Lec 16 : Shannon's Channel Capacity

16.10.2017

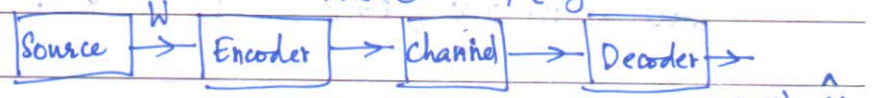
Recap

* C for BEC, BSC, Symmetric channel.
* Setting of Coding theorem

TODAY

- Joint Typicality
- Coding Theorem

Setting:



$W \in \{1, 2, \ldots M\}$

Rate $R = \frac{\log M}{n}$, Channel Capacity C.

DMC:

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

No feedback

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

Definition:

$A_\epsilon^{(n)}$ = set of jointly typical sequences

$$= \left\{ (x^n, y^n) \;\middle|\; \left|H(X) - \frac{1}{n}\log\frac{1}{p(x^n)}\right| < \epsilon, \left|H(Y) - \frac{1}{n}\log\frac{1}{p(y^n)}\right| < \epsilon, \left|H(X,Y) - \frac{1}{n}\log\frac{1}{p(x^n,y^n)}\right| < \epsilon \right\}$$

where $p(x^n, y^n) = \prod_{k=1}^{n} p(x_k, y_k)$, i.e., $\{(X^n, Y^n)\}$ is an iid source.

Consider

$$p(y^n|x^n) = \frac{p(x^n y^n)}{p(x^n)} = \frac{\prod_{k=1}^{n} p(x_k y_k | x^{k-1} y^{k-1})}{p(x^n)}$$

$$= \frac{1}{p(x^n)} \prod_{k=1}^{n} p(x_k | x^{k-1} y^{k-1}) \, p(y_k | x^k y^{k-1})$$

$$= \frac{1}{p(x^n)} \prod_{k=1}^{n} p(x_k | x^{k-1}) \, p(y_k | x_k)$$

$$= \frac{1}{p(x^n)} \, p(x^n) \prod_{k=1}^{n} p(y_k | x_k) = \prod_{k=1}^{n} p(y_k | x_k).$$

Properties of $A_\epsilon^{(n)}$

1) $\Pr\left((X^n, Y^n) \in A_\epsilon^{(n)}\right) \to 1$ as $n \to \infty$

Proof: By the WLLN

$$\frac{1}{n}\log\frac{1}{p(X^n)} \longrightarrow H(X) \quad \text{(in probability)}$$

Hence given $\epsilon > 0$, $\exists\, n_1$ s.t.

$$\Pr\left\{\left|H(X) - \frac{1}{n}\log\frac{1}{p(X^n)}\right| \geq \epsilon\right\} < \epsilon/3 \quad \forall\, n \geq n_1 \quad \to (1)$$

Similarly for $n \geq n_2$

$$\Pr\left\{\left|H(Y) - \frac{1}{n}\log\frac{1}{p(Y^n)}\right| \geq \epsilon\right\} < \epsilon/3 \quad \forall\, n \geq n_2 \quad \to (2)$$

and for $n \geq n_3$

$$\Pr\left\{\left|H(X,Y) - \frac{1}{n}\log\frac{1}{p(X^n,Y^n)}\right| \geq \epsilon\right\} \leq \epsilon/3 \quad \to (3)$$

For $n \geq \max(n_1, n_2, n_3)$ (1), (2), (3) hold.

$$P((x^n, y^n) \notin A_\epsilon^{(n)}) \leq \epsilon/3 + \epsilon/3 + \epsilon/3,$$

It follows by union bound that $P((x^n, y^n) \in A_\epsilon^{(n)}) \geq 1 - \epsilon$

Thus as $n \to \infty$, $P((x^n, y^n) \in A_\epsilon^{(n)}) \to 1$

Bounds on size of $A_\epsilon^{(n)}$

Upper bound

$$1 \geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p(x^n y^n) \geq |A_\epsilon^{(n)}| \, 2^{-n(H(X,Y)+\epsilon)}$$

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

Lower bound

For sufficiently large n.

$$P((x^n, y^n) \in A_\epsilon^{(n)}) \geq 1 - \epsilon$$

$$\downarrow = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P(x^n, y^n) \leq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} 2^{-n(H(X,Y)-\epsilon)}$$

$$= |A_\epsilon^{(n)}| \, 2^{-n(H(X,Y)-\epsilon)}$$

$$\Rightarrow |A_\epsilon^{(n)}| \geq (1-\epsilon) \, 2^{n(H(X,Y)-\epsilon)}$$

Next introduce $\tilde{X}^n$, $\tilde{Y}^n$ s.t. $P_{\tilde{X}^n \tilde{Y}^n}(x^n, y^n) = P_{X^n}(x^n) \, P_{Y^n}(y^n)$

$\Rightarrow P_{\tilde{X}^n}(x^n) = P_{X^n}(x^n)$ and $P_{\tilde{Y}^n}(y^n) = P_{Y^n}(y^n)$

Qn: What can we say about $P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)})$?

$$\sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{\tilde{X}^n, \tilde{Y}^n}(x^n, y^n) = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n}(x^n) \, P_{Y^n}(y^n)$$

$$\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} \, 2^{-n(H(Y)+\epsilon)} = |A_\epsilon^{(n)}| \, 2^{-n(H(X)+H(Y)+2\epsilon)}$$

$$\geq (1-\epsilon) \, 2^{n(H(X,Y)-\epsilon) - n(H(X)+H(Y)+2\epsilon)} \quad \text{[for sufficiently large } n\text{]}$$

$$= (1-\epsilon) \, 2^{n H(X,Y) - H(X) - H(Y) - 3\epsilon} = (1-\epsilon) \, 2^{-n(I(X,Y)+3\epsilon)}$$

$$\sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{(\tilde{X}^n, \tilde{Y}^n)}(x^n, y^n) = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n}(x^n) \, P_{Y^n}(y^n) \leq |A_\epsilon^{(n)}| \, 2^{-n(H(X)-\epsilon)} \, 2^{-n(H(Y)-\epsilon)}$$

$$\leq 2^{n(H(X,Y)+\epsilon)} \, 2^{-n(H(X)-\epsilon)} \, 2^{-n(H(Y)-\epsilon)}$$

$$\therefore = 2^{-n(I(X,Y)-3\epsilon)}$$

for sufficiently large n. $\leftarrow$

$$\therefore (1-\epsilon) \, 2^{-n(I(X,Y)+3\epsilon)} \leq P((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X,Y)-3\epsilon)}$$

The Shannon random Coding argument

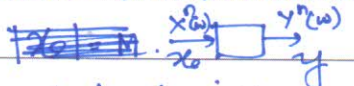
Shannon uses an ensemble of codes. The codebook

$$\begin{bmatrix} x_1(1), & \dots & x_n(1) \\ \vdots & & \\ x_1(w), & \dots & x_n(w) \\ \vdots & & \\ x_1(M), & \dots & x_n(M) \end{bmatrix}$$

(All codewords are equally likely to be transmitted)

Given rate R and length n, set $M = \lceil 2^{nR} \rceil$

Let $p(x)$ be some distribution on $\mathcal{X}$.



Choose this codebook with probability

$$\prod_{w=1}^{M} \prod_{i=1}^{n} p(x_i(w))$$

Show that this ensemble contains a code for any $R < C$ with $Pr(\mathcal{E}) \to 0$.

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

$$= \sum_{\mathcal{C}} Pr(\mathcal{C}) \frac{1}{M} \sum_{w=1}^{M} \underbrace{Pr(\mathcal{E}|\mathcal{C}, W=w)}_{\lambda_w(\mathcal{C})} \quad \text{[all codewords are equally likely]}$$

$$= \frac{1}{M} \sum_{w=1}^{M} \sum_{\mathcal{C}} Pr(\mathcal{C})\, \lambda_w(\mathcal{C}) = \frac{1}{M} \sum_{w=1}^{M} \sum_{\mathcal{C}} Pr(\mathcal{C})\, \lambda_1(\mathcal{C})$$

(this is independant of w!!)

$$= \sum_{\mathcal{C}} Pr(\mathcal{C})\, \lambda_1(\mathcal{C}) = Pr(\mathcal{E}|W=1)$$

Thus we are in a setting where the 1st codeword is transmitted and we want the average probability of incorrect decoding of this codeword averaged over all codebooks.

Define $E_i = \{(x^n(i), y^n) \in A_\epsilon^{(n)}\}$

Then $Pr(\mathcal{E}|w=1) = Pr(E_1^c \cup E_2 \cup E_3 \dots \cup E_M)$

$Pr(E_1^c) = Pr((x^n(1), y^n) \notin A_\epsilon^{(n)}) < \epsilon$

$Pr(E_w) = Pr((\tilde{x}^n, \tilde{y}^n) \in A_\epsilon^n) \le 2^{-n(I(X;Y)-2\epsilon)}$

$2 \le w \le M$

$$\therefore Pr(\mathcal{E}) \le Pr(E_1^c) + (M-1)\, Pr(E_w)$$
$$< \epsilon + (M-1)\, 2^{-n(I(X;Y)-3\epsilon)}$$
$$< \epsilon + 2^{nR}\, 2^{-n(I(X;Y)-3\epsilon)}$$
$$= \epsilon + 2^{-n(I(X;Y) - R - 3\epsilon)}$$

Decoding rule:

Given y^n look for $\hat{w}$ st $(x^n(\hat{w}), y^n) \in A_\epsilon^{(n)}$. declare a decoding failure if there is no $\hat{w}$ or if there is more than one such $\hat{w}$. Decoding error: decoding failure or $\hat{w} \neq w$.

It follows that if $I(X;Y) - R - 3\epsilon > 0$ can make error small. Hence if $R < I(X;Y)$ can make $Pr(\mathcal{E}) \to 0$ by suitably selecting ϵ, n.