2(Y)$ for the hash function to approximate the uniform distribution over m bits. Of course, we're conditioned on $C_1(X)$.

$$H(X | C_1(X)) = m - H(C_1(X))$$

so we have in total

$$H(X, C_2(Y) | C_1(X)) = m - H(C_1(X)) + H(C_2(Y)).$$

From this it is clear that these distributions will be far apart if $H(C_1)$ is far from $H(C_2)$. Otherwise the hash function makes the output distributions similar. This completes the reduction.

Exercise 3. Show how we can flatten the distribution of C_1, C_2 by repeating independent copies. Use Hoeffding's inequality, a generalization of Chernoff bound, which says for X_1, \dots, X_n independent with mean μ and taking values in $[a, b]$, and $X = \frac{1}{n} \sum_{i=1}^n X_i$,

$$\Pr[|X - \mu| \geq \Delta] \leq 2 \cdot \exp\left(-\frac{2n\Delta^2}{(b-a)^2}\right)$$

for any $\Delta > 0$.

Solution. First, sorry for the impreciseness of the statement of this exercise. There should be several definitions here. This proof follows [Vad99].

For a distribution X , we say x is Δ -typical if $2^{-H(X)-\Delta} \leq \Pr[X = x] \leq 2^{-H(X)+\Delta}$.

Then we say a distribution X is Δ -flat if for all $t > 0$,

$$\Pr_{x \sim X}[x \text{ is } t\Delta \text{ typical}] \geq 1 - 2^{-t^2+1}.$$

With these definitions in place, we'll show that if X is a distribution such that for all $x \in \text{Supp}(X)$, $\Pr[X = x] \geq 2^{-m}$, then X^k (k independent copies of X) is $\sqrt{k} \cdot m$ -flat.

For x to be Δ typical, we want $H(X) - \Delta \leq \log \frac{1}{\Pr[X=x]} \leq H(X) + \Delta$. Therefore we can write

$$\begin{aligned} & \Pr_{(x_1, \dots, x_k) \sim X^k} [(x_1, \dots, x_k) \text{ is not } t\Delta \text{ typical}] = \\ & \Pr_{(x_1, \dots, x_k) \sim X^k} [H(X^k) - t\Delta \leq \log \frac{1}{\Pr[X^k = (x_1, \dots, x_k)]} \leq H(X^k) + t\Delta] = \\ & \Pr_{(x_1, \dots, x_k) \sim X^k} \left[\left| \log \frac{1}{\Pr[X^k = (x_1, \dots, x_k)]} - H(X^k) \right| \geq t\Delta \right] = \\ & \Pr_{(x_1, \dots, x_k) \sim X^k} \left[\left| \sum_{i=1}^k \log \frac{1}{\Pr[X = x_i]} - H(X^k) \right| \geq t\Delta \right]. \end{aligned}$$

But $H(X^k) = k \cdot H(X)$ and $\mathbb{E}_{x \sim X} \left[\log \frac{1}{\Pr[X=x]} \right] = H(X)$. Also, since $\Pr[X = x] \geq 2^{-m}$ for all x in the support of X , we know $\log \frac{1}{\Pr[X=x]}$ takes values in $[0, m]$. Therefore we can apply Hoeffding's inequality to say that

$$\Pr_{(x_1, \dots, x_k) \sim X^k} [(x_1, \dots, x_k) \text{ is not } t\Delta \text{ typical}] \leq 2 \cdot \exp\left(-\frac{2k(t\Delta/k)^2}{m^2}\right).$$

When we put in $\Delta = \sqrt{k}/m$, we get a bound on the failure probability of $\exp(-2t^2 + 1)$, which is sufficient for a bound of 2^{-t^2+1} as desired. \square

References

- [GV99] Oded Goldreich and Salil Vadhan. Comparing entropies in statistical zero knowledge with applications to the structure of SZK. In *Proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity*, 1999.
- [Lyn] Ben Lynn. Pseudo-random number generators.
- [Vad99] Salil Vadhan. *A Study of Statistical Zero-Knowledge Proofs*. PhD thesis, Massachusetts Institute of Technology, 1999.