EE7330: Network Information Theory **2021**

Lecture Notes 10: CONVERSE OF CHANNEL CODING THEOREM

Instructor: Shashank Vatedka *Scribe: EE20RESCH14005*

10.1 GIST OF LECTURE 9

In the last lecture, we have seen the Fano's Inequality given as below.

10.1.1 Fano's Inequality

If M and $\hat{M}$ are jointly distributed and $P_e = Pr[M \neq \hat{M}]$, then

$$\mathrm{H}(\mathrm{M}|\hat{M}) \leqslant H_2(P_e) + P_e log_2|\mathcal{M}| \tag{10.1}$$

The Fano's Inequality provides a lower bound on the entropy of the modulo-2 sum of two binary random vectors.

10.1.2 Channel Coding Theorem

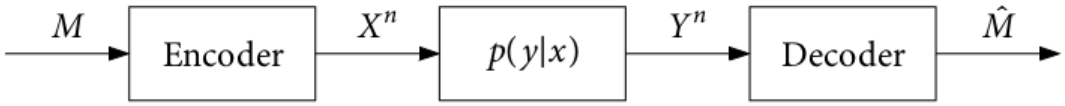

Figure 10.1: Point - to - Point Communication System

For a discrete memory-less channel (DMC), all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a set of (ENC_n, DEC_n) with maximum probability of error $P_e \to 0$.

$$C = \underset{\mathbf{P_x}}{\mathrm{Max}}\, I(X;Y) \tag{10.2}$$

10.1.3 Converse of Channel Coding Theorem

Conversely, any set of (ENC_n, DEC_n) with $P_e \to 0$ must have $R \leqslant C$.

In other way,

Consider any sequence (ENC_n, DEC_n) such that

$$\begin{aligned} \liminf_{n\to\infty} \quad & \frac{k_n}{n} \geqslant C + \epsilon \\ & \text{Then,} \\ \liminf_{n\to\infty} \quad & P_e = Pr[M \neq \hat{M}] \geqslant \epsilon \end{aligned} \tag{10.3}$$

Tools Required

1. Fano's Inequality
2. For any P_{X^n}, if Y^n is obtained by passing X^n through a DMC with channel probability transition matrix $P_{Y|X}$,

$$I(X^n; Y^n) \leqslant \Sigma_{i=1}^{n} \quad I(X_i; Y_i) \leqslant nC \tag{10.4}$$

   This is true for a DMC. It provides an upper bound and is dependent on

$$P(Y^n|X^n) = \pi_{i=1}^{n} P(Y_i|X_i) \tag{10.5}$$

10.2 Proof of the Converse of Channel Coding Theorem

10.2.1 Proof of 2^{nd} Tool

For a DMC,

$$I(X^n; Y^n) \leqslant \Sigma_{i=1}^{n} \quad I(X_i; Y_i) \leqslant nC \tag{10.6}$$

1. Proof of

$$\begin{aligned} & H(Y_i|Y_1.....Y_{i-1}, X^n) \leqslant H(Y_i|X_i)(\text{in general}) \\ & \text{else} \\ & H(Y_i|Y_1.....Y_{i-1}, X^n) = H(Y_i|X_i)(\text{only for DMC}) \\ & \text{using} \\ & P_{(Y^n|X^n)} = \pi_{i=1}^{n} P_{(Y_i|X_i)} \end{aligned} \tag{10.7}$$

   Given

$$H(Y_i|Y_1.....Y_{i-1}, X^n) = H(Y_i|X_i)(\text{only for DMC}) \tag{10.8}$$

We know that, for a DMC

$$\begin{aligned} P(y_i|x_1.....x_n) &= P(y_i|x_i) \\ \because P(y_1....y_i|x_1.....x_n) &= P(y_1|x_1)P(y_2|x_2)....P(y_i|x_i) \end{aligned} \tag{10.9}$$

Using the above property,

$$\begin{aligned} H(Y_i|Y_1.....Y_{i-1}, X^n) \\ = \Sigma_{X^n,y_1,....y_{i-1}} P(y_i, y_1.....y_{i-1}, X^n) log_2 \frac{1}{P(y_i|y_1.....y_{i-1}, X^n)} \\ = \Sigma_{X^n,y_1,....y_{i-1}} P(y_i, x_i, y_1....y_{i-1}, x_1....x_{i-1}, x_{i+1}....x_n) log_2 \frac{1}{P(y_i|y_1.....y_{i-1}, X^n)} \\ \because P(Y^n|X^n) = \pi_{i=1}^n P(Y_i|X_i) \\ = \Sigma_{X^n,Y_1,....Y_{i-1}} P(Y_i, X_i, Y_1....Y_{i-1}, X_1....X_{i-1}, X_{i+1}....X_n) log_2 \frac{1}{P(Y_i|X_i)} \\ = \Sigma_{x_i,y_i} P(y_i x_i) log_2 \frac{1}{P(y_i|x_i)} = H(Y_i|X_i) \end{aligned} \tag{10.10}$$

Here if the channel is any other such as Fading channel or memory channel then the above condition is not true since Y_i is dependent on other Y_i.

2. Using the Relation between Mutual Information, Entropy and Conditional Entropy,

$$I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n)$$

$$\begin{aligned} \text{Using Chain rule of Conditional Entropy} \\ \Rightarrow I(X^n; Y^n) = \Sigma_{i=1}^n [H(Y_i|Y_1.....Y_{i-1}) - H(Y_i|Y_1.....Y_{i-1}, X^n)] \\ \text{On conditioning} \\ \Rightarrow I(X^n; Y^n) \leqslant \Sigma_{i=1}^n [H(Y_i) - H(Y_i|Y_1.....Y_{i-1}, X^n)] \\ \Rightarrow I(X^n; Y^n) = \Sigma_{i=1}^n [H(Y_i) - H(Y_i|X_i)] \text{(from above)} \\ \Rightarrow I(X^n; Y^n) = \Sigma_{i=1}^n I(X_i; Y_i) \end{aligned} \tag{10.11}$$

From Channel coding Theorem

$$C = \underset{\mathbf{P_X}}{\text{Max}}\, I(X; Y) \tag{10.12}$$

and $\because$ n distributions

$$\Rightarrow I(X^n; Y^n) = \Sigma_{i=1}^n I(X_i; Y_i) \leqslant nC \tag{10.13}$$

3. Proof of Converse
Assuming k_n input message bits which are iid unif $\approx$ [0,1]. The entropy of k_n is given by $H(M^{k_n})$. Use Fano's Inequality for obtaining a lower bound for P_e. For a good code, P_e will be vanishing and for a bad code P_e may equal 1.

$$\begin{aligned} k_n = H(M^{k_n}) &= H(M^{k_n}|\hat{M}^{k_n}) + I(M^{k_n}; \hat{M}^{k_n}) \\ &\leqslant H_2(P_e) + P_e log_2|\mathcal{M}| + I(M^{k_n}; \hat{M}^{k_n}) \\ &\leqslant H_2(P_e) + P_e k_n + I(M^{k_n}; \hat{M}^{k_n}) \end{aligned} \tag{10.14}$$

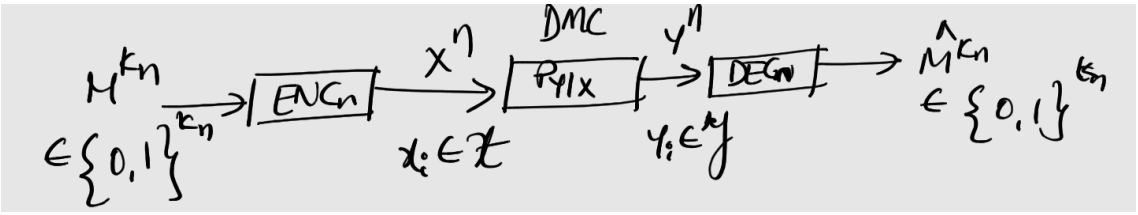

Figure 10.2: Point - to - Point Communication System

By using data processing inequality and tool 2, $I(M^{k_n}; \hat{M}^{k_n})$ may be written as

$$\begin{aligned} I(M^{k_n}; \hat{M}^{k_n}) &\leqslant nC \\ \because I(M^{k_n}; \hat{M}^{k_n}) &\leqslant I(X^n;^n) \text{ from block diagram above} \\ &\Rightarrow k_n \leqslant H_2(P_e) + P_e k_n + nC \\ &\Rightarrow k_n - nC - P_e k_n \leqslant H_2(P_e) \end{aligned} \tag{10.15}$$

Now divide on both sides by n,

$$\begin{aligned} \Rightarrow \frac{k_n - nC - P_e k_n}{n} &\leqslant \frac{H_2(P_e)}{n} \\ \Rightarrow \frac{k_n(1-P_e)}{n} - C &\leqslant \frac{H_2(P_e)}{n} \end{aligned} \tag{10.16}$$

Now apply limits on both sides.
Suppose, in any DMC, operation is running at rate less than Capacity, then

$$\begin{aligned} \Rightarrow \limsup_{n\to\infty}\left(\frac{k_n(1-P_e)}{n} - C - \frac{H_2(P_e)}{n}\right) &\leqslant 0 \\ \text{Assume } R = \lim_{n\to\infty} \tfrac{k_n}{n} & \\ \text{and} \limsup_{n\to\infty} \tfrac{H_2(P_e)}{n} &= 0 \\ \Rightarrow R(1 - \limsup_{n\to\infty} P_e) - C - 0) &\leqslant 0 \\ \Rightarrow \limsup_{n\to\infty} P_e \geqslant \frac{R-C}{R} &\leqslant 0 \end{aligned} \tag{10.17}$$

Inference This gives trivial lower bound for P_e which is always negative and is not possible. Assuming the rate of operation is greater than capacity,i.e if $R > C$

$$\text{Assume } R - C = \epsilon \text{and} \Rightarrow \limsup_{n\to\infty} P_e \geqslant \tfrac{\epsilon}{R} \tag{10.18}$$

Inference If $R > C$, the probability of error is bounded away from 0 for sufficiently large n. This proves that there exists an error and always bounded by ϵ. This converse is sometimes called *weakconverse* to the channel coding theorem. It is also possible to prove a *strongconverse*, which states that for rates above capacity, the probability of error goes exponentially to 1. The sudden error transition is called Phase transition phenomenon as shown in figure.
This sharp transition can be proved and will be will be decreasing at $2^{-\alpha n}$ for $R < C$.

10.2.2 Next Class

The next class will cover about Mrs. Greber's Lemma.

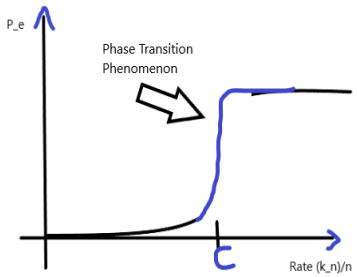

Figure 10.3: P_e vs Rate Curve