Lec 18 : Converse to Coding theorem (Contd) 25·10·2017

$$\text{Source} \xrightarrow{W} \text{Enc} \xrightarrow{X^n \in \mathcal{X}^n} \text{channel} \xrightarrow{Y^n \in \mathcal{Y}^n} \text{Decoder} \to \hat{W} = g(Y^n)$$

DMC $\quad P(y_k \mid x^k\, y^{k-1}) = P(y_k \mid x_k)$

No feedback $\quad P(x_k \mid y^{k-1} x^{k-1}) = P(x_k \mid x^{k-1})$

Thm (Channel coding theorem)

Given a DMC with information capacity C and every rate $R < C$ there exists a sequence of $(2^{nR}, n)$ codes with max prob of error $\lambda^{(n)} \to 0$. Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

Proof: We proved achievability earlier - the converse remains.

$W \in \{1, 2, \dots M\}$, $M = 2^{nR}$, equally likely

$W \to X^n \to Y^n \to \hat{W}$

$$\begin{aligned}
nR = H(W) &= I(W;\hat{W}) + H(W \mid \hat{W}) \\
&\leq I(X^n; Y^n) + H(W \mid \hat{W}) \\
&\leq I(X^n; Y^n) + 1 + P_e \log(2^{nR}) \quad \text{[Fano's inequality]} \\
&= H(Y^n) - H(Y^n \mid X^n) + 1 + P_e(nR) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) + 1 + P_e(nR) \quad \text{DMC w/o feedback} \\
&= \sum_{i=1}^{n} I(X_i; Y_i) + 1 + P_e(nR) \\
&\leq nC + 1 + nR P_e^{(n)}
\end{aligned}$$

$E = 1$ if $W \neq \hat{W}$, 0 Else

$$H(W, E \mid \hat{W}) = H(W \mid \hat{W}) + \cancel{H(E \mid W, \hat{W})} = H(E \mid \hat{W}) + H(W \mid E, \hat{W})$$

$$\Rightarrow H(W \mid \hat{W}) \leq 1 + P_e^{(n)} \log 2^{nR} = 1 + P_e^{(n)}(nR)$$

$$\Rightarrow \boxed{R \leq C + \frac{1}{n} + R P_e^{(n)}} \to (1)$$

By making n large enough can make the last two terms on the right as small as possible. It follows that

$R \leq C$ must hold due to (2)

Note: $\lambda^{(n)} \to 0$, $\quad P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$, $\quad \lambda^{(n)} = \max_{i=1}^{M} \{\lambda_i\}$, $M = 2^{nR}$.

Note that (1) $\Rightarrow$

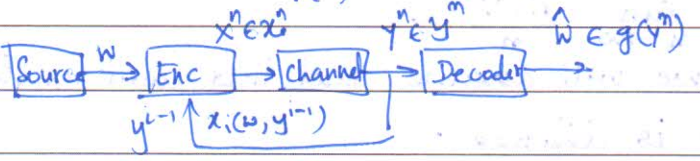
$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR} \to (2)$$

Hence $R > C \Rightarrow P_e^{(n)}$ will be bounded from below, But $P_e^{(n)} \to 0$.

$\therefore$ Any achievable rate is upper bounded by the information capacity $C_I = \max_{P(x)} I(X;Y)$. On the other hand we can get as close to C_I as possible. $C_{Op} =$ sup of all achievable rates. $\Rightarrow C_{Op} = C_I$

Thm (feedback capacity) DMC

$C_{FB} = C = \max_{p(x)} I(X;Y)$ where C_{FB} = capacity with feedback ie $(2^{nR}, n)$ code, codeword is a function of w, y^{i-1} ie, $x_i(w, y^{i-1})$



A code with feedback could of course disregard the feedback, hence we know $\exists$ a sequence of codes with rate $\to C$ from below. $\Rightarrow C_{FB} \geq C$. $\to$ (1)

To prove that any achievable rate $R \leq C$

$W \to X^n \to Y^n \to \hat{W}$

$$nR = H(W) = I(W;\hat{W}) + H(W|\hat{W})$$
$$\leq I(W;Y^n) + 1 + P_e \log 2^{nR}$$
$$= H(Y^n) - H(Y^n|W) + 1 + P_e(nR)$$
$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}W) + 1 + P_e(nR)$$
$$= \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}W X_i) + 1 + P_e(nR)$$

(as X_i is a function of W, Y^{i-1})

$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) + 1 + P_e(nR)$$

$(Y^{i-1}, W) \to X_i \xrightarrow{\text{DMC}} Y_i$ [due to DMC] $\Rightarrow$ given X_i, Y_i is independent of $Y^{i-1}W$.

$$= \sum_{i=1}^{n} I(X_i;Y_i) + 1 + P_e(nR)$$
$$\leq nC + 1 + P_e(nR).$$

$\Rightarrow P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$.

$\therefore$ If $R > C$.

then $P_e^{(n)}$ is lower bounded.

$\therefore R \leq C$. $\Rightarrow C_{FB} = \sup R \leq C$ $\to$ (2)

From (1), (2) $C_{FB} = C$. (19)

Thm (Source Channel Coding theorem) If $V_1, \ldots V_n$ is a finite alphabet stochastic process that satisfies the AEP and $H(V) < C$, there exists a source-channel code with $Pr(\hat{V}^n \neq \hat{V}^n) \to 0$. Conversely for any stationary process if $H(V) > C$, the prob of error is bounded away from zero and it is not possible to send the process over the channel with arbitrarily low probability of error.

Recall $\quad \mathcal{H}(\mathcal{V}) = \lim_{n\to\infty} \frac{H(V_1, V_2, \ldots V_n)}{n} = \lim_{n\to\infty} \frac{H(V^n)}{n}$.

(we will assume that this limit exists)

"process satisfies the AEP" $\Rightarrow \mathcal{H}(\mathcal{V})$ is defined.

$$A_\epsilon^{(n)} = \left\{ v^n \;\middle|\; \left|\mathcal{H}(\mathcal{V}) - \frac{1}{n}\log\frac{1}{P(v^n)}\right| < \epsilon \right\}.$$

$$\Pr(v^n \notin A_\epsilon^{(n)}) \le \epsilon, \qquad |A_\epsilon^{(n)}| \le 2^{n(\mathcal{H}(\mathcal{V})+\epsilon)}$$

$$V^n \to X^n \to Y^n \to \hat{V}^n$$

$$\boxed{\text{Source}} \xrightarrow{V^n} \boxed{\text{Enc}} \xrightarrow{X^n \in \mathcal{X}^n} \boxed{\text{Channel}} \xrightarrow{Y^n \in \mathcal{Y}^n} \boxed{\text{Decoder}} \xrightarrow{\hat{V}^n = g(Y^n)}$$

We next only code the typical sequences:

$$2^{n(\mathcal{H}(\mathcal{V})+\epsilon)} \longrightarrow W \in \{1, \ldots, 2^{n\mathcal{H}(\mathcal{V})+\epsilon}\}$$

and transmit this over the DMC.

Can transmit it with probability (Infact $\lambda^{(n)} \le \epsilon$) of error $P_e^{(n)} \le \epsilon$ provided that

$$\frac{n(\mathcal{H}(\mathcal{V})+\epsilon)}{n} < C.$$

i.e., $\mathcal{H}(\mathcal{V}) < C - \epsilon$

Converse:

$$\mathcal{H}(\mathcal{V}) \le \frac{H(V^n)}{n}$$

assuming stationarity this is a decreasing sequence (as proved in assignment)

$$= \frac{1}{n}\left\{ I(V^n; \hat{V}^n) + H(V^n \mid \hat{V}^n) \right\}$$

$$\le \frac{1}{n}\left\{ I(X^n; Y^n) + 1 + n\Pr(V^n \ne \hat{V}^n)\log|\mathcal{V}| \right\}$$

$$\le \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i) + \frac{1}{n} + \Pr(V^n \ne \hat{V}^n)\log|\mathcal{V}|$$

$$\le C + \frac{1}{n} + \Pr(V^n \ne \hat{V}^n)\log|\mathcal{V}|$$

Hence $\mathcal{H}(\mathcal{V}) \le C$

Probability (average) of error in transmitting the source V^n is $\le 2\epsilon$

- ϵ if the source is atypical
- ϵ when it is typical.

$$P(V^n \ne \hat{V}^n) \le P(V^n \notin A_\epsilon^{(n)}) + P(V^n \ne \hat{V}^n \mid V^n \in A_\epsilon^{n})$$

$$\le \epsilon + \epsilon = 2\epsilon.$$