CS 229r Information Theory in Computer Science	March 7, 2019
Lecture 12	
Instructor: Madhu Sudan	*Scribe: Albert Chalom*

1 Book Keeping

1.1 Admin

- Project link on Canvas.
- Express interest

1.2 Today

- Set disjointness
- Information complexity

1.3 References

We'll focus on:

- [Bar-Yossef, Jayram, Kumar, Sivakumar]

Previous work:

- [Babai, Frankl, Simon]
- [Kalyanasundaram, Schnitger]
- [Razborov]

2 Disjointness

We will first consider the Disj^n problem, which asks on two n-length strings X and Y, is there some index i, such that $X_i = Y_i = 1$. This is equivalent to $X_i \cap Y_i = 1$. We formally define the problem below:

Definition 1. Disj^n(X,Y) = 1 if $\exists i$ st $X_i = Y_i = 1$ and 0 otherwise.

Exercise 2. *$\forall independent X and Y, \forall \mu = \mu_x \times \mu_y$ show a protocol with error $\leq \varepsilon$ and $\tilde{O}(\sqrt{n})$*

This implies that hardness needs $X \not\perp Y$.

3 Conditional Mutual Information

Definition 3. For (X, Y, Z) jointly distributed, $I(X, Y|Z)$ is the information about X from Y conditioned on Z.

We can rigorously measure this as $I(X, Y|Z) = E_{Z \sim P_z}[I(X|_{Z=z}, Y|_{Z=z}] = H(X|Z) - H(X|Y, Z)$.

Recall that with entropy we had a property that $H(X|Z) \leq H(X)$. However, there is no definitive relationship between information and conditional information (e.g $I(X, Y)$ and $I(X, Y|Z)$).

Example 4. *Consider the distribution,* $X = Y = Z$ *with* $Z \in \{0,1\}^n$
$I(X,Y) = n, I(X,Y|Z) = 0$ *so here conditioning reduced information.*

Example 5. *Consider* $X \perp Y, Z = X \oplus Y$*, with* $X, Y \in Unif\{0,1\}^n$
There here $I(X,Y) = 0, I(X,Y|Z) = n$ *so here conditioning increased information.*

Example 6. *Consider* $X \leftrightarrow Y \leftrightarrow Z$ *as a Markov Chain such (so X and Z are independent given Y), then* $I(X,Y) \geq I(X,Y|Z)$ *and* $I(X,Z|Y) = 0$*.*

Exercise 7. *Prove the above example. Hint use that* $H(X|Y,Z) = H(X|Y)$

3.1 Motivation

Fix a randomized protocol Π that *epsilon*-computes f, if it computes f with error at most epsilon for all inputs (x,y).

Goal: How much does an observer learn about the inputs from watching the interaction?

3.2 Example protocol

Consider the following protocol with R as public randomness.

Alice		*Bob*
x, R		y, R
	$\xrightarrow{R \oplus X}$	
	$\xleftarrow{f(x,y)}$	

In this case, $I((X;Y); R \oplus X, f(X,Y)) \leq H(f(X,Y))$ so the observer learns little because they can't see the randomness that Alice and Bob both see.

Therefore we should condition on public Randomness R, but not on any private randomness R_A or R_B

4 Information Complexity

Definition 8. Information complexity is defined as the amount of information an observer learns about (X,Y) from the transcript of communication, over the given distribution mu. We assume the observer has access to public randomness but not private randomness.

Mathematically, for a protocol, $IC_\mu(\Pi) = I(XY, \Pi|R)$.
For a function $IC_\mu(f) = \min_{\Pi st.\Pi\ \varepsilon\text{-computes f}}(IC_\mu(\pi))$

If Π is a k-bit protocol that ε-computes f, $IC_\mu(f) \leq k$

4.1 Plan

$IC_{\mu_n}(Disj^n) = \Omega(n)$ (we will prove)

The following statements are true, but we won't prove them here. Instead we will prove something analagous for conditional IC (to be defined below).

- $IC_{\mu_n}(Dsij^n) \geq nIC_{\mu_1}(Disj^1)$
- $IC_{\mu_1}(Disj^1) = \Omega(1)$

4.2 One dimensional binary disjointness

$Disj^1(u,v) = u \wedge v$

Example 9. *An intuitive protocol for computing And would be*

$$\begin{array}{ccc} Alice & & Bob \\ u & & v \\ & \xrightarrow{u} & \\ & \xleftarrow{u \wedge v} & \end{array}$$

If $u = 0$ then an observer only learns one bit (u), but if $u = 1$ then both bits are revealed to an observer, so on average $\frac{3}{2}$ bits are revealed.

This raises the question can we do better? If $u = v = 1$ then both bits are revealed, so ideal is when u or v are zero, the ideal case is we don't learn anything about the other bit.

Example 10. *Now consider the following randomized protocol.*

Alice picks $t_a \in [0,1]$ at random, and Bob picks $t_b \in [0,1]$ at random. Then at time t_a Alice sends 0 to Bob if $U = 0$, and at time t_b Bob sends 0 to Alice if $V = 0$. The protocol ends after the first bit is sent, so if $U = V = 0$, Alice will only send her bit if $t_a < t_b$ and Bob will only send his bit if $t_b < t_a$. We assume that bits are sent instantly and since we sample from a continuous distribution t_a will never equal t_b.

The idea here is if $(uv) = 00, 01,$ or 10 then we only learn one of u or v, but if $(uv) = 11$ we learn both u and v, so on average $\frac{5}{4}$ bits are learned.

This analysis is a bit loose because after we wait for longer, we would bias the other bit to be more likely to be 1.

Exercise 11. *Come up with a tight bound for the protocol.*

4.3 Proof of $IC_\mu(Disj^n) = \Omega^n$

Let μ be the following distribution with (X_i, Y_i) iid with

$$(X_i, Y_i) = \begin{cases} 00 & \text{with prob } 1/2 \\ 01 & \text{with prob } 1/4 \\ 10 & \text{with prob } 1/4 \end{cases}$$

Next consider the following way of sampling this distribution with (X, Y, Z) with $Z \sim Unif(\{0,1\}^n$

```
for i = 1 to n do
    if Z[i] = 0 then X[i] = 0, Y[i] ~ Unif{0,1}
    if Z[i] = 1 then Y[i] = 0, X[i] ~ Unif{0,1}
```

4.3.1 CIC (Conditional Information Cost)

$CIC_\mu(\Pi) = I((X,Y); \Pi | R, Z)$.

We will prove the following two statements

1. $CIC_\mu(Disj^n) \geq n \times CIC_\mu(Disj^1)$ (today)
2. $CIC_\mu(Disj^1) = \Omega(1)$ (next class, non-trivial)

Observation 12. *Consider a Markov Chain $\Pi \leftrightarrow (X,Y) \leftrightarrow Z$), then $\Pi | X, Y \perp Z | X, Y$.*

Then $IC_\mu(\Pi) \geq CIC_\mu(\Pi)$

To see this we know $I((X,Y),\Pi)|R) \geq I((X,Y),\Pi|R,Z)$ and $IC_\mu(\Pi) = I((X,Y),\Pi)|R)$ and $CIC_\mu(\Pi) = I((X,Y),\Pi|R,Z)$

$I((X,Y),\Pi|R,Z) = H(X,Y|R,Z) - H(X,Y|\Pi,R,Z)$

$H(X,Y|R,Z) = \sum_{i=1}^{n} H(X_i,Y_i|R,Z,X_{<i},Y_{<i} = \sum_{i=1}^{n} = H(X_i,Y_i|Z_i) = \sum_{i=1}^{n} H(X_i,Y_i|R,Z)$

$H(X,Y|\Pi,R,Z) = \sum_{i=1}^{n} H(X_i,Y_i|\Pi,R,Z,X_{<i},Y_{<i}) \leq \sum_{i=1}^{n} H(X_i,Y_i|\Pi,R,Z)$

$I((X,Y),\Pi|R,Z)) \geq \sum_{i=1}^{n} H(X_i,Y_i|R,Z) - H(X_i,Y_i|\Pi,R,Z) = \sum_{i=1}^{n} I((X_i,Y_i),\Pi|R,Z)$

We now want to show that $I((X_i,Y_i),\Pi|R,Z) \geq CIC(Disj^1)$

Let us now consider the following two protocols

4.3.2 Protocol A

Consider both Alice and Bob to have access to $w \sim Bern(.5)$ and R', and private randomness R_a, R_b. Alice will create a random variable U, and Bob will create a random variable V according to the following distribution:

```
if w = 0 then U= 0, V is random
if w = 1 then V = 0, U is random
```

The goal of this protocol, Π' is to compute $U \wedge V$

Alice		*Bob*
w, R', R_a		w, R', R_b
computes U,		computes V
	$\rightarrow$	
	$\leftarrow$	
	.	
	.	
	.	
	$\xrightarrow{U \wedge V}$	

This protocol reveals $I((U,V),\Pi'|R',W)$.

4.3.3 Protocol B

Now let Z, R, be shared randomness for Alice and Bob, and again give them private randomness R_a, R_b. Using Z Alice and Bob can compute X and Y according to the distribution μ using their shared randomness, and consider the following protocol Π.

Alice		*Bob*
Z, R, R_a		Z, R, R_b
computes $X_1, \ldots X_n$,		computes $Y_1, \ldots, Y_n$
	$\rightarrow$	
	$\leftarrow$	
	.	
	.	
	.	
	$\xrightarrow{Disj^n(X,Y)}$	

Then this protocol reveals information $I((X_i,Y_i),\Pi|R,Z)$

4.3.4 Combining Protocols

We now want to show $I((X_i,Y_i),\Pi|R,Z) \geq I((U,V),\Pi'|R',W) = CIC(Disj^1)$ by showing how we can reduce protocol A to protocol B, by using protocol B to achieve the task of protocol A.

We can let $X_i = U, Y_i = V$ and use R' to generate Z and R, allowing Alice and Bob to generate their remaining X_j and Y_js. Then because for all $j \neq i, X_j \wedge Y_j = 0$ by construction, this will output $X_i \wedge Y_i$ computing $Disj^1$.

Therefore we have shown $I((X_i, Y_i), \Pi | R, Z) \geq CIC(Disj^1)$, which shows $CIC_\mu(Disj^n) \geq n \times CIC_1(Disj^1)$