**6.895 Essential Coding Theory** September 13, 2004

# Lecture 2

*Lecturer: Madhu Sudan* *Scribe: Joungkeun Lim*

## 1 Overview

We consider the problem of communication in which a source wish to transmit information to a receiver. The transmission is conducted through a channel, which may generate errors in the information depending on the options. In this model, we will introduce Shannon's coding theorem, which shows that depending on the properties of the source and the channel, the probability of the receiver's restoring the original data varies with a threshold.

## 2 Shannon's theory of information

In this section we will discuss the main result from Shannon's paper which was introduced in 1948 and founded the theory of information.

There are three entities in Shannon's model:

- Source : The party which produces information by a probabilistic process.
- Channel : The means of passing information from source to receiver. It may generate errors while transporting the information.
- Receiver : The party which receives the information and tries to figure out information at source's end.

There are two options for channel, "Noisy" and "Noiseless"

- Noisy channel : A channel that flips some bits of information sent across them. The bits that flips are determined by a probabilistic process.
- Noiseless channel : A channel that perfectly transmits the information from source to receiver without any error.

The source will generate and encode its message, and send it to receiver through the channel. When the message arrives, the receiver will decode the message. We want to find the encoding-decoding scheme which makes it possible for a receiver to restore the exact massage which a source sent. Shannon's theorem states the conditions with which a restoration can be conducted with high probability.

### 2.1 Shannon's coding theorem

**Theorem 1** *(Shannon's coding theorem)*
*There exist positive real values capacity $C$ and rate $R$ satisfying the followings. If $R < C$ then information transmission is feasible(coding theorem.) If $R > C$ then information transmission is not feasible(Converse of coding theorem.)*

Capacity C and rate R are the values associated with a source and a channel respectively. The general way to compute this two values are a bit complicated. To get a better understanding, we will start with simple examples of Shannon's model one in noiseless model and one in noisy model.

## 2.2 preliminaries

Before studying the examples, we study the property of the binary entropy function and Chernoff bounds which make crucial roles in the analyses in later chapters.

**Definition 2** *For $p \in [0,1]$, the binary entropy function is defined as follows.*

$$H(p) = plog_2\frac{1}{p} + (1-p)log_2\frac{1}{1-p}.$$

H(p) is a concave function and has a maximum values 1 where p=1/2. The following property of H(P) is used in later chapters.

- Let $B_n(0,r)$ be a ball of radius r(in Hamming distance) and center 0 in $\{0,1\}^n$. $V(r,n) = Vol(B_n(0,r)) = \sum_{i=0}^{r} \left( \begin{array}{c} n \\ i \end{array} \right) \approx 2^{H(r/n)\cdot n}$. Hence $Vol(pn,n) \approx 2^{H(p)\cdot n}$.

**Lemma 3** *(Chernoff Bounds)*
*If $\eta_1, \eta_2, \cdots, \eta_n$ are independent random variables in [0,1] with $EXP(\eta_i) = p$, then*

$$Pr[|\frac{\sum_{i=1}^{n}\eta_i}{n} - p| > \epsilon] \leq 2^{-\epsilon^2\cdot n}.$$

## 2.3 An example of noiseless model

Source produces a sequence of bits such that each bits are 0 with probability 1-p and 1 with probability p, where $p \leq 1/2$. Source produces one bit per unit of time. For it is a noiseless channel, the channel transmits exactly same bits to a receiver as the bits given from the source. The channel is allowed to transmit C bits per unit of time.. In this case, the rate of source is given as the entropy function H(p) and the capacity value is the number of bits transmitted through channel per unit of time. When n is the amount of time we used the channel, the Shannon's coding theorem is expressed as follows.

**Theorem 4** *(Shannon's noiseless coding theorem)*
*If $C > H(p)$, then there exist encoding function $E_n$ and decoding function $D_n$ such that Pr[Receiver figures out what the source produced]$\geq 1 - exp(-n)$.*
*Also if $C > H(p)$, then there exist encoding function $E_n$ and decoding function $D_n$ such that Pr[Receiver figures out what the source produced]$\leq exp(-n)$.*

## 2.4 An example of noisy model

The source produces a sequence of bits such that each bits are 0 with probability 1/2 and 1 with probability 1/2. Source produceds R bits per unit of time, where $R < 1$. For it is a noisy channel, the channel filps each bit with a probabilistic process. In this example, channel flips each bit with probability p. Also the channel transmits one bit per unit of time. In this case, the rate R is the number of bits produced in the source per unit of time and the capacity C is given as 1-H(p). Then shannon's coding theorem is expressed as follows.

**Theorem 5** *(Shannon's noisy coding theorem)*
*If $R < 1-H(p)$ then there exist encoding function $E_n$ and decoding function $D_n$ such that Pr[Receiver figures out what the source produced]$\geq 1 - exp(-n)$.*
*Also If $R > 1 - H(p)$ then there exist encoding function $E_n$ and decoding function $D_n$ such that Pr[Receiver figures out what the source produced]$\leq exp(-n)$.*

We prove the first part of theorem using probabilistic method and give an idea of the proof for the second part of theorem.

**Proof** (First part)

Let k be the number of bits produced by source, then $k = R \cdot n$. For $R < 1 - H(p)$, there exists $\epsilon > 0$ such that $R < 1 - H(p + \epsilon)$. For this $\epsilon$, let $r = n(p + \epsilon) = n \cdot p'$. Now we can restate the theorem as follows.

If $R = k/n < 1 - H(p)$, then there exists functions $E : \{0,1\}^k \to \{0,1\}^n, D : \{0,1\}^n \to \{0,1\}^k$ such that $Pr_{\eta \in BSC_p, m \in U_k}[m \neq D(E(m) + \eta)] \leq exp(-n)$, where $U_k$ is uniform distribution on k bit strings and $BSC_{p,n}$ is distribution on n bit strings with each bits to be 0 with probability 1-p and 1 with probability p.

Pick the encoding function $E : \{0,1\}^k \to \{0,1\}^n$ at random, and the decoding function $D : \{0,1\}^n \to \{0,1\}^k$ works as follows. Given a string $y \in \{0,1\}^n$, we find the $m \in \{0,1\}^k$ such that $\Delta(y, E(m))$ is minimized. This m is the value of $D(y)$. Fix $m \in \{0,1\}^k$ and fix $E(m)$ also. For E is randomly chosen, $E(m')$ is still random when $m' \neq m$. Let $y$ be the value that the receiver acquires. In order for $D(y) \neq m$ at least one of the following two events must occur:

- There exists some $m' \neq m$ such that $E(m') \in B(y, r)$.
- $y \notin B(E(m), r)$

If neither of above events happen, then m is the unique message such that E(m) is within a distance of r from $y$ and so $D(y) = m$.

We prove that the events above happen with low probability. For the first event to happen, the error $\eta = y - E(m)$ has more than $n(p + \epsilon)$ of 1 bits. By Chernoff bounds we will have

$$Pr[y \notin B(E(m), r)] \leq 2^{-(\epsilon^2/2)n}.$$

For the second event happen to happen, fix y and an $m' \neq m$ and consider the event that $E(m') \in B(y, r)$. For $E(m')$ is random, the probability of this event is exactly $Vol(B(y, r))/2^n$. Using

$$Vol(B(y, p'n)) \approx 2^{H(p')n},$$

we have

$$Pr[E(m') \in B(y, r)] \approx 2^{H(p')n-n}.$$

Using union bound, we get $Pr[\exists m' \in \{0,1\}^k \ s.t. \ E(m') \in B(y, r)] \leq 2^{k+H(p')n-n}$

For $R = k/n < 1 - H(p')$, $2^{k+H(p')n-n} = exp(-n)$. Therefore the probability that second event happens is also bounded by $exp(-n)$.

Hence the probability of at least one of above two events happens is bounded by $exp(-n)$ where m and E(m) is fixed. Therefore for the random E and associated D, the probability is still bounded. Using probabilistic method, we see that there exists a encoding E and associated decoding D such that the probability that any of two events happen is still bounded by $exp(-n)$.

■

Here we give the brief sratch of the proof for second part of theorem. Decoding function partitions universe to $2^k$ regions. By Chernoff bounds, Pr[number of 1 bits in error <pn] is low. Hence when E(m) was transmited from source, the corrupted value y that arrives at receiver will have spread-out distribution around E(m). It means the region that covers most of possible y value has much larger size than one of the $2^k$ region that contains E(m). It will make the decoding inaccurate.