Coding and Information Theory
Chapter 5
Using an Unreliable Channel - A

Xuejun Liang

Fall 2022

Quick Review of Last Lecture (1)

- System Entropies for the Binary Symmetric Channel

$$H(\mathcal{B}) \geq H(\mathcal{A})$$

$$H(\mathcal{B} \mid \mathcal{A}) = H(P)$$

$$H(\mathcal{A} \mid \mathcal{B}) \leq H(\mathcal{A})$$

$$H(\mathcal{A} \mid \mathcal{B}) = H(p) + H(P) - H(q).$$

$$H(\mathcal{B} \mid \mathcal{A}) \leq H(\mathcal{B})$$

- Extension of Shannon's First Theorem to Information Channels

$$\frac{L_n}{n} \to H(\mathcal{A} \mid \mathcal{B}) \quad \text{as} \quad n \to \infty$$

- Mutual Information

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) - H(\mathcal{A} \mid \mathcal{B})$$

$$I(\mathcal{B}, \mathcal{A}) = H(\mathcal{B}) - H(\mathcal{B} \mid \mathcal{A})$$

$$I(\mathcal{A}, \mathcal{B}) = I(\mathcal{B}, \mathcal{A})$$

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B})$$

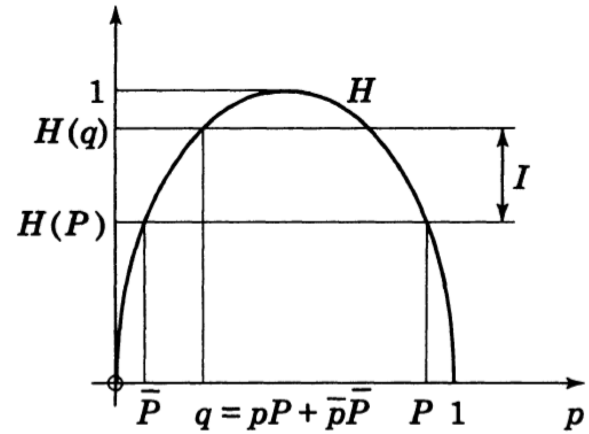
$$I(\mathrm{A}, \mathrm{B}) \geq 0$$

Quick Review of Last Lecture (2)

- Mutual Information for the Binary Symmetric Channel

$$I(\mathcal{A},\mathcal{B}) = H(pP + \overline{p}\overline{P}) - H(P)$$

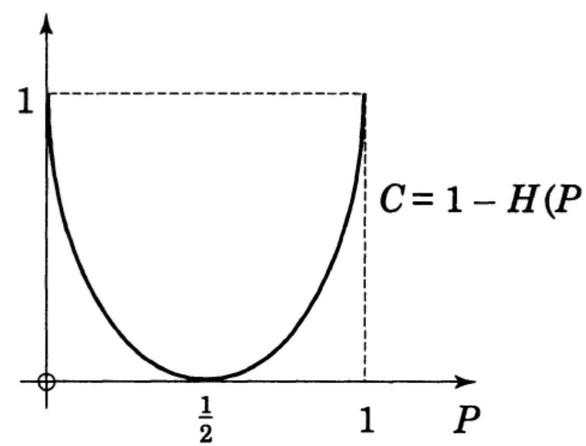
$$0 \leq I(\mathcal{A},\mathcal{B}) \leq 1 - H(P)$$



- Channel Capacity C

$$\mathrm{C} = \max\{I(\mathrm{A},\mathrm{B}) : A\ is\ input\ of\ \Gamma\}$$

  - The BSC has channel capacity **C = 1 - H(P)** attained when the input satisfies p = 1/2



Chapter 5
Using an Unreliable Channel

1. Decision Rules
2. An Example of Improved Reliability
3. Hamming Distance
4. Statement and Outline Proof of Shannon's Theorem
5. The Converse of Shannon's Theorem
6. Comments on Shannon's Theorem

The aim of this chapter

- Shannon’s Fundamental Theorem states that
  - the capacity C of Γ is the least upper bound for the rates at which one can transmit information accurately through Γ.
- We will look at a simple example of how this accurate transmission might be achieved.

5.1 Decision Rules

- A decision rule, or a decoding function $\Delta: B \rightarrow A$
  - $b_j \rightarrow \Delta(b_j) = a_{j^*}$
  - Meaning: receiver sees b_j and decides $a_i = a_{j^*}$ was sent

Example 5.1

Let Γ be the BSC, so that A = B = Z_2. If the receiver trusts this channel, then Δ should be the identity function.

The average probability Pr_C of correct decoding is

$$\mathrm{Pr}_C = \sum_j q_j Q_{j^*j} = \sum_j R_{j^*j} \qquad (5.1)$$

where $\mathrm{Pr}(a = a_{j^*} \mid b = b_j) = Q_{j^*j}$. and $R_{ij} = q_j Q_{ij}$

Decision Rules (Cont.)

- The error probability Pr_E (the average probability of incorrect decoding) is

$$\mathrm{Pr_E} = 1 - \mathrm{Pr_C} = 1 - \sum_j R_{j^*j} = \sum_j \sum_{i \neq j^*} R_{ij} \qquad (5.2)$$

- Ideal observer rule
  - Minimizes Pr_E, or equivalently, which maximizes Pr_C
- How to maximize Pr_C
  - For each j, we choose i = j* to maximize the backward probability $\mathrm{Pr}(a_i | b_j) = Q_{ij}$. Or
  - For each j, we choose i = j* to maximize the joint probability $R_{ij} = q_j Q_{ij}$.

Decision Rules (Cont.)

- Example 5.2
  - Γ is the BSC, compute the Ideal observer rule Δ.

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \overline{p} \end{pmatrix} \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix} = \begin{pmatrix} pP & p\overline{P} \\ \overline{p}\overline{P} & \overline{p}P \end{pmatrix}$$

$$\Delta(0) = \begin{cases} 0 & \text{if } pP > \overline{p}\overline{P} \\ 1 & \text{if } pP < \overline{p}\overline{P}, \end{cases} \quad \text{and} \quad \Delta(1) = \begin{cases} 1 & \text{if } \overline{p}P > p\overline{P} \\ 0 & \text{if } \overline{p}P < p\overline{P}, \end{cases}$$

- A maximum likelihood rule
  - For each j, we choose $i = j^*$ to maximize the forward probability $\Pr(b_j|a_i) = P_{ij}$.

Example 5.3

- Let us apply the maximum likelihood rule Δ to the BSC, where $P > 1/2$ and compute Pr_C and Pr_E. (input probabilities $p, \bar{p}$)

Example 5.4

- For a specific illustration, let us return to Example 4.5, where P = 0.8 and p = 0.9.
- Compare the maximum likelihood rule and the ideal observer rule
  - Maximum likelihood rule
  - Ideal observer rule

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \overline{p} \end{pmatrix} \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix} = \begin{pmatrix} pP & p\overline{P} \\ \overline{p}\overline{P} & \overline{p}P \end{pmatrix}$$

$$= \begin{pmatrix} 0.9 \times 0.8 & 0.9 \times 0.2 \\ 0.1 \times 0.2 & 0.1 \times 0.8 \end{pmatrix} = \begin{pmatrix} 0.72 & 0.18 \\ 0.02 & 0.08 \end{pmatrix}$$

Example 5.5

- Let Γ be the binary erasure channel (BEC) in Example 4.2, with P > 0. Compute the maximum likelihood rule, and compute Pr_C and Pr_E. (input probabilities $p, \bar{p}$)

$$(P_{i,j}) = \begin{pmatrix} P & 0 & \bar{P} \\ 0 & P & \bar{P} \end{pmatrix}$$

$$(R_{i,j}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} \begin{pmatrix} P & 0 & \bar{P} \\ 0 & P & \bar{P} \end{pmatrix} = \begin{pmatrix} pP & 0 & p\bar{P} \\ 0 & \bar{p}P & \bar{p}\bar{P} \end{pmatrix}$$

5.2 An Example of Improved Reliability

- Given an unreliable channel, how can we transmit information through it with greater reliability?
- Considering BSC with 1 > P > $1/2$.
    1) Compute the maximum likelihood rule

    2) Compute the mutual information I(A, B), assuming $p = 1/2$

    3) Compute the error-probability Pr_E

An Example of Improved Reliability (Cont.)

- Now, sending each input symbol a = 0 or 1 three times in succession. So
  - The input consists of two binary words 000 and 111.
  - the output consists of eight binary words 000, 001, 010, 100, 011, 101, 110, and 111.
  - Transmission rate is $1/3$
  - The forward probabilities for this new input and output

	000	001	010	100	011	101	110	111
000	P^3	P^2Q	P^2Q	P^2Q	PQ^2	PQ^2	PQ^2	Q^3
111	Q^3	PQ^2	PQ^2	PQ^2	P^2Q	P^2Q	P^2Q	P^3

  - The maximum likelihood rule, called majority decoding

$$\Delta : \begin{cases} 000, 001, 010, 100 \mapsto 000, \\ 011, 101, 110, 111 \mapsto 111. \end{cases}$$

An Example of Improved Reliability (Cont.)

- The forward probabilities for this new input and output

	000	001	010	100	011	101	110	111
000	P^3	P^2Q	P^2Q	P^2Q	PQ^2	PQ^2	PQ^2	Q^3
111	Q^3	PQ^2	PQ^2	PQ^2	P^2Q	P^2Q	P^2Q	P^3

- The maximum likelihood rule, called majority decoding

$$\Delta : \begin{cases} 000, 001, 010, 100 \mapsto 000, \\ 011, 101, 110, 111 \mapsto 111. \end{cases}$$

- A new binary symmetric channel Γ'

$$M' = \begin{pmatrix} P^3 + 3P^2Q & 3PQ^2 + Q^3 \\ 3PQ^2 + Q^3 & P^3 + 3P^2Q \end{pmatrix}$$

$$\begin{matrix} 0 \\ 1 \end{matrix} \longrightarrow \begin{matrix} 000 \\ 111 \end{matrix} \longrightarrow \Gamma \longrightarrow \begin{matrix} 000 \\ 001 \\ 010 \\ 100 \\ 011 \\ 101 \\ 110 \\ 111 \end{matrix} \longrightarrow \begin{matrix} 0 \\ 1 \end{matrix}$$

$$\Pr_C = P^3 + 3P^2Q$$

$$\Pr_E = 3PQ^2 + Q^3 = Q^2(3 - 2Q) \approx 3Q^2$$

Generalized Idea

- If Γ is a channel with an input A having an alphabet A of r symbols, then any subset $C \subseteq A^n$ can be used as a set of code-words which are transmitted through Γ
  - For instance, the repetition code R^n over A consists of all the words w = $aa \ldots a$ of length n such that $a \in A$.
  - In this case, $|C| = r = r^1$. So the rate is $1/n$.
  - In general, $|C| = r^k$. So the rate is k/n.
- The transmission rate can be defined as

$$R = \frac{\log_r |\mathcal{C}|}{n} \qquad (5.3)$$