Dec 15: Channel Coding Theorem (towards)

DATE 11.10.17

Recap

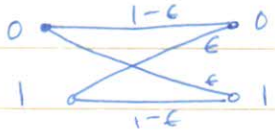
- Arithmetic coding
- Channel Capacity
  $C = \max_{P_X} I(X;Y)$
- Some simple examples

TODAY

- Further examples
  - BEC, BSC
  - Symmetric ch
- Discrete Memoryless channel set up
- Joint typicality
- Coding Theorem

Further examples of channels

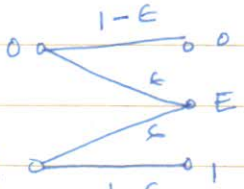
1. BSC



$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\epsilon, 1-\epsilon)$
$\leq 1 - H(\epsilon, 1-\epsilon)$

(By picking X uniform, Y is uniform) $C = 1 - H(\epsilon, 1-\epsilon)$ bits/channel use

2. BEC



$I(X,Y) = H(X) - H(X|Y) = H(X) - \epsilon H(X)$
$= (1-\epsilon) H(X) \leq (1-\epsilon)$

$C = 1 - \epsilon$

(Reed Muller Codes achieve capacity)

$\{ P_R(Y=E) H(X|Y=E)$
$+ P_R(Y=0) H(X|Y=0)$ (0)
$+ P_R(Y=1) H(X|Y=1) \}$ (0)

3. Symmetric channel

Defn: A channel is said to be symmetric if the rows of the channel transition matrix $[P(y|x)]$ are permutations of each other and likewise the columns are permutations of one another.

A channel is said to be weakly symmetric if all the rows are permutations of each other and the column sums are equal

Eg: $[p(y|x)] \Rightarrow$ (columns are y →, rows are x ↓)

$$\begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Symmetric

where $\underline{r}$ is row vector in $[p(y|x)]$

$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\underline{r})$
$\leq \log|\mathcal{Y}| - H(\underline{r})$ $\therefore C = \log|\mathcal{Y}| - H(\underline{r})$

By picking X uniform, Y is uniform

DATE

Eg: $[P(y|x)] = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{bmatrix}$ $\Rightarrow$ weakly symmetric channel

$\underline{s} \rightarrow$ row vector in $[P(y|x)]$

Similarly $I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\underline{s})$

$\leq \log|3| - H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$. WHY?

$C = \log|3| - H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$

achievable by picking X to be uniform.

$\Rightarrow Y$ is uniform. (as column sum is same)

$$P(Y=y) = \sum_x P(X=x)\,P(Y=y|X=x).$$

$$= \sum_x \frac{1}{|\mathcal{X}|} P(Y=y|X=x)$$

$$= \frac{C_y}{|\mathcal{X}|} = \frac{C_{y'}}{|\mathcal{X}|} = P(Y=y').$$

$$= \frac{1}{|\mathcal{Y}|}$$

$C_{y'}|\mathcal{Y}| = |\mathcal{X}|$. By summing up all elements in $[P(y|x)]$ columnwise & rowwise

Proposition: $C = \max_{P(x)} I(X;Y)$.

$\rightarrow C \geq 0$; $\rightarrow C \leq \log|\mathcal{X}|$, $\rightarrow C \leq \log|\mathcal{Y}|$

Setting of Channel Coding Theorem:

$W \in \{1, 2, \cdots M\}$ one of the M messages

$X^n(w) = X_1(w), X_2(w), \ldots X_M(w)$ (notation)

Channel: $P_{Y^n|X^n}(y^n|x^n)$ (characterization)

Discrete channel: $\mathcal{X}$, $\mathcal{Y}$ finite

Defn: The nth extension of the discrete memoryless channel (DMC) is the channel

$(\mathcal{X}^n, P(y^n|x^n), \mathcal{Y})$ where

$$P(y_k|x^k y^{k-1}) = P(y_k|x_k) \quad \text{--- (1)} \quad k = 1, 2, \cdots n$$

If the DMC is used without feedback

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x_i) \rightarrow (2)$$

DATE

unless otherwise specified, we will assume DMC w/o feedback ie, assume ② holds.

Defn: An (M, n) code for channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of a) an index set $\{1, 2, .. M\}$ ⓑ Encoding function $X^n: \{1, 2, .. M\} \longrightarrow \mathcal{X}^n$ yielding codewords $\{x^n(1), x^n(2), .. x^n(M)\}$. The set of codewords is called the codebook. ⓒ A decoding function: $g: \mathcal{Y}^n \longrightarrow \{1, 2, .. M\}$ which is a deterministic rule assigning a guess for each possible Y^n.

ERROR PROB (definition) Conditional probability of error

Let $\lambda_i = P_R(g(Y^n) \neq i \mid X^n = x^n(i))$

= Conditional probability of error given ith codeword was transmitted.

Defn: $\lambda^{(n)} = \max_{i \in [M]} \lambda_i \qquad [M] = \{1, 2, .. M\}$

(maximum probability of error for an (M, n) code)

Defn: $P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$ = {average probability of error.

(note: if all codewords are used with equal likelihood, then $P_e^{(n)}$ is the (overall) probability of error).

clearly $P_e^{(n)} \leq \lambda^{(n)}$

Rate of Code = $\frac{\log M}{n}$, A rate R is said to be achievable over the channel if $\exists$ a sequence of $(\lceil 2^{nR} \rceil, n)$ codes with $\lambda^{(n)} \to 0$ as $n \to \infty$

C = sup {achievable rate}

operational capacity