Information Theory with Applications, Math6397
Lecture Notes from October 7th, 2014

taken by Nandini Deka

Last Time

- AEP for Discrete memoryless channels.
- Proof for Shannon's channel coding theorem.

3.1.14 Channel Coding Theorem (proof continued)

Remarks from last time

- The strings for the channel code are generated randomly according to $\mathbb{P}_{X^{\otimes n}}$, with X such that $I(X;Y)$ is maximal. Despite this independence among the symbols for one codeword *when the codebook is generated*, the distribution of of symbols in the transmitted codewords will not be i.i.d. This is because the transmitted codewords are uniformly distributed on a *sample* of size m_n drawn from the i.i.d. distribution, not on all of $\mathbb{A}^n$.
- The decoder was chosen such that if source alphabet is $C_n = \{1, 2, ..., m_n\}$, the decoder maps y to :

$$\psi_n(y) = \begin{cases} j : if \ \ \phi(j) = c_j, (c_j, y) \in F_\delta^n \ \ and \ \ no \ \ (c_{j'}, y) \in F_\delta^n \ \ for \ \ any \ \ j^{'} \neq j \\ 1 : (arbitrary) \ \ else \end{cases}$$

Source of errors in encoding/decoding

1. From not being typical input-output behavior for the channel.
2. From being not one-to-one for typical set, i.e $\exists m^{'} (\neq m)$ such that $(c_{m'}, y) \in F_\delta^n$.

We had, the expected error probability for code sequence, C_m, randomly generated as:

$$E_{X^{\otimes n}}[\Lambda(c_m)] \leq \mathbb{P}_{(X,Y)^{\otimes n}}\Big((F_\delta^n)^c\Big) + \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)$$

Remark: Since the $c_{m'}$ are generated randomly, the value on the right hand side is still a random variable!

Step 3

Bound the expected error probability (removing any randomness in the above expression):

$$E[P_e] = \frac{1}{m_n}\sum_{m=1}^{m_n} E[\Lambda(c_m)]$$

where the terms inside the summation are expectations wrt $c_1, c_2, ..., c_{m_n}$.

Using the result in Step 2, we get:

$$E[P_e] \leq \mathbb{P}_{(X,Y)^{\otimes n}}\Big((F_\delta^n)^c\Big) + \frac{1}{m_n}\sum_{m=1}^{m_n}\sum_{m\neq m'} E\Big[\sum_{\substack{y\in\mathbb{B}^n \\ (c_{m'},y)\in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)\Big]$$

Observe that the distribution of each $c_{m'}$, which was generated randomly (independently) is the same as that of c_m.

Thus averaging wrt $c_{m'}$, we get:

$$E[P_e] \leq \mathbb{P}_{(X,Y)^{\otimes n}}\Big((F_\delta^n)^c\Big) + \frac{1}{m_n}\sum_{m=1}^{m_n}\sum_{m'\neq m} \underline{\sum_{c_m\in\mathbb{A}^n} \mathbb{P}_{X^{\otimes n}}(c_m) \sum_{\substack{y\in\mathbb{B}^n \\ (c_{m'},y)\in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)}$$

(the underlined portion gives the sum over all typical pairs)

Now averaging over F_δ^n ,which has size bounds, in the second term we get:

$$E[P_e] \leq \mathbb{P}_{(X,Y)^{\otimes n}}\Big((F_\delta^n)^c\Big) + \frac{1}{m_n}\sum_{m=1}^{m_n}\sum_{m'\neq m} \sum_{\substack{(a,y)\in\mathbb{A}^n\times\mathbb{B}^n \\ (a,y)\in F_\delta^n}} \mathbb{P}_{X^{\otimes n}}(a)\mathbb{P}_{Y^{\otimes n}}(y)$$

Assuming 'n' sufficiently large, set typical sequences have size

$$|F_\delta^n| \leq exp(n(H(X_1, Y_1) + \delta)$$

and the probabilities being summed above are bounded by:

$$\substack{\mathbb{P}_{X^{\otimes n}}(a)\leq exp(-n(H(X_1)-\delta)) \\ \mathbb{P}_{Y^{\otimes n}}(y)\leq exp(-n(H(Y_1)-\delta))}$$

We conclude,

$$E\Big[\frac{1}{m_n}\sum_{m=1}^{m_n}\sum_{m'\neq m}\sum_{\substack{y\in\mathbb{B}^n \\ (c_{m'},y)\in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)\Big] \leq \underline{\frac{1}{m_n}\sum_{m=1}^{m_n}}\sum_{m'\neq m} |F_\delta^n| exp(-n(H(X_1)-\delta))exp(-n(H(Y_1)-\delta))$$

(underlined portion averages out to give 1)

$$\leq (m_n - 1)exp(n(H(X_1, Y_1) + \delta)exp(-n(H(X_1) - \delta)exp(-n(H(Y_1) - \delta))$$

For the next step, we recall:

- We have: $C - \tau < \frac{1}{n} ln(m_n) \leq C - \tau/2 = C - 4\delta$, with $\delta = \tau/8$
  or : $m_n \leq exp(n(C - 4\delta))$

- $H(X_1) + H(Y_1) - H(X_1, Y_1) = I(X_1, Y_1)$ and the random variables X_1 were chosen so that $C = I(X_1; Y_1)$.

Inserting this, we get

$$E\Big[\frac{1}{m_n} \sum_{m=1}^{m_n} \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)\Big] \leq exp(n(C - 4\delta)).exp(-n(I(X_1; Y_1) - 3\delta)$$

$$= exp(-n\delta)$$

Thus for sufficiently large n :

- $E\Big[\frac{1}{m_n} \sum_{m=1}^{m_n} \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)\Big] < \delta$

- $\mathbb{P}_{(X,Y)^{\otimes n}}\Big((F_\delta^n)^c\Big) < \delta$

Collecting terms, we have:

$$E[P_e] \leq \mathbb{P}_{(X,Y)^{\otimes n}}\Big((F_\delta^n)^c\Big) + exp(-n\delta)$$
$$\Rightarrow E[P_e] \leq 2\delta = \tau/4 < \epsilon$$

[**Note:** this is for a randomly chosen channel code]

We have thus found that a random choice of channel code gives an expected error probability $E[P_e] < \epsilon$. So, among all possible (random) choices for codewords, there exists one choice for which

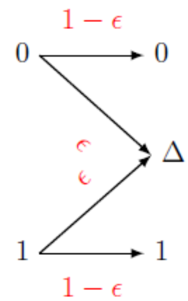
$$P_e(\psi_n \circ \gamma \circ \phi_n \neq id) < \epsilon \,.$$

This concludes the proof. □

Remark: While source coding, we had a rate of code symbols used for source symbols, here we have a **transmission rate**, number of source symbols "carried" by one channel input symbol.

3.1.15 Example

- Binary erasure channel (BEC)



To compute capacity of the BEC, we have to define the mutual information of the conditional probability measure:

Given input symbol $x \in \mathbb{A}$, let:

$$I(X = x; Y) = \sum_{y \in Y} \mathbb{W}(y|x) ln \frac{\mathbb{W}(y|x)}{\mathbb{P}_Y(y)}$$

Which means:

$$\begin{aligned} I(X;Y) &= \sum_{x,y} \mathbb{P}_{X,Y}(x,y) ln \frac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_X(x)\mathbb{P}_Y(x)} \\ &= \sum_{x \in X} \mathbb{P}_X(x) \sum_{y \in Y} \mathbb{W}(y|x) ln \frac{\mathbb{W}(y|x)}{\mathbb{P}_Y(y)} \\ &= \sum_{x \in X} \mathbb{P}_X(x) I(X = x; Y) \end{aligned}$$

Proposition: $\mathbb{P}_X$ achieves the capacity if and only if:

$$I(X = x; Y) = \begin{cases} C & : \mathbb{P}_X(x) > 0 \\ \leq C & : \mathbb{P}_X(x) = 0 \end{cases}$$

Proof: comes from the definition of capacity of a channel :

$$C = \max_X I(X;Y)$$

□

For BEC we need $\mathbb{P}_X > 0$ and $\mathbb{P}_Y > 0$ to achieve 'C', so:

$$\begin{aligned} C &= \max_Y I(X = 0; Y) = \max_Y I(X = 1; Y) = \max_{X,Y} I(X;Y) \\ &= \max_{X,Y} (H(Y) - H(Y|X)) \end{aligned}$$