6.441 Transmission of Information Mar 16, 2005

Lecture 11

Lecturer: Madhu Sudan *Scribe: Sang Joon Kim*

Today we will talk about coding theorem for symmetric channel.

1 Admin

- I will be out of town next week. Chung is in charge.
- Midterm is in 7 days.

2 Review

First, we review what we did last time.

2.1 DMC(Discrete Memoryless Channel)

DMC(Discrete Memoryless Channel) is the channel that can be repeatedly used for each transmitting information. The detail description is as following:

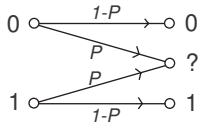
$$\begin{cases} \Omega_x - \text{ input alphabet} \\ \Omega_y - \text{ output alphabet} \\ P_{y|x}(y|x)_{y\in\Omega_y, x\in\Omega_x} - \text{ transition probability matrix} \end{cases}$$

Definition 1

$$\textit{Capacity} \triangleq \max_{X\sim P_X, Y\sim P_{Y|X}} I(X;Y)$$

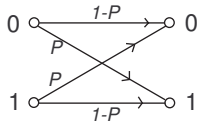
For maximizing the channel capacity, we choose the set X and determine the distribution of X,(P_X). Y is the received data set that is correlated random variable of X by $P_{Y|X}$.

2.2 BEC(Binary Erasure Channel)



BEC is like above and Capacity $= 1 - P$. If $P = 0.9$ then, we lose 90 % of information in this channel, i.e. for one successful transmission, we should try average 10 times transmission.

2.3 BSC(Binary Symmetric Channel)



Capacity of BSC $= 1 - H(P)$. BSC is less useful channel than BEC but more reasonable channel. For example, if $P = 0.1$ than, $H(0.1)$ is much larger than 0.1 and the channel capacity of BSC is less than that of BEC. Also, if we assume that $P = 0.5$ than, we don't need to send any information because we get no information of transmitted data from received data. One more example is that if $P = 0.49$ then, capacity $\approx 10^{-4}$. It means that we should retransmit 10000 times for one sending. This is obviously not very reliable situation and our question is that "how can we achieve a good capacity for this channel?".

2.4 Symmetric Channel

Symmetric channel is the channel that satisfies the following properties:

$$P_{Y|X} \text{ is } \begin{cases} \text{all rows are permutation of each other.} \\ \text{all columns are permutation of each other.} \end{cases}$$

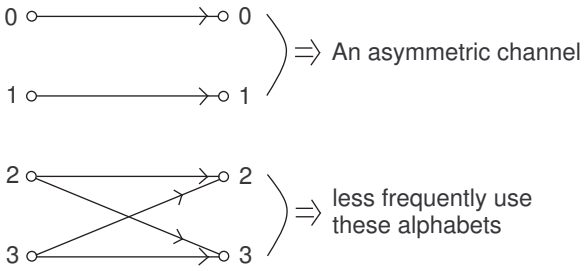
Example

$$P_{Y|X} = \begin{bmatrix} .21 & .21 & .21 & .21 & .16 \\ .16 & .21 & .21 & .21 & .21 \\ .21 & .16 & .21 & .21 & .21 \\ .21 & .21 & .16 & .21 & .21 \\ .21 & .21 & .21 & .16 & .21 \end{bmatrix}$$

i^{th} row represents the probability of Y given $X = x_i$.

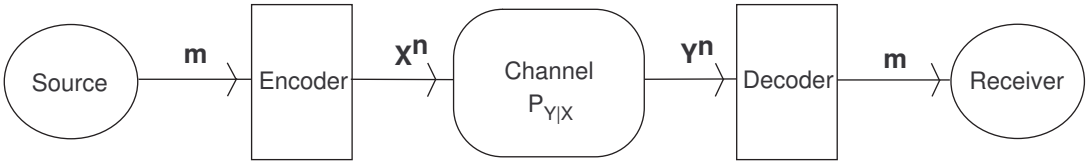
Capacity of symmetric channel $= \log |\Omega_y| - H(\text{first row})$. When X is uniformly distributed we achieve $H(Y) = \log |\Omega_y|$. We don't have a clue which message is transmitted if the row is uniformly distributed.

Example



In the above channel, if there are some x_i that are rarely transmitted then we don't use the uniform distribution. Also, there may be an asymmetric channel but we are only looking at symmetric case in the class.

3 Coding theorem



$$\begin{cases} E & : \{0,1\}^k \to \Omega_x^n \\ D & : \Omega_y^n \to \{0,1\}^k \end{cases}$$

$$m = D(\bar{Y}) \leftarrow \bar{Y} \leftarrow P_{Y|X}(\bar{X}) \leftarrow \bar{X} \leftarrow E(m)$$

The purpose of communication is not to guarantee absolutely the successful transmission, but to increase the reliability of the channel.

Definition 2

$$\textit{Decoding Error Rate} \triangleq Pr_{m\in\{0,1\}^k,\bar{Y}}\left[m \neq D(\bar{Y})\right]$$
$$\textit{where } \bar{Y} = \textit{channel}(E(m))$$

We assume that X is uniformly distributed because X is in the typical set. What we want to do is to make the decoding error rate go to 0 very fast and it is the same as increasing the reliability of the channel. If we have an enough time to decode $\bar{Y}$, looking all message and finding the best one - most likely probability is the optimum.

Optimal Decoding Algorithm D for fixed E

$$D(y^n) \rightarrow m_0 = arg\max_m \{Pr[E(m)] \cdot Pr[y^n|E(m)]\}$$
$$\text{(if we assume the uniform distribution for } E(m))$$
$$= arg\max_m \left\{\frac{1}{2^k}\prod_{i=1}^{n} P_{Y|X}(y_i|E(m)_i)\right\}$$

This is the same as maximum a posterior probability.

Our goal is to find the encoding function that allow to achieve following property:

$$\boxed{\frac{k}{n} \rightarrow \text{Capacity}}$$

3.1 BEC

In BEC, $Pr[\bar{Y} = ?^n] = P^n > 0$. This means that there is always probability that we fail to decode the received information.

$$m_1, \ldots, m_k \Rightarrow X^n = (x_1, \ldots, x_n) \Rightarrow (x_1, x_2, ?, x_4, ?, ?, x_7, \ldots, x_n)$$
$$\text{Channel sends} \{x_j\}_{j\in S}, \quad S \subseteq \{1, 2, \ldots, n\}$$

In above situation, message $\{m_i\}$ is encoded to X^n and transmitted. The received information has several ?s those are erased data. What we want to do is as following:

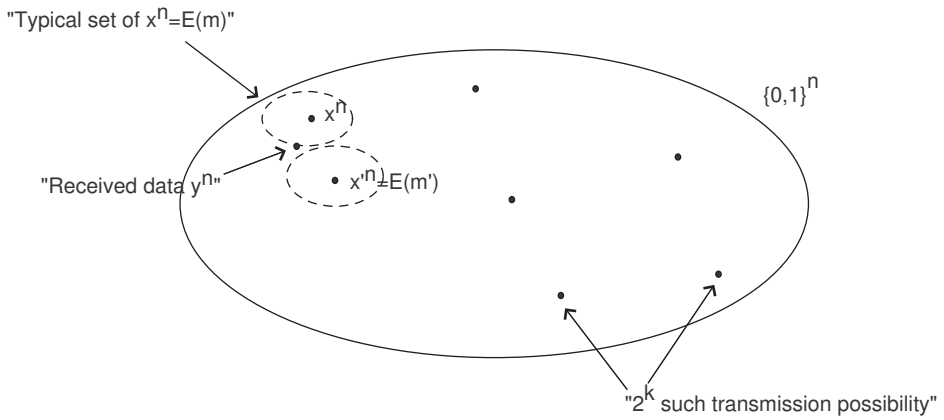
$$I(m_1, \ldots, m_k; \{x_j\}_{j\in S}) \approx k$$

That mutual information between $\{m_i\}$ and received data is approximately k bits means we can decode almost exactly whatever the received data is. This is our goal.

$$|S| \approx (1-P)n$$
$$\frac{k}{n} \approx \text{Capacity} \Leftrightarrow k \approx (1-P)n$$

Challenge : How to encode $m_1, \ldots, m_k \rightarrow x_1, \ldots, x_n$ so that almost every subset of size $(1+\epsilon)k$ give k bits of information about $m_1, \ldots, m_k$?

3.2 BSC



In BSC, if $m = (m_1, m_2, \ldots, m_k)$ is transmitted via the encoded data $x^n = (x_1, x_2, \ldots, x_n)$, the typical set of the received data $y^n = (y_1, y_2, \ldots, y_n)$ forms a geometric circle in $\{0,1\}^n$ space which has x as its origin. (The geometric distance between $x, y \in \{0,1\}^n$ is defined as the number of coordinate where they differ.) Therefore, our challenge is following.

Challenge : How to design the encoding function E so that $E(m)$ and $E(m')$ is far for most pair m, m'?

3.3 Shannon Encoding Function

Pick the Encoding Funtion $E : \{0,1\}^k \to \Omega_x^n$ as follows,

For every $m \in \{0,1\}^k$,
$E(m)$ is chosen uniformly from Ω_x^n and independently from $E(m')$ for all $m' \neq m$

Then, the following lemma holds.

Lemma 3 *If $k = R \cdot n$ for some $R < C$(capacity of the channel), then*

$$\lim_{n\to\infty} \Pr_{E,m,\bar{y}=channel(E(m))}[m \neq D(\bar{y})] = 0$$

The lemma implies the following.

$$\lim_{n\to\infty} \min_{E}\{\Pr_{m,\bar{y}=channel(E(m))}[m \neq D(\bar{y})]\} = 0$$

In other words, the lemma says all rates below capacity of the channel are achievable. Now, define the following typical set.

Definition 4 *For $\bar{x}^n \in \Omega_x^n$, define the set $A_{\epsilon,\bar{x}}^{(n)}$ as follows,*

$$A_{\epsilon,\bar{x}}^{(n)} \triangleq \{\bar{y}^n \in \Omega_y^n | \Pr[\bar{y} \text{ received } | \bar{x} \text{ transmitted }] \geq 2^{(-H(r)-\epsilon)n}\}$$

,where r is the first row of the transition matrix of the channel.

Then, we can easily check the following claim.

Claim 5 $\forall \bar{x}$, $|A_{\epsilon,\bar{x}}^{(n)}| \leq 2^{(H(r)+\epsilon)n}$.

Now, we think the following two events. The first event is the event of receiving $\bar{y} \notin A^{(n)}_{\epsilon,\bar{x}}$ when transmitting $\bar{x} = E(m)$. AEP tells the probability that this event happens goes to 0 as $n \to \infty$. The second event is the event of receiving $\bar{y} \in A^{(n)}_{\epsilon,E(m')}$ such that $\exists m' \neq m$ when transmitting $\bar{x} = E(m)$. The probability that this event happens is greatest when $\bar{y}$ distributes uniformly in Ω_y^n, because this channel is symmetric. Therefore, the probability that this event happens is

$$\begin{aligned}\sum_{m'} Pr(y \in A^{(n)}_{\epsilon,E(m')}) &\leq \sum_{m'} \frac{|A^{(n)}_{\epsilon,E(m')}|}{|\Omega_y|^n} \leq \sum_{m'} \frac{2^{(H(r)+\epsilon)n}}{|\Omega_y|^n} \\ &\leq 2^k 2^{(H(r)-\log|\Omega_y|)n+\epsilon n} = 2^k 2^{-Cn+\epsilon n} = 2^{-\epsilon n} \text{ if } k = Rn, \epsilon = \frac{C-R}{2}\end{aligned}$$

This implies that the probability that this event happens goes to 0 as $n \to \infty$.