CS 229r Information Theory in Computer Science Apr 5, 2016

Lecture 19

Lecturer: Madhu Sudan *Scribe: Themis Gouleakis*

1 Preliminaries

In order to continue the proof, we will need to use the following claims:

Claim 1 *Let E be an event in a probability space. Then,*

$$D(p(x|E\|p(x))) \leq \log(1/p(E))$$

where $D(p\|q)$ denotes the Kullback-Leibler divergence.

Proof For the Kullback-Leibler divergence, we have that

$$D(p(x|E\|p(x))) = E_{x\sim p(x|E)} \log \frac{p(x|E)}{p(x)}$$

Conditioning on the event E, makes the probability mass outside of E equal to 0 and blows up the probability of each element in E by $\frac{1}{p(E)}$. Thus, we get the result. ■

By using convexity arguments, we get the following generalization of claim 1:

Claim 2 *Let E be an event and A, X be random variables with support of size k. Then,*

$$\mathbb{E}_{A|E}[D(p(x|A,E)\|p(x))] \leq \log(k/p(E))$$

Claim 3 *Let E be an event and U, A, X be random variables with support of size k. Then,*

$$\mathbb{E}_{A,U|E}[D(p(x|A,U,E)\|p(x|U))] \leq \log(k/p(E))$$

Claim 4 *Let $p(x,y)$ and $q(x,y) = q(x) \cdot q(y)$ (i.e $q(x,y)$ defines a product measure) be two probability distributions. Then*

$$D(p(x,y)\|q(x,y)) \geq D(p(x)\|q(x)) + D(p(y)\|q(y))$$

2 Raz Lemma

We will now see the proof of the parallel repetition theorem by Raz [3] presenting a simplified proof due to Holenstein [4].

The theorem is stated as follows:

Theorem 5 *Let G be a 2-prover-1-round game. If $val(G) < 1 - \alpha$, then*

$$val(G^t) < 2^{-\Omega_{\alpha,k}(t)}$$

where k is the number of possible responses for each player.

Recall here from the previous lecture that we use the notation w_i for the event where $V(x_i, y_i, a_i, b_i) = 1$ and w_S for the event where $V(x_i, y_i, a_i, b_i) = 1$ for all $i \in S \subset \{1, \ldots, t\}$.

The main ingredient for the proof of the above theorem is Raz lemma [3] which is stated below:

Lemma 6 (Raz's Lemma) *If* $\text{val}(G) < 1 - \alpha$ *then there exists a constant $\gamma(k, \alpha)$ such that $\forall S \subseteq \{1, \ldots, t\}$ with $|S| \leq \gamma t$, then either*

1. $\Pr[w_S] \leq 2^{-\gamma t}$
2. *$\exists i \notin S$ such that* $\Pr[w_i | w_S] \leq 1 - \alpha/2$

Note that the above lemma implies theorem 5 and the upper bound is $val(G^t) < \max\{2^{-\gamma t}, (1 - \frac{\alpha}{2})^{\gamma t}\}$. Roughly, $\gamma \sim \frac{\alpha^2}{\log k}$.

Proof For the sake of contradiction we assume that the conclusion of the above lemma does not hold.

That is,

$$\exists S \subseteq \{1, \ldots, t\} : |S| < \gamma t$$

Such that both:

1. $\Pr[w_S] \geq 2^{-\gamma t}$
2. $\forall i \notin S : \Pr[w_i | w_S] > 1 - \frac{\alpha}{2}$

Now, let S be the last $t - r$ coordinates. We would like to use the to above conditions in order to contradict the fact that $val(G) < 1 - \alpha$. However, the second condition above is about a distribution conditional on the event W_S, while $val(G)$ is the probability of winning an one shot game, which is unconditional. So, our strategy would be to prove the existence of an index i such that $|\Pr[w_i | w_S] - \Pr[w_i]|$ is sufficiently small.

Indeed, using claim 1 and the fact that $p(w_S) \geq 2^{-\gamma t}$, we get:

$$D\left(p\left(\begin{array}{c} x_1 \ldots x_r \\ y_1 \ldots y_r \end{array} \middle| w_s\right) \middle\| p\left(\begin{array}{c} x_1 \ldots x_r \\ y_1 \ldots y_r \end{array}\right)\right) \leq \gamma t$$

By claim 4:

$$\frac{1}{r}\sum_{i=1}^{r} D\ \left(p\left(\begin{array}{c} x_i \\ y_i \end{array}\ \middle|\ w_s\right)\middle\| p\left(\begin{array}{c} x_i \\ y_i \end{array}\right)\right) \leq \frac{\gamma t}{r}$$

Since $D(p\|q) \geq |p-q|^2$:

$$\mathbb{E}_r \left| p\left(\begin{array}{c} x_i \\ y_i \end{array}\ \middle|\ w_s\right) - p\left(\begin{array}{c} x_i \\ y_i \end{array}\right)\right|^2 \leq \frac{\gamma t}{r}$$

$$\mathbb{E}_r \left| p\left(\begin{array}{c} x_i \\ y_i \end{array}\ \middle|\ w_s\right) - p\left(\begin{array}{c} x_i \\ y_i \end{array}\right)\right| \leq \sqrt{2\gamma}$$

So,

$$i : \left| p\left(\begin{array}{c} x_i \\ y_i \end{array}\ \middle|\ w_s\right) - p\left(\begin{array}{c} x_i \\ y_i \end{array}\right)\right| \leq \sqrt{2\gamma} \tag{1}$$

We now want to argue that $val(G) > 1-\alpha$ if the 2 conditions above hold. For that it would be helpful if X^t, Y^t were independent. However, we have to condition on w_S to prove this lemma, and unfortunately, $X^t|w_S, Y^t|w_S$ are not independent. Our goal is to find an auxiliary random variable U such that X^t, Y^t become conditionally independent with respect to w_S, U.

The auxiliary random variable U is defined as follows:

$$U = \left(\begin{array}{c} V_1 \dots V_r\ X_S \\ T_1 \dots T_r\ Y_S \end{array}\right)$$

where

$$V_i = \left\{\begin{array}{ll} 0 & w.p\ \ 1/2 \\ 1 & w.p\ \ 1/2 \end{array}\right.$$

$$T_j = \left\{\begin{array}{ll} X_j & if\ \ V_j = 0 \\ Y_j & if\ \ V_j = 1 \end{array}\right.$$

We define U_{-i} as follows:

$$U_{-i} = \left(\begin{array}{c} V_1 \dots V_{i-1}\ V_{i+1} \dots V_r\ X_S \\ T_1 \dots T_{i-1}\ T_{i+1} \dots T_r\ Y_S \end{array}\right)$$

As we said, desirable property we would like the random variable U to have is that: $X^t \perp Y^t | W_S, U, A_S$, where A_S denotes the set of answers from Alice. Indeed this is true for the above choice of U.

We will now show that Alice and Bob can use shared randomness in order to sample from the distribution $p(u, i, A_S|X_i = x, Y_i = y, W_S)$ without communicating. After doing that, they can sample privately the variables X_t, Y^t conditioned on those variables (U, W_S, A_S) and return their answers A_i, B_i for the i-th game. We also have that:

$$\Pr[success] = \Pr[W_i] = \mathbb{E}_i[\Pr[W_i]] \geq \mathbb{E}_i[\Pr[W_i|W_S]] - \sqrt{2\gamma} \geq 1 - \frac{\alpha}{2} - \sqrt{2\gamma} > 1 - \alpha$$

This is a contradiction to the fact that since the X_t, Y_t are independent, their probability of success can be at most the value of an one-shot game (for which $val(G) < 1 - \alpha$). That finishes the proof of lemma 6.

■

It now remains to show that Alice and Bob can indeed sample from the distribution $p(u, i, A_S | X_i = x, Y_i = y, W_S)$ without communicating. So, we assume that they use their shared randomness so that Alice samples from $p(u, i, A_S | X_i = x, W_S)$ and Bob samples from $p(u, i, A_S | Y_i = y, W_S)$. Even though they sample from different distributions, we can show that the distributions are close enough so that Alice and Bob can use correlated sampling and get the same sample most of the time. More specifically, we will use the following two lemmas:

Lemma 7 *There exists some $\gamma(\alpha, k)$ such that*

$$p(i, x_i, y_i) \cdot p(U_{-i}, A_S | W_S, i, x_i) \overset{\varepsilon}{\approx} p(i, x_i, y_i, A_S, U_{-i} | W_S) \overset{\varepsilon}{\approx} p(i, x_i, y_i) \cdot p(U_{-i}, A_S | W_S, i, y_i)$$

where $p(x) \overset{\varepsilon}{\approx} q(x) \Leftrightarrow |p(x) - q(x) \leq \varepsilon|$ and $\varepsilon = (\alpha - \gamma)/10$ in our case.

Proof *[sketch]*

Using claim 3, we can show that

$$\begin{aligned} p(i, x_i, y_i, A_S, U_{-i} | W_S) &= p(A_S, U_{-i} | W_S) \cdot p(i, x_i, y_i | A_S, U_{-i}, w_S) && (2) \\ &\overset{2\varepsilon}{\approx} p(i, x_i, A_S, U_{-i} | W_S) \cdot p(y_i | i, x_i) && (3) \\ &= p(i, x_i | w_S) \cdot p(U_{-i}, A_S | w_S, i, x_i) \cdot p(y_i | i, x_i) && (4) \\ &\overset{\varepsilon}{\approx} p(i, x_i, y_i) \cdot p(U_{-i}, A_S | w_S, i, x_i) && (5) \end{aligned}$$

For the last step, we also used the fact that conditioning on w_S does not change much, as equation 1 suggests.

The second approximation:

$$p(i, x_i, y_i, A_S, U_{-i} | W_S) \approx p(i, x_i, y_i) \cdot p(U_{-i}, A_S | W_S, i, y_i)$$

follows by symmetry.

■

Lemma 8 (correlated sampling) *There is a protocol for Alice and Bob to used shared randomness to sample a random variable such that Alice gets value $x \sim p$ and Bob value $y \sim q$ and the probability that their values differ is:* $\Pr[x \neq y] \leq 2|p - q|$.

Proof Alice and Bob can use their shared randomness to sample an infinite sequence of tuples: $\{(x_i, \rho_i)\}$, where each x_i is distributed uniformly on the sample space and each rho_i is a uniformly distributed real number in $[0, 1]$. Alice will pick the x_i with the smallest index i such that $p(x_i) \geq \rho_i$, while Bob

will pick the x_i with the smallest index i such that $q(x_i) \geq \rho_i$. It is easy to see that $\forall i, j : \frac{\Pr[Alice\ \ picks\ \ x_i]}{\Pr[Alice\ \ picks\ \ x_j]} = \frac{p(x_i)}{p(x_j)}$ and similarly for Bob. So, they sample exactly from the distributions p, q respectively, and also the only way they get a different sample is if for some rho_i it holds that: $p(x_i) < \rho_i < q(x_i)$ (or vice versa), which happens with probability at most $2|p - q|$. ■

References

[1] Feige, Uriel. On the success probability of the two provers in one-round proof systems. In *Structure in Complexity Theory Conference, Proceedings of the Sixth Annual.* IEEE, 1991.

[2] Verbitsky, Oleg. Towards the parallel repitition conjecture. In *Structure in Complexity Theory Conference, Proceedings of the Sixth Annual.* IEEE, 1994.

[3] Raz, Ran. A parallel repetition theorem. *SIAM Journal on Computing 27.3*, 1998.

[4] Holenstein, Thomas. Parallel repetition: simplifications and the no-signaling case. *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing.* ACM, 2007.