LECTURE 11

Key definitions

μ distribution on finite set Ω. $X \sim \mu$

ENTROPY $H(X) = \sum_{\omega \in \Omega} \mu(\omega) \log \frac{1}{\mu(\omega)}$

$= -\sum_{\omega \in \Omega} \mu(\omega) \log \mu(\omega)$

CONDITIONAL ENTROPY

$$H(X|Y) = \sum_y \Pr[Y=y] \cdot H(X|Y=y)$$

MUTUAL INFORMATION

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

CONDITIONAL MUTUAL INFORMATION

$$I(X;Y|Z) = \sum_z \Pr[Z=z] \cdot I(X;Y|Z=z)$$

INFORMATION COST of π w.r.t μ

$I(X,Y; \pi(X,Y))$ for $(X,Y) \sim \mu$.

INFORMATION COMPLEXITY $IC_{\mu,\delta}(f)$

Lowest cost/info of any δ-error protocol for f where mutual info measured w.r.t μ.

Last time we (sort of) proved that

1) Info cplx of $f \leq$ comm cplx of f.

2) For disjointness, want to compute OR of n bitwise ANDs

3) Look at mixtures (averages) of product distributions over X and Y. Get product distributions when conditioning on choice made by mixing variable.

4) Can define conditional info cplx in this setting and prove that info revealed about (X, Y) by Π $\geq \sum_j$ info revealed about (X_j, Y_j).

5) For the right distribution, can show that Π has to compute $AND(x_j, y_j)$ for every j. Hence Π reveals at least as much about (X_j, Y_j) as best protocol for one-bit AND

Proof idea: want to compute u AND v. Insert u and v into jth coordinates of X and Y. Generate dummies for other coordinates. Run protocol Π. Important to get distributions exactly right!

So $\boxed{R_\delta^{priv}(DISJ_n) \geq n \cdot CIC_{\zeta,\delta}(AND)}$

Want to prove conditional information cplx lower bound $CIC_{\zeta,\delta}(AND) = \Omega(1)$

for $AND(x,y) = 1$ iff $x = y = 1$ and 0 otherwise

and ζ mixture of product distributions

X	Y	D	ζ		
0	0	A	1/4	} (0,0)	50%
0	0	B	1/4		
0	1	A	1/4	(0,1)	25%
0	1	B	0		
1	0	A	0		
1	0	B	1/4	(1,0)	25%
1	1	A	0	} (1,1)	never happens!
1	1	B	0		

Question 1 Why is this not totally obvious?

Clearly some amount $\varepsilon > 0$ of info must be revealed!?

Not so clear.. Note that protocol will always get inputs for which 0 is correct answer!

Question 2 So why is this not impossible then?!

Protocol can always say just "0" and be done without revealing any info!

No, it can't. Because protocol must get all inputs right with prob $\geq 1-\delta$ including (1,1).

So although we will never run protocol on this data, it would have to get it right if we did... Use this to argue that different 0-result inputs must yield different transcripts.

Let P be any randomized protocol that computes x AND y correctly on all inputs $(x,y) \in \{0,1\}^2$ with prob $\geq 1-\delta$

Let Z be random variable distributed uniformly in $\{0,1\}$.

Using definition of ζ and expanding on values of D, we want to lower-bound

$$I(X,Y;\ P(X,Y) \mid D) =$$

$$= \frac{1}{2}\ I(X,Y;\ P(X,Y) \mid D=A) + \frac{1}{2}\ I(X,Y;\ P(X,Y) \mid D=B)$$

$$= \frac{1}{2}\left(I(Z;\ P(0,Z)) + I(Z;\ P(Z,0)) \right) \qquad (1)$$

For if $D=A$ we know $X=0$ and if $D=B$ then $Y=0$
(Formally, this follows since entropy of independent random variables sum up; a random coin flip is independent from the constant 0, and a constant has zero entropy)

Summands in (1) are on form $I(Z;\ \Phi(Z))$ where Z uniform in $\{0,1\}$ and $\Phi(z)$ for $z \in \{0,1\}$ are random variables.

Want to study this mutual info by help of metrics on probability distributions

In particular, Hellinger distance turns out to be a magic metric with just the right properties

DEFINITION 1 (HELLINGER DISTANCE)

Let P and Q be probability distributions on a domain Ω. Then their HELLINGER DISTANCE $h(P,Q)$ is defined by

$$h(P,Q) = \sqrt{1 - \sum_{\omega \in \Omega} \sqrt{P(\omega) Q(\omega)}}$$

Yes, this is a metric! Recall that d is a metric on some set if

$$d(x,y) = d(y,x) \qquad \text{symmetry}$$
$$d(x,y) \geq 0 \qquad \text{with equality iff } x = y$$
$$d(x,z) \leq d(x,y) + d(y,z) \qquad \text{triangle inequality}$$

MAGIC LEMMA 2

Let Z be random variable uniform on $\{z_1, z_2\}$. Let $\Phi(z_1)$ and $\Phi(z_2)$ be random variables such that $\Phi(z)$ is independent from Z for each $z \in \{z_1, z_2\}$. Let Φ_z denote distribution of $\Phi(z)$. Then it holds that

$$I(Z; \Phi(Z)) \geq h^2(\Phi_{z_1}, \Phi_{z_2}).$$

For us, $\Phi(z)$ will be protocol transcripts Z corresponds to coin flips $D = A/B$.

Note that protocol transcript is independent of this transcript, since it simply runs on the inputs given without any external info.

Denote P_{ij} = protocol transcript on input (i,j)

We will also need:

LEMMA 3 CAUCHY-SCHWARTZ INEQUALITY

For $x, y \in \mathbb{R}^n$ $\quad \|x\| \, \|y\| \geq |\langle x, y \rangle|$

or squared and written out:

$$\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2 \geq \left(\sum_{i=1}^{n} x_i y_i \right)^2$$

Common special case when all $y_i = 1$:

$$n \cdot \sum_{i=1}^{n} x_i^2 \geq \left(\sum_{i=1}^{n} x_i \right)^2 \qquad (2)$$

Combining (1), (2) and Magic Lemma 2 we get

$$I(X, Y; P(X,Y) \mid D) \geq$$

$$\frac{1}{2} \left(h^2(P_{00}, P_{01}) + h^2(P_{00}, P_{10}) \right) \geq \quad [\text{Lemma 2}]$$

$$\geq \frac{1}{4} \left(h(P_{00}, P_{01}) + h(P_{00}, P_{10}) \right)^2 \geq \quad [(2)]$$

$$\frac{1}{4} \left\{ h^2(P_{01}, P_{10}) \right. \quad [\text{triangle inequality}] \qquad (3)$$

Look at four protocol distributions

$P_{00}, P_{01}, P_{10}, P_{11}$

Clearly P_{11} should be different from the rest.

But why should P_{01} and P_{10} be different?

Because of magic "rectangular properties" of Hellinger distance!

Recall that for deterministic protocols, if $\Pi(x,y) = \Pi(x',y') = \tau$, then $\Pi(x,y') = \Pi(x',y) = \tau$.
An analogue holds for Hellinger distance for randomized protocols.

MAGIC LEMMA 4 (CUT-AND-PASTE LEMMA)
For any randomized protocol Π and for any $x,x' \in \mathcal{X}$ and $y,y' \in \mathcal{Y}$, it holds that

$$h(\Pi_{x,y}, \Pi_{x',y'}) = h(\Pi_{xy'}, \Pi_{x'y}).$$

Plugging this into (3) we get

$$I(X,Y; P(X,Y)|D) \geq \frac{1}{4} h^2(P_{01}, P_{10}) = \frac{1}{4} h^2(P_{00}, P_{11}) \quad (4)$$

But now we're in good shape! Since AND(0,0) ≠ AND(1,1), [P_{00} and P_{11} distributions] should look different.

MAGIC LEMMA 5
For any δ-error randomized protocol Π for a function f, and for any (x,y) and (x',y') such that $f(x,y) \neq f(x',y')$, we have $h^2(\Pi_{xy}, \Pi_{x'y'}) \geq 1 - 2\sqrt{\delta}$

[according to BJKS – should maybe be $\geq 1 - \sqrt{2\delta}$]

Plugging this into (4), we get

$$CIC_{\varepsilon,\delta}(AND) \geq I(X,Y; P(X,Y) \mid D)$$

[for best protocol P]

$$\geq \frac{1}{4} h^2(P_{00}, P_{11})$$

$$\geq \frac{1}{4}\left(1 - 2\sqrt{\delta}\right)$$

Now the set disjointness lower bound follows.

$$\boxed{R^{priv}_{\delta}(DISJ_n) \geq n \cdot CIC_{\varepsilon,\delta}(AND) \geq \frac{n}{4}\left(1 - 2\sqrt{\delta}\right)}$$

Now let us verify the (magic) lemmas top-down.

LEMMA 6 Let Π protocol over $\mathcal{X} \times \mathcal{Y}$ with possible transcripts $\mathcal{T}$. Then there are mappings

$$g_A : \mathcal{T} \times \mathcal{X} \to \mathbb{R}$$
$$g_B : \mathcal{T} \times \mathcal{Y} \to \mathbb{R}$$

such that for every $x \in \mathcal{X}, y \in \mathcal{Y}, \tau \in \mathcal{T}$ it holds that

$$P_r[\Pi(x,y) = \tau] = g_A(\tau; x) \cdot g_B(\tau; y)$$

Proof Recall that for any deterministic protocol, the set of inputs x, y that generate a transcript τ is a combinatorial rectangle $A \times B$ for $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$.

Given random strings a, b of Alice & Bob, if we consider (x,a) and (y,b) to be "inputs" then protocol is deterministic.

Let $\mathcal{A}(\tau) \times \mathcal{B}(\tau)$ be combinatorial rectangle corresponding to ~~[illegible]~~ transcript τ in this "extended deterministic" protocol.

$$\text{let } \mathcal{A}(\tau, x) = \{ (x,a) \text{ for all random strings } a \text{ leading to } \mathcal{A}(\tau) \}$$

$$\mathcal{B}(\tau, y) = \{ (y,b) \text{ for all random } b \text{ leading to } \mathcal{B}(\tau) \}$$

$$\text{Let } \mathcal{X}(x) = \{ \text{all pairs } (x,a) \}$$
$$\mathcal{Y}(y) = \{ \text{all pairs } (y,b) \}$$

$$\text{Let } q_A(\tau; x) = |\mathcal{A}(\tau, x)| \,/\, |\mathcal{X}(x)|$$
$$q_B(\tau, y) = |\mathcal{B}(\tau, y)| \,/\, |\mathcal{Y}(y)|$$

On input (x,y) Alice chooses (x,a) uniformly at random from $\mathcal{X}(x)$ and Bob chooses (y,b) unif at random from $\mathcal{Y}(y)$. The transcript will be τ if and only if $(x,a) \in \mathcal{A}(\tau,x)$ and $(y,b) \in \mathcal{B}(\tau,y)$.

Hence by independence

$$\Pr[\pi(x,y) = \tau] = q_A(x;\tau) \cdot q_B(\tau; y). \quad \square$$

Proof of Lem 4 (cut-and-paste)

$$1 - h^2(\pi_{xy}, \pi_{x'y'})$$

$$= \sum_\tau \sqrt{\Pr[\pi(x,y)=\tau]\,\Pr[\pi(x',y')=\tau]} \quad [\text{by def}]$$

$$= \sum_\tau \sqrt{q_A(\tau;x)\cdot q_B(\tau;y)\cdot q_A(\tau,x')\,q_B(\tau,y')} \quad [\text{Lem 6}]$$

$$= \sum_\tau \sqrt{q_A(\tau;x)\cdot q_B(\tau;y')\cdot q_A(\tau,x')\cdot q_B(\tau,y)}$$

$$= \sum_\tau \sqrt{\Pr[\pi(x,y')=\tau]\cdot \Pr[\pi(x',y)=\tau]}$$

$$= 1 - h^2\left(\pi_{xy'}, \pi_{x'y}\right) \qquad \square$$

For Lemma 5, we need another definition

DEF 7 (TOTAL VARIATION DISTANCE)
The TOTAL VARIATION DISTANCE $V(P,Q)$ between two probability distributions P and Q on Ω is

$$V(P,Q) = \frac{1}{2}\sum_{\omega\in\Omega} |P(\omega) - Q(\omega)|$$

FACT 8

$$V(P,Q) = \max_{\Omega'\subseteq\Omega}\{P(\Omega') - Q(\Omega')\}$$

Proof Exercise

LEMMA 9 [Le Cam & Yang 1990]
If P and Q are probability distributions on the same domain, then

$$V(P,Q) \leq h(P,Q)\sqrt{2 - h^2(P,Q)}$$

Proof of Magic Lemma 5:
Let $\mathcal{T}$ be all transcripts on which Π outputs ~~$f(x,y)$~~ f. Since Π computes $f(x,y)$ on input (x,y) with prob $\geq 1-\delta$,
$\Pr[\Pi(x,y) \in \mathcal{T}] \geq 1-\delta$. Since f errs on (x',y') with prob $\leq \delta$, $\Pr[\Pi(x',y') \in \mathcal{T}] \leq \delta$.
Using Fact 8, $V(\Pi_{xy}, \Pi_{x'y'}) \geq 1-2\delta$.
Now apply Lemma 9 and do the necessary calculations ☐

Now all that remains is Magic Lemma 2.
For which... More definitions!

Recall Kullback-Leibler divergence

$$KL(P\|Q) = \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)}$$

DEF 10 JENSEN-SHANNON DIVERGENCE

$$JS(P,Q) = \frac{1}{2}\left(KL\left(P \,\|\, \frac{P+Q}{2}\right) + KL\left(Q \,\|\, \frac{P+Q}{2}\right)\right)$$

LEMMA 11 [Lin 1991]

For distributions P and Q on the same domain, $JS(P,Q) \geq h^2(P,Q)$

Proof of Magic Lemma 2:

Sufficient to prove $I(Z; \Phi(Z)) = JS(\Phi_{z_1}, \Phi_{z_2})$ by Lemma 11. First write

$$I(X;Y) = I(Y;X)$$

$$= H(Y) - H(Y|X)$$

$$= -\sum_y Pr[Y=y] \log Pr[Y=y]$$

$$- \left(- \sum_x Pr[X=x] \sum Pr[Y=y|X=x] \log Pr[Y=y|X=x] \right)$$

$$= \sum_x Pr[X=x] \cdot \sum_y Pr[Y=y|X=x] \log \left(\frac{Pr[Y=y|X=x]}{Pr[Y=y]} \right) \quad (5)$$

Let μ be distribution of Y and μ_x distribution of Y conditioned on $X=x$. Then we get

$$(5) = \sum_{x \in \mathcal{X}} Pr[X=x] \cdot KL(\mu_x \| \mu) \quad (6)$$

Now set $X = Z$ and $Y = \Phi(Z)$. For each $z \in \{z_1, z_2\}$ $\Phi(z)$ is independent of Z.

Hence conditioned on $Z=z$ the distribution of $\Phi(Z)$ equals Φ_z

Since Z is uniform over $\{z_1, z_2\}$,

$\Phi(Z) \sim (\Phi_{z_1} + \Phi_{z_2})/2$

Using (5) and (6) we get

$$I(Z; \Phi(Z)) =$$

$$\sum_{z \in \{z_1, z_2\}} Pr[Z=z] \, KL\left(\Phi_z \,\|\, \frac{\Phi_{z_1} + \Phi_{z_2}}{2}\right)$$

$$= JS(\Phi_{z_1}, \Phi_{z_2})$$

$\square_{4}$