CS 229r Information Theory in CS **April 18, 2019**

Lecture 22

Instructor: Madhu Sudan *Scribe: Jayshree Sarathy*

1 Administrative Notes

- Project check-in on Canvas and email or in person. Look at Piazza post for details.
- Project presentations should be less than 15 minutes each, including the time for transitions, Q&A, and projector trouble.
- Think carefully about what really needs to be in the presentation (ie. proofs should be skippable) and the division within the team.
- Try to link your presentations to the course material.

2 Today's Agenda

- Distributions in theoretical CS
- Sampleable distributions
- Statistical distance problem
- Statistical distance reduces to its complement
- Motivation and context for statistical distance problem

3 Distributions in TCS

The story of derandomization and randomness extraction starts with Von Neumann [8]. He considered the question: how can we conduct an unbiased coin toss using a biased coin? For example, consider the following biased coin that outputs

$$\begin{cases} \text{Heads} & \text{w.p. p} \\ \text{Tails} & \text{w.p. 1-p} \end{cases}$$

where p is unknown. The insight is that the event of getting heads and tails on consecutive tosses has probability $p(1-p)$, and the event of getting tails and heads on consecutive tosses also has probability $(1-p)p$. Thus, we can toss the biased coin repeatedly, only recording when we get one of these two events, which gives us an unbiased outcome.

Next, Blum [2] considered the problem of generating a large number of independent bits from a Markovian source. Instead of the strong independence in the previous problem, the independence here is weaker, as we have some hidden state.

Santha and Vazirani [5] took this idea further to consider "slightly random sources."

$$X_n | X_1, \ldots, X_{n-1}$$

Finally, Srinivasan and Zuckerman [6] considered min-entropy sources. Min-entropy is similar to the Shannon entropy, except that we replace the expectation with a minimum.

$$H_\infty(x) \triangleq \min_{w \in \{0,1\}^n} \log \frac{1}{\Pr[x = w]}$$

Since the expectation of a random variable is always greater than or equal to its minimum, we have that

$$H_\infty(x) \leq H(x)$$

Even if $H(x)$ is large, it might be insufficient for randomness extraction. For example, if we define x to be

$$x = \begin{cases} y & \text{w.p. } 1/2 \\ \text{Unif}(\Omega) & \text{w.p. } 1/2 \end{cases}$$

then the entropy is high, but the min-entropy is low, and thus x is not conducive for extracting randomness.

Exercise 1. *Show a second example of a random variable or distribution for which the min-entropy is low but the entropy is high.*

Therefore, Srinivasan and Zuckerman consider min-entropy sources where

$$H_\infty(x) \geq \varepsilon \cdot n$$

The ideas of randomized algorithms we have discussed are also used in cryptography, distributed algorithms, and game theory. For example, in cryptography, randomness is the key idea that allows us to encrypt a message with low entropy, such as a single bit. In distributed algorithms, randomness is required for n players to select a leader amongst themselves.

4 Sampleable Distributions

To define sampleable distributions, we consider Boolean circuits that take as input random bits and encode a sampling algorithm.

Definition 2 (Distributions encoded as circuits)**.** *Let $C : \{0,1\}^m \to \{0,1\}^n$ be a circuit with m inputs and n outputs, where $m, |C| = poly(n)$. The distribution associated with C is $Z = C(Y)$ when $Y \sim Bern(1/2)^m$.*

Note that circuits can be evaluated in time polyonomial to their size and express general computations. Since the circuits defined above are polynomial in size, they capture the notion of efficiently sampleable distributions.

5 Statistical Distance

Definition 3 (Promise Problem)**.** *A promise problem* Π *is a pair* (Π_Y, Π_N) *of disjoint sets of strings corresponding to YES and NO instances. Given a string* x *that is promised to be in* $\Pi_Y \cup \Pi_N$*, the problem is to decide whether* $x \in \Pi_Y$ *or* $x \in \Pi_N$*.*

Definition 4 (Statistical Difference)**.** *If* P *and* Q *are distributions (or random variables) supported on* Ω*, the statistical difference between* P *and* Q *is defined in two ways.*

$$\begin{aligned} \delta(P,Q) &\triangleq \frac{1}{2}\sum_{w\in\Omega} |P(w) - Q(w)| \\ &\triangleq \max_{T:\Omega\to\{0,1\}} E_{X\sim P}[T(x)] - E_{x\sim Q}[T(Y)] \end{aligned}$$

The statistical difference problem is a promise problem $\mathrm{SD}^{c,f}$ of deciding whether a pair of efficiently sampleable distributions is statistically close or far, as measured by statistical difference. The parameters for the problem are c and f, where $0 \le c \le f \le 1$, which are used to define the sets CLOSE and FAR.

$$\begin{aligned} \mathrm{SD}_Y^{c,f} &= \mathrm{CLOSE}^c = \{(c_1,c_2) : \delta(c_1,c_2) \le c\} \\ \mathrm{SD}_N^{c,f} &= \mathrm{FAR}^f = \{(c_1,c_2) : \delta(c_1,c_2) > f\} \end{aligned}$$

Definition 5 (Statistical Difference Problem)**.** *Given input* $(c_1,c_2) \in \mathrm{CLOSE}^c \cup \mathrm{FAR}^f$*, the statistical difference problem* $\mathrm{SD}^{c,f}$ *is to decide whether the input is in* CLOSE^c *or whether it is in* FAR^f*.* SD *returns YES if the input is in* CLOSE^c *and NO if the input is in* FAR^f*.*

We do not know how to solve $\mathrm{SD}^{1/3,2/3}$ efficiently. Only recently have we found a near-poly-time algorithm for the problem of Graph Isomorphism (GI), which is contained in $\mathrm{SD}^{1/3,2/3}$. We suspect that there do not exist efficient algorithms for $\mathrm{SD}^{1/3,2/3}$.

Next, we ask whether we can amplify $\mathrm{SD}^{1/3,2/3}$, which has relatively close statistical difference, into $\mathrm{SD}^{2^{-\sqrt{n}},1-2^{-\sqrt{n}}}$, which has very large statistical distance. That is, can we give a Karp reduction from $\mathrm{SD}^{1/3,2/3}$ to $\mathrm{SD}^{2^{-\sqrt{n}},1-2^{-\sqrt{n}}}$? The answer is yes, from a result by Sahai and Vadhan [4]. Whether we can amplify $\mathrm{SD}^{.49,.5}$ into $\mathrm{SD}^{1/3,2/3}$, however, is still an open problem. A good reference for this is the survey by Goldreich and Vadhan [3].

Definition 6 (Complement of Statistical Difference Problem)**.** $\overline{\mathrm{SD}}^{c,f}$ *returns NO if* (c_1,c_2) *are close and YES if they are far.*

It turns out that SD is Karp-reducible to its complement ($\mathrm{SD} \equiv \overline{\mathrm{SD}}$).

Theorem 7 ($SD \equiv \overline{\mathrm{SD}}$)**.** *There exists a poly-time reduction* R *such that, when given* (c_1,c_2)*, returns* (d_1,d_2) *such that*

- *if* (c_1,c_2) *are close, then* (d_1,d_2) *are far.*
- *if* $(c_1,c_2$ *are far, then* (d_1,d_2) *are close.*

6 Motivation and Context

Where does the notion of statistical distance come up? One important motivation is from work in cryptography and zero-knowledge by Goldwasser, Micali, and Rackoff. They define a complexity class called statistical zero knowledge (SZK) and show that Graph Non-Isomorphism (GNI) is in SZK. Before describing this result, we define zero knowledge proofs, the class SZK, and the problems Graph Isomorphism (GI) and Non-Isomorphism (GNI).

6.1 Statistical Zero Knowledge

Definition 8 (Interactive Proof)**.** *Let (P, V) be an interactive protocol and let Π be a promise problem. (P, V) is an interactive proof system for Π if the following conditions hold. For input x and security parameter k:*

1. *(Efficiency) (P, V) is polynomially bounded and V is polynomial time computable.*
2. *(Completeness) If $x \in \Pi_Y$, then V accepts with probability at least $1 - c(k)$*
3. *(Soundness) If $x \notin \Pi_Y$, then for all P^*, V rejects with probability at least $s(k)$*

where $c(k)$ and $s(k)$ are computable in time $poly(k)$ and $1 - c(k) > s(k) + 1/poly(k)$.

The completeness and soundness of the protocol can be amplified to $1 - 2^{-k}$ through repetition. Now, we will define simulators and zero-knowledge.

Definition 9 (Simulator)**.** *A simulator S is an algorithm that simulates the verifier's view of the interaction with the prover. Let (P, V) be an interactive protocol. V's view of the interaction on common input x is the random variable $\langle P, V\rangle(x, 1^k) = (m_1, \ldots, m_t; r)$ containing all messages $m_1, \ldots, m_t$ exchanged between P and V as well as r_V, the substring of r of the random bits V has read through the interaction.*

Statistical zero knowledge requires that the statistical difference between the simulator's output distribution and the verifier's view is negligible. We will allow the simulator to fail, and only measure the quality of the simulation on non-failure. We call the simulator useful if it fails less than half of the time, and we let $\tilde{S}(x, 1^k)$ be the output distribution of S on x, conditioned on non-failure of S.

Definition 10 (Honest-Verifier Statistical Zero Knowledge)**.** *An interactive proof system (P, V) for a promise problem Π is honest-verifier statistical zero knowledge if there exists a useful probabilistic poly-time algorithm S and a negligible function $\mu(\cdot)$ such that for all $x \in \Pi_Y$ and all $k > 0$:*

$$\mathit{StatDiff}(\tilde{S}(x, 1^k), \langle P, V\rangle(x, 1^k)) \leq \mu(k)$$

6.2 Graph Non-Isomorphism

Definition 11 (Isomorphic graphs)**.** *Let $G_0 = (V, E)$ be an undirected graph, where $V = [n]$. Let $\pi : [n] \to [n]$ be a permutation on V, and let $\pi(G_0)$ denote the graph (V, E'), where $E' = \{(\pi(u), \pi(v)) : (u, v) \in E)\}$. Let $G_1 = (V, F)$ be a second graph that has the same vertex set as G_0. If there exists a π such that $G_1 = \pi(G_0)$, then G_0 and G_1 are isomorphic, denoted by $G_0 \cong G_1$.*

Definition 12 (GI and GNI)**.** *Graph isomorphism is the language* $\mathrm{GI} = \{G_0, G_1 : G_0 \cong G_1\}$. *Graph non-isomorphism is the complement of* GI.

GI $\in NP$, so graph isomorphism can be proved easily; the certificate is simply the isomorphism, which can be checked efficiently. Graph non-isomorphism, however, is harder. There are no classical proofs known for GNI.

However, there is an efficient interactive proof for GNI, which relies on two observations. Let S_0 and S_1 be the set of graphs that are isomorphic to G_0 and G_1, respectively.

1. If $G_0 \ncong G_1$, then S_0 is disjoint from S_1.
2. If $G_0 \cong G_1$, then a uniform draw from S_0 is indistinguishable from a uniform draw from S_1.

Thus, the interactive proof for GNI below tests whether the prover can distinguish between a uniform draw from S_0 versus S_1. Note that instead of testing all possible permutations of G_b, the verifier only has to test one randomly sampled permutation π to get a convincing answer.

Verifier		**Prover**
$b \leftarrow_\$ \{0,1\}$ $\pi \leftarrow_\$ \Sigma_n$ $K = \pi(G_b)$		
	$\xrightarrow{\quad K \quad}$	
		If $G_0 \cong K$, set $\hat{b} = 0$ Else, set $\hat{b} = 1$
	$\xleftarrow{\quad \hat{b} \quad}$	
Accept if $\hat{b} = b$. Otherwise, reject.		

As defined in the previous section, SZK is the class of interactive protocols for promise problems with a probabilistic, poly-time verifier which satisfy completeness, soundness, and statistical zero-knowledge. Below, we informally outline these properties in the context of GNI. To show that GNI $\in$ SZK, one must show that the following properties hold.

- Completeness: If $G_0 \ncong G_1$, then an unbounded prover can find $\hat{b} = b$, and the verifier will accept with probability 1.
- Soundness: If $G_0 \cong G_1$, then $I(K; b | G_0, G_1) = 0$ which implies that $\{K | b = 0\} \cong_d \{K | b = 1\}$. The verifier should accept with probability negligibly close to 1/2.
- Zero-Knowledge: The verifier should learn nothing from the interaction other than the statement that $G_0 \ncong G_1$.

Exercise 13. *Show (informally) that the interactive* GNI *protocol fulfills the first two properties.*

- Completeness: If $G_0 \ncong G_1$, then S_0 is disjoint from S_1. An unbounded prover can enumerate through S_0 and S_1 to check which of the two sets K lies in. Thus, the prover will correctly find $\hat{b} = b$ with probability 1, so the verifier will accept with probability 1.

- Soundness: If $G_0 \cong G_1$, let us consider how much information is given by K about b. First, we can compute the mutual information.

$$I(K; b) = H(K) - H(K|b) = H(K) - H(K) = 0$$

  Since $S_0 \cong S_1$, the distribution of K does not change based on b. Thus, the mutual information between K and b is 0. By symmetry, this implies that $H(b|K) = H(b) = 1$, so the prover can only guess $\hat{b} = b$ with probabilitity negligibly close to $1/2$.

We will now describe some main results in this area. First, $\mathrm{SD}^{1/3,2/3} \in \mathrm{SZK}$, and is in fact SZK-complete. Second, $\mathrm{GNI} \leq \overline{SD}^{0,1}$, and by a simple negation, $\mathrm{GI} \leq SD^{0,1}$. By the equivalence of SD and $\overline{\mathrm{SD}}$, the first result implies that $\overline{\mathrm{SD}}^{1/3,2/3}$ also has an SZK proof. Since $\mathrm{GI} \leq \mathrm{SD}^{1/3,2/3}$, GI has an SZK proof as well.

Exercise 14. *Show (informally) that* $\overline{SD}^{2^{-\sqrt{n}},1-2^{-\sqrt{n}}}$ *has an* SZK *proof.*

Let X_0, X_1 be the distributions encoded by input circuits $c_0, c_1 \in \{0,1\}^m \to \{0,1\}^n$. The SZK proof protocol is illustrated below:

Verifier		**Prover**
$b \leftarrow_\$ \{0,1\}$ $r \leftarrow_\$ \{0,1\}$ $x = X_b(r)$		
	$\xrightarrow{x}$	
		If $\Pr[X_0 = x] \geq \Pr[X_1 = x]$, set $\hat{b} = 0$ Else, set $\hat{b} = 1$
	$\xleftarrow{\hat{b}}$	
Accept if $\hat{b} = b$. Otherwise, reject.		

- Completeness: If $c_1, c_2 \in \mathsf{FAR}^{1-2^{-\sqrt{n}}}$, then the mass of X_0 and X_1 at x will be drastically different, so the prover will guess $\hat{b} = b$ with overwhelming probability.

- Soundness: If $c_1, c_2 \in \mathsf{CLOSE}^{2^{-\sqrt{n}}}$, then the mass of X_0 and X_1 at x will be very similar, so the prover can only guess $\hat{b} = b$ with probability negligibly close to $1/2$.

- Zero-knowledge: The only information that is revealed to the verifier by the interaction is $\hat{b}$, which is a guess for a value b that the verifier already knows.

References

[1] Goldwasser, S., Micali, S. and Rackoff, C., 1989. The knowledge complexity of interactive proof systems. SIAM Journal on computing, 18(1), pp.186-208.

[2] Blum, M., 1986. Independent unbiased coin flips from a correlated biased source—a finite state Markov chain. Combinatorica, 6(2), pp.97-108.

[3] Goldreich, O. and Vadhan, S., 2011. On the complexity of computational problems regarding distributions. In Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation (pp. 390-405). Springer, Berlin, Heidelberg.

[4] Sahai, A. and Vadhan, S., 2003. A complete problem for statistical zero knowledge. Journal of the ACM (JACM), 50(2), pp.196-249.

[5] Santha, M. and Vazirani, U.V., 1986. Generating quasi-random sequences from semi-random sources. Journal of computer and system sciences, 33(1), pp.75-87.

[6] Srinivasan, A. and Zuckerman, D., 1999. Computing with very weak random sources. SIAM Journal on Computing, 28(4), pp.1433-1459.

[7] Vadhan, S.P., 1999. A study of statistical zero-knowledge proofs (Doctoral dissertation, Massachusetts Institute of Technology).

[8] Von Neumann, J., 1951. 13. various techniques used in connection with random digits. Appl. Math Ser, 12(36-38), p.5.