

ECE253/CSE208 Introduction to Information Theory

Lecture 10: Channel Capacity

Dr. Yu Zhang

ECE Department

University of California, Santa Cruz

- Chap 7 of *Elements of Information Theory (2nd Edition)* by Thomas Cover & Joy Thomas

Memoryless Discrete Channels (DMC)

We are seeking the answers to the following two core questions in information theory:

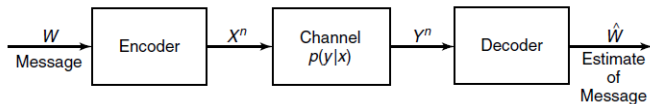
- The fundamental limit of data compression is $H(\mathcal{X})$: Chap 5
- The fundamental limit of data transmission: Chap 7

There is a duality between data compression and transmission. That is, we want to reduce redundancy as much as possible for data compression while adding controlled redundancy to combat errors for data transmission.

Definition

A **discrete channel** to be a system of an input alphabet \mathcal{X} and output alphabet \mathcal{Y} and a probability transition matrix $p(y|x)$ that expresses the probability of observing the output symbol y given that we send the symbol x . The channel is **memoryless** if the probability distribution of the output depends only on the input at the time and is conditionally independent of previous channel inputs or outputs.

Definitions of Channel Capacity



Definition

"Information" channel capacity of a DMC is defined as $C = \max_{p(x)} I(X; Y)$.

Definition

"Operational" channel capacity is the number of bits we can transmit reliably (with an arbitrarily low probability of error) per channel usage.

$C = \log(\text{number of identifiable inputs by passing through the channel with low error})$.

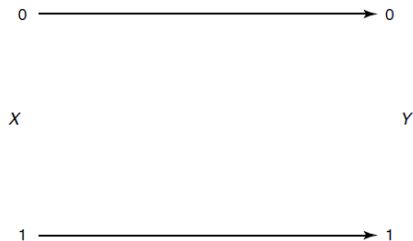
Shannon's second theorem (Noisy-channel coding theorem):

"Information" channel capacity = "operational" channel capacity

Noiseless Binary Channel

Example (Noiseless binary channel (NBC))

Operational-wise, we can transmit one bit without any error per channel use. Hence $C = 1$ bit. Meanwhile, we can calculate the *information* capacity as

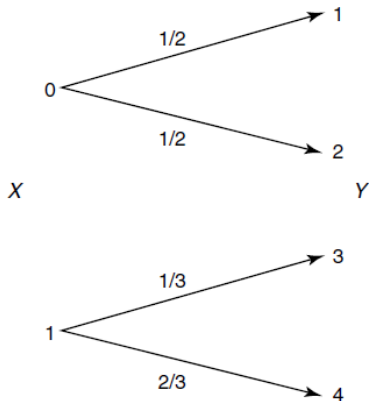


$$\begin{aligned} C &= \max_{p(x)} I(X; Y) \\ &= \max_{p(x)} [H(X) - H(X|Y)] \\ &= \max_{p(x)} H(X) \\ &= 1 \text{ bit} \quad \text{if } p(x) = (1/2, 1/2) \end{aligned}$$

Noisy Channel with Non-overlapping Outputs

Example (Noisy channel with non-overlapping outputs)

Similar to the previous example, we can transmit one bit without any error per channel use. Hence $C = 1$ bit.



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

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

$$= \max_{p(x)} H(X)$$

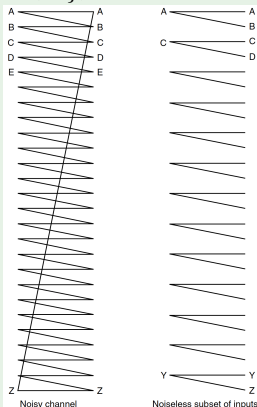
$$= 1 \text{ bit} \quad \text{if } p(x) = (1/2, 1/2)$$

[Exercise: find the solution by an alternative way $C = \max_{p(x)} (H(Y) - H(Y|X)).$]

Noisy Typewriter

Example (Noisy Typewriter)

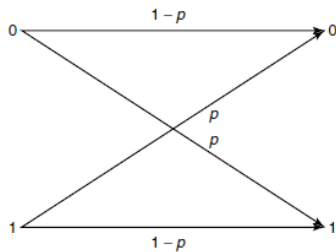
The channel input is received correctly with probability 0.5, or is transformed into the next letter with probability 0.5. Note that if we transmit every other symbols $\{A, C, E, \dots, Z\}$, there will be no error in the communication. Therefore, $C = \log(13)$ bits.



$$\begin{aligned} C &= \max_{p(x)} I(X; Y) \\ &= \max_{p(x)} [H(Y) - H(Y|X)] \\ &= \max_{p(x)} \left[H(Y) - \sum p(x) H(Y|X = x) \right] \\ &= \max_{p(x)} [H(Y) - 1] \\ &= \log(26) - 1 \\ &= \log(13) \text{ bits} \end{aligned}$$

Binary Symmetric Channel

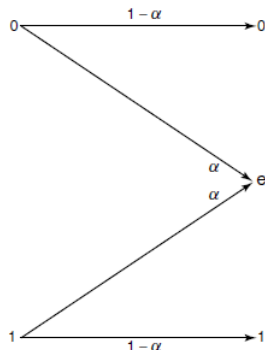
Example (Binary Symmetric Channel (BSC))



$$\begin{aligned} C &= \max_{p(x)} I(X; Y) \\ &= \max_{p(x)} [H(Y) - H(Y|X)] \\ &= \max_{p(x)} [H(Y) - H(p)] \quad H(Y|X = i) = H(p), i = 1, 2 \\ &= 1 - H(p) \text{ bits} \quad \text{if } p(x) = (1/2, 1/2) \Rightarrow p(y) = (1/2, 1/2) \end{aligned}$$

Binary Erasure Channel

Example (Binary Erasure Channel (BEC))



$$\begin{aligned} C &= \max_{p(x)} I(X; Y) \\ &= \max_{p(x)} [H(Y) - H(Y|X)] \\ &= \max_{p(x)} [H(Y) - H(\alpha)] \\ &= \max_{\pi} (1 - \alpha)H(\pi) \\ &= (1 - \alpha) \text{ bits} \quad \text{when } \pi = 1/2 \end{aligned}$$

Define $p(X = 1) := \pi$, we have

$$p(Y = 0) = (1 - \pi)(1 - \alpha), p(Y = 1) = \pi(1 - \alpha), p(Y = e) = (1 - \pi)\alpha + \pi\alpha = \alpha \implies$$

$$H(Y) = H((1 - \pi)(1 - \alpha), \pi(1 - \alpha), \alpha) = H(\alpha) + (1 - \alpha)H(\pi).$$

Symmetric Channel

Example (Symmetric Channel (SC))

Given the channel, i.e., the doubly stochastic matrix collects all conditional probabilities:

$$p(y|x) = \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}.$$

All rows and columns are permutations of each other, the channel $p(y|x)$ is symmetric.

$$\begin{aligned} C &= \max_{p(x)} H(Y) - H(Y|X) \\ &= \max_{p(x)} H(Y) - H(\mathbf{r}) \\ &= \log |\mathcal{Y}| - H(\mathbf{r}) \quad [\mathbf{r} \text{ is any row of the matrix}] \end{aligned}$$

$$\begin{aligned} p(y) &= \sum_x p(y|x)p(x) = \frac{1}{|\mathcal{X}|} \sum_x p(y|x) \\ &= \frac{c}{|\mathcal{X}|} = \frac{1}{|\mathcal{Y}|} \quad [\text{uniform input} \rightarrow \text{uniform output}] \end{aligned}$$

Weakly Symmetric Channel

Example (Weakly Symmetric Channel (WSC))

If all rows are permutations of each other, while the sums of all columns $\sum_x p(y|x)$ are equal (not necessarily equal to 1), then the channel is called *weakly symmetric*; e.g.,

$$p(y|x) = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}.$$

Theorem

For a weakly symmetric channel, we have $C = \log |\mathcal{Y}| - H(\mathbf{r})$. This is achieved by a uniform distribution over \mathcal{X} .

Channel capacity has the following properties:

- Lower and upper limits: $0 \leq C \leq \min(\log |\mathcal{X}|, \log |\mathcal{Y}|)$
- $I(X; Y)$ is a continuous concave function of $p(x)$; see Theorem 2.7.4 in the book.
- $C = \max_{p(x)} I(X; Y)$. Generally, no closed-form solution for this nonlinear optimization problem that can be solved by e.g., 1) KKT conditions; 2) Frank-Wolf gradient search algorithm; and 3) Other iterative algorithms.

Thank You!

Email: <zhangy@ucsc.edu>

Homepage: <https://people.ucsc.edu/~yzhan419/>