Lec 17: Converse to Coding Theorem — 25.10.17

Recap

- $C = \max_{p(x)} I(X;Y)$

Examples

- DMC

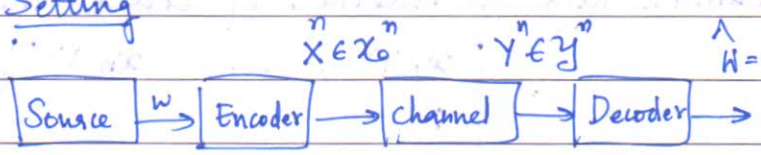
$P(y_k | x^k, y^{k-1}) = P(y_k | x_k)$

No feedback

$P(x_k | x^{k-1}, y^{k-1}) = p(x_k | x^{k-1})$

- Jointly Typical
  - $A_\epsilon^{(n)}$ → bounds on size
  - $Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)})$
  - $P_{\tilde{X}^n, \tilde{Y}^n}(x^n, y^n) = P_{X^n}(x^n) P_{Y^n}(y^n)$

Setting



Random codebook

$X_1(1)$	$X_2(1)$	...	$X_n(1)$
$X_1(2)$	$X_2(2)$		$X_n(2)$
$X_1(M)$	$X_2(M)$		$X_n(M)$

all are iid with distribution $P_X(x)$

λ_w = prob of error, when w-th message is transmitted

$\lambda^{(n)} = \max_w \lambda_w$, $P_e^{(n)} = \frac{1}{M} \sum_{w=1}^{M} \lambda_w$.

Decoding

$\hat{W} = w$ if $(X^n(w), Y^n) \in A_\epsilon^{(n)}$

$(X^n(w'), Y^n) \notin A_\epsilon^{(n)} \quad \forall w' \neq w$

$$Pr(\mathcal{E}) = \sum_{\mathcal{C}} Pr(\mathcal{C}) Pr(\mathcal{E}|\mathcal{C}) = \sum_{\mathcal{C}} Pr(\mathcal{C}) \frac{1}{M} \sum_{w=1}^{M} Pr(\mathcal{E}|\mathcal{C}, W=w)$$

$$= \frac{1}{M} \sum_{w=1}^{M} \sum_{\mathcal{C}} Pr(\mathcal{C}) \underbrace{Pr(\mathcal{E}|\mathcal{C}, W=w)}_{\lambda_w(\mathcal{C})} = \sum_{\mathcal{C}} Pr(\mathcal{C}) Pr(\mathcal{E}|\mathcal{C}, 1)$$

$$= Pr(\mathcal{E}|W=1)$$

$Pr(\mathcal{E}|W=1) = Pr(E_1^c \cup E_2 \cup \cdots E_M)$

where $E_i = Pr((X_L^n(i), Y^n) \in A_\epsilon^{(n)})$

$M = \lceil 2^{nR} \rceil$

$\therefore Pr(\mathcal{E}|W=1) \leq Pr(E_1^c) + \sum_{i=2}^{M} Pr(E_i) \leq \epsilon + 2^{-n(I(X;Y) - R - 3\epsilon)}$

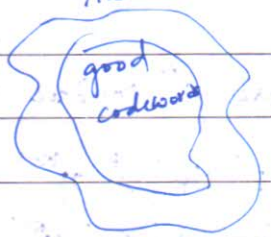
$\leq 2\epsilon$ (by choosing n large enough)

① In summary, for sufficiently large n, the probability of error $Pr(\mathcal{E})$ can be made $\leq 2\epsilon$. Since this is an average over all codebooks, there must exist a codebook $\mathcal{C}_0$ s.t.

$Pr(\mathcal{E}|\mathcal{C} = \mathcal{C}_0) \leq 2\epsilon$. Here the requirement is that $R < I(X;Y)$

② let $p(x)$ be the distribution achieving 'information capacity' C. Then it suffices that $R < C$.

Since the $P_e(\varepsilon)$ averaged over all M messages is $\leq 2\epsilon$, no more than half the codewords can have $\lambda_w \geq 4\epsilon$.



Throw away these.

Then we are left with a codebook with $\geq \frac{M}{2} = 2^{nR-1}$ codewords.

$\lambda^{(n)} \leq 4\epsilon$

note $\frac{\log 2^{nR-1}}{n} = R - \frac{1}{n}$; can be made arbitrarily close to R, hence to C.

Conclusion: Any rate $R < C$ is achievable.

R achievable $\Rightarrow$ sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ as $n \to \infty$

Converse: zero-error case

Consider the setting where $\lambda^{(n)} = 0 \Rightarrow \lambda_w = 0$ for all messages w. Let a rate R be given. Let W be a random variable uniformly distributed over $\{1, 2, \ldots nR\}$. Let the setting now be same as earlier

$$nR = H(W) = I(W; Y^n) + H(W|Y^n)$$
$$= I(W; Y^n) \quad [\text{as this is zero error case}]$$
$$\leq I(X^n; Y^n) \quad [\text{Data processing inequality } X^n = X^n(W)]. \hookrightarrow (i)$$

~~$\leq n(I(X;Y))$ [DMC] - (1)~~

$\therefore H(Y^n) - H(Y^n|X^n)$

$$H(Y^n|X^n) = \sum_{(x^n,y^n)} p(x^n y^n) \log \frac{1}{P(y^n|x^n)} = \sum_{(x^n,y^n)} P(x^n, y^n) \sum_{i=1}^{n} \log \frac{1}{P(y_i|x_i)} \quad \to \text{DMC}$$

$$= \sum_{i=1}^{n} \sum_{(x^n,y^n)} p(x^n, y^n) \log \frac{1}{P(y_i|x_i)} = \sum_{i=1}^{n} \sum_{(x_i,y_i)} P(x_i, y_i) \log \frac{1}{P(y_i|x_i)} = \sum_{i=1}^{n} H(Y_i|X_i)$$

$\Rightarrow (i) \leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n} I(X_i; Y_i) \leq nC$

Thus $R \leq C$ if zero error probability is desired.