

## Lecture Notes 11: Tools for proving converse results

Instructor: Shashank Vatedka

Scribe: Ritesh Kumar

**Disclaimer:** These notes have not been subjected to the usual scrutiny reserved for formal publications. Please email the course instructor in case of any errors.

## 11.1 Fano's inequality

Consider  $M$  and  $\hat{M}$  are jointly distributed and probability of error  $P_e = \Pr[M \neq \hat{M}]$ . And  $M$  &  $\hat{M} \in \mathbb{M}$ . Then

$$H(M/\hat{M}) \leq H_2(P_e) + P_e \log |\mathbb{M}| \quad (11.1)$$

Proof :

Consider a indicator function  $E$  as :

$$E = \begin{cases} 1 & \text{if } \hat{M} \neq M \\ 0 & \text{if } \hat{M} = M \end{cases} \quad (11.2)$$

So, we have  $P_E(1) = P_e$ ,  $P_E(0) = 1 - P_e$ , and we can write  $H(E) = H_2(P_e)$ . Now

$$H(M, E|\hat{M}) = H(M|\hat{M}) + H(E|M, \hat{M}) \quad (\text{chain rule of entropy}) \quad (11.3)$$

$$H(M|\hat{M}) = H(E|\hat{M}) + H(M|\hat{M}, E) - H(E|M\hat{M}) \quad (11.4)$$

Since  $E$  is independent from  $M$  and  $\hat{M}$ , we can write,

$$H(M|\hat{M}) = H(E|\hat{M}) + H\left(\frac{M}{E}, \hat{M}\right) \quad (11.5)$$

$$\leq H(E) + H(M|E, \hat{M}) \quad (11.6)$$

$$= H_2(P_e) + H(M|\hat{M}, E=0) P_E(0) + H(M|\hat{M}, E=1) P_E(1) \quad (11.7)$$

Here,  $H(M|\hat{M}, E=0) = 0$  because for  $E=0$ ,  $M = \hat{M}$ , and hence entropy = 0.

$$= H_2(P_e) + H(M|\hat{M}, E=1) P_E(1) \quad (11.8)$$

$$= H_2(P_e) + H(M|\hat{M}, E=1) P_e \quad (11.9)$$

$$\leq H_2(P_e) + H(M) P_e \quad (11.10)$$

$$\leq H_2(P_e) + P_e \log_2 |\mathbb{M}| \quad \text{Proved.} \quad (11.11)$$

## 11.2 Proof of converse channel coding theorem

**Theorem :** Consider any sequence with  $(ENC_n, DEC_n)$  for DMC with transition probability  $P_{Y|X}$  such that ,

$$\liminf_{n \rightarrow \infty} \frac{K_n}{n} \geq C + \epsilon$$

then,

$$P_e = \limsup_{n \rightarrow \infty} Pr \left[ \hat{M}_i \neq M_i \right] \geq \frac{\epsilon}{R}$$

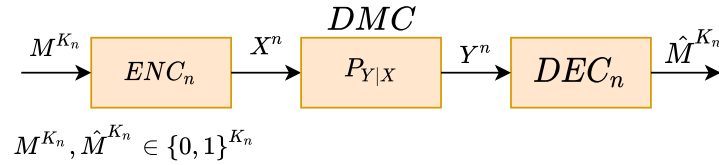


Figure 11.1: Single user over DMC

**Proof:**

We need some well-known inequalities to prove this. 1) Fano's inequality and 2) bound on mutual information. Fano's inequality we have discussed just above and now let us discuss bound on mutual information.

**Lemma :** For any  $P_{X^n}$ , if  $Y^n$  is obtained by passing  $X^n$  through Discrete memoryless channel having transition probability  $P_{Y|X}$  then,

$$I(X^n; Y^n) \leq \sum_{i=1}^n I(X_i; Y_i) \leq nC \quad (11.12)$$

**Proof of lemma :**

Consider the expression of mutual information,

$$I(X^n; Y^n) = H(Y^n) - H(Y^n | X^n) \quad (11.13)$$

$$= \sum_{i=1}^n [H(Y_i | Y_1, Y_2 \dots Y_{i-1}) - H(Y_i | Y_1, Y_2 \dots Y_{i-1}, X^n)] \quad (11.14)$$

(Chain rule of entropy)

$$\leq \sum_{i=1}^n [H(Y_i) - H(Y_i | Y_1, Y_2 \dots Y_{i-1}, X^n)] \quad (11.15)$$

(Since conditioning decreases entropy)

Now,

$$H(Y_i | Y_1, Y_2 \dots Y_{i-1}, X^n) = \sum_{x^n, y_1, \dots, y_n} P_{Y|X}(y_i | y_1, \dots, y_{i-1}, x^n) \times \log \left( \frac{1}{P_{Y|X}(y_i | y_1, \dots, y_{i-1}, x^n)} \right) \quad (11.16)$$

Since for discrete memoryless channel present output depends only on the present input, hence we can write.

$$P_{Y|X}(y_i|y_1, \dots, y_{i-1}, x^n) = P_{Y|X}(y_i|x_i, y_1, x_i, \dots, y_{i-1}, x_{i-1}) = P_{Y|X}(y_i|x_i) \quad (11.17)$$

Using eq (11.17) in eq (11.16), we can write,

$$H(Y_i|Y_1, Y_2, \dots, Y_{i-1}, X^n) = \sum_{x^n, y_1, \dots, y_n} P_{Y|X}(y_i|x_i) \times \log_2 \left( \frac{1}{P_{Y|X}(y_i|x_i)} \right) = H(Y_i|X_i) \quad (11.18)$$

By combining all i.e using eq(11.18) in eq11.15 we can write,

$$= \sum_{i=1}^n [H(Y_i) - H(Y_i|X_i)] \quad (11.19)$$

And we have,

$$\sum_{i=1}^n [H(Y_i) - H(Y_i|X_i)] = \sum_{i=1}^n I(X_i; Y_i) \quad (11.20)$$

From eq(11.13), eq(11.15) and eq(11.20) we have,

$$I(X^n; Y^n) \leq \sum_{i=1}^n I(X_i; Y_i) \leq nC \quad (11.21)$$

This conclude the proof of eq (11.12).

### Proof of converse:

We have message, which is i.i.d and uniform so we can write,

$$K_n = H(M^{K_n}) \quad (11.22)$$

$$= H(M^{K_n}|\hat{M}^{K_n}) + I(M^{K_n}; \hat{M}^{K_n}) \quad (11.23)$$

$$\leq H_2(P_e) + P_e \log_2(2^{M^{K_n}}) + I(M^{K_n}; M'^{K_n}) \quad (11.24)$$

( using Fano's inequality)

$$\leq H_2(P_e) + P_e K_n + I(M^{K_n}; \hat{M}^{K_n}) \quad (11.25)$$

$$\leq H_2(P_e) + P_e K_n + I(X^n; Y^n) \quad (11.26)$$

( using data processing inequality)

$$\leq H_2(P_e) + P_e K_n + nC \quad (\text{From eq 11.21}) \quad (11.27)$$

$$\begin{aligned} K_n \leq H_2(P_e) + P_e K_n + nC &\Rightarrow \frac{H_2(P_e)}{n} \geq \frac{K_n(1 - P_e)}{n} - C \\ \lim_{n \rightarrow \infty} \left( \frac{K_n(1 - P_e)}{n} - C \right) &\leq \lim_{n \rightarrow \infty} \frac{H_2(P_e)}{n} \end{aligned}$$

Where,  $P_e = \limsup_{n \rightarrow \infty} Pr[\hat{M}_i^{K_n} \neq M_i^{K_n}]$

Let  $\frac{K_n}{n} = R$ ,

$$\lim_{n \rightarrow \infty} P_e \geq \frac{R - C}{R} \quad (11.28)$$

$$\limsup_{n \rightarrow \infty} P_e \geq \frac{R - C}{R} \quad (11.29)$$

If we operate at the rate more than capacity of the channel say  $R = C + \epsilon$

$$\limsup_{n \rightarrow \infty} P_e \geq \frac{\epsilon}{R} \quad (11.30)$$

That is probability of error is always non-zero. But if we operate the channel below the capacity and for

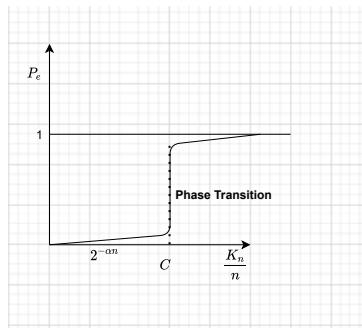


Figure 11.2: Probability of error with capacity constraint

large  $n$ , in (11.29) we can observe  $P_e$  approaches to 0.

### 11.3 Mrs Gerber's Lemma:

Suppose we have  $X = (X_1, \dots, X_n)$  where  $X_i \in \mathbb{M}^n \in \{0, 1\}$  be a binary random  $n$ -vector. Let  $P_X(x) = P(X = x)$ , where  $x \in \mathbb{M}^n$  define its probability distribution. Let say  $X \sim \text{Ber}(q)$ . Let us consider that the random vector  $X$  is the input to a binary symmetric channel with crossover probability  $p$ , where  $0 < p < \frac{1}{2}$ . Let be  $Y = (Y_1, \dots, Y_n)$  where  $Y_i \in \mathbb{M}^n \in \{0, 1\}$  the corresponding channel output  $n$ -vector. The probability

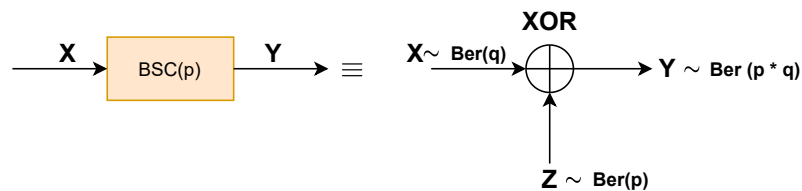


Figure 11.3: BSC channel

distribution of  $Y$  is defined using transition probability of channel and  $X$  (described in figure 11.3). We will use notation  $\mathbf{p} \star \mathbf{q} = \mathbf{p}(\mathbf{1}-\mathbf{q}) + \mathbf{q}(\mathbf{1}-\mathbf{p})$ . Then ,

$$H(Y) \geq H_2(p \star H_2^{-1}(H(X))) \quad (11.31)$$

with equality if and only if the  $\{X_i\}_1^n$  are independent. and  $H\{X^k\} = kp$

**Lemma :** Suppose we are generating  $X$  which depends on some distribution  $U$ , and it passes through BSC( $p$ ) (with transition probability  $p$ ) with output distribution  $Y$ . Then,

$$H\left(\frac{Y}{U}\right) \geq H_2\left(H_2^{-1}\left(H\left(\frac{X}{U}\right)\right) \star p\right) \quad (11.32)$$

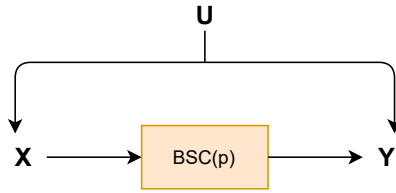


Figure 11.4: Binary symmetric channel

In vector form,

$$\frac{H(Y^n|U)}{n} \geq H_2\left(H_2^{-1}\left(H\left(\frac{X^n|U}{n}\right)\right) \star p\right) \quad (11.33)$$

*Proof.* We have claim ,

$$H\left(\frac{Y}{U}\right) \geq H_2\left(H_2^{-1}\left(H\left(\frac{X}{U}\right)\right) \star p\right) \quad (11.34)$$

Considering R.H.S. of the inequality we have,

$$H_2\left(H_2^{-1}\left(H\left(\frac{X}{U}\right)\right) \star p\right) \quad (11.35)$$

Suppose we have  $f(u) = H_2(H_2^{-1}(u) \star p)$ .

Now, eq (11.35) can be written as,

$$f\left(\sum_u P_U(u) \cdot H(X|U=u)\right) \quad (11.36)$$

Since  $f$  is convex function, we can write,

$$f\left(\sum_u P_U(u) \cdot H(X|U=u)\right) \leq \sum_u P_U(u) \cdot f(H(X|U=u)) \quad (11.37)$$

Again using the expression of  $f(u)$  we can have,

$$f(H(X|U=u)) = H_2(H_2^{-1}(H(X|U=u)) \star p) \quad (11.38)$$

For  $U = u$ ,  $X \sim \text{Ber}(q_u)$  and  $Y \sim (q_u \star p)$ .

$$H(X|U=u) = H_2(q - u \star p) = H_2(H_2^{-1}(H(X|U=u)) \star p) = f(H(X|U=u)) \quad (11.39)$$

Combining all above expressions, R.H.S. becomes,

$$\text{R.H.S} \leq \sum_u P_U(u) \cdot H(X|U=u) \quad (11.40)$$

Hence,

$$H\left(\frac{Y}{U}\right) \geq H_2\left(H_2^{-1}\left(H\left(\frac{X}{U}\right)\right) \star p\right) \quad (11.41)$$

and this concludes the proof.  $\square$