Coding and Information Theory
Chapter 4
Information Channels - C

Xuejun Liang

Fall 2022

Chapter 4: Information Channels

1. Notation and Definitions
2. The Binary Symmetric Channel
3. System Entropies
4. System Entropies for the Binary Symmetric Channel
5. Extension of Shannon's First Theorem to Information Channels
6. Mutual Information
7. Mutual Information for the Binary Symmetric Channel
8. Channel Capacity

Quick Review of Last Lecture (1)

- The Binary Symmetric Channel
  - The channel relationships for BSC
  - Bayes' formula for BSC
  - Examples
- System Entropies
  - $H(\mathrm{A}), H(\mathrm{B}), H(\mathrm{A}|\mathrm{B}), H(\mathrm{B}|\mathrm{A}), and\ H(\mathrm{A}, \mathrm{B})$

$$H(\mathcal{A} \mid \mathcal{B}) = \sum_i \sum_j R_{ij} \log \frac{1}{Q_{ij}}$$

$$H(\mathcal{B} \mid \mathcal{A}) = \sum_i \sum_j R_{ij} \log \frac{1}{P_{ij}}$$

$$H(\mathcal{A}, \mathcal{B}) = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}}$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) \quad (4.5)$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) \quad (4.6)$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}) \quad (4.7)$$

Quick Review of Last Lecture (2)

- System Entropies for BSC

$$H(\mathcal{A}) = -p\log p - \bar{p}\log\bar{p} = H(p),$$
$$H(\mathcal{B}) = -q\log q - \bar{q}\log\bar{q} = H(q),$$

H(p) is strictly convex function

$$\boxed{H(pP + \bar{p}\bar{P}) \geq PH(p) + \bar{P}H(\bar{p})}$$

$$\boxed{H(pP + \bar{p}\bar{P}) \geq pH(P) + \bar{p}H(\bar{P})}$$

$$H(\mathcal{B}) \geq H(\mathcal{A}), \qquad (4.8)$$

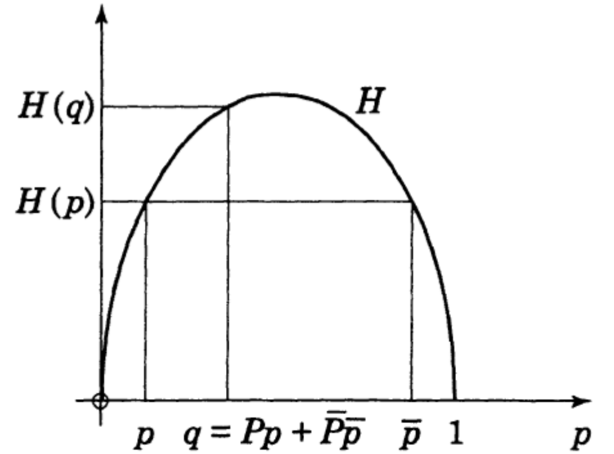


Figure 4.5

For the BSC, we have $H(\mathcal{B} \mid \mathcal{A}) = H(P)$

- the sender's uncertainty about the output is equal to the uncertainty as to whether symbols are transmitted correctly

- The equivocation for the BSC is

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

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

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) \qquad (4.6)$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}) \qquad (4.7)$$

- The BSC satisfies

$$H(\mathcal{B} \mid \mathcal{A}) \leq H(\mathcal{B}), \quad (4.9)$$

the uncertainty about B generally decreases when A is known

$$H(\mathcal{A} \mid \mathcal{B}) \leq H(\mathcal{A}), \quad (4.10)$$

the uncertainty about A generally decreases when B is known

with equality if and only if $P = 1/2$ or p = 0, 1.

$$\boxed{H(pP + \bar{p}\bar{P}) \geq pH(P) + \bar{p}H(\bar{P})}$$

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

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

4.5 Extension of Shannon's First Theorem to Information Channels

- Extension of Shannon's First Theorem
  - The greatest lower bound of the average word-lengths of uniquely decodable encodings of the input A of a channel, given knowledge of its output B, is equal to the equivocation $H(\mathrm{A}|\mathrm{B})$.
- Interpretation
  - the receiver knows B but is uncertain about A; the extra information needed to be certain about A is the equivocation $H(\mathrm{A}|\mathrm{B})$, and
  - this is equal to the least average word-length required to supply that extra information (by some other means, separate from Γ).

Extension of Shannon's First Theorem

- Theorem 4.8
  - If the output B of a channel is known, then by encoding A^n with n sufficiently large, one can find uniquely decodable encodings of the input A with average word-lengths arbitrarily close to the equivocation $H(A|B)$.

$$H(\mathcal{A} \mid b_j) \le L_{(j)} \le 1 + H(\mathcal{A} \mid b_j)$$

$$H(\mathcal{A} \mid \mathcal{B}) \le L \le 1 + H(\mathcal{A} \mid \mathcal{B})$$

$$H(\mathcal{A}^n \mid \mathcal{B}^n) = nH(\mathcal{A} \mid \mathcal{B})$$

$$H(\mathcal{A}^n \mid \mathcal{B}^n) \le L_n \le 1 + H(\mathcal{A}^n \mid \mathcal{B}^n)$$

$$nH(\mathcal{A} \mid \mathcal{B}) \le L_n \le 1 + nH(\mathcal{A} \mid \mathcal{B})$$

$$H(\mathcal{A} \mid \mathcal{B}) \le \frac{L_n}{n} \le \frac{1}{n} + H(\mathcal{A} \mid \mathcal{B})$$

use Shannon-Fano coding of extensions A^n of A

use the conditional probabilities $\Pr(a_i|b_j)$ for A

$$A^n \longrightarrow \Gamma^n \longrightarrow B^n$$

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

4.6 Mutual Information

- If Γ is a channel with input A and output B, then the entropy H(A) of A has three equivalent interpretations:
  1. it is the uncertainty about A when B is unknown;
  2. it is the information conveyed by A when B is unknown;
  3. it is the average word-length needed to encode A when B is unknown.
- Similarly, the equivocation H($\mathrm{A}|\mathrm{B}$) has three equivalent interpretations:
  1. it is the uncertainty about A when B is known;
  2. it is the information conveyed by A when B is known;
  3. it is the average word-length needed to encode A when B is known.

Mutual Information (Cont.)

- The mutual information is defined as the difference between these two numbers:

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

- This also has three equivalent interpretations:
  1. it is the amount of uncertainty about A resolved by knowing B;
  2. it is the amount of information about A conveyed by B;
  3. it is the average number of symbols, in the code-words for A, which refer to B.

$I(\mathrm{A}, \mathrm{B})$ represents how much information A and B have in common

Examples

- Example 4.9
  - For a rather frivolous example, let Γ be a film company, A a book, and B the resulting film of the book. Then $I(\mathrm{A}, \mathrm{B})$ represents how much the film tells you about the book.
- Example 4.10
  - Let A be a lecture, Γ a student taking notes, and B the resulting set of lecture notes. Then $I(\mathrm{A}, \mathrm{B})$ measures how accurately the notes record the lecture.

Mutual Information (Cont.)

- Interchanging the roles of A and B, we can define

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

- We have

$$I(\mathcal{A}, \mathcal{B}) = I(\mathcal{B}, \mathcal{A}) \quad (4.15)$$

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}) \quad (4.16)$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) \quad (4.6)$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A} \mid \mathcal{B}) \quad (4.7)$$

- Theorem 4.11
  - For every channel Γ we have $I(\mathrm{A}, \mathrm{B}) \geq 0$, with equality if and only if the input A and the output B are statistically independent.

Corollary 3.9

$$\boxed{\sum_{i=1}^{q} x_i \log_r \frac{1}{x_i} \leq \sum_{i=1}^{q} x_i \log_r \frac{1}{y_i}\,,}$$

$$\begin{aligned} p_i &= \textstyle\sum_j R_{ij} \\ q_j &= \textstyle\sum_i R_{ij} \\ \textstyle\sum_i \sum_j R_{ij} &= \textstyle\sum_i \sum_j p_i q_j = 1 \end{aligned}$$

- Corollary 4.12
  - For every channel Γ we have

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

  - in each case, there is equality if and only if the input A and the output B are statistically independent.

4.7 Mutual Information for the Binary Symmetric Channel

- Let us take the channel Γ to be the BSC, we have

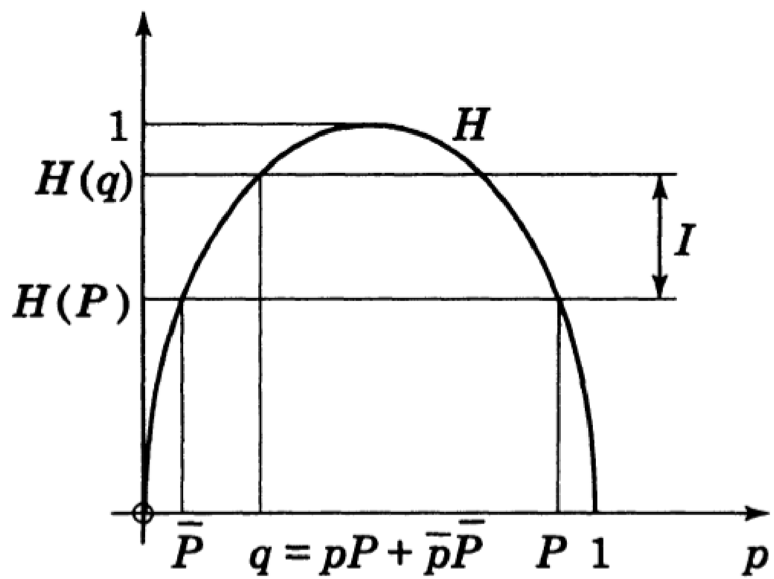
$$I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B} \mid \mathcal{A})$$

$H(\mathcal{B}) = H(q)$ and $H(\mathcal{B} \mid \mathcal{A}) = H(P)$ where $q = pP + \overline{p}\overline{P}$

- So that

$$\begin{aligned} &I(\mathcal{A},\mathcal{B}) \\ &= H(q) - H(P) \\ &= H(pP + \overline{p}\overline{P}) - H(P) \end{aligned}$$

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



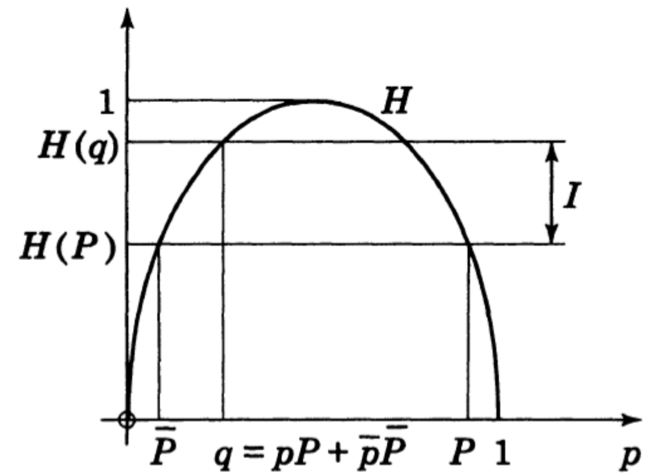
4.8 Channel Capacity

- The mutual information $I(\mathrm{A}, \mathrm{B})$ for a channel Γ represents how much of the information in the input A is emerging in the output B.
  - This depends on both Γ and A
- The capacity C of a channel Γ is defined to be the maximum value of the mutual information $I(\mathrm{A}, \mathrm{B})$, where A ranges over all possible inputs for Γ.
  - This depends on Γ alone, represents the maximum amount of information which the channel can transmit

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

Example 4.13

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



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

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

- Figure shows C as a function of P
  - C is greatest when P is 0 or 1
  - C is least when P = 1/2