CS 229r Information Theory in Computer Science	March 26, 2019

Lecture 15

Instructor: Madhu Sudan *Scribe: Duncan Rheingans-Yoo*

1 Topics For Today:

- Compressing Interactive Communication
- (Along the way): Correlated Sampling

2 Review of Protocol

Setup:

- Two communicators Alice, Bob
- Input $(x, y) \sim \mu$: x observed by Alice, y observed by Bob

Protocol:
Alice and Bob want a protocol in place to communicate about their inputs. They are allowed randomness:

- Public randomness R observed by both
- Private randomness R_A observed by Alice and R_B observed by Bob

Alice sends first bit of transcript $\pi_1(x, R, R_A) \in \{0, 1\}$. Bob observes π_1 and sends second bit $\pi_2(y, R, R_B, \pi_1) \in \{0, 1\}$. In general, π_i is a function of:

- $x, R, R_A, \pi_{<i}$, i odd (Alice sends)
- $y, R, R_B, \pi_{<i}$, i even (Bob sends)

Such an interaction is illustrated by the diagram below:
INSERT DIAGRAM

3 Internal Information Complexity

Definition 1 (Internal Info Complexity)**.** The *internal information complexity* of a protocol π is given by

$$\begin{aligned} IC_\mu^{int}(\pi) =& I(x; \pi | y, R) + I(y; \pi | x, R) \\ =& \sum_{i=1}^{k} I(\pi_i; x | y, R, \pi_{<i}) + I(\pi_i; y | x, R, \pi_{<i}) = \sum_{i=1}^{k} V_i \end{aligned}$$

Intuitively, it is the amount of information the protocol conveys to Alice and Bob about each others' inputs.

Definition 2 (External Info Complexity)**.** The *external information complexity* of a protocol π is given by

$$IC_\mu^{ext}(\pi) = I(xy; \pi | R)$$

Intuitively, it is the amount of information the protocol conveys to an outside observer about x and y.

Exercise 3. *Using the fact that only one of $I(\pi_i; x|y, R, \pi_{<i})$ and $I(\pi_i; y|x, R, \pi_{<i})$ can be nonzero for a given i, show that $IC_\mu^{int}(\pi) \leq IC_\mu^{ext}(\pi)$*

Exercise 4. *Show that $I(x; \pi, R|y) + I(y; \pi, R|x) = I(x; \pi|y, R) + I(y; \pi|x, R)$*

Definition 5 (Protocol Simulation)**.** Protocol π', consisting of public randomness R', private randomness R'_A, R'_B, functions π'_i defined as before, and output functions O_A, O_B is said to *simulate* protocol π if:

- $O_A = O_B = (R, \pi)$
- The distributions of x, y, π, R are preserved

4 Today's Main Compression Theorem

Theorem 6 (BBCR)**.** *$\forall\ \pi$ with $CC(\pi) = k$ and $IC(\pi) = I$, $\exists\ \pi'$ simulating π with $CC(\pi') = O(\sqrt{I \cdot k} \cdot \log k)$*

As an aside, there are a couple more theorems we will present without proof:

Theorem 7. *$\exists\ \pi'$ with $CC(\pi') = 2^{O(I)}$*

Theorem 8. *The above are tight.*

The rest of today will be about building to a proof of Theorem 6.
Aside:

- $CC(\pi^{\otimes n}) \geq CC(\pi)\sqrt{n}$
- $CC(\pi^{\otimes n}) = h \cdot IC(\pi)(1 + o(1))$ (I don't really understand what these two lines are; I just copied them)

5 Protocols, Priors, Information Cost

We now take a closer look at protocols, to get us to the point where we cam prove Theorem 6. Assume π has no common randomness. We can conceptualize the interactive communication as progressing from the root of a tree to a leaf, as visualized below:
INSERT TREE DIAGRAM
The position in the tree encodes $\pi_{<i}$. For a node u, let $P_u^A = \pi_i|\pi_{<i}, x$ be what Alice thinks will happen next and let $P_u^B = \pi_i|\pi_{<i}, y$ be what Bob thinks will happen next.

Let's begin by thinking about the special case of $I = 0$.

Claim 9. *$I = 0$ only if $\forall\ u$, we have $P_u^A = P_u^B$. In this case, we can simulate the entire path from root to leaf using common randomness R and zero communication*

Now let's think about $I \neq 0$.
Goal: Sample root to leaf path according to the $\{P_u\}_u$, or equivalently, sample leaf according tyo the right distribution on leaves.

6 Correlated Sampling

Consider the following setting:

- Alice observes the realization of a r.v. P, and Bob observes the realization of a r.v. Q.
- Alice and Bob can utilize some public randomness R

- Without communicating, Alice must create output some ω_A an Bob must create some output ω_B
- **Goal:** Want $\omega_A \sim P$, $\omega_B \sim Q$, $\min Pr[\omega_A \neq \omega_B]$

This is illustrated in the diagram below:
NEED DIAGRAM
If P and Q have disjoint supports, $Pr[\omega_A \neq \omega_B] = 1$ necessarily. In the previous section, we saw that if P and Q have the same distribution, common randomness gives $\min Pr[\omega_A \neq \omega_B] = 0$. We are interested in some interpolation between these special cases, where P and Q share support Ω but may not be exactly the same.

Definition 10. $\delta(P,Q) = \frac{1}{2}\sum_{\omega\in\Omega}|P(\omega) - Q(\omega)|$

Exercise 11. *Show that* $\forall\ P,Q,\ \min Pr[\omega_A \neq \omega_B] \geq \delta(P,Q)$

Lemma 12. $\exists$ *protocol s.t.*

$$Pr[\omega_A \neq \omega_B] \leq \frac{2\delta(P,Q)}{1+\delta(P,Q)} \leq 2\delta(P,Q)$$

Proof Idea: Let the public randomness R sample from the $\Omega \times [0,1]$ grid many times. Each point given by (a_i, b_i) for $a_i \in \Omega$ and $b_i \in [0,1]$. Alice outputs $\omega_A = a_i$ for i the first point for which $b_i \leq P(a_i)$. Bob outputs $\omega_B = a_j$ for j the first point for which $b_j \leq Q(a_j)$. This is illustrated by the diagram below:
NEED DIAGRAM
Note that if P, Q are very similar, $\omega_A = a_i = a_j = \omega_B$ with high probability. In fact, the first inequality of Lemma 12 will follow from this protocol.$\blacksquare$

Let $P_u^A = Bern(\frac{1}{2} - \delta)$ and $P_u^B = Bern(\frac{1}{2})$ for all u. Then by Union-Bound, the total variation distance $TVD(leaf^A, leaf^B) \leq O(k\delta)$ because there is at most δ difference for each node. Now assume $k\delta$ is tiny. So we get the same leaf except w.p. $O(k\delta) = O(\sqrt{k}\cdot\sqrt{k\delta^2})$. We have $V_i = I(\pi_i; x|y, \pi_{<i}) + I(\pi_i; y|x, \pi_{<i}) = \delta^2$ (in our case $P_j^* = Bern(\frac{1}{2} - \delta)$), which implies that $I = k\delta^2$. So our probability of error is $O(\sqrt{k}\cdot\sqrt{k\delta^2}) = O(\sqrt{I}\sqrt{k})$. This is not quite where we need to be to prove Theorem 6, so we'll pick this up next lecture.

7 Ideas for Exercises

Exercise 13. *Verify the claim in the proof idea of Lemma 12 that this protocol achieves the first inequality of Lemma 12*

Exercise 14. *Find an example of protocol π' simulating some π where the length of the simulated transcript π' is less than the entropy of the transcript $H(\pi)$. Morally, why is this possible? (As we said, when $I = 0$ the length of π' can be 0, so a non-trivial π will give us this result. Morally, the shared randomness R' is doing our communication for us)*