Information Theory & Coding

Oct 27th 2020

Reminder: Today 5pm: Midterm assigned
due: Sunday night (23:59).

Yesterday: properties of p_X that maximize $I(X;Y)$.

Today: proof of the "coding theorem"!

Recall: given a channel $P(y|x)$, memoryless,
we say that a rate R is achievable if
$\forall \varepsilon > 0$ $\exists$ enc, dec with
$$\hat{P}_e(\text{enc}, P, \text{dec}) < \varepsilon \quad \&$$
$$\text{rate}(\text{enc}) \geq R$$

$$\text{enc}: \underbrace{\{1, \dots, M\}}_{\equiv \log_2 M \text{ bits}} \to \mathcal{X}^n \qquad \text{rate} = \frac{\log_2 M}{n}$$

$$\text{dec}: \mathcal{Y}^n \to \{1, \dots, M\}$$

$$P_{e,m} = \Pr(\text{dec}(Y^n) \neq m \mid X^n = \text{enc}(m))$$

$$\bar{P}_e(\text{enc}, P, \text{dec}) = \frac{1}{M} \sum_{m=1}^{M} P_{e,m}$$

$$\hat{P}_e(\;\;\text{''}\;\;) = \max_{1 \leq m \leq M} P_{e,m} \geq \bar{P}_e(\text{enc}, P, \text{dec}).$$

Coding Thm: given a channel P, every $R < C(P)$ is achievable.

Thm A Given channel $\underline{P}$, $R < C(P)$, $\varepsilon > 0$.

there exists enc, dec s.t

$$\text{enc} : \{1 .. M\} \to \mathcal{X}^n \quad \text{has} \quad M \geq 2 \cdot 2^{nR},$$

$$\text{dec} : \mathcal{Y}^n \to \{1 .., M\} \quad \text{and}$$

$$\overline{P_e}(\text{enc}, P, \text{dec}) < \varepsilon.$$

Proof: (later today).

Recall: We are given $P(y|x)$ as the description of a memoryless channel $\to \square \to$

$$\Pr(Y^n = y^n \mid X^n = \underline{x}^n) = \prod_{i=1}^{n} P(y_i | x_i).$$

Observe that the Coding Theorem is a corollary of Thm A:

Why?: Suppose Thm A is true. So fix any $\varepsilon > 0$ we can find

enc, dec with $M \geq 2 \cdot 2^{nR}$ and

$$\overline{P_e}(\text{enc}, P, \text{dec}) < \varepsilon/2$$

$$\parallel$$

$$\frac{1}{M} \sum_{i=1}^{M} P_{e,m}(\text{enc}, P, \text{dec}) < \varepsilon/2$$

Q: for how many m's can

$$P_{e,m} \geq \varepsilon \ ?$$

we know $\displaystyle\sum_{m=1}^{M} P_{e,m} \leq M \cdot \frac{\varepsilon}{2}$ (*) if $> \frac{M}{2}$ of m's

had $P_{e,m} \geq \varepsilon$ then $\displaystyle\sum_{m=1}^{M} P_{e,m} > \frac{M}{2} \cdot \varepsilon$, contradicts

(*). So we find that

for at least $\geq \frac{M}{2}$ of m's $P_{e,m} < \varepsilon$

If we restrict the encoder to these m's we

have

$$\mathrm{enc}' : \{1, .., M'\} \to \mathcal{X}^n, \quad M' \geq M/2$$

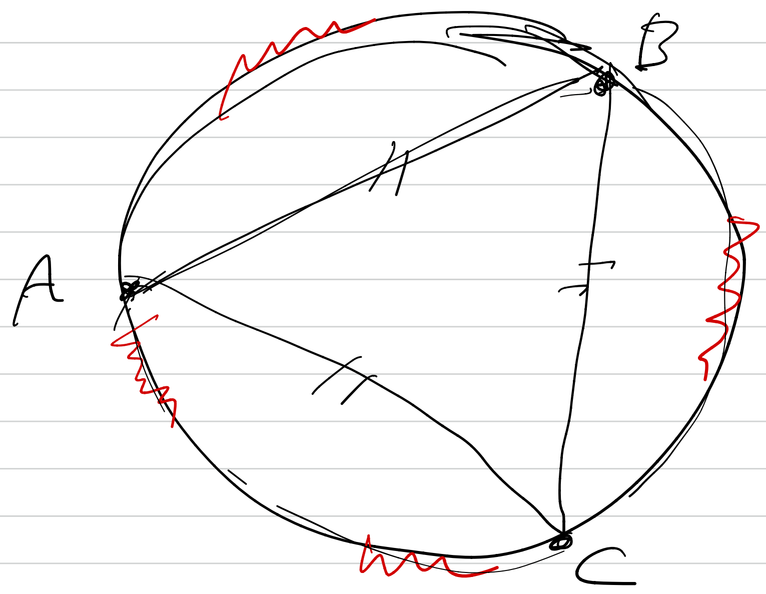
$$(\hat{P}_e(\mathrm{enc}', P, \mathrm{dec}) < \varepsilon).$$

$$\mathrm{rate}(\mathrm{enc}') = \frac{1}{n} \log M' \geq \frac{1}{n} \log \frac{M}{2}$$

$$\geq \frac{1}{n} \log \frac{1}{\cancel{2}} \cancel{2} 2^{nR} = R$$

So, it suffices to prove Thm A.

Aside on Probabilistic Method:



total (red) $< \frac{1}{3}$ of the circumference

aim: find equilateral $\triangle$ ABC with no red vertices.

Claim: no matter the red pattern, such a $\triangle$ exists.

Proof: pick the point A randomly and uniformly on the circle (this determines B and C too.)

let $R(A)$: # of red vertices among A, B, C.

$$= \mathbb{1}\{A \text{ is red}\} + \mathbb{1}\{B \text{ is red}\} + \mathbb{1}\{C \text{ is red}\}$$

$$E[R(A)] = E[\mathbb{1}\{A \text{ is red}\}] \leftarrow \Pr(A \text{ is red})$$
$$+ E[\mathbb{1}\{B \text{ is red}\}] \leftarrow \Pr(B \text{ " " })$$
$$+ E(\mathbb{1}\{C \text{ is red}\}) \leftarrow \Pr(C \text{ " " })$$

$$= \Pr(A \text{ is red}) \leftarrow < 1/3$$
$$+ \Pr(B \text{ is red}) \leftarrow < 1/3$$
$$+ \Pr(C \text{ " " }) \leftarrow < 1/3$$

$$< 1$$

$\Rightarrow \underline{E(R(A)) < 1} \Rightarrow \underline{Pr(R(A) < 1) > 0}.$

$\Rightarrow Pr(R(A) = 0) > 0.$

$\Rightarrow \exists (A, B, C)$ firms A, B, C s.t no vertex is red. //

We will use the probabilistic method to prove Thm A.

We will compute

$E[\bar{P}_e(ENC, P, DEC)]$ and upper bound it by ε. $\Rightarrow \exists$ enc, dec $\bar{P}_e(enc, P, dec) < \varepsilon$.

Proof of Thm A: Given P, $R < C(P)$, $\varepsilon > 0$,

fix n (to be chosen later), set

$M = \lceil 2 \cdot 2^{nR} \rceil \geq 2 2^{nR}$ and chose an encoder

$(ENC : \{1 .., M\} \to \mathcal{X}^n$ by letting

each $\{ENC(m)_i : m = 1 .., M, \ i = 1 .., n\}$

to be iid $\sim P_X$. where P_X is a

distribution s.t $I(X;Y) > R$.

For the decoder:

Recall the concept of Typicality: given a p_U, we $T(n, \varepsilon, p_U) = \{(u_1, \dots, u_n) :$

$$\frac{\#\{i : u_i = u\}}{n} = (1 \pm \varepsilon)\, p_U(u).\}$$

take $\mathcal{U} = \mathcal{X} \times \mathcal{Y}$ let

$$\boxed{p_U(x,y) = p_X(x)\, P(y|x).}$$ let

$$T(n, \varepsilon) = \{(x^n, y^n) : \frac{\#\{i : (x_i, y_i) = (x,y)\}}{n} = (1 \pm \varepsilon)\, p_X(x) P(y|x) \quad \forall x, y\}.$$

$dec(y^n)$: for each $m = 1, \dots, M = \lceil 2 \cdot 2^{nR} \rceil$

test ~~check~~ if $(enc(m), y^n) \in T(n, \varepsilon')$.

if exactly one m ~~satisfies the~~ passes the test

let $dec(y^n) = m$. otherwise choose

$dec(y^n)$ uniformly at random from $\{1, \dots, M\}$.

We now compute

$$E[\bar{P}_e(\underline{ENC}, P, \underline{DEC})]$$

$$= \frac{1}{M} \sum_{i=1}^{M} \underbrace{E[P_{e,i}(ENC, P, DEC)]}$$

$$M \left\{ \underbrace{\begin{bmatrix} enc(1)_1 & \cdots & enc(1)_n \\ \vdots & & \vdots \\ enc(M)_1 & \cdots & enc(M)_n \end{bmatrix}}_{n} \right.$$

encoder, decoder "symmetric" in the message

$$= \frac{1}{M} \sum_{i=1}^{M} E[P_{e,1}(ENC, P, DEC)]$$

$$= \underline{E[P_{e,1}(ENC, P, DEC)]}.$$

$$E(P_{e,1}(ENC, P, DEC))$$

$$= \sum_{enc} \Pr(ENC = enc) \underbrace{\sum_{y^n} P(y^n | enc(1)) \, \mathbb{1}\{dec(y^n) \neq 1\}}_{P_{e,1}(enc, P, dec)}$$

$$= \sum_{enc, y^n} \Pr(ENC = enc, Y^n = y^n \mid m = 1)\ \mathbb{1}\{dec(y^n) \neq 1\}$$

$$\mathbb{1}\{dec(y^n) \neq 1\}$$

$$\leq \mathbb{1}\{(enc(1), y^n) \notin T\}$$

$$+ \mathbb{1}\{enc(2), y^n) \in T\}$$

$$+ \mathbb{1}\{(enc(3), y^n) \in T\} + \dots \mathbb{1}\{(enc(M), y^n) \in T\}$$

$$\Uparrow$$

$$\mathbb{1}\{(enc(1), y^n) \in T,\ (enc(2), y^n) \notin T,\ \dots (enc(M), y^n) \notin T\}$$

$$\leq \mathbb{1}\{dec(y^n) = 1\}$$

$$\Rightarrow E[P_{e,1}(ENC, P, DEC)]$$

$$\leq E[\mathbb{1}\{(ENC(1), Y^n) \notin T\}]$$

$$+ \sum_{m=2}^{M} E[\mathbb{1}\{(ENC(m), Y^n) \in T\}]$$

$$= \Pr((ENC(1), Y^n) \notin T)$$

$$+ (M-1) \Pr((ENC(2), Y^n) \in T)$$

$(M-1) \le 2 \cdot 2^{nR}$ so

$$E(\bar{P}_e(ENC, P, DEC))$$

$$\le Pr_1((ENC(1), Y^n) \notin T)$$

$$+ 2 \cdot 2^{nR}\, Pr_1((ENC(2), Y^n) \in T)$$

$$Pr((ENC(1), Y^n) = (x^n, y^n))$$

$$= \left(\prod_{i=1}^{n} p_X(x_i)\right) \prod_{i=1}^{n} P(y_i \mid x_i) = \prod_{i=1}^{n} p_{XY}(x_i, y_i)$$

Thus $Pr_1((ENC(1), Y^n) \notin T) \searrow 0$ as n gets large.

$$Pr_1((ENC(2), Y^n) = (x^n, y^n))$$

$$= \left(\prod_{i=1}^{n} p_X(x_i)\right)\left(\prod_{i=1}^{n} p_Y(y_i)\right) = \prod_{i=1}^{n} q_{XY}(x_i, y_i)$$

$q_{XY} = p_X \cdot p_Y$.

So $\Pr\big(\underbrace{(\mathrm{ENC}(2), Y^n)}_{\sim q} \in \underset{p}{T} \big)$

$$\leq 2^{-n\left[\underbrace{D(p\|q)}\, - \varepsilon\text{-Junk}\right]}.$$

$$D(p\|q) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = I(X;Y) > R$$

$$\Pr\big((\mathrm{ENC}(2), Y^n) \in T \big) \leq 2^{-n(I(X;Y) - \varepsilon\text{Junk})},$$

So:

$$E\left[\bar{P}_e (\mathrm{ENC}, P, \mathrm{DEC}) \right]$$

$$\leq (\searrow 0 \text{ with } n \nearrow \infty)$$

$$+ 2 \cdot \underbrace{2^{nR} \cdot 2^{-n[I(X;Y) - \varepsilon\text{ Junk}]}}_{2^{-n(I(X;Y) - R - \varepsilon\text{ Junk}]}}.$$

Since $I(X;Y) > R$, we can find ε' s.t.

$\underbrace{I(X;Y) - R - \varepsilon' \text{ Junk}} > 0$. With this choice of ε', we have

$$\underline{E(\bar{P}_e(ENC, P, DEC))} \leq \underline{(\text{something} \searrow 0 \text{ as } n \nearrow)}$$

$$\underbrace{+ \, 2 \cdot 2^{-n(\overbrace{I - R - \varepsilon' \cdots}^{>0})}}_{\searrow 0 \text{ as } n \nearrow}$$

Now choose n sufficiently large s.t the right hand side is $< \varepsilon$. So, for large enough n,

$$E(\bar{P}_e(ENC, P, DEC)) < \varepsilon$$

$$\Rightarrow \exists \text{ enc, dec s.t } \bar{P}_e(\text{enc}, P, \text{dec}) \underline{< \varepsilon}$$

$$\text{enc}: \{1, \dots, M\} \to \mathcal{X}^n$$
$$\text{dec}: \mathcal{Y}^n \to \{1, \dots, M\} \quad \text{with } \underline{M \geq 2 \cdot 2^{nR}}.$$

//

<u>Proving Thm A:</u>