7. Distributional Communication Complexity

Lecturer: Prahladh Harsha *Scribe: Pranabendu Misra*

In today's lecture, we will introduce distributional complexity and discrepancy, methods used to bound randomized communication complexity. The main reference for today's lecture are §3.3–3.5 of Kushilevitz and Nisan's book [KN97].

7.1 Public coins vs. private coins

In randomized communication complexity, public coin protocols are at least as efficient as private coin protocols. This is because any private coin protocol can be simulated by a public coin protocol. However, public coin protocol can have lower cost than private coin ones, as seen from the equality function EQ. We have seen that

$$\begin{aligned} \mathrm{R}^{\mathrm{pub}}_{1/3}(\mathsf{EQ}_n) &= \Theta(1), \\ \mathrm{R}_{1/3}(\mathsf{EQ}_n) &= \Theta(\log n). \end{aligned}$$

Thus, there is a gap of $\Omega(\log n)$ between public coin protocols and private coin protocols. The following theorem show that this is the largest possible gap between the two.

Theorem 7.1 (public vs. private)**.** *For any function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $\varepsilon, \delta > 0$,*

$$\mathrm{R}_{\varepsilon+\delta}(f) \leq \mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) + O\left(\log n + \log\left(\frac{1}{\delta}\right)\right).$$

See [KN97, §3.3] for a proof of this statement.

7.2 Distributional communication complexity

In this section, we introduce a new communication cost called the distributional communication complexity, which will become useful while obtaining lower bounds on randomized communication complexity.

Here we consider a probability distribution μ over the set of inputs $\mathcal{X} \times \mathcal{Y}$. We are interested in the best (least communication) deterministic protocol P that has an error at most ε when the input is drawn randomly from the distribution μ. In other words, the deterministic protocol P errs on at most ε fraction of the inputs, when the inputs are weighted according to the distribution μ. I.e.,

$$\Pr_{(X,Y)\sim\mu}[P(X,Y) \neq f(X,Y)] \leq \varepsilon.$$

Definition 7.2 (distributional communication complexity)**.** *The distributional communication complexity of a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ over the distribution μ and with error at most ε, denoted by $\mathrm{D}^{\mu}_{\varepsilon}(f)$, is the cost of the best deterministic protocol that gives the correct answer on at least $(1-\varepsilon)$ fraction of all the inputs, weighted by μ.*

Here, as before, the cost of a protocol is the worst case (over all inputs) number of bits exchanged by the protocol.

Thus $\mathrm{D}^{\mu}_{\varepsilon}(f) = \min_{P \in \mathcal{P}} \mathrm{cost}(P)$ where $\mathcal{P}$ is the collection of protocols P satisfying

$$\Pr_{(X,Y)\sim\mu}[P(X,Y) \neq f(X,Y)] \leq \varepsilon$$

and for any protocol P, $\mathrm{cost}(P) = \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ communication on input (x,y).

Example: greater-than (GT) function

$$\mathsf{GT}(x,y) = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{otherwise} \end{cases},$$

where $x, y \in \{0,1\}^n$ are viewed as the binary representations of numbers in $\{0,\ldots,2^n-1\}$. Let unif be the uniform distribution on the inputs. Consider the protocol where Alice send only the the most significant bit of x, $MSB(x)$, to Bob, and Bob accepts if $MSB(x) \geq MSB(y)$. This protocol gives the correct answer on at least $\frac{3}{4}$ fraction of the inputs. Hence, $\mathrm{D}^{\mathrm{unif}}_{\frac{1}{4}}(\mathsf{GT}) = O(1)$.

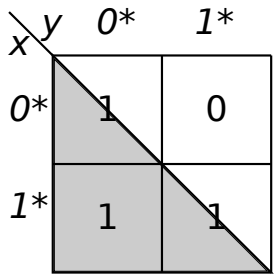

Figure 1: The GT function.

It is easy to check that the randomized communication complexity of a function $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)$ upper bounds the distributional complexity $\mathrm{D}^{\mu}_{\varepsilon}(f)$ for every distribution μ. The following theorem shows that the other direction is true as well. i.e. there exists a distribution μ such that $\mathrm{D}^{\mu}_{\varepsilon}(f) \geq \mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)$. Thus, distributional complexity completely characterizes randomized complexity.

Theorem 7.3. $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) = \max_{\mu} \mathrm{D}^{\mu}_{\varepsilon}(f)$.

Proof. It is easy to see that for all distributions μ, $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) \geq \mathrm{D}^{\mu}_{\varepsilon}(f)$. Let μ be any distribution on the inputs $\mathcal{X}\times\mathcal{Y}$. Let P be a randomized protocol for f, with cost $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)$, and let R denote the random string used by the protocol P. Then for all pairs of strings (x,y), we have,

$$\Pr_{R}[P(x,y;R) = f(x,y)] \geq 1-\varepsilon.$$

In particular, we have

$$\Pr_{R,(X,Y)\sim\mu}[P(X,Y;R) = f(X,Y)] \geq 1-\varepsilon.$$

Therefore, there exists a fixed string r such that

$$\Pr_{(X,Y)\sim\mu}[P(X,Y;r) = f(X,Y)] \geq 1-\varepsilon.$$

So we run the protocol P while fixing the above r, to get a deterministic protocol which is correct in at least $(1-\varepsilon)$ fraction of the cases, relative to distribution μ. Hence, $\mathrm{D}^{\mu}_{\varepsilon}(f) \leq \mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f)$.

Now, for the other direction. Suppose $\max_{\mu}\{\mathrm{D}^{\mu}_{\varepsilon}(f)\} \leq c$, we will show that $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) \leq c$. Consider the following 2 player zero-sum game. Player 1 chooses a deterministic protocol P for f of cost c (and whatever error), and Player 2 chooses an input $(x,y) \in \mathcal{X}\times\mathcal{Y}$. Both players make their choices in parallel, so that neither is aware of the other's choice.

The payoff for Player 1 is $\begin{cases} 1 & \text{if } P(x,y) = f(x,y) \\ 0 & \text{otherwise} \end{cases}$.

Now, the fact that $\mathrm{D}^{\mu}_{\varepsilon}(f) \leq c$ for every μ implies that for every randomized strategy of Player 2 (i.e., for every probability distribution μ over $\mathcal{X}\times\mathcal{Y}$), Player 1 can obtain expected payoff $(1-\varepsilon)$ using the protocol P of cost $\mathrm{D}^{\mu}_{\varepsilon} \leq c$. By the min-max theorem for zero-sum games (see Appendix A for details), Player 1 has a randomized strategy, with an expected payoff of $(1-\varepsilon)$ for every choice of inputs of Player 2. Now note that a randomized strategy for Player 1 is a distribution over cost c deterministic protocols, ie., a public coins protocol of cost at most c. Thus, there exists a public coin protocol of cost at most c that is correct on every input with probability at least $1-\varepsilon$. Hence, $\mathrm{R}^{\mathrm{pub}}_{\varepsilon}(f) \leq \max_{\mu}\mathrm{D}^{\mu}_{\varepsilon}(f)$. □

The above theorem can be used to prove lower bounds on randomized communication complexity. To do so, we need to come up with a suitable probability distribution μ and then prove lower bounds on $\mathrm{D}^{\mu}_{\varepsilon}(f)$.

7.3 Discrepancy

Recall the rectangle method for showing lower bounds for deterministic communication complexity $\mathrm{D}(f)$. By showing that there are no large monochromatic rectangles in M_f, we showed that the number of rectangles in a monochromatic rectangle cover of M_f must be large. Hence, the deterministic communication complexity must be large.

We can do something similar in the case of distributional communication complexity. Let μ be a distribution on the inputs $\mathcal{X}\times\mathcal{Y}$. First consider rectangles $S\times T$, where $S\subseteq\mathcal{X}$, $T\subseteq\mathcal{Y}$. We call a rectangle unbalanced, if it is mostly 1s or 0s (with respect to μ). To show that the communication complexity must be large, it suffices to show that any unbalanced rectangle must be small, or equivalently, any large rectangle must be balanced. This implies that most rectangles must be small, since otherwise the protocol will make an error on a large fraction of inputs. Most rectangles being small implies that there is a large number of rectangles and hence the distributional complexity is large. This notion of "large unbalanced rectangles" is captured in the following definition.

Definition 7.4 (discrepancy)**.** *Let μ be a distribution on $\mathcal{X}\times\mathcal{Y}$. Then the discrepancy of f on a rectangle $R = S\times T$ under distribution μ, denoted by* $\mathrm{disc}_{\mu}(f;R)$*, is defined as follows:*

$$\mathrm{disc}_{\mu}(f;R) = \left|\Pr_{\mu}[(X,Y)\in R \text{ and } f(X,Y)=1] - \Pr_{\mu}[(X,Y)\in R \text{ and } f(X,Y)=0]\right|.$$

The discrepancy of f under μ is defined as

$$\mathrm{disc}_\mu(f) = \max_R \mathrm{disc}_\mu(f; R).$$

The following theorem shows how discrepancy can be used to lower bound distributional complexity.

Theorem 7.5. *For all distributions μ and $\forall \varepsilon \in (0, \frac{1}{2})$,*

$$\mathrm{D}^\mu_{\frac{1}{2}-\varepsilon}(f) \geq \log_2 \left(\frac{2\varepsilon}{\mathrm{disc}_\mu(f)} \right).$$

Proof. Let P be a deterministic protocol for f of cost $c = \mathrm{D}^\mu_{1/2-\varepsilon}(f)$ that is correct on at least $(\frac{1}{2} + \varepsilon)$-fraction of the inputs, weighted according to μ. Let L be the set of leaves of the protocol tree of P, and for $l \in L$, let R_l be the corresponding rectangle. We can assume that for each leaf l, the protocol P labels the leaf R_l with 0 if the weight of 0s in R_l exceeds that of 1s (ie. $\mu(R_l \cap f^{-1}(0)) \geq \mu(R_l \cap f^{-1}(1))$), and with 1 otherwise. (Otherwise, flipping the label only reduces the error.)

$$\begin{aligned}
2\varepsilon &\leq \Pr_\mu [P(X,Y) = f(X,Y)] - \Pr_\mu [P(X,Y) \neq f(X,Y)] \\
&= \sum_{l \in L} \left(\Pr_\mu [P(X,Y) = f(X,Y) \text{ and } (X,Y) \in R_l] - \Pr_\mu [P(X,Y) \neq f(X,Y) \text{ and } (X,Y) \in R_l] \right) \\
&\leq \sum_{l \in L} \left| \Pr_\mu [P(X,Y) = 1 \text{ and } (X,Y) \in R_l] - \Pr_\mu [P(X,Y) = 0 \text{ and } (X,Y) \in R_l] \right| \\
&= \sum_{l \in L} \mathrm{disc}_\mu(f; R_l) \\
&\leq \sum_{l \in L} \mathrm{disc}_\mu(f) \\
&\leq 2^c \, \mathrm{disc}_\mu(f)
\end{aligned}$$

Now taking $\log_2$ on both sides, we get the result. □

7.4 The Inner Product function

The inner product function is defined as follows:

$$\mathsf{IP}(x, y) = \langle x, y \rangle \pmod 2 = \sum_{i=1}^{n} x_i y_i \pmod 2.$$

Let H be a $2^n \times 2^n$ matrix defined as $H = J - 2M_{\mathsf{IP}}$, where J is the all-1s matrix. In other words, $H(x, y) = 1$ if $\langle x, y \rangle = 0 \pmod 2$ and $H(x, y) = -1$ otherwise. It is easy to check that the matrix H satisfies $HH^T = H^T H = 2^n I$. Hence the spectral norm of H, $\|H\|$, is $\sqrt{2^n}$. We will now use this fact to bound the discrepancy of IP with respect to the uniform distribution unif.

Let $S \times T$ be a rectangle. Then,

$$\begin{aligned}
\mathrm{disc}_{\mathrm{unif}}(\mathsf{IP}; S \times T) &= \frac{1}{2^{2n}} \left| \sum_{x \in S, y \in T} H(x,y) \right| \\
&= \frac{1}{2^{2n}} \left| \mathbb{1}_S \cdot H \cdot \mathbb{1}_T \right| \qquad [\text{where } \mathbb{1}_S, \mathbb{1}_T \text{ are characteristic vectors of } S, T] \\
&\leq \frac{1}{2^{2n}} \|\mathbb{1}_S\| \cdot \|H\| \cdot \|\mathbb{1}_T\| \\
&\leq \frac{1}{2^{2n}} \sqrt{|S| \cdot 2^n \cdot |T|} \\
&\leq \frac{2^{3n/2}}{2^{2n}} \\
&= 2^{-n/2}
\end{aligned}$$

Now using Theorem 7.3 and Theorem 7.5 we have,

$$\mathrm{R}^{\mathrm{pub}}_{\frac{1}{2}-\varepsilon}(\mathsf{IP}) \geq \mathrm{D}^{\mathrm{unif}}_{\frac{1}{2}-\varepsilon}(\mathsf{IP}) \geq \frac{n}{2} - \log_2\left(\frac{1}{2\varepsilon}\right).$$

References

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997. doi:10.2277/052102983X.

A von Neumann's min-max theorem

Consider a 2-player game as follows. There is an $M \times N$ payoff matrix A. The M rows index ways in which player 1 can make a move, and the N rows index the ways in which player 2 can make a move. Both players simultaneously choose a move, say row i and column j. Then player 1 has to give player 2 the amount $A_{i,j}$.

Now allow the players to be randomized. A strategy for player 1 (player 2) is a distribution σ on the rows (μ on the columns). Each player chooses a move according to his/her strategy. The expected payoff is $\sigma^T A \mu$. The goal of player 1 is to choose a σ that minimizes, over the worst choice of μ, the expected payoff $\sigma^T A \mu$. The goal of player 2 is to choose a μ that maximizes, over the worst choice of σ, the expected payoff $\sigma^T A \mu$. The celebrated min-max theorem says that if a player announces his/her strategy and allows the other player to make an adversarial choice, the expected payoff is the same no matter which player announces the strategy. That is,

$$\min_{\sigma} \max_{\mu} \sigma^T A \mu = \max_{\mu} \min_{\sigma} \sigma^T A \mu.$$

In the setting we consider, the rows of A index deterministic protocols with cost at most c. The columns index inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Thus μ is a distribution on inputs, and σ is a distribution on deterministic cost c protocols ie. a randomised cost c protocol. $A(P, xy)$ is 1 if $P(x, y) = f(x, y)$, 0 otherwise.

Since $\forall \mu$, $\mathrm{D}^{\mu}_{\varepsilon}(f) \leq c$, we see that $\max_{\mu} \min_{\sigma} \sigma^T A \mu \leq \varepsilon$. (For each μ, find the deterministic cost c protocol P that works for it, and set σ to be 1 on P and 0 elsewhere. Then $\sigma^T A \mu$ is exactly the error of P with respect to μ.) By the min-max theorem, $\min_{\sigma} \max_{\mu} \sigma^T A \mu \leq \varepsilon$. Consider the distribution σ that achieves this minimum. Then the corresponding randomised protocol P_σ has error at most ε on every distribution. In particular, for any input (x, y), let μ_{xy} be the distribution that is 1 at (x, y) and 0 elsewhere; P_σ has error at most ε on μ_{xy}. Thus for every input (x, y), P_σ has error at most ε.

B Spectral norm of a matrix

Let A be a real matrix. The spectral norm of A, denoted $\|A\|_2$, is defined as follows:

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x \neq 0} \frac{\sqrt{\langle Ax, Ax \rangle}}{\sqrt{\langle x, x \rangle}}$$

(In the main text, we drop the subscript for notational convenience.)

For a square matrix X, $\sigma_{\max}(X)$ denotes the largest singular value of X, and $\lambda_{\max}(X)$ denotes the largest eigenvalue of X. If X is symmetric, then $X^T X$ is positive semidefinite (that is, $\forall x$, $x^T A x \geq 0$). If A is symmetric, then it can be shown that

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} = \sigma_{\max}(A).$$