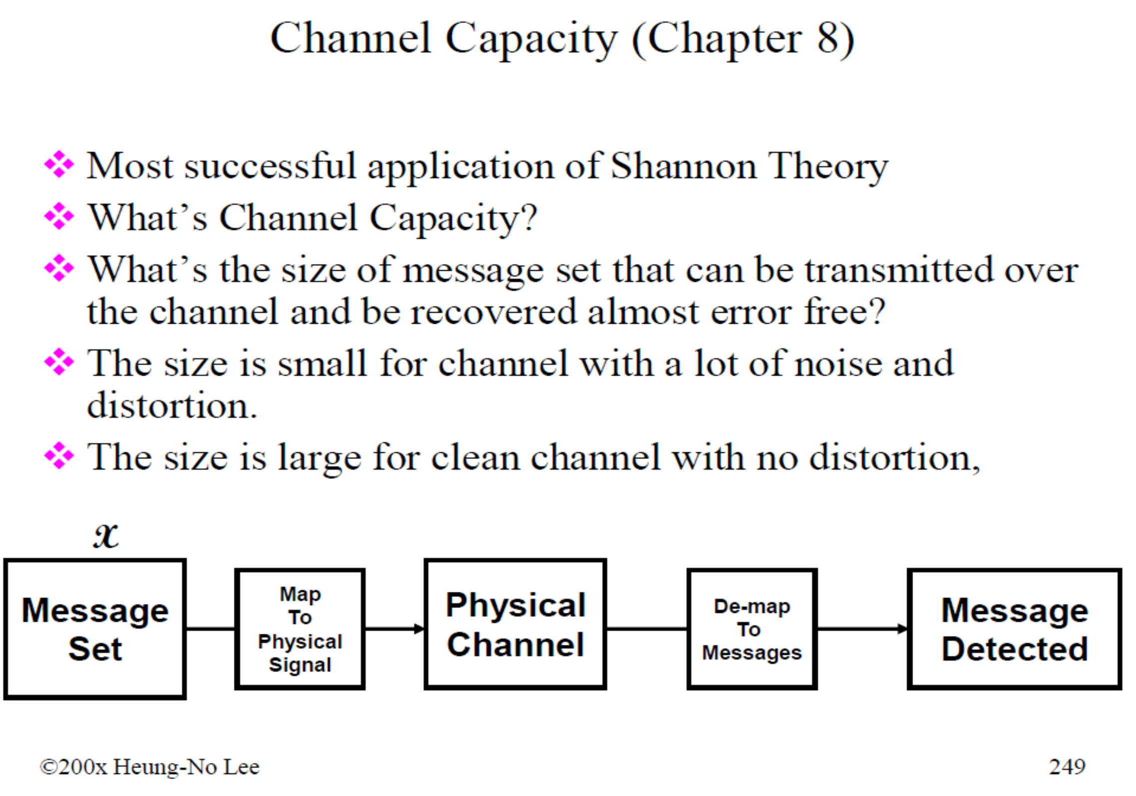
Channel Capacity

By Heung-No Lee
Date: Nov. 1st, 2020

Chapter 7 (Cover and Thomas, 2nd Edition or Chapter 8 in the 1st Edition) discusses the channel coding theory. The channel here is to mean a conditional probability distribution $p(y|x)$. Given an input $x \in \mathcal{X}$ fed to a channel, what is the range $\mathcal{Y}$ of output y from the channel? What is the probability of an output $p(y|x)$ for each $y \in \mathcal{Y}$. Answering this questions for all $x \in \mathcal{X}$ will define the channel. This input-output model is general enough to accommodate many if not all input-output relations successfully. This you will see soon.

Given a channel $p(x|y)$ one can define the capacity of the channel. The capacity is the maximum amount of information, [bits/channel-use], that can be sent over the channel almost error-free.



Physical channels: Statistical errors due to physical limitation can be introduced into the statistical channel.

- A digital symbol sent at a transmitter vs. A digital symbol decided at a receiver
- The word prepared to utter inside your head vs. the word actually spoken out.
- Reading from a memory device (magnetic recording tape, Hard discs, memory cells)
- Writing into a memory device: a word to be written vs. a word actually written
- Input picture with a class label to a neural network classifier vs. a classified label

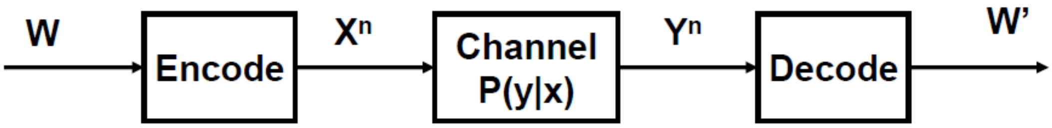
Channel Capacity

- Capacity1 = log(Size of the Message Set)
- T = Time required to transmit a message in the set.
- Capacity = Capacity1/T, bits/sec
- Capacity = Capacity in n-use of channel / n = *bits/channel use*
- Capacity is the number of bits that can be transferred over a given channel with almost no error.

 250

Why almost no error?
It has to do with the law of large numbers. Recall convergence in probability. Recall the surface hardening argument. As the dimension increases, the probability of the set of samples on which the discrepancy between the converging random variable and to what it is converging is getting smaller and smaller. As the dimension is enlarged, the magic occurs and the error gets smaller and smaller. We will see that in the proofs of channel capacity theorems.

Discrete Channel



- Input alphabet X, output alphabet Y, and the probability transition matrix p(y|x) ~ prob. of observing y given x transmitted .
- The channel is memoryless if $P(y^n|x^n) = \prod p(y|x)$.
- We can transform the physical channel into the discrete channel.
- Example) A cable with a certain bandwidth F_{max} and noise power spectral density → Sample at twice the bandwidth and obtain the discrete channel (sample IN sample OUT) .

"*Information*" Channel Capacity

- $C = \max_{p(x)} I(X; Y)$
  - The maximum is taken over all input distr. p(x).
  - Compared with the "*operational*" channel capacity ~ the highest rate in bits/channel-use, at which information can be sent with an arbitrary small probability of error.
  - Shannon proved that the *operational* channel capacity is equal to the *information* channel capacity.

 252

Operational channel capacity is the capacity of a code in practice. A code in practice has the following form:

1. A codebook is used.
2. A codebook is a collection of codewords.
3. The size of the codebook is 2^{nR} where R [bits/channel-use] is the rate of the code. This is to mean there are 2^{nR} codewords. For example, suppose $R = 0.5$ [bits/channel-use] and $n = 6$ be the length of each codeword. Then, there are $2^{6(0.5)} = 8$ binary codewords of length 6.
4. Encoding is a function from a set of messages to a codebook. It maps a message to a codeword. Suppose the same example given in 3. There are 8 messages in the message set. Each can be enumerated by 000, 001, 010, …, 111 without loss of generality. Each message is mapped to a codeword of length 6 in the codebook.
5. The term, a single channel-use, is to mean that the channel $p(y|x)$ is used once. For example, to send a vector of input symbols $(x_1, x_2, x_3, x_4, x_5, x_6)$, the channel has to be used six times. The output of the channel is a vector $(y_1, y_2, y_3, y_4, y_5, y_6)$. If each channel use is independent and identical, the vector channel can be given as $\prod_{i=1}^{6} p(y_i | x_i)$.
6. Decoding is a function from the set of outputs of the channel to the set of messages. It maps an output vector $(y_1, y_2, y_3, y_4, y_5, y_6)$ to an input vector $(x_1, x_2, x_3, x_4, x_5, x_6)$ which is a codeword. From a codeword, the transmitted message can be de-mapped.
7. An error occurs when the transmitted message and the decoded message are not the same.

Noiseless Binary Channel

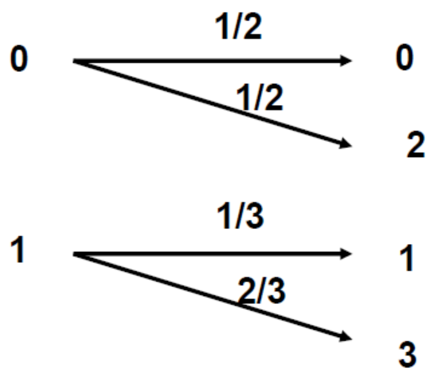
- $C = \max_{p(x)} I(X; Y) = ?$

$$0 \longrightarrow 0$$

$$1 \longrightarrow 1$$

Noisy Channel with Non-Overlapping Output

- C = ?



Noisy Typewriter

- The typewriter writes the input letter with prob. 1/2 or the next letter in the alphabet with prob. 1/2
- $C = ?$
- $C = \max_{p(x)} I(X ; Y)$

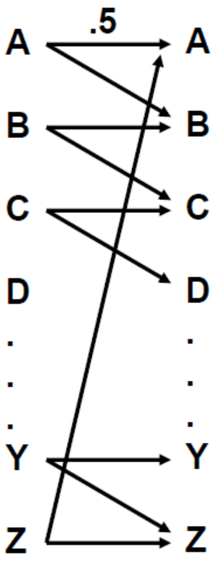
  $= \max_{p(x)} [H(Y) - H(Y|X)]$

  $= \max_{p(x)} H(Y) - 1$

  $= \log_2(26) - 1$

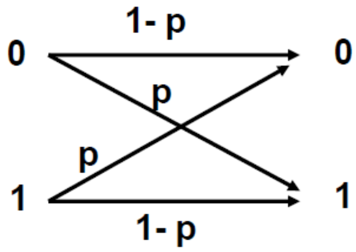
  $= \log_2(26/2)$

  $= \log_2(13)$

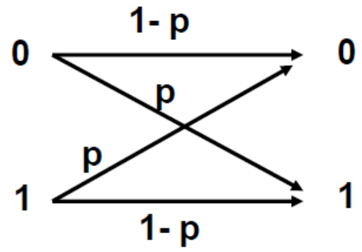


Binary Symmetric Channel

- $I(X; Y) = H(Y) - H(Y|X)$
- $C = ?$



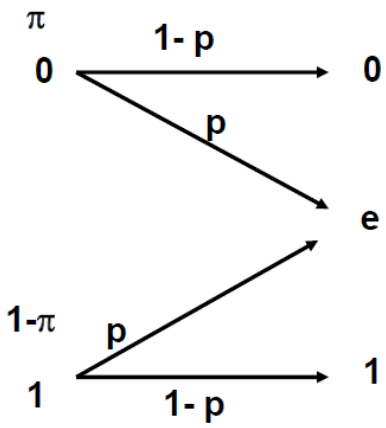
Binary Symmetric Channel (Off)



- $I(X; Y) = H(Y) – H(Y|X)$
  $= H(Y) - \sum p(x) H(Y|X=x)$
  $= H(Y) – H(p)$
  $\leq 1 – H(p)$
- The equality is when $H(Y) = 1$.
- Y is uniform when X is uniform.
- $p(x) \sim$ uniform.
- $C = 1 – H(p)$ bits.

Binary Erasure Channel

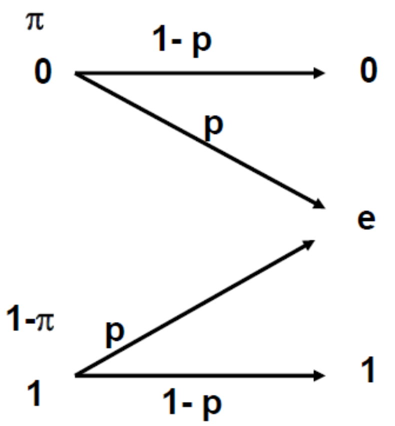
- Some bits are lost (no decision)
- $C = \max_{p(x)} I(X ; Y)$
  $= \max H(Y) - H(Y|X)$
  $= ?$



- Let $E = \{Y = e\}$; $P(E)=p$; $P(E^c)=1-p$
- $H(Y) = H(Y, E) = H(Y) + H(E|Y)$
  $= H(E) + H(Y|E) = H(p)+(1-p)H(\pi)$

Binary Erasure Channel (Off)

- Some bits are lost (no decision)
- $C = \max_{p(x)} I(X ; Y)$
  $= \max H(Y) - H(Y|X)$
  $= \max H(Y) - H(p)$
  $= \max_{\pi} \{H(\pi(1\text{-}p), p, (1\text{-}\pi)(1\text{-}p)) - H(p)\}$
  $= \max_{\pi} (1\text{-}p)H(\pi)$
  $\leq 1 - p$



- Let $E = \{Y = e\}$; $P(E)=p$; $P(E^c)=1\text{-}p$
- $H(Y) = H(Y, E) = H(Y) + H(E|Y)$
  $= H(E) + H(Y|E) = H(p)+(1\text{-}p)H(\pi)$

Binary Erasure Channel

- $C = 1 - p$
- With prob. p, the bit is lost.
- Thus, we could recover at most 1-p percentages of bits.

Symmetric Channels

- p(y | x) = [0.5 0.1 0.4; 0.4 0.5 0.1; 0.1 0.4 0.5]
  - x-th row, y-th column
  - All the rows are permutations of each other and so are the columns

- $I(X ; Y) = H(Y) – H(Y| X)$

  $= H(Y) – H(\text{any row})$

  $\leq \log|\mathcal{Y}| – H(0.1, 0.4, 0.5)$
- The equality, when?
- Thus, the answer is?

The channel is given by

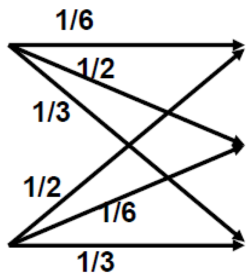
$$p(y|x) = \begin{pmatrix} 0.5 & 0.1 & 0.4 \\ 0.4 & 0.5 & 0.1 \\ 0.1 & 0.4 & 0.5 \end{pmatrix},$$

Or more descriptively by

$p(y \mid x)$	$y = 0$	$x = 1$	$x = 2$
$x = 0$	0.5	0.1	0.4
$x = 1$	0.4	0.5	0.1
$x = 2$	0.1	0.4	0.5

Weakly Symmetric

- p(y | x) = [1/6 1/2 1/3; 1/2 1/6 1/3]
  - Rows are permutation of each other
  - Column sums are equal
- C = log|$\mathcal{Y}$| - H(any row), which is achieved when?

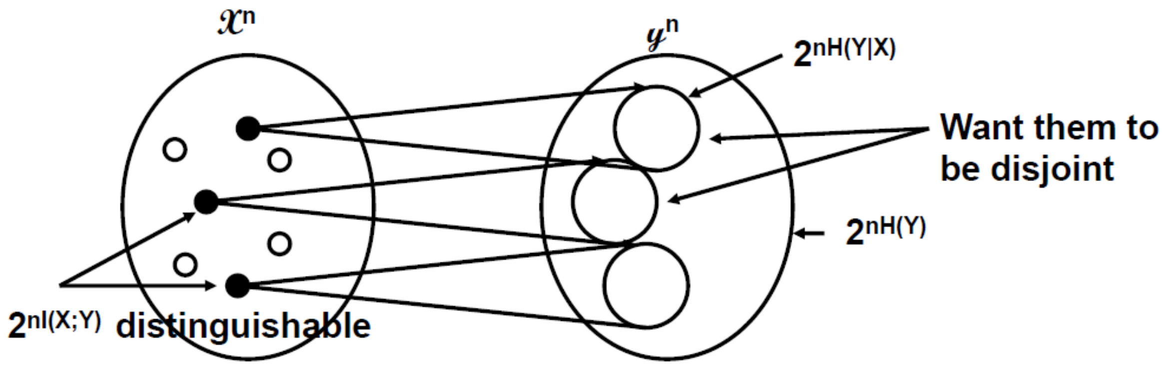


Properties of Channel Capacity

1. $C \geq 0$, since $I(X; Y) \geq 0$
2. $C \leq \log|X|$, why?
3. $C \leq \log|Y|$,
4. $I(X; Y)$ is a continuous function of $p(x)$.
5. $I(X; Y)$ is a concave $\bigcap$ function of $p(x)$ for fixed $p(y|x)$.
   - Or $I(X; Y)$ is a convex $\bigcup$ function of $p(y|x)$ for fixed $p(x)$.

Channel Coding Theorem (Idea)

- Recall the typical set argument
- Recall the noisy type writer problem
- Consider the set of n symbol sequences
- $2^{n[H(Y) - H(Y|X)]} = 2^{n\,I(X;\,Y)}$, **how many distinguishable seq.'es?**



Channel Coding Theorem (Idea) (2)

- The total number of disjoint sets $\leq 2^{nI(X;\,Y)}$
- That's the maximum number of distinguishable sequences you can send at the transmitter while satisfying the close-to-zero decision error requirement.
- What's the role of a large n?

Homework #5

1. Do the set of problems given at the end Hamming code note.
   A. Complete the standard array of the (7, 4) Hamming code.
   B. Do the Bit Flip channel problems.
   C. What is the capacity of the Bit Flip channel? Assume iid channels.
   D. What is the rate of the (7, 4) Hamming code?
   E. What is the length of a codeword? How many channel-uses should you use to send a codeword?
   F. What is the size of the codebook?
   G. What is the length of a message?
2. Cover and Thomas P7.2
3. Cover and Thomas P7.4
4. Cover and Thomas P7.3
5. Cover and Thomas P7.8
6. Cover and Thomas P7.9
7. Cover and Thomas P7.13