Mathematical methods in communication — May 26th, 2009

Lecture 5

Lecturer: Haim Permuter — *Scribe: Elad Perez and Uriel Odes*

I. Challenge:

Exercise 1 *Finding an optimal code for the infinite case:*. Consider the set $\{p_i\}_{i=1}^{\infty}$, such that for any $i \epsilon N : p_i > 0, p_i \geq p_{i+1}, \sum_{i=1}^{\infty} p_i = 1$. Find an optimal code, i.e. a prefix code such that $E[L]$ is minimal (L is the length of the code word assigned to each value of X, so L is equal to the length of the code word assigned to x_i with probability p_i). As is for today, this is an unsolved question!

II. Channel Coding

Definition 1 A discrete memoryless channel (DMC) is a channel with the following properties:

1) Discrete time -$[1, 2, 3, ...]$.
2) Memoryless channel- current output Y_i depends on the history (X^i, Y^{i-1}) only through the current input X_i. $(X^{i-1}, Y^{i-1}) - X_i - Y_i \Leftrightarrow P_{Y_i|X^i,Y^i-1} = P_{Y_i|X_i}$

We consider the following channel coding problem:

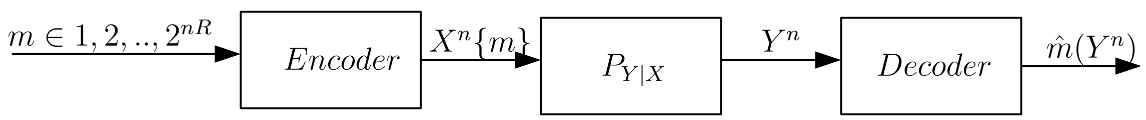

Fig. 1. Communication system

We define two types of errors, **Perror:**

1) $\lambda_{av}^{(m)} = P(M \neq \hat{M})$
2) $\lambda_{max}^{(m)} = \max_m P(m \neq \hat{m} : M = m)$

Definition 2 A $(2^{nR}, n)$ code is:

1) A message set of size $\{0,1\}^{nR}$
2) Enc: $\{0,1\}^{nR} \to \chi^n$
3) Dec: $\chi^n \to \{0,1\}^{nR}$

Definition 3 (Achievable rate) A rate R is achievable if there exists a sequence of codes $(2^{nR}, n)$ s.t. $\lambda_{max}^{(m)} \to 0$ when $n \to \infty$.

Definition 4 (Capacity) The capacity of a channel, C, is the suprimum of all R such that R is an achievable rate (as defined above).

Definition 5 we define C^I=$\max_{P_X} I(X;Y)$.

Theorem 1 the capacity C is equal to C^I, as defined above.

Example 1 (Erasure Channel) *The binary erasure channel:*

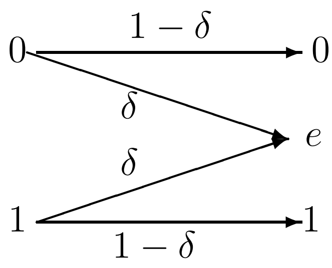

Fig. 2. Binary erasure channel (BEC) with erasure probability δ.

This is a model of a channel where a 0 symbol has a probability of δ to become e(symbol for error) and a probability of $1-\delta$ to become 0. and a 1 symbol has a probability of δ to become e, and a probability of $1-\delta$ to become 1. now to calculate the capacity of the channel:

$$\begin{aligned}
C^I &= \max_{P_X} I(X;Y) = \max_{P_X}(H(X) - H(X|Y)) \\
&= \max_{P_X}(H(X) - P_Y(1)H(X|Y=1) - P_Y(0)H(X|Y=0) - P_Y(e)H(X|Y=e)) \\
&\overset{(a)}{=} \max_{P_X}(H(X) - P_Y(e)H(X|Y=e)) \\
&\overset{(b)}{=} \max_{P_X}(H(X) - P_Y(e)H(X)) = \max_{P_X}(H(X) - \delta H(X))
\end{aligned}$$

$$= \max_{P_X}(1-\delta)H(X)$$
$$= 1-\delta \tag{1}$$

where

(a) Follows from given $Y = 1$ we know with a probability of 1 that $X = 1$ thus $H(X|Y = 1) = 0$ and given $Y = 0$ we know with a probability of 1 that $X = 0$ thus $H(X|Y = 0) = 0$

(b) Follows from given $Y = e$ we have no information on the value of X thus $H(X|Y = e) = H(X)$

Example 2 (Binary symmetric channel) *The binary symmetric channel:*

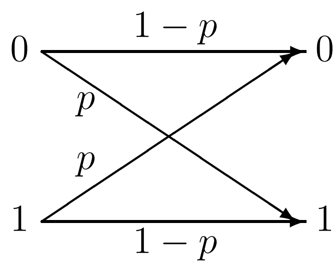

Fig. 3. Binary symmetric channel (BSC) with probability p.

This is a model of a channel where a 0 symbol has a probability of p to become 1 and a probability of $1-p$ to become 0. and a 1 symbol has a probability of p to become 0, and a probability of $1-p$ to become 1.

now to calculate the capacity of the channel:

We can write $Y = X \oplus Z$ where Z is distributed Bernoulli(p), Z is independent of X:

$$I(X;Y) = H(Y) - H(Y|X) \overset{(a)}{=} H(Y) - H(Z \oplus X|X) = H(Y) - H(Z|X)$$
$$\overset{(b)}{=} H(Y) - H(Z) \leq 1 - H(p) \tag{2}$$

Where

(a) Follows from If X is given than $H(Y|X) = H(f(Z,X)|X)$ where f is a function.

(b) Follows from Z and X are independent.

Notice: If X is distributed Bernoulli($\frac{1}{2}$), then $I(X;Y) = 1-H(p)$ therefore $C^I = 1-H(p)$

Lemma 1 For a memoryless channel without feedback
i.e. $P(x_i|x^{i-1}, y^{i-1}) = P(x_i|x^{i-1})$, we have: $P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x_i)$

Proof:

$$\begin{aligned}
P(y^n|x^n) &= \frac{P(y^n, x^n)}{P(x^n)} \\
&\overset{(a)}{=} \frac{\prod_{i=1}^{n} P(y_i, x_i|y^{i-1}, x^{i-1})}{P(x^n)} \\
&= \frac{\prod_{i=1}^{n} P(x_i|x^{i-1}, y^{i-1}) P(y_i|y^{i-1}, x^i)}{P(x^n)} \\
&\overset{(b)}{=} \frac{\prod_{i=1}^{n} P(x_i|x^{i-1}) P(y_i|x_i)}{P(x^n)} \\
&= \frac{P(x^n) \prod_{i=1}^{n} P(y_i|x_i)}{P(x^n)} \\
&= \prod_{i=1}^{n} P(y_i|x_i) \qquad (3)
\end{aligned}$$

Where

(a) Follows from chain-rule.

(b) Follows from a memoryless channel without feedback.