E2 205 Error-Control Coding
Lecture 13: Noisy channel-Random Coding Exponent

Shreya Shrestha Meel

September 25, 2019

1 Goal:

To show that it is possible to communicate reliably across a discrete memoryless channel at all rates $R < C$ where:

$$C = \max_{p(x)} I(X;Y)$$
$$= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

2 Setting:

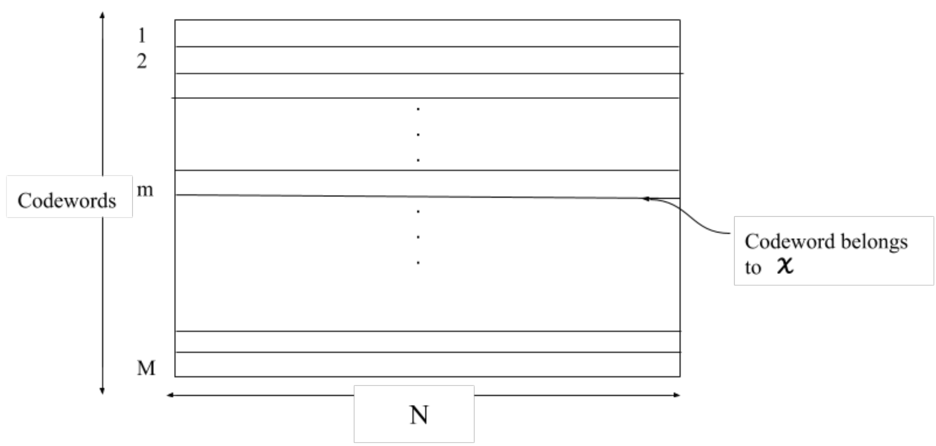
To begin with, we do not assume a memoryless channel. We will communicate

across the channel using a random(!) block code of length N, containing M codewords with,

$$M = \lceil \exp(NR) \rceil, 0 < R < 1$$

for some rate R (in *nats* per symbol).

3 Random Code construction



Let $Q_N(\underline{x})$ be a given probability distribution on N-tuples over $\mathcal{X}$. Each codeword is chosen independently of the other codewords using the same distribution $Q_N(\underline{x})$. (codeword and 1 codeword 2 may be same).

Let m be a specific integer in the range, $1 \le m \le M$. We will upper bound the probability of error incurred when the m^{th} message is transmitted, averaged over all codes.
We will use $\overline{P}_{e,m}$ to denote this probability.

$$\begin{aligned}\overline{P}_{e,m} &= Pr(error \mid m);\ m^{th}\text{message was transmitted(ML decoder).}\\ &= \sum_{\underline{x}_m \in \mathcal{X}^N} Q_N(\underline{x}_m) Pr(error \mid m, \underline{x}_m)\\ &= \sum_{\underline{x}_m} Q_N(\underline{x}_m) \sum_{\underline{y} \in \mathcal{Y}^{\mathcal{N}}} p(\underline{y} \big| \underline{x}_m)) Pr(error \big| m, \underline{x}_m, \underline{y})\end{aligned}$$

$$Pr(error \big| m, \underline{x}_m, \underline{y}) \leq P(\bigcup_{m' \neq m} A_{m'})\ ; 1 \leq m' \leq M,$$

and, $A_{m'}$ is the event that $\underline{x}_m$ is the m'^{th} codeword and $p(\underline{y} \mid \underline{x}'_m) \geq p(\underline{y} \mid \underline{x}_m)$.

$$\therefore P(A_{m'}) = \sum_{\underline{x}_{m'}} Q_N(\underline{x}_{m'}) \mathbb{1}_{\left\{p(\underline{x} | \underline{x}_{m'}) \geq p(\underline{y} | \underline{x}_m)\right\}}$$

$$\mathbb{1}_E = \begin{cases} 1, & \text{if } E \text{ holds} \\ 0, & \text{otherwise} \end{cases}$$

$$P(A_{m'}) \leq \sum_{\underline{x}_{m'}} Q_N(\underline{x}_{m'}) \left[\frac{p(\underline{y} \mid \underline{x}_{m'})}{p(\underline{y} \mid \underline{x}_m)}\right]^s ; \text{for any } s, 0 \leq s \leq 1, \tag{1}$$

(s is used to tighten the bound).

Also,

$$P(\bigcup_{m' \neq m} A_{m'}) \leq \left[\sum_{m' \neq m} P(A_{m'})\right]^\rho ; \text{for any } \rho, 0 \leq \rho \leq 1, \tag{2}$$

(ρ is used to tighten the union bound).
$\therefore$ From (1), (2) and (3), we get:

$$\overline{P}_{e,m} \leq \sum_{\underline{x}_m} Q_N(\underline{x}_m) \sum_{\underline{y}} p(\underline{y} \mid \underline{x}_m) \left(\sum_{m' \neq m} P(A_{m'})\right)^\rho$$

$$\leq \sum_{\underline{x}_m} Q_N(\underline{x}_m) \sum_{\underline{y}} p(\underline{y} \mid \underline{x}_m) \left(\sum_{m' \neq m} \sum_{\underline{x}_{m'}} Q_N(\underline{x}_{m'}) \left[\frac{p(\underline{y} \mid \underline{x}_{m'})}{p(\underline{y} \mid \underline{x}_m)}\right]^s\right)^\rho$$

The index, $\underline{x}'_m$ is dummy, and $Q_N(\underline{x}'_m)$ is identical $\forall m' \neq m$. We get $(M-1)$ such terms.

$$= (M-1)^{\rho} \sum_{\underline{y}} \sum_{\underline{x}_m} Q_N(\underline{x}_m) p(\underline{y} \mid \underline{x}_m)^{(1-s\rho)} \left(\sum_{\underline{x}} Q_N(\underline{x}) p(\underline{y} \mid \underline{x})^s \right)^{\rho}$$

Turns out this expression is minimised by setting $s = \dfrac{1}{1+\rho}, \therefore 1-s\rho = \dfrac{1}{1+\rho}$.

$$= (M-1)^{\rho} \sum_{\underline{y}} \sum_{\underline{x}_m} Q_N(\underline{x}_m) p(\underline{y} \mid \underline{x}_m)^{\frac{1}{1+\rho}} \left(\sum_{\underline{x}} Q_N(\underline{x}) p(\underline{y} \mid \underline{x})^{\frac{1}{1+\rho}} \right)^{\rho}$$

$$= (M-1)^{\rho} \sum_{\underline{y}} \left[\sum_{\underline{x}_m} Q_N(\underline{x}_m) p(\underline{y} \mid \underline{x}_m)^{\frac{1}{1+\rho}} \right]^{\rho+1}$$

(This result is Gallager's Theorem, 5.6.1, Information Theory and Reliable communication.)

4 Specialisation to the DMC

For Discrete channel,

$$\overline{P}_e, m \leq (M-1)^{\rho} \sum_{\underline{y}} \left[\sum_{\underline{x}} Q_N(\underline{x}) p(\underline{y} \mid \underline{x})^{\frac{1}{1+\rho}} \right]^{\rho+1}$$

Over the Discrete Memoryless Channel, we have:

$$p(\underline{y} \mid \underline{x}) = \prod_{i=1}^{N} p(y_i \mid x_i).$$

We set $\mathcal{Y} = \{0, 1, ...J-1\}$. Also, we choose:

$$Q_N(\underline{x}) = \prod_{i=1}^{N} Q(x_i).$$

Q is some pmf on $\mathcal{X} = \{0, 1, ...K-1\}$.

$$\overline{P}e, m \leq (M-1)^{\rho} \sum_{y_1} ... \sum_{y_N} \left(\sum_{x_1} ... \sum_{x_N} \prod_{i=1}^{N} Q(x_i) \big[p(y_i \mid x_i)\big]^{\frac{1}{1+\rho}} \right)^{\rho+1}$$

$$= (M-1)^{\rho} \sum_{y_1} \dots \sum_{y_N} \left(\prod_{i=1}^{N} \sum_{x_i} Q(x_i) \big[p(y_i \mid x_i) \big]^{\frac{1}{1+\rho}} \right)^{\rho+1}$$

$$= (M-1)^{\rho} \sum_{y_1} \dots \sum_{y_N} \prod_{i=1}^{N} \left(\sum_{x_i} Q(x_i) \big[p(y_i \mid x_i) \big]^{\frac{1}{1+\rho}} \right)^{\rho+1}$$

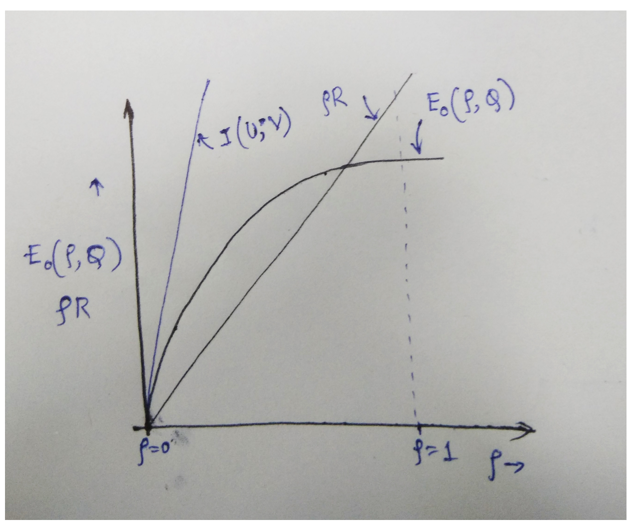
$$= (M-1)^{\rho} \prod_{i=1}^{N} \sum_{y_i} \left(\sum_{x_i} Q(x_i) \big[p(y_i \mid x_i) \big]^{\frac{1}{1+\rho}} \right)^{\rho+1}$$

$$< \exp(\rho N R) \left(\sum_{j=0}^{J-1} \left(\sum_{k=0}^{K-1} Q(k) \big[p(j \mid k) \big]^{\frac{1}{1+\rho}} \right)^{\rho+1} \right)^{N}$$

$$\left(\lceil \exp(NR) \rceil = M \ \therefore M - 1 \leq \exp(NR) \leq M \right)$$

$$= \exp(-N(E_0(\rho, Q) - \rho R)), \text{where:}$$

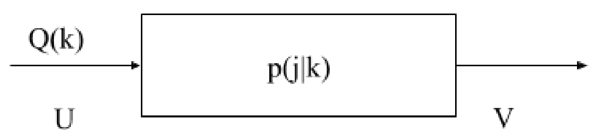
$$E_0(\rho, Q) \triangleq -\ln \left[\sum_{j=0}^{J-1} \left(\sum_{k=0}^{K-1} Q(k) \big[p(j \mid k) \big]^{\frac{1}{1+\rho}} \right)^{\rho+1} \right]$$



$$\frac{\partial}{\partial \rho}(E_0(\rho, Q) - \rho R) = 0 \implies R = \frac{\partial}{\partial \rho} E_0(\rho, Q)$$

We are going to show that, the value of R corresponding to the slope of the blue line on the graph is the **mutual information** associated with U and V over channel $p(j \big| k)$.

$$R = \sum_j \sum_k Q(k) p(j \big| k) \ln \frac{p(j \big| k)}{p(j)} (\text{in } nats)$$



5 Finding capacity over the DMC

To find the capacity, we find the partial derivative of $E_0(\rho, Q)$ w.r.t ρ and setting ρ to 0.

$$\left.\frac{\partial E_0(\rho, Q)}{\partial \rho}\right|_{\rho=0} = \left.\frac{-1}{\sum_j \left[\sum_k Q(k) \left[p(j \mid k)\right]^{\frac{1}{1+\rho}}\right]^{\rho+1}}\right|_{\rho=0} \frac{\partial}{\partial \rho} \sum_j \left.\left[\sum_k Q(k) \left[p(j \mid k)\right]^{\frac{1}{1+\rho}}\right]^{\rho+1}\right|_{\rho=0}$$

Set:

$$z(\rho) = \sum_k Q(k) \left[p(j \mid k)\right]^{\frac{1}{1+\rho}}$$

Note that $z(0) = p(j)$.

$$\left.\frac{\partial z(\rho)}{\partial \rho}\right|_{\rho=0} = \sum_k Q(k) \frac{\partial}{\partial \rho} \exp\left.\left(\frac{1}{1+\rho} \ln p(j \big| k)\right)\right|_{\rho=0}$$

$$= \sum_k Q(k) \left[p(j \big| k)\right]^{\frac{1}{1+\rho}} \left(\frac{-1}{(1+\rho)^2}\right) \ln p(j \big| k)\Big|_{\rho=0}$$

$$= \sum_k Q(k)p(j|k) \ln \frac{1}{p(j|k)}$$

$$\left.\frac{\partial E_0(\rho, Q)}{\partial \rho}\right|_{\rho=0} = (-1) \sum_j \left.\frac{\partial}{\partial \rho}[z(\rho)]^{\rho+1}\right|_{\rho=0}$$

$$=(-1)\textstyle\sum_j \frac{\partial}{\partial \rho} \exp\left.\Big((\rho+1) \ln z(\rho)\Big)\right|_{\rho=0} = (-1) \sum_j \Big(z(\rho)^{\rho+1}\Big) \frac{\partial}{\partial \rho}\left.\Big((\rho+1) \ln z(\rho)\Big)\right|_{\rho=0}$$

$$= (-1) \sum_j \Big(z(\rho)^{\rho+1}\Big) \left.\left[(\rho+1)\frac{1}{z(\rho)} \frac{\partial}{\partial \rho} z(\rho) + \ln(z(\rho))\right]\right|_{\rho=0}$$

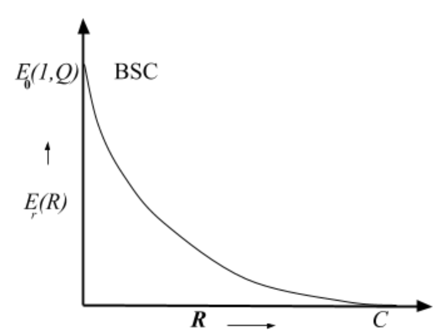
$$= (-1) \sum_j p(j) \left[\frac{1}{p(j)} \sum_k Q(k)p(j|k) \ln \frac{1}{p(j|k)} + \ln \Big(p(j)\Big)\right]$$

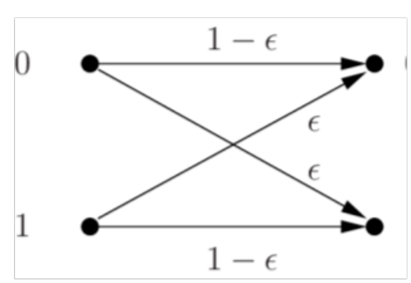
$$= \sum_j \sum_k Q(k)p(j|k) \ln \frac{p(j|k)}{p(j)} = \mathbf{I(Q;P)}.$$

Thus, we have shown that:

$$\begin{aligned} \overline{P}_{e,m} &\leq \exp(-N \max_{\rho,Q}(E_0(\rho, Q) - \rho R)) \\ &= \exp(-N E_r(R)), (\text{where } E_r(R) \text{ is the random coding exponent}). \end{aligned}$$

By setting $Q()$ to be the distribution that achieves capacity over the DMC, we see that for all rates $R < C$, $E_r(R) > 0$ (by choosing ρ optimally such that $R = \frac{\partial E_0(\rho,Q)}{\partial \rho}$). This shows that we can communicate reliably across the DMC at all rates $R < C$.





BSC

This is tight for larger R, can be tightened for smaller R.