Note Title 25-Jun-19

$(X_1, X_2, \ldots, X_n) \in \{0,1\}^n$

$X_1 = 1$ w.p. p

$\quad\; 0 \qquad 1-p \qquad\qquad p \in (0,1)$

$B^{(n)} \subseteq \{0,1\}^n$

1. $P_e^{(n)}(B^{(n)}) = Pr((B^{(n)})^c) = 1 - Pr(B^{(n)})$ prob. of error (we would like this value to be small)
2. $\text{Ambiguity}(B^{(n)}) = \dfrac{|B^{(n)}|}{|\{0,1\}^n|}$ (we would like ambiguity to be small)

Suppose that we require $P_e^{(n)}(B^{(n)}) \leq \varepsilon$ for some fixed $\varepsilon > 0$, $\varepsilon < 1$

Now we will ask, how small can the ambiguity be?

Find a set $\boxed{B_\varepsilon^{(n)} = \arg\min_{B^{(n)} :\, P_e^{(n)}(B^{(n)}) \leq \varepsilon} \text{Ambiguity } B^{(n)}}$

Question : How does $\dfrac{|B_\varepsilon^{(n)}|}{|\{0,1\}^n|}$ behave?

$$\frac{|B_\varepsilon^{(n+1)}|}{|\{0,1\}^{n+1}|} \leq \frac{|B_\varepsilon^{(n)}|}{|\{0,1\}^{n}|}$$

Proof: Consider the set $B_\varepsilon^{(n)} \times \{0,1\} = B^{(n+1)}$

$$P_e^{(n)}(B^{(n+1)}) = \sum_{x^{n+1} \notin B^{(n+1)}} p(x^n) = \sum_{x^n \notin B_\varepsilon^{(n)}} \left[p(x^n)p(0) + p(x^n)p(1) \right]$$

$$= \sum_{x^n \notin B_\varepsilon^{(n)}} p(x^n) \left[\underbrace{p(0) + p(1)}_{=1} \right] = P_e^{(n)}\left((B_\varepsilon^{(n)})^c \right) \leq \varepsilon$$

$$\overbrace{\frac{|B_\varepsilon^{(n+1)}|}{|\{0,1\}^{n+1}|}} \leq \frac{|B^{(n+1)}|}{|\{0,1\}^{n+1}|} = \frac{|B_\varepsilon^{(n)}| \times 2}{2^{n+1}} = \frac{|B_\varepsilon^{(n)}|}{2^n} = \overbrace{\text{Ambiguity}(B_\varepsilon^{(n)})}$$

$\square$

Ambiguity $(B_\varepsilon^{(n)})$ is non-increasing in n.

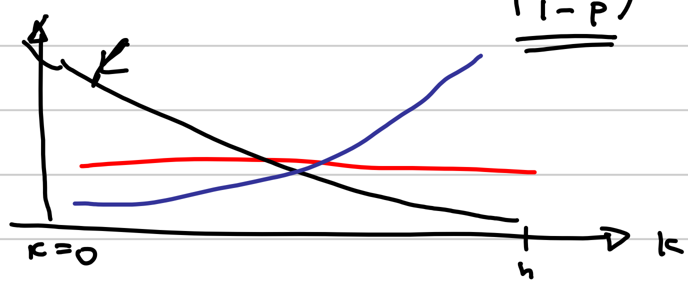
Ambiguity $(B_\varepsilon^{(n)}) \geq 0$

Let's try to reason it out.

Strings X^n with k ones and $(n-k)$ 0's all have probability

$$(1-p)^{n-k}\, p^k = (1-p)^n \left(\frac{p}{1-p}\right)^k$$

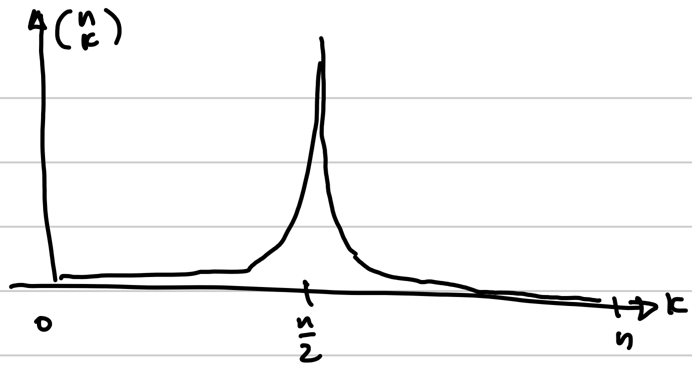
$p < \frac{1}{2}$

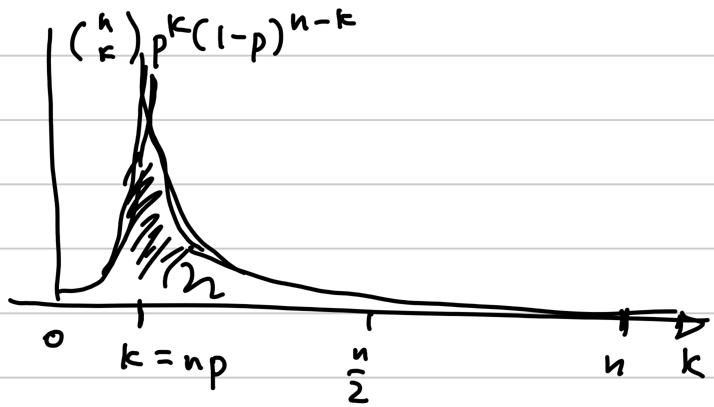


$p > \frac{1}{2}$

$\cancel{p = \frac{1}{2}}$

There are $\binom{n}{k}$ strings with $(n-k)$ 0's and k 1's.





WLLN

$\frac{1}{n} \sum_{i=1}^{n} X_i \to E[X]$ in prob

$= p$

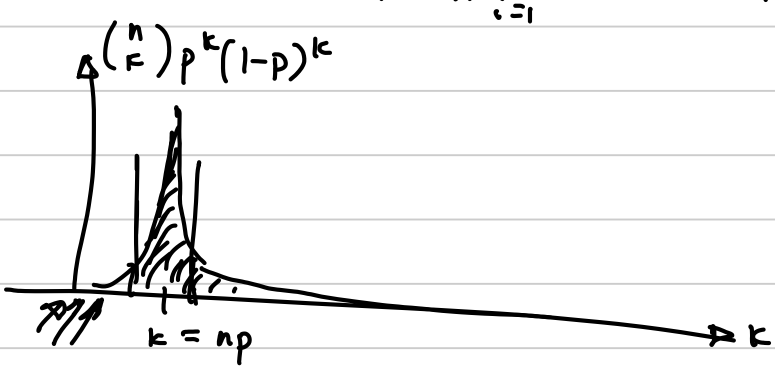
$\sum_{i=1}^{n} X_i = \#$ 1s in string

$\to np$

Weak Law of Large Numbers (WLLN)

Given $X_1, X_2, \ldots \sim \text{iid } p(x)$ on $\mathcal{X}$

$$\Pr\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - E[X] \right| \geq \sqrt{a} \right) \leq \frac{\operatorname{var}(X)}{a \cdot n} \to 0 \quad \text{as } n \to \infty$$



Given that with high probability we expect to see a string with np '1's an $n(1-p)$ '0's, it should also be true that

$$\underline{p(X^n)} \cong \underline{p^{np}(1-p)^{n(1-p)}}$$

Is that true??

$$\underline{p(X^n)} = \prod_{i=1}^{n} p(x_i) = 2^{\log \prod_{i=1}^{n} p(x_i)} = 2^{\sum_{i=1}^{n} \log p(x_i)} = 2^{n \cdot \frac{1}{n} \sum_{i=1}^{n} \log \underbrace{p(x_i)}}$$

$$= \underline{2^{-n \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{p(x_i)}}}$$

$$p(x^n) = 2^{-n \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{p(x_i)}\right)}}$$

$$\frac{1}{n}\sum_{i=1}^{n} \log \frac{1}{p(x_i)} \longrightarrow \underbrace{E\left[\log \frac{1}{p(x_i)}\right]} \text{ in probability}$$

$$H(X) \triangleq E\left[\log \frac{1}{p(x)}\right] = \text{"entropy"}$$

$$p(x^n) = 2^{-n \cdot \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{p(x_i)}\right)} \longrightarrow 2^{-nH(X)} \text{ in probability}$$

$$\Pr\left(2^{-n(H(X)+\varepsilon)} \le p(x^n) \le 2^{-n(H(X)-\varepsilon)}\right) \to 1 \text{ as } n \to \infty.$$

$$p^{np}(1-p)^{n(1-p)} = 2^{\log[p^{np}(1-p)^{n(1-p)}]}$$

$$= 2^{np\log p + n(1-p)\log(1-p)}$$

$$= 2^{-n[p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}]} = 2^{-nH(X)}.$$

$$A_\varepsilon^{(n)} = \{ x^n : 2^{-n(H(X)+\varepsilon)} \le p(x^n) \le 2^{-n(H(X)-\varepsilon)} \}$$ "typical set"

$\Pr(A_\varepsilon^{(n)}) \longrightarrow 1$ as $n \to \infty$ by WLLN

As long as our $P_e^{(n)}$ constraint was greater than 0, this set will meet that error constraint for n sufficiently large.

$$P_e^{(n)}(B_\varepsilon^{(n)}) \le \varepsilon$$

Let's bound $|A_\varepsilon^{(n)}|$

$$1 \ge \Pr(A_\varepsilon^{(n)}) = \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \ge |A_\varepsilon^{(n)}| \, 2^{-n(H(X)+\varepsilon)} \Rightarrow |A_\varepsilon^{(n)}| \le \frac{1}{2^{-n(H(X)+\varepsilon)}} = 2^{n(H(X)+\varepsilon)}$$

$$1-\varepsilon \le \Pr(A_\varepsilon^{(n)}) \le |A_\varepsilon^{(n)}| \, 2^{-n(H(X)-\varepsilon)} \Rightarrow |A_\varepsilon^{(n)}| \ge (1-\varepsilon)\, 2^{n(H(X)-\varepsilon)}$$

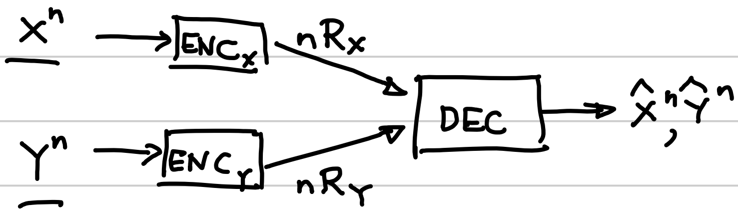
for n suff. large

for n suff. large

$$\text{Ambiguity}(A_\varepsilon^{(n)}) \le \frac{2^{n(H(X)+\varepsilon)}}{|\mathcal{X}|^n} = \frac{2^{n(H(X)+\varepsilon)}}{2^{n\log|\mathcal{X}|}} = 2^{-n(\log|\mathcal{X}|-H(X))}$$

$H(X) \le \log|\mathcal{X}|$ with equality iff $X \sim \text{Unif}(\mathcal{X})$.

$\therefore$ $\text{Ambiguity}(A_\varepsilon^{(n)}) \longrightarrow 0$ as $n \to \infty$ provided $X \nsim \text{Unif}(\mathcal{X})$.



$(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n) \sim p(x,y)$

$\exists\,(x,y)\in\mathcal{X}\times\mathcal{Y}$ s.t. $p(x,y) \neq p(x)p(y)$

A <u>multiple access source code</u> is defined by a pair of encoders

$$\alpha_{X,n} : \mathcal{X}^n \longrightarrow \{1,\ldots,2^{nR_X}\}$$
$$\alpha_{Y,n} : \mathcal{Y}^n \longrightarrow \{1,\ldots,2^{nR_Y}\}$$

represents all possible binary descriptions of length nR_X, nR_Y respectively

and a single decoder

$$\beta_n : \{1,\ldots,2^{nR_X}\}\times\{1,\ldots,2^{nR_Y}\} \longrightarrow \mathcal{X}^n\times\mathcal{Y}^n$$

Note Title 25-Jun-19

$$P_e^{(n)} = \Pr\left(\beta_n \left(\underbrace{\alpha_n(X^n)}, \underbrace{\alpha_n(Y^n)} \right) \neq (X^n, Y^n) \right)$$

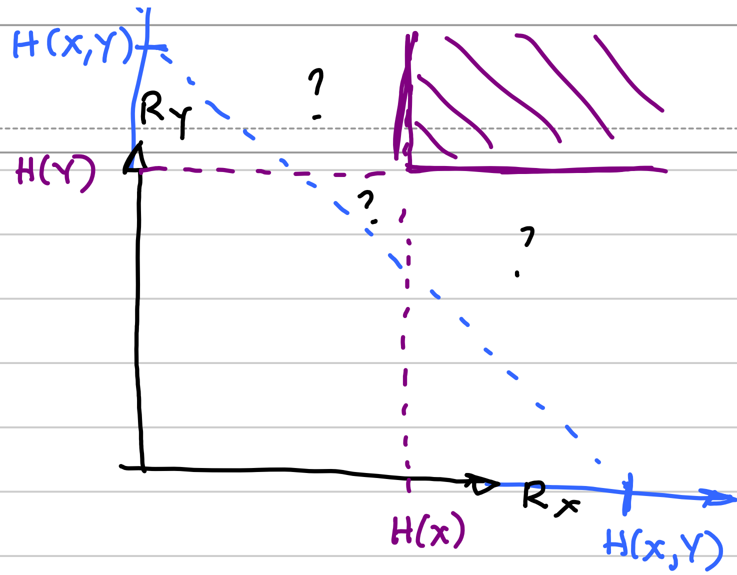
We say that a rate pair (R_X, R_Y) is <u>achievable</u> if $\exists$ a seq. of $((2^{nR_X}, 2^{nR_Y}), n)$-multiple access source codes

$$\{((\alpha_{X,n}, \alpha_{Y,n}), n)\}_{n=1}^{\infty}$$

with $P_e^{(n)} \longrightarrow 0$ as $n \to \infty$.

The source coding region is the closure of the set of achievable rates.

Note Title 25-Jun-19



$$X^n \rightarrow \boxed{\alpha_{X,n}} \xrightarrow{R_X} \boxed{\beta_{X,n}} \rightarrow \hat{X}^n$$

$$Y^n \rightarrow \boxed{\gamma_{Y,n}} \xrightarrow{R_Y} \boxed{\beta_{Y,n}} \rightarrow \hat{Y}^n$$

$$X^n, Y^n \rightarrow \boxed{\gamma_{(X,Y),n}} \xrightarrow{R_X+R_Y} \boxed{\beta_{(X,Y),n}} \rightarrow (\hat{X}^n, \hat{Y}^n)$$

Random code design:

For each $x^n \in \mathcal{X}^n$, choose

$\alpha_{X,n}(x^n) \sim \text{Unif}(\{1, \dots, 2^{nR_X}\})$ independently of all other descriptions

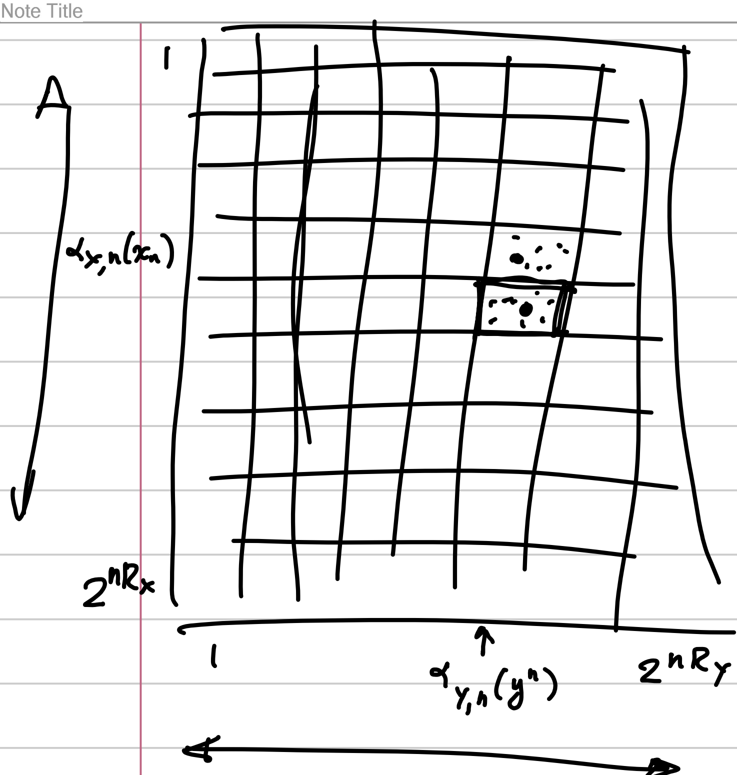
For each $y^n \in \mathcal{Y}^n$

$\alpha_{Y,n}(y^n) \sim \text{Unif}(\{1, \dots, 2^{nR_Y}\})$ independently of all other descriptions

Decoder design: We saw in Lecture 1 that

$$\left.\begin{bmatrix} X^n \in A_\varepsilon^{(n)}(X) \\ Y^n \in A_\varepsilon^{(n)}(Y) \\ (X^n, Y^n) \in A_\varepsilon^{(n)}(X,Y) \end{bmatrix}\right\} \text{ with probability approaching 1 as } n \to \infty$$

We will call any $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ that satisfies these constraints a "jointly typical pair"



For any $(w_x, w_Y) \in \{1, \dots, 2^{nR_x}\} \times \{1, \dots, 2^{nR_Y}\}$

$$\beta_n(w_x, w_Y) \triangleq \begin{cases} (x^n, y^n) & \text{if } \alpha_{x,n}(x^n) = w_x \\ & \quad \alpha_{Y,n}(y^n) = w_Y \\ & \text{AND } \underline{(x^n, y^n) \in A_\varepsilon^{(n)}} \\ & \text{AND } \nexists\, (\hat{x}^n, \hat{y}^n) \neq (x^n, y^n) \text{ s.t.} \\ & \quad \alpha_{x,n}(\hat{x}^n) = \alpha_{x,n}(x^n) = w_x \\ & \quad \alpha_{Y,n}(\hat{y}^n) = \alpha_{Y,n}(y^n) = w_Y \\ & \quad (\hat{x}^n, \hat{y}^n) \in A_\varepsilon^{(n)} \\ \text{"error"} & \text{otherwise} \end{cases}$$

jointly typical set

wrt distribution on codes

$E[P_e^{(n)}]$

Note Title 25-Jun-19

Find $E[P_e^{(n)}]$:

When does an error occur with our code?

① $E_0 = \{ (X^n, Y^n) \notin A_\varepsilon^{(n)} \}$

② $E_1 = \{ \exists\ (\hat{x}^n, \hat{y}^n) \neq (X^n, Y^n)$ s.t. $\alpha_{X,n}(\hat{x}^n) = \alpha_{X,n}(X^n)$, $\alpha_{Y,n}(\hat{y}^n) = \alpha_{Y,n}(Y^n)$, $(\hat{x}^n, \hat{y}^n) \in A_\varepsilon^{(n)} \}$

$= E_{1A} \cup E_{1B} \cup E_{1C}$

$E_{1A} = \{ \exists\ \hat{y}^n \neq Y^n$ s.t. $\alpha_{Y,n}(\hat{y}^n) = \alpha_{Y,n}(Y^n)$, $(X^n, \hat{y}^n) \in A_\varepsilon^{(n)} \}$

$E_{1B} = \{ \exists\ \hat{x}^n \neq X^n$ s.t. $\alpha_{X,n}(\hat{x}^n) = \alpha_{X,n}(X^n)$, $(\hat{x}^n, Y^n) \in A_\varepsilon^{(n)} \}$

$E_{1C} = \{ \exists\ \hat{x}^n \neq X^n, \hat{y}^n \neq Y^n$ s.t. $\alpha_{X,n}(\hat{x}^n) = \alpha_{X,n}(X^n)$, $\alpha_{Y,n}(\hat{y}^n) = \alpha_{Y,n}(Y^n)$, $(\hat{x}^n, y^n) \}$

Note Title 25-Jun-19

$$E[P_e^{(n)}] = \underline{Pr}(E_0 \cup E_{1A} \cup E_{1B} \cup E_{1C})$$

$$\le Pr(E_0) + Pr(E_{1A}) + Pr(E_{1B}) + Pr(E_{1C})$$

$Pr(E_0) \to 0 \qquad n \to \infty$

$$Pr(E_{1A}) = \sum_{\substack{x^n \in \mathcal{X}^n \\ y^n \in \mathcal{Y}^n}} p(x^n, y^n)\, Pr(\exists \hat{y}^n \ne y^n : \alpha_{Y,n}(\hat{y}^n) = \alpha_{Y,n}(y^n),\ (x^n, \hat{y}^n) \in A_\varepsilon^{(n)})$$

$$= \sum_{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n} p(x^n, y^n) \sum_{\hat{y}^n \ne y^n : (x^n, \hat{y}^n) \in A_\varepsilon^{(n)}} Pr(\alpha_{Y,n}(\hat{y}^n) = \alpha_{Y,n}(y^n))$$

$$= \sum_{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n} p(x^n, y^n) \sum_{\hat{y}^n \ne y^n : (x^n, \hat{y}^n) \in A_\varepsilon^{(n)}} 2^{-nR_Y}$$

$$= \sum_{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n} p(x^n, y^n)\, |\{\hat{y}^n \ne y^n : (x^n, \hat{y}^n) \in A_\varepsilon^{(n)}\}|\, 2^{-nR_Y}$$

Note Title 25-Jun-19

$$A_\varepsilon^{(n)}(Y \mid X^n = x^n)$$

$$|\{\underbrace{\hat{y}^n \neq y^n}: (x^n, y^n) \in A_\varepsilon^{(n)}\}| \leq |\underbrace{\{\hat{y}^n : (x^n, \hat{y}^n) \in A_\varepsilon^{(n)}\}}|$$

For any $(x^n, y^n) \in A_\varepsilon^{(n)}$: $2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$

$$2^{-n(H(X)+\varepsilon)} \leq p(y^n) \leq 2^{-n(H(Y)-\varepsilon)}$$

$$\underbrace{2^{-n(H(X,Y)+\varepsilon)} \leq p(x^n, y^n)} \leq 2^{-n(H(X,Y)-\varepsilon)}$$

$$1 \geq Pr(A_\varepsilon^{(n)}(Y \mid X^n = x^n) \mid X^n = x^n)$$

$$= \sum_{\hat{y}^n \in \underbrace{A_\varepsilon^{(n)}(Y \mid X^n = x^n)}} \underbrace{p(y^n \mid x^n)}$$

$$p(y^n \mid x^n) = \frac{p(x^n, y^n)}{p(x^n)}$$

$$\geq |A_\varepsilon^{(n)}(Y \mid X^n = x^n)| \frac{2^{-n(H(X,Y)+\varepsilon)}}{2^{-n(H(X)-\varepsilon)}} = \underbrace{2^{-n(\underbrace{H(Y \mid X)}+2\varepsilon)}} |A_\varepsilon^{(n)}(Y \mid X^n = x^n)|$$

$$H(Y \mid X) \triangleq H(X,Y) - H(X)$$

Note Title 25-Jun-19

$$|A_\varepsilon^{(n)}(Y|X^n=x^n)| \le 2^{n(H(Y|X)+2\varepsilon)}$$

$$\Pr(E_{1A}) \le 2^{n(H(Y|X)+2\varepsilon)}\, 2^{-nR_Y}$$
$$= 2^{-n(R_Y-(H(Y|X)+2\varepsilon))}$$
$$\to 0 \quad \text{as} \quad n\to\infty \quad \text{provided} \quad \boxed{R_Y > H(Y|X)+2\varepsilon}$$

$$\Rightarrow \underbrace{R_Y > H(Y|X)}$$

Similarly,

$$\Pr(E_{1B}) \le 2^{-n(R_X-(H(X|Y)+2\varepsilon))}$$
$$\to 0 \quad \text{as } n\to\infty \quad \text{provided} \quad \boxed{R_X > H(X|Y)+2\varepsilon}$$

$\Pr(E_{1C})$

Note Title 25-Jun-19

$$\Pr(E_{1c}) = \sum_{x^n, y^n} p(x^n, y^n) \Pr(\exists\ \hat{x}^n \neq x^n, \hat{y}^n \neq y^n \text{ s.t } \alpha_{x,n}(\hat{x}^n) = \alpha_{x,n}(x^n),\ \alpha_{y,n}(\hat{y}^n) = \alpha_{y,n}(y^n),\ (\hat{x}^n, \hat{y}^n) \in A_\varepsilon^{(n)} \})$$

$$= \sum_{x^n y^n} p(x^n y^n) \sum_{\substack{\hat{x}^n \neq x^n \\ \hat{y}^n \neq y^n : \\ (\hat{x}^n, \hat{y}^n) \in A_\varepsilon^{(n)}}} \underbrace{\Pr(\alpha_{x,n}(\hat{x}^n) = \alpha_{x,n}(x^n))}_{2^{-nR_x}} \cdot \underbrace{\Pr(\alpha_{y,n}(\hat{y}^n) = \alpha_{y,n}(y^n))}_{2^{-nR_y}}$$

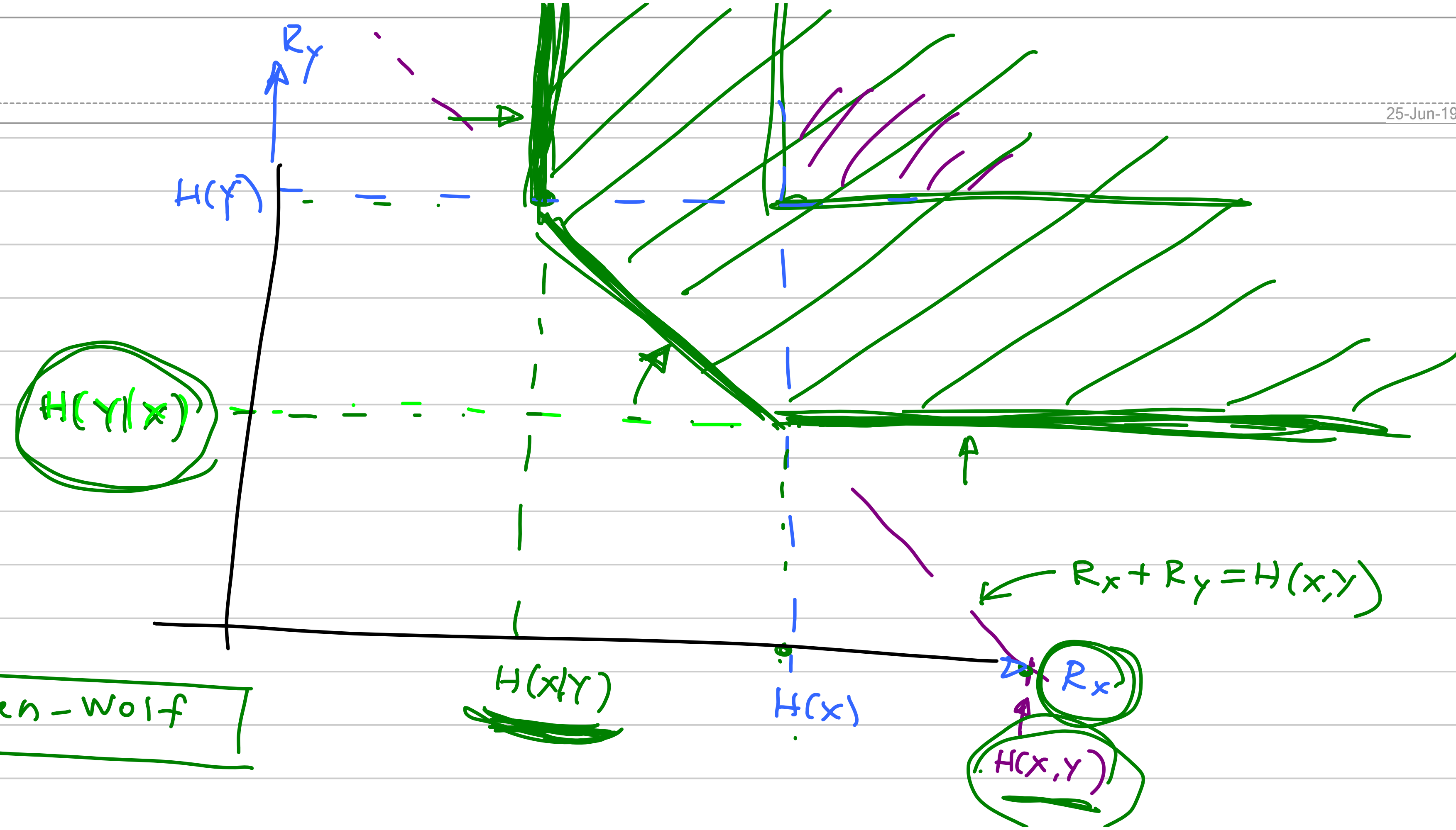
$$\leq |A_\varepsilon^{(n)}|\ 2^{-n(R_x + R_y)}$$

$$|A_\varepsilon^{(n)}| \leq 2^{n(H(X,Y)+\varepsilon)}$$

$$\leq 2^{-n(R_x + R_y - (H(X,Y)+\varepsilon))} \to 0 \text{ as } n \to \infty \text{ provided}$$

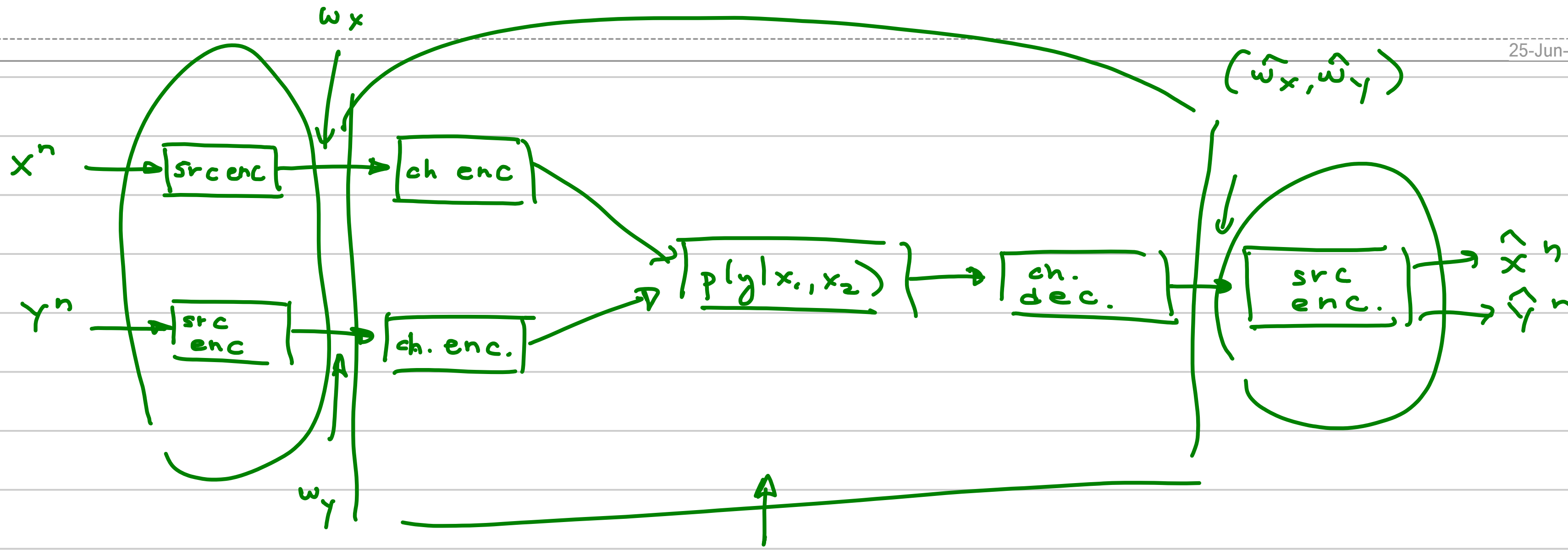
$$\boxed{R_x + R_y > H(X,Y) + \varepsilon}$$

Note Title 25-Jun-19



Slepian-Wolf

Note Title 25-Jun-19



Note Title 25-Jun-19

Note Title 25-Jun-19

Note Title 25-Jun-19

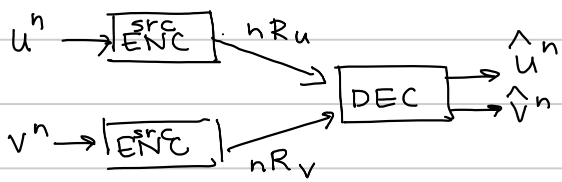
Note Title 25-Jun-19

Note Title 25-Jun-19

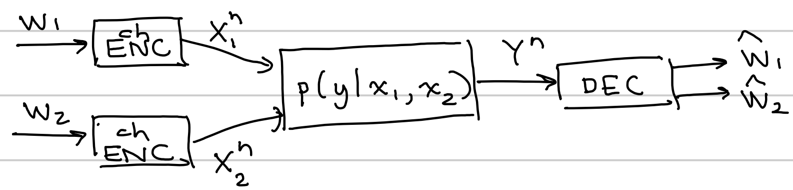
Note Title 25-Jun-19

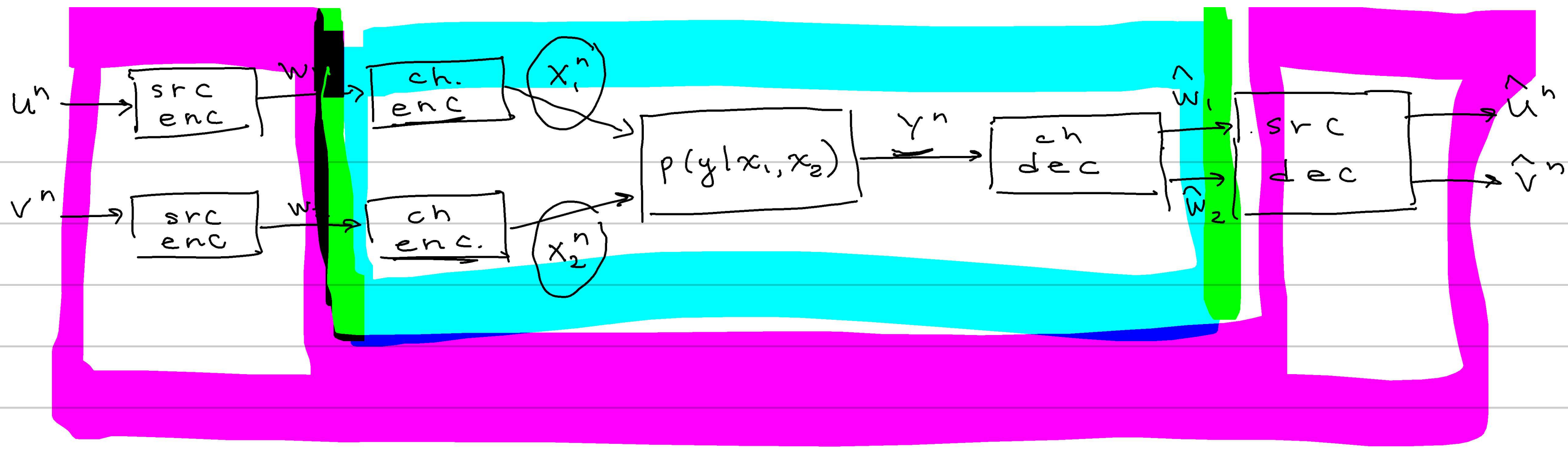
Network Information Theory : Lecture 3

Last time : Multiple Access Source Coding

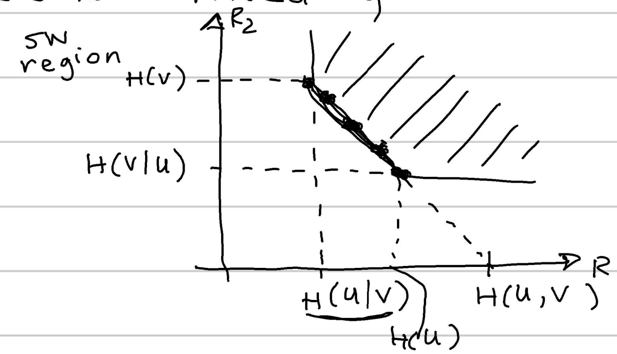


Today Multiple Access Channel Coding





A $((2^{nR_1}, 2^{nR_2}), n)$ multiple access channel (MAC) code is defined by



a pair of encoders

$$f_{1,n} : \{1, \dots, 2^{nR_1}\} \longrightarrow \mathcal{X}_1^n$$

$$f_{2,n} : \{1, \dots, 2^{nR_2}\} \longrightarrow \mathcal{X}_2^n$$

and a single decoder

$$g_n : \mathcal{Y}^n \longrightarrow \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}$$

We assume that the messages W_1 and W_2 are independent & uniformly distr

The probability of error for this code is

$$P_e^{(n)} = \Pr\left(g_n(Y^n) \neq (W_1, W_2) \mid X_1^n = f_{1,n}(W_1), X_2^n = f_{2,n}(W_2) \right).$$

A rate vector (R_1, R_2) is achievable if $\exists$ a seq. of $((2^{nR_1}, 2^{nR_2}), n)$ MAC codes $\{((f_{1,n}, f_{2,n}), g_n)\}_{n=1}^{\infty}$ with $P_e^{(n)} \to 0$ as $n \to \infty$.

The capacity region for the MAC is the closure of the set of all achievable rates.

To derive the capacity, we again use an argument in 2 parts:

Achievability

Converse

Achievability

Random code Design: Fix $p_1(x_1)$ on alphabet $\mathcal{X}_1$, $p_2(x_2)$ on alphabet $\mathcal{X}_2$.

Draw the codewords $f_{1,n}(1), f_{1,n}(2), \ldots, f_{1,n}(2^{nR_1}) \sim$ iid $\prod_{i=1}^{n} p_1(x_{1i})$

$f_{2,n}(1), f_{2,n}(2), \ldots, f_{2,n}(2^{nR_2}) \sim$ iid $\prod_{i=1}^{n} p_2(x_{2i})$.

Design the decoder : For each $y^n \in \mathcal{Y}^n$

$$g_n(y^n) = \begin{cases} (w_1, w_2) & \text{if } (f_{1,n}(w_1), f_{2,n}(w_2), y^n) \in A_\varepsilon^{(n)} \\ & \text{and } \nexists\ (\hat{w}_1, \hat{w}_2) \neq (w_1, w_2) \text{ s.t.} \\ & (f_{1,n}(\hat{w}_1), f_{2,n}(\hat{w}_2), y^n) \in A_\varepsilon^{(n)} \\ \text{"error"} & \text{otherwise} \end{cases}$$

$$E[P_e^{(n)}] = E\left[\frac{1}{2^{nR_1}} \cdot \frac{1}{2^{nR_2}} \sum_{(w_1,w_2)\in\{1,\dots,2^{nR_1}\}\times\{1,\dots,2^{nR_2}\}} Pr\left(g_n(Y^n) \neq (w_1,w_2) \mid X_1^n = f_{1,n}(w_1), X_2^n = f_{2,n}(w_2)\right)\right]$$

$$= \frac{1}{2^{nR_1}} \frac{1}{2^{nR_2}} \sum_{(w_1,w_2)} \underbrace{E\left[Pr\left(g_n(Y^n) \neq (w_1,w_2) \mid X_1^n = f_{1,n}(w_1), X_2^n = f_{2,n}(w_2)\right)\right]}_{\text{the same } \forall\ w_1, w_2 \text{ by symmetry of code design}}$$

$$= \underbrace{E\left[Pr\left(g_n(Y^n) \neq (1,1) \mid X_1^n = f_{1,n}(1), X_2^n = f_{2,n}(1)\right)\right]}$$

Events of interest: $E_{ij} = \{(\underline{f_{1,n}(i)}, \underline{f_{2,n}(j)}, Y^n) \in A_\varepsilon^{(n)}\}$

Error events: E_{11}^c : codewords transmitted are not jointly typical with Y^n received

$(i,j) \neq (1,1)$ E_{ij} : codewords (i,j) are jointly typical with Y^n but (i,j) was not the message pair sent.

$$E[P_e^{(n)}] = E[\Pr(g_n(Y^n) \neq (1,1) \mid (X_1^n, X_2^n) = (f_{1,n}(1), f_{2,n}(1)))]$$

$$= \Pr(E_{11}^c \cup \bigcup_{(i,j) \neq (1,1)} E_{ij} \mid \underbrace{(X_1^n, X_2^n) = (f_{1,n}(1), f_{2,n}(1))}_{\text{"(1,1) sent"}})$$

$$\leq \Pr(E_{11}^c \mid (1,1)\text{ sent}) + \sum_{j\neq 1} \Pr(E_{1j} \mid (1,1)\text{ sent}) + \sum_{i \neq 1} \Pr(E_{i1} \mid (1,1)\text{ sent})$$

$$+ \sum_{i\neq 1}\sum_{j \neq 1} \Pr(E_{ij} \mid (1,1)\text{ sent}) \quad \text{by union bound}$$

$\Pr(E_{11}^c \mid (1,1)\text{ sent}) \longrightarrow 0$ as $n \to \infty$ by AEP (WLLN)

$$\sum_{j\neq 1} \Pr(E_{1j} \mid (1,1)\text{ sent}) = (2^{nR_2}-1)\Pr(E_{12} \mid (1,1)\text{ sent}) \quad \text{by symmetry of code design}$$

$$\le 2^{nR_2} \sum_{(x_1^n,\hat{x}_2^n,y^n)\in A_\varepsilon^{(n)}} \underbrace{p_1(x_1^n)\,p_2(\hat{x}_2^n)\,p(y^n \mid x_1^n)}$$

$$\le 2^{nR_2} \sum_{(x_1^n,\hat{x}_2^n,y^n)\in A_\varepsilon^{(n)}} \underbrace{p_2(\hat{x}_2^n)}\,\underbrace{p(x_1^n,y^n)}$$

$$\le 2^{nR_2}\,|A_\varepsilon^{(n)}|\;2^{-n(H(X_2)-\varepsilon)}\;2^{-n(H(X_1,Y)-\varepsilon)}$$

$$\le 2^{nR_2}\;2^{n(H(X_1,X_2,Y)+\varepsilon)}\;2^{-n(H(X_2)-\varepsilon)}\;2^{-n(H(X_1,Y)-\varepsilon)}$$

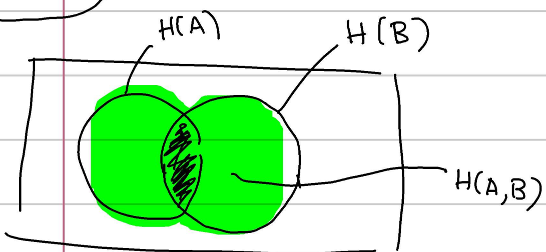
$$= 2^{-n(I(X_1,Y;X_2)-3\varepsilon-R_2)}$$

$\xrightarrow[n\to\infty]{} 0$ provided $R_2 < \underbrace{I(X_1,Y;X_2)} - 3\varepsilon$

$$= \underbrace{\cancel{I(X_1;X_2)}}_{=0 \text{ since } X_1 \perp\!\!\!\perp X_2} + \underbrace{I(Y;X_2 \mid X_1)}_{= I(X_2;Y \mid X_1)}$$

$$\boxed{R_2 < I(X_2;Y \mid X_1)}$$

$\underline{I(A;B)} = H(A)+H(B) - H(A,B)$



Similarly $\sum_{i \neq 1} \Pr(E_{i1} \mid (1,1) \text{ sent}) \longrightarrow 0$ as $n \to \infty$ provided

$$\boxed{R_1 < I(X_1; Y \mid X_2)}$$

Finally $\sum_{i \neq 1} \sum_{j \neq 1} \Pr(E_{ij} \mid (1,1) \text{ sent}) \leq \underbrace{2^{nR_1} 2^{nR_2}} \underbrace{\sum_{(\hat{x}_1^n, \hat{x}_2^n, y^n) \in A_\varepsilon^{(n)}}} \underbrace{p_1(\hat{x}_1^n)} \underbrace{p_2(\hat{x}_2^n)} \underbrace{p(y^n)}$

$$\leq 2^{n(R_1+R_2)} \left| A_\varepsilon^{(n)} \right| 2^{-n(H(X_1)-\varepsilon)} 2^{-n(H(X_2)-\varepsilon)} 2^{-n(H(Y)-\varepsilon)}$$

$$\leq 2^{n(R_1+R_2)} 2^{n(H(X_1,X_2,Y)+\varepsilon)} 2^{-n(\underbrace{H(X_1)+H(X_2)}_{H(X_1,X_2) \text{ since } X_1 \perp\!\!\!\perp X_2}+H(Y)-3\varepsilon)}$$

$$= 2^{-n(H(X_1,X_2)+H(Y)-H(X_1,X_2,Y)-4\varepsilon-(R_1+R_2))}$$

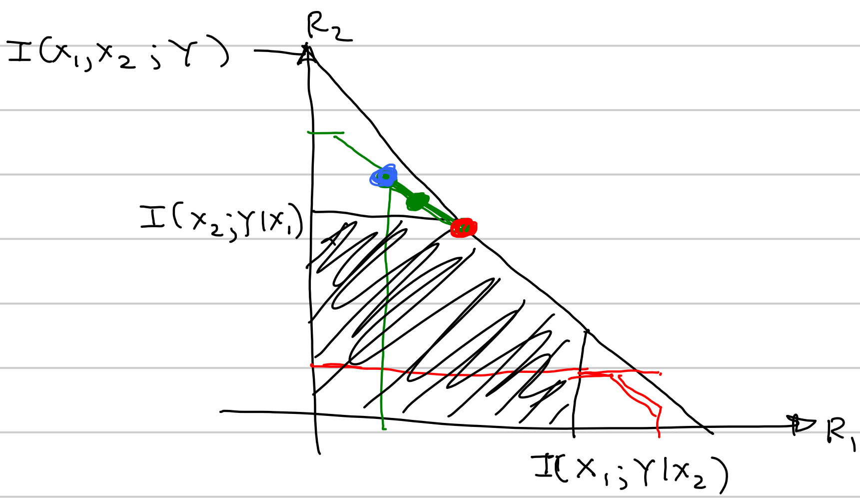
$$= 2^{-n(I(X_1,X_2;Y)-4\varepsilon-(R_1+R_2))}$$

$\longrightarrow 0$ as $n \to \infty$ provided $\boxed{R_1 + R_2 < I(X_1, X_2; Y)}$

$E[P_e^{(n)}] \to 0$ as $n \to \infty$ provided

$$\boxed{\begin{aligned} R_1 &< I(X_1;Y|X_2) \\ R_2 &< I(X_2;Y|X_1) \\ R_1+R_2 &< I(X_1,X_2;Y) \end{aligned}}$$

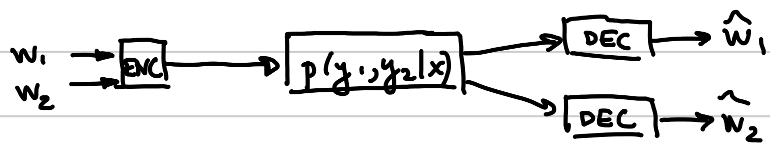
taking the union over all $p(x_1)p(x_2)$ and then the convex hull and the closure turns out to give the complete capacity region for MAC.



Network Information Theory : Lecture #4 June 26, 2019

Note Title 26-Jun-19

Broadcast Channel



$W_1 \perp\!\!\!\perp W_2$

We look at a special case of broadcast channels:

Physically degraded broadcast channels

$$p(y_1, y_2 | x) = \underbrace{p(y_1 | x) p(y_2 | y_1)}$$

A $((2^{nR_1}, 2^{nR_2}), n)$ broadcast channel (BC) code is defined by

a single encoder

$$f_n : \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\} \longrightarrow \mathcal{X}^n$$

and a pair of decoders

$$g_{1,n} : \mathcal{Y}_1^n \longrightarrow \{1, \ldots, 2^{nR_1}\}$$
$$g_{2,n} : \mathcal{Y}_2^n \longrightarrow \{1, \ldots, 2^{nR_2}\}.$$

The average probability of error for the code is

$$P_e^{(n)} = \Pr\left((g_{1,n}(Y_1^n), g_{2,n}(Y_2^n)) \neq (W_1, W_2) \mid X^n = f_n(W_1, W_2)\right)$$
$$(W_1, W_2) \sim \text{Unif}\left(\{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\}\right)$$

A rate pair (R_1, R_2) is <u>achievable</u> if $\exists$ a seq. of $((2^{nR_1}, 2^{nR_2}), n)$ BC codes $\{(f_n, (g_{1,n}, g_{2,n}))\}_{n=1}^{\infty}$ with $P_e^{(n)} \to 0$ as $n \to \infty$.

The <u>capacity region</u> is the closure of the set of achievable rates.

<u>Achievability</u> :

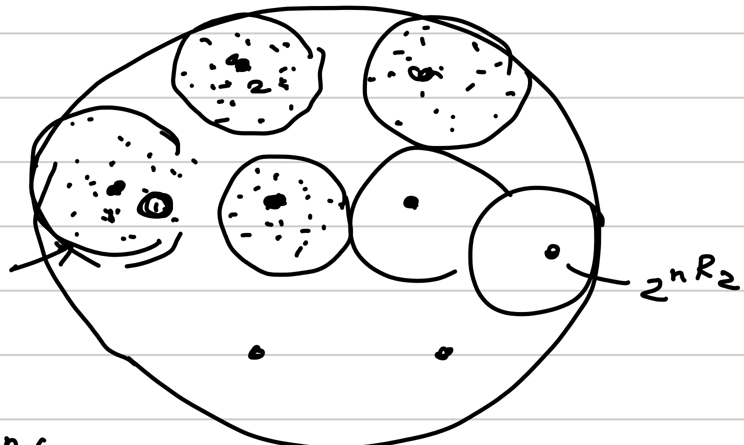
Random code design: Fix $p(u)$. Fix $p(x|u)$

Draw $\underbrace{U^n(1), U^n(2), \ldots, U^n(2^{nR_2})}$ $\sim$ iid $\prod_{i=1}^{n} p(u_i)$.

For each w_2, draw

$$\underbrace{f_n(1, w_2), f_n(2, w_2), \ldots, f_n(2^{nR_1}, w_2)}_{\sim \text{ iid } \prod_{i=1}^{n} p(x_i | u_i)}$$

where $u^n = U^n(w_2)$.



Decoder design: For each $y_2^n \in \mathcal{Y}_2^n$

$$g_{2,n}(y_2^n) = \begin{cases} w_2 & \text{if } (U^n(w_2), Y_2^n) \in A_\varepsilon^{(n)} \\ & \text{and } \nexists\ \hat{w}_2 \neq w_2 \text{ s.t } (U^n(\hat{w}_2), Y_2^n) \in A_\varepsilon^{(n)} \\ \text{"error"} & \text{otherwise.} \end{cases}$$

For each $y_1^n \in \mathcal{Y}_1^n$,

$$g_{1,n}(y_1^n) = \begin{cases} w_1 & \text{if } \exists \text{ a unique } w_2 \text{ s.t. } (u^n(w_2), f_n(w_1,w_2), Y_1^n) \in A_\varepsilon^{(n)} \\ & \text{and } \nexists\ (\hat{w}_1,\hat{w}_2) \neq (w_1,w_2) \text{ s.t. } (u^n(\hat{w}_2), f_n(\hat{w}_1,\hat{w}_2), Y_1^n) \in A_\varepsilon^{(n)} \\ \text{"error"} & \text{otherwise} \end{cases}$$

$$E[P_e^{(n)}] = E\Big[\frac{1}{2^{nR_1}} \cdot \frac{1}{2^{nR_2}} \sum_{(w_1,w_2) \in \{1,\dots,2^{nR_1}\} \times \{1,\dots,2^{nR_2}\}} Pr\big((g_{1,n}(Y_1^n), g_{2,n}(Y_2^n)) \neq (w_1,w_2) \,\big|\, X^n = f_n(w_1,w_2)\big)\Big]$$

$$= E\Big[Pr\big((g_{1,n}(Y_1^n), g_{2,n}(Y_2^n)) \neq (1,1) \,\big|\, X^n = f_n\underbrace{(1,1)}\big)\Big]$$

Events of interest at receiver 2:

$$E_j^{(2)} = \{ (\underbrace{u^n(j)}, \underbrace{Y_2^n}) \in A_\varepsilon^{(n)} \}$$

Events of interest at receiver 1:

$$E_j^{(1)} = \{ (u^n(j), Y_1^n) \in A_\varepsilon^{(n)} \}, \quad E_{ij}^{(1)} = \{ (u^n(j), f_n(i,j), Y_1^n) \in A_\varepsilon^{(n)} \}$$

$$E[P_e^{(n)}] = \Pr\Big((E_1^{(2)})^c \cup \bigcup_{j\neq 1} E_j^{(2)} \cup \underbrace{(E_{11}^{(1)})^c}_{} \cup \underbrace{\bigcup_{j\neq 1} E_j^{(1)}}_{} \cup \bigcup_{i\neq 1} E_{i1}^{(1)} \,\Big|\, \underbrace{x^n = f_n(1,1)}_{\text{"(1,1) sent"}} \Big)$$

$$\le \Pr\big((E_1^{(2)})^c \mid (1,1) \text{ sent}\big) + \Pr\big((E_{11}^{(1)})^c \mid (1,1) \text{ sent}\big)$$

$$+ \underbrace{\sum_{j\neq 1} \Pr\big(E_j^{(2)} \mid (1,1) \text{ sent}\big)}_{} + \sum_{j\neq 1} \Pr\big(E_j^{(1)} \mid (1,1) \text{ sent}\big)$$

$$+ \underbrace{\sum_{i\neq 1} \Pr\big(E_{i1}^{(1)} \mid (1,1) \text{ sent}\big)}_{}$$

$$\Pr\big((E_1^{(2)})^c \mid (1,1) \text{ sent}\big) + \Pr\big((E_1^{(1)})^c \mid (1,1) \text{ sent}\big) \longrightarrow 0 \quad \text{as } n\to\infty \quad \text{AEP}$$

$$\sum_{j\neq 1} \Pr\big(E_j^{(2)} \mid (1,1) \text{ sent}\big) \le 2^{nR_2} \sum_{(\hat{u}^n, y_2^n) \in A_\varepsilon^{(n)}} p(\hat{u}^n)\, p(y_2^n)$$

$$\le \underbrace{2^{nR_2}}_{}\, 2^{n(H(\underbrace{U,Y_2}_{})+\varepsilon)}\, 2^{-n(H(\underbrace{U}_{})-\varepsilon)}\, 2^{-n(H(\underbrace{Y_2}_{})-\varepsilon)}$$

$$= 2^{-n(I(U;Y_2) - 3\varepsilon - R_2)} \xrightarrow[n\to\infty]{} 0 \quad \text{provided } R_2 < I(U;Y_2)$$

similarly

$$\sum_{j\neq 1} \Pr\left(E_j^{(1)} \mid (1,1)\text{ sent}\right) \le 2^{-n(I(U;Y_1)-3\varepsilon-R_2)} \xrightarrow[n\to\infty]{} 0 \quad \text{provided } R_2 < I(U;Y_1)$$

Finally :

$$\sum_{i\neq 1} \Pr\left(E_{i1}^{(1)} \mid (1,1)\text{ sent}\right) = \left(2^{nR_1}-1\right)\Pr\left(E_{21}^{(1)} \mid (1,1)\text{ sent}\right)$$

$$\le 2^{nR_1} \sum_{(u^n,\hat{x}^n,y_1^n)\in A_\varepsilon^{(n)}} \underbrace{p(u^n)\,p(\hat{x}^n|u^n)}\, p(y_1^n|u^n)$$

$$\le 2^{nR_1}\, 2^{n(H(U,X,Y_1)+\varepsilon)}\, 2^{-n(H(U)-\varepsilon)}\, \frac{2^{-n(H(X,U)-\varepsilon)}}{2^{-n(H(U)+\varepsilon)}} \cdot \frac{2^{-n(H(U,Y_1)-\varepsilon)}}{2^{-n(H(U)+\varepsilon)}}$$

$$= 2^{-n(\underbrace{-H(U,X,Y_1)+H(U)}+H(X|U)+H(Y_1|U)-5\varepsilon-R_1)}$$

$$= 2^{-n(-H(X,Y_1|U)+H(X|U)+H(Y_1|U)-5\varepsilon-R_1)}$$

$$= 2^{-n(I(X;Y_1|U) - 5\varepsilon - R_1)}$$

$\longrightarrow 0$ as $n \to \infty$ provide $R_1 < I(X;Y_1|U)$

$$\left.\begin{aligned} R_1 &< I(X;Y_1|U) \\ R_2 &< I(U;Y_1) \\ R_2 &< I(U;Y_2) \end{aligned}\right\} \to R_2 < \underbrace{\min\{I(U;Y_1), I(U;Y_2)\}}_{\text{by data processing inequality}} = I(U;Y_2)$$

$$\boxed{\begin{aligned} R_1 &< I(X;Y_1|U) \\ R_2 &< I(U;Y_2) \end{aligned}}$$

This turns out to describe the capacity region of the broadcast channel in the physically degraded case

It turns out that for any pair of channels

$$\underbrace{p(y_1, y_2 | x)}$$

$$\underbrace{\hat{p}(y_1, y_2 | x)}$$

for which
$$p(y_1 | x) = \hat{p}(y_1 | x)$$
$$p(y_2 | x) = \hat{p}(y_2 | x)$$

the capacity regions for BCs $p(y_1, y_2 | x)$ & $\hat{p}(y_1, y_2 | x)$ are the same.

We say that channel $\hat{p}(y_1, y_2 | x)$ is <u>stochastically</u> <u>degraded</u> if $\exists$ a physically degraded channel $p(y_1, y_2 | x)$ different from $\hat{p}$

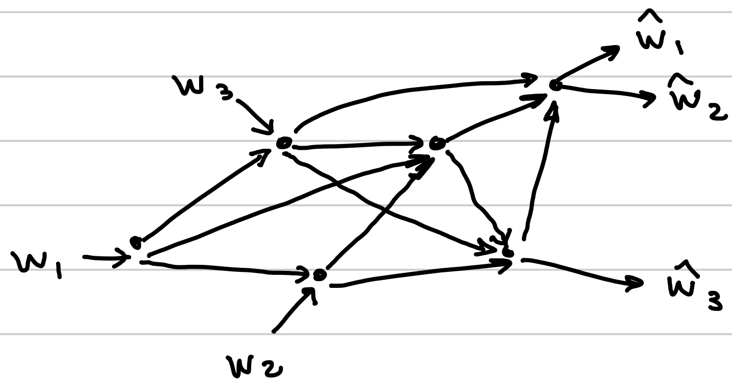
s.t. $\hat{p}(y_1|x) = p(y_1|x) \quad \forall\ x, y_1$

$$\hat{p}(y_2|x) = p(y_2|x) \quad \forall\ x, y_2 .$$

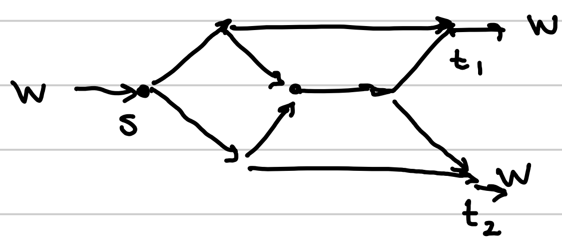
We have solved the capacity region for all stochastically degraded channels.

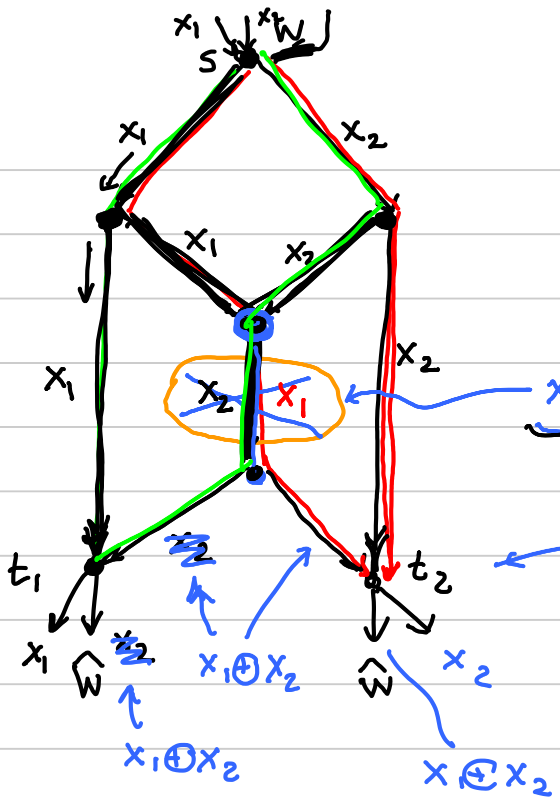
Lecture 5 - Network Coding

Note Title 27-Jun-19



Multicast Network Coding





$x_1 \oplus x_2$

$\mathcal{N}$: $G = (V, E)$

each edge $e \in E$ has a fixed capacity C_e

$\Rightarrow$ edge e can carry $n C_e$ bits over n channel uses

$(x_2, x_1 \oplus x_2) \rightarrow (x_1, x_2)$

$$x_1, \; x_1 \oplus x_2 \Rightarrow \begin{cases} x_1 \oplus (x_1 \oplus x_2) = x_2 \\ x_1 \end{cases}$$

A $(2^{nR}, n)$ multicast network code for an acyclic network is defined by:

A collection of encoders:

For each edge $e = (\underline{u}, v)$

the encoder $f_{n,e} : \prod_{(w,\underline{u}) \in E} \{1, \ldots, \underbrace{2^{n C(w,u)}}\} \longrightarrow \{1, \ldots, 2^{n C_e}\}$

For each terminal node $t \in T$, where T = set of terminal nodes $\subseteq V$

$$g_{n,t} : \prod_{(w,t) \in E} \{1, \ldots, 2^{n C(w,t)}\} \longrightarrow \{1, \ldots, 2^{nR}\}$$

The probability of error for this code (averaged over all possible messages $W \in \{1, \ldots, 2^{nR}\}$, $W \sim \text{Uniform}\{1, \ldots, 2^{nR}\}$)

$$\boxed{P_e^{(n)} = \Pr\left(\bigcup_{t \in T} \{ g_{n,t}(Y_t^n) \neq W \} \right)}$$

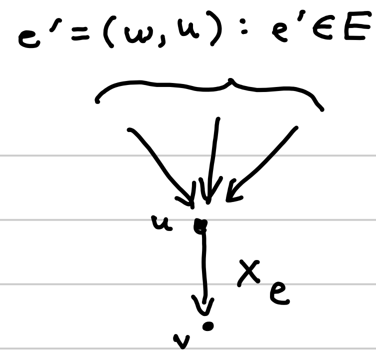
For $e=(u,v)\in E$:

Where X_e^n = information carried by edge

$= f_{n,e}(X_{e'}^n : e'=(w,u),\ e'\in E)$

For any $v\in V$:

$Y_v^n = (X_{(u,v)}^n : (u,v)\in E)$.



Notice that X_e^n is some deterministic function of W that is being calculated in a step-by-step process across this network.

The step-by-step functions are sometimes called "local encoding functions".

Each of those functions can also be represented by a single deterministic function of the source information W.

For example, for $e=(u,v)$

$$X_e^n = f_{n,e}(X_{e'}^n : e'=(w,u),\ e'\in E) = f_{n,e}(Y_u^n)$$

can also be represented by some function $X_e^n = F_{n,e}(W)$.

Achievability:

Random encoder design:

For each edge $e=(u,v)\in E$ and $y_u^n \in \underbrace{\mathcal{Y}_u^n}_{} = \prod_{(w,u)\in E} \{1,\dots,2^{nC_{(w,u)}}\}$,

choose $f_{n,e}(y_u^n) \sim$ iid Uniform$\{1,\dots,2^{nC_e}\}$.

Fix this code. Let $F_{n,e}(w)$ represent the end-to-end operation that chooses X_e^n.

Design the decoder:

For each $t\in T$ and each output $\overbrace{y_t^n \in \underbrace{\mathcal{Y}_t^n}_{}}$,

$$g_{n,t}(\underbrace{y_t^n}_{}) = \begin{cases} w & \text{if } (F_{n,(u,t)}(w) : (u,t)\in E) = \underbrace{y_t^n}_{} \\ & \text{and } \nexists\ \hat{w}\neq w\ (F_{n,(u,t)}(\hat{w}) : (u,t)) = y_t^n \\ \text{"error"} & \text{otherwise} \end{cases}$$

$$F_{n,t}(w) \triangleq (F_{n,(u,t)}(w) : (u,t)\in E)$$

$$E[P_e^{(n)}] = E\left[\frac{1}{2^{nR}}\sum_{\omega=1}^{2^{nR}} \Pr\left(\bigcup_{t\in T} g_{n,t}(F_{n,t}(\omega)) \neq \omega\right)\right]$$

$$= E\left[\Pr\left(\bigcup_{t\in T} g_{n,t}(F_{n,t}(1)) \neq 1\right)\right]$$

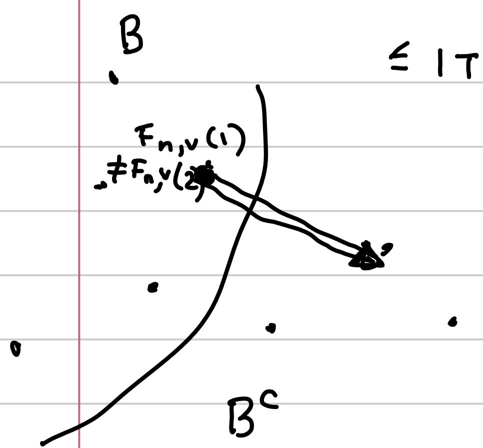
$$\leq \sum_{t\in T} E\left[\Pr\left(g_{n,t}(F_{n,t}(1)) \neq 1\right)\right]$$

$$\leq |T| \max_{t\in T} \underline{E\left[\Pr\left(g_{n,t}(F_{n,t}(1)) \neq 1\right)\right]}$$

$$\leq |T| \max_{t\in T} \sum_{\hat{\omega}\neq 1} \Pr\left(F_{n,t}(\underline{\hat{\omega}}) = F_{n,t}(1)\right)$$

$$\lesssim |T| \max_{t\in T} 2^{nR} \Pr\left(\underline{F_{n,t}(2) = F_{n,t}(1)}\right)$$

$$= |T| \max_{t\in T} 2^{nR} \Pr\left(\bigcup_{\substack{B\subset V :\, s\in B\\ t\notin B}} \{F_{n,v}(2) \neq F_{n,v}(1)\ \forall v\in B,\ F_{n,v}(2) = \underline{F_{n,v}(1)}\ \forall v\notin B\}\right)$$



$$\le |T| \max_{t \in T} 2^{nR} \sum_{\substack{B \subset V :\, s \in B \\ t \in B^c}} \underbrace{\Pr\big(F_{n,v}(2) \neq F_{n,v}(1) \ \forall v \in B, \ F_{n,v}(2) = F_{n,v}(1) \ \forall v \notin B \big)}$$

$$\underbrace{\Pr\big(F_{n,v}(2) \neq F_{n,v}(1) \ \forall v \in B \big) \Pr\big(F_{n,v}(2) = F_{n,v}(1) \ \forall v \in B^c \ \big| \ F_{n,v}(2) \neq F_{n,v}(1) \ \forall v \in B \big)}$$

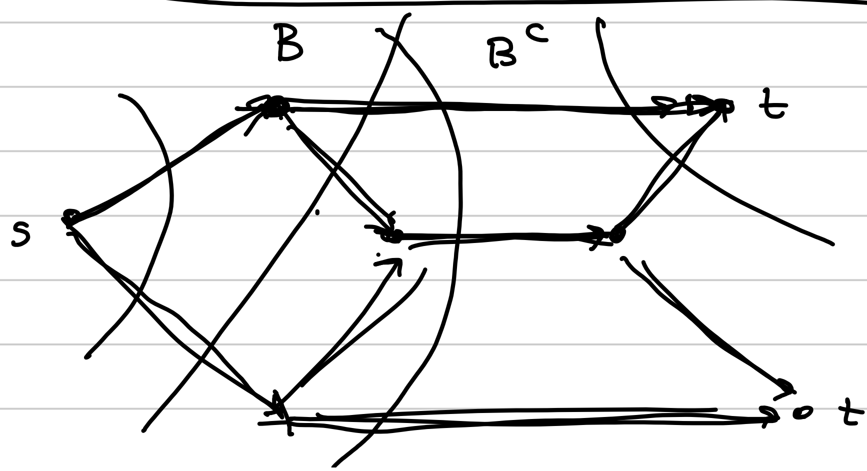
$$\le \Pr\big(F_{n,v}(2) = F_{n,v}(1) \ \forall v \in B^c \ \big| \ F_{n,v}(2) \neq F_{n,v}(1) \ \forall v \in B \big)$$

$$\le |T| \max_{t \in T} \underbrace{2^{nR}} \ 2^{|V|} \max_{\substack{B \subset V :\, s \in B \\ t \in B^c}} \prod_{\substack{e \in E :\, e = (i,j) \\ i \in B \\ j \in B^c}} 2^{-nC_e}$$

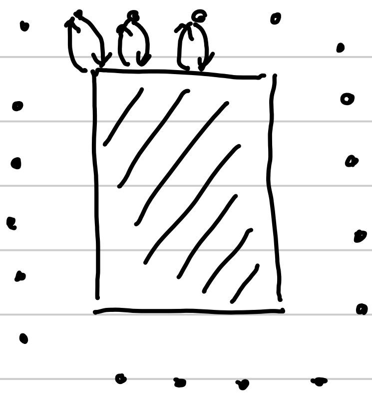
$$= \underbrace{|T|\, 2^{|V|}}\ \overbrace{\underbrace{\max_{t \in T}\ \max_{\substack{B \subset V :\, s \in B \\ t \in B^c}}}\ 2^{-n\left(\sum_{\substack{e \in E :\, e=(i,j) \\ i \in B \\ j \notin B}} C_e - R\right)}}$$

$\longrightarrow 0$ as $n \to \infty$ provided that

$$\boxed{R < \min_{t \in T}\ \min_{\substack{B \subset V :\, s \in B \\ t \in B^c}}\ \sum_{\substack{(i,j) \in E :\, i \in B \\ j \in B^c}} C_e}$$



Lecture # 6 : General Multi-terminal Networks



$$\underbrace{p(y^{(1)}, y^{(2)}, \ldots, y^{(K)} \mid x^{(1)}, \ldots, x^{(K)})}$$

Let $W^{(ij)}$ = msg. from node i intended for node j

$\in \{1, \ldots, 2^{nR^{(ij)}}\}$

Note Title 27-Jun-19

A $((2^{nR_{ij}})_{i,j\in\{1,\dots,k\}}, n)$ code for our network is defined by a family of encoders, where for each node v, at each time $t\in\{1,\dots,n\}$, the encoder for node v at time i

$$f^{(v)}_{n,t} : \left(\prod_{l=1}^{K} \{1,\dots,2^{nR_{vl}}\}\right) \times (\mathcal{Y}^{(v)})^{t-1} \longrightarrow \underline{\mathcal{X}^{(v)}}.$$

and a family of decoders:

$$\underbrace{g^{(v)}_n} : \left(\prod_{l=1}^{K} \{1,\dots,2^{nR_{vl}}\}\right) \times (\mathcal{Y}^{(v)})^{n} \longrightarrow \prod_{l=1}^{K} \{1,\dots,2^{nR_{lv}}\}$$

we will use the notation

$$\underbrace{g_n^{(j,v)} : \left(\prod_{l=1}^{k} \{1,\dots,2^{nR_{vl}}\}\right) \times (\mathcal{Y}^{(v)})^{n} \longrightarrow \{1,\dots,2^{nR_{jv}}\}}$$

to represent the component (j,v) from the decoder

Note Title 27-Jun-19

The error probability for this code is

$$\Pr\left(\bigcup_{i=1}^{K} \bigcup_{j=1}^{K} \left\{ g_n^{(i,j)}\left(W^{(j,1)}, W^{(j,2)}, \ldots, W^{(j,K)}, (Y^{(j)})^n \right) \neq W^{(i,j)} \right\} \right).$$

We will say that a rate vector $(R_{ij} : i \in \{1, \ldots, K\}, j \in \{1, \ldots, K\})$ is achievable if $\exists$ a seq. of $(2^{nR_{ij}} : i, j \in \{1, \ldots, K\})$ with $P_e^{(n)} \to 0$ as $n \to \infty$.

We will try to derive the following converse:

For any subset $S \subset \{1, \ldots, K\}$

$$\boxed{\sum_{i \in S, j \in S^c} R_{ij} \leq I\left(\underline{X^{(S)}} ; Y^{(S^c)} \mid \underline{X^{(S^c)}} \right)}$$

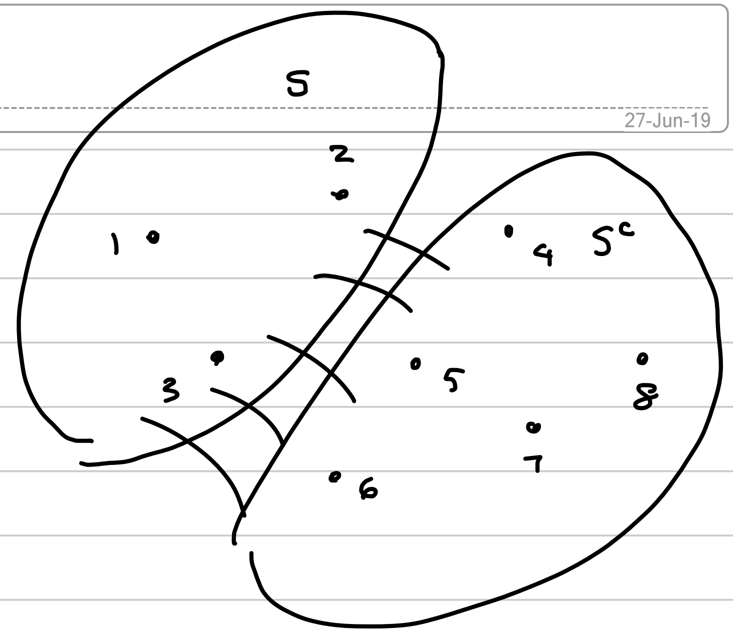
Note Title 27-Jun-19

where $X^{(S)} = (X^{(v)} : v \in S)$

$X^{(S^c)} = (X^{(v)} : v \in S^c = \{1, \dots, k\} \setminus S)$

$Y^{(S)} = (Y^{(v)} : v \in S)$

$Y^{(S^c)} = (Y^{(v)} : v \in S^c)$



Let $T = \{(i,j) : i \in S, j \in S^c\}$.

Let $T^c = \{(i,j) : i,j \in \{1, \dots, k\}\} \setminus T$.

Let $W^{(T)} = (W^{(ij)} : (ij) \in T)$

$W^{(T^c)} = (W^{(ij)} : (ij) \in T^c)$

Note Title 27-Jun-19

$$\sum_{(ij)\in T} nR_{ij} = \sum_{(ij)\in T} H(W^{(ij)}) \quad \text{since } W^{(ij)} \sim \text{Unif}(\{1,\dots,2^{nR_{ij}}\})$$

$$= H(W^{(T)}) \quad \text{since all } W^{(ij)} \text{ are indep.}$$

$$= H(W^{(T)} \mid W^{(T^c)}) \quad \text{since } W^{(T)} \perp\!\!\!\perp W^{(T^c)}$$

$$= I(W^{(T)}; (Y^{(S^c)})^n \mid W^{(T^c)}) + H(W^{(T)} \mid (Y^{(S^c)})^n, W^{(T^c)})$$

$$\le I(W^{(T)}; (Y^{(S^c)})^n \mid W^{(T^c)}) + n\varepsilon_n$$

for some $\varepsilon_n \to 0$ by Fano's inequality since $((Y^{(S^c)})^n, W^{(T^c)})$ contains all the decoder inputs for all nodes in S^c used to reconstruct the msg.s $W^{(T)}$.

$$= H((Y^{(S^c)})^n \mid W^{(T^c)}) - H((Y^{(S^c)})^n \mid W^{(T)}, W^{(T^c)}) + n\varepsilon_n$$

Note Title 27-Jun-19

$$= \sum_{t=1}^{n} \left[H\left(Y_t^{(S^c)} \mid W^{(T^c)}, (Y^{(S^c)})^{t-1} \right) - H\left(Y_t^{(S^c)} \mid W^{(T)}, W^{(T^c)}, (Y^{(S^c)})^{t-1} \right) \right] + n\varepsilon_n$$

chain rule for entropy

$$= \sum_{t=1}^{n} H\left(Y_t^{(S^c)} \mid X_t^{(S^c)}, W^{(T^c)}, (Y^{(S^c)})^{t-1} \right) - H\left(Y_t^{(S^c)} \mid X_t^{(S^c)}, W^{(T)}, W^{(T^c)}, (Y^{(S^c)})^{t-1} \right) \right] + n\varepsilon_n$$

since $X_t^{(S^c)}$ is a det. func. of the values $W^{(T^c)}$ and $(Y^{(S^c)})^{t-1}$!

$$\leq \sum_{t=1}^{n} \left[H\left(Y_t^{(S^c)} \mid X_t^{(S^c)} \right) - H\left(Y_t^{(S^c)} \mid W^{(T)}, W^{(T^c)}, (Y^{(S^c)})^{t-1}, X_t^{(S^c)}, X_t^{(S)} \right) \right] + n\varepsilon_n$$

by conditioning reduces entropy

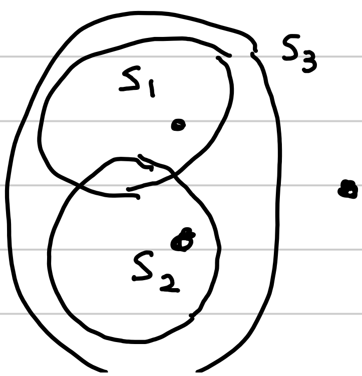
$$= \sum_{t=1}^{n} \left[H\left(Y_t^{(S^c)} \mid X_t^{(S^c)} \right) - H\left(Y_t^{(S^c)} \mid X_t^{(S)}, X_t^{(S^c)} \right) \right] + n\varepsilon_n$$

since $\left(W^{(T)}, W^{(T^c)}, (Y^{(S^c)})^{t-1} \right) \to \left(X_t^{(S)}, X_t^{(S^c)} \right) \to Y_t^{(S^c)}$ form a Markov chain

$$= \sum_{t=1}^{n} I\left(X_t^{(S)} ; Y_t^{(S^c)} \mid X_t^{(S^c)}\right) + n\varepsilon_n$$

With a bit more work, this gives

$$\boxed{\sum R_{ij} \le I\left(X^{(S)} ; Y^{(S^c)} \mid X^{(S^c)}\right)}$$



$$\begin{array}{ll} S_1 : & R_1 \le I(X_1 ; Y \mid X_2) \\ S_2 : & R_2 \le I(X_2 ; Y \mid X_1) \\ S_3 : & R_1 + R_2 \le I(X_1, X_2 ; Y) \end{array}$$

$$(w^{(T)}, w^{(T^c)}, (Y^{(S^c)})^{t-1}) \to (X_t^{(S)}, X_t^{(S^c)}) \to Y_t^{(S^c)}$$

$$p(y_t^{(S^c)} \mid (x_t^{(S)}, x_t^{(S^c)}), (w^{(T)}, w^{(T^c)}, (Y^{(S^c)})^{t-1}))$$

$$= p(y_t^{(S^c)} \mid (x_t^{(S)}, x_t^{(S^c)}))$$