Coding and Information Theory
Chapter 4
Information Channels

Xuejun Liang

2019 Fall

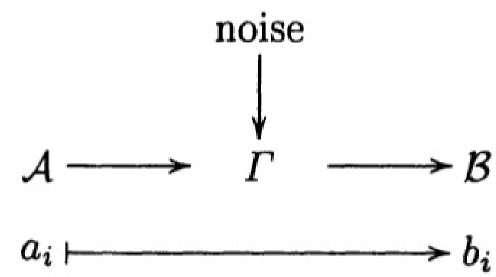
Chapter 4: Information Channels

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

The aim of this chapter

- We Consider
  - a source sending messages through an unreliable (or noisy) channel to a receiver
- Our aim here is
  - to measure how much information is transmitted, and how much is lost in this process, using several different variations of the entropy function, and then
  - to relate this to the average word-length of the code used.

4.1 Notation and Definitions



- Information channel Γ
- Input of Γ: Source A,
  - with finite alphabet A of symbols $a = a_1, \dots, a_r$, having probabilities

$$p_i = \Pr(a = a_i)$$ where

$$0 \le p_i \le 1 \quad \text{and} \quad \sum_{i=1}^{r} p_i = 1$$

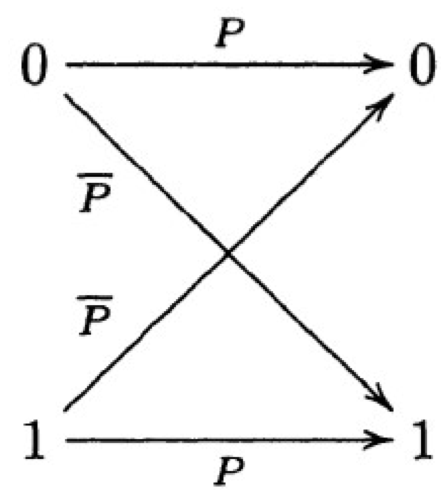
- Output of Γ: Source B,
  - with a finite alphabet B of symbols $b = b_1, \dots, b_s$, having probabilities

$$q_j = \Pr(b = b_j)$$ where

$$0 \le q_j \le 1 \quad \text{and} \quad \sum_{j=1}^{s} q_j = 1$$

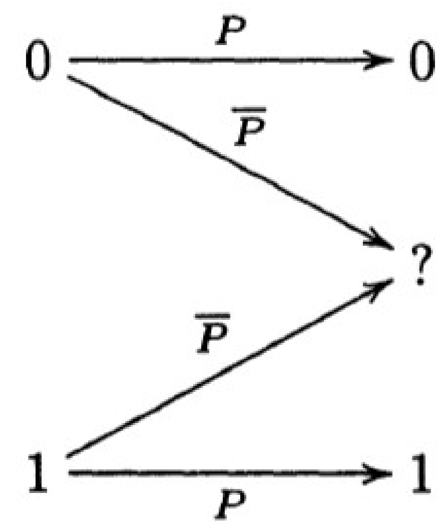
Example 4.1

- Binary symmetric channel (BSC)
  - $\mathrm{A} = \mathrm{B} = Z_2 = \{0, 1\}$.
  - Each input symbol $a = 0$ or 1 is correctly transmitted with probability P, and is incorrectly transmitted (as $\bar{a} = 1 - a$) with probability $\bar{P} = 1 - P$, for some constant P $(0 \leq P \leq 1)$.



Example 4.2

- Binary erasure channel (BEC)
  - $\mathrm{A} = Z_2 = \{0, 1\}$.
  - $\mathrm{B} = \{0, 1, ?\}$.
  - Each input symbol $a = 0$ or 1 is correctly transmitted with probability P, and is erased (or made illegible) with probability $\bar{P}$, indicated by an output symbol $b = ?$



Forward Probabilities

- Forward probabilities of Γ

$$P_{ij} = \Pr(b = b_j \mid a = a_i) = \Pr(b_j \mid a_i)$$

- We have $$\sum_{j=1}^{s} P_{ij} = 1$$

- The channel matrix $$M = (P_{ij}) = \begin{pmatrix} P_{11} & \dots & P_{1s} \\ \vdots & & \vdots \\ P_{r1} & \dots & P_{rs} \end{pmatrix}$$

- For instance, if Γ is the BSC or BEC we have $$M = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} P & 0 & \overline{P} \\ 0 & P & \overline{P} \end{pmatrix}$$

Combining two channels

- **Sum** $\Gamma + \Gamma'$
  - If Γ and Γ' have disjoint input alphabets A and A', and disjoint output alphabets B and B', then the **sum** $\Gamma + \Gamma'$ has input and output alphabets $A \cup A'$ and $B \cup B'$.
  - Each input symbol is transmitted through Γ or Γ', so the channel matrix is a block matrix

$$\begin{pmatrix} M & O \\ O & M' \end{pmatrix}$$

where M and M' are the channel matrices for Γ and Γ'.

Combining two channels

- **Product** $\Gamma \times \Gamma'$
  - The input and output alphabets are A x A' and B x B’
  - The sender transmits a pair $(a, a') \in$ A x A' by simultaneously sending a through Γ and a' through Γ'
  - A pair $(b, b') \in$ B X B' is received
  - Thus the forward probabilities are

$$\Pr((b, b') \mid (a, a')) = \Pr(b \mid a) \,.\, \Pr(b' \mid a')$$

  - So the channel matrix is the **Kronecker product** $M \otimes M'$ of the matrices M and M' for Γ and Γ'.
    - if $M = (P_{ij})$ and $M' = (P'_{kl})$ are $r \times s$ and $r' \times s'$ matrices, then $M \otimes M'$ is an $rr' \times ss'$ matrix, with entries $P_{ij}P'_{kl}$

Example

- If Γ and Γ' are binary symmetric channels, with channel matrices

$$M = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix} \quad \text{and} \quad M' = \begin{pmatrix} P' & \overline{P'} \\ \overline{P'} & P' \end{pmatrix}$$

- then $\Gamma + \Gamma'$ and $\Gamma \times \Gamma'$ have channel matrices

$$\begin{array}{c} \\ 0 \\ 1 \\ 0' \\ 1' \end{array} \begin{array}{c} \begin{array}{cccc} 0, & 1, & 0', & 1' \end{array} \\ \begin{pmatrix} P & \overline{P} & 0 & 0 \\ \overline{P} & P & 0 & 0 \\ 0 & 0 & P' & \overline{P'} \\ 0 & 0 & \overline{P'} & P' \end{pmatrix} \end{array} \quad \text{and} \quad \begin{array}{c} \begin{array}{cccc} (0,0'), & (1,0'), & (0,1'), & (1,1') \end{array} \\ \begin{pmatrix} PP' & \overline{P}P' & P\overline{P'} & \overline{P}\,\overline{P'} \\ \overline{P}P' & PP' & \overline{P}\,\overline{P'} & P\overline{P'} \\ P\overline{P'} & \overline{P}\,\overline{P'} & PP' & \overline{P}P' \\ \overline{P}\,\overline{P'} & P\overline{P'} & \overline{P}P' & PP' \end{pmatrix} \end{array} \begin{array}{c} \\ (0,0') \\ (1,0') \\ (0,1') \\ (1,1') \end{array}$$

The channel relationships

- The channel relationships

$$\sum_{i=1}^{r} p_i P_{ij} = q_j \qquad (4.2)$$

Where $p_i = \Pr(a = a_i)$, $q_j = \Pr(b = b_j)$ and

$$P_{ij} = \Pr\left(b = b_j \middle| a = a_i\right) = \Pr(b_j|a_i)$$

(4.2) can be written as $\mathbf{p}M = \mathbf{q}\,.\quad (4.2')$

- The backward probabilities

$$Q_{ij} = \Pr\left(a = a_i \mid b = b_j\right) = \Pr\left(a_i \mid b_j\right)$$

- The joint probabilities

$$R_{ij} = \Pr\left(a = a_i \text{ and } b = b_j\right) = \Pr\left(a_i, b_j\right)$$

Bayes' Formula

- **Bayes' Formula**

$$Q_{ij} = \frac{p_i}{q_j} P_{ij} \qquad (4.3)$$

provided $q_j \neq 0$.

- Combining this with (4.2) we get

$$Q_{ij} = \frac{p_i P_{ij}}{\sum_{k=1}^{r} p_k P_{kj}} \qquad (4.4)$$

4.2 The Binary Symmetric Channel

- Binary symmetric channel (BSC)
  - $\mathrm{A} = \mathrm{B} = Z_2 = \{0, 1\}$.
  - the channel matrix has the form

$$M = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix}$$

    for some P where $0 \leq P \leq 1$
  - The input probabilities have the form

$$p_0 = \Pr(a = 0) = p,$$

$$p_1 = \Pr(a = 1) = \overline{p},$$

    for some p such that $0 \leq p \leq 1$
  - The channel relationships = ? And Bayes' formula = ?

Examples

- Example 4.4
  - Let the input A be defined by putting $p = 1/2$
  - Probabilities of the output symbols: q_0 = ? And q_1 = ?
  - The backward probabilities: Q_{00}, Q_{01}, Q_{10}, Q_{11}, = ?
- Example 4.5
  - Suppose that P = 0.8 and p = 0.9
  - Probabilities of the output symbols: q_0 = ? And q_1 = ?
  - The backward probabilities: Q_{00}, Q_{01}, Q_{10}, Q_{11}, = ?
  - Necessary and sufficient conditions on p and P for

$$Q_{00} > Q_{10} \text{ and } Q_{01} > Q_{11}$$

4.3 System Entropies

- The input A and the output B of a channel Γ
  - the input entropy $H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i}$
  - the output entropy $H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$
  - If $b = b_j$ is received, there is a conditional entropy

$$H(\mathcal{A} \mid b_j) = \sum_i \Pr(a_i \mid b_j) \log \frac{1}{\Pr(a_i \mid b_j)} = \sum_i Q_{ij} \log \frac{1}{Q_{ij}}$$

  - the equivocation (of A with respect to B)

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

System Entropies (Cont.)

- Similarly, if a_i is sent then the uncertainty about B is the conditional entropy

$$H(\mathcal{B} \mid a_i) = \sum_j \Pr(b_j \mid a_i) \log \frac{1}{\Pr(b_j \mid a_i)} = \sum_j P_{ij} \log \frac{1}{P_{ij}}$$

- the equivocation of B with respect to A

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

- the joint entropy

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

System Entropies (Cont.)

- If A and B are statistically independent, then

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

- In general, A and B are related, rather than independent, then

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

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

- We call $H(\mathrm{A}), H(\mathrm{B}), H(\mathrm{A}|\mathrm{B}), H(\mathrm{B}|\mathrm{A}), and\ H(\mathrm{A},\mathrm{B})$ the system entropies.

4.4 System Entropies for the Binary Symmetric Channel

- The input and output entropies for BSC are

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

  where $q = pP + \bar{p}\bar{P}$.

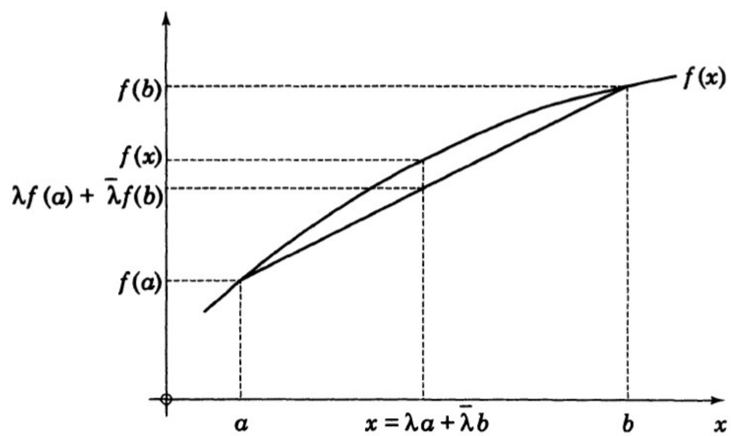
- Definition: A function f: [0,1] $\rightarrow R$ is strictly convex, if for $a, b \in [0,1]$ $and\ x = \lambda a + \bar{\lambda} b\ with\ 0 \leq \lambda \leq 1$,

$$f(x) \geq \lambda f(a) + \overline{\lambda} f(b)$$

  with equality if and only if x = a or b, that is, $a = b$ or $\lambda = 0\ or\ 1$.

System Entropies for BSC (Cont.)

- Lemma 4.6
  - If a function f : [0,1] $\rightarrow R$ is continuous on the interval [0,1] and twice differentiable on (0,1) , with f"(x) < 0 for all $x \in$ (0,1), then f is strictly convex.
- Corollary 4.7
  - The entropy function $H(p)$ is strictly convex on [0,1].



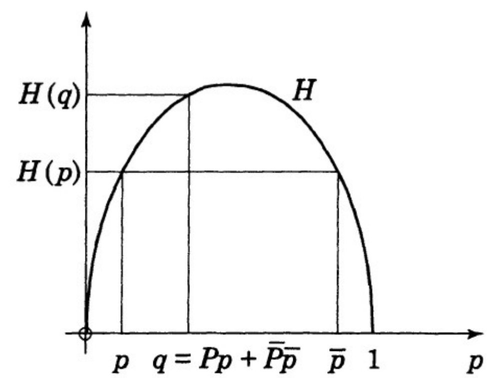


Figure 4.5

System Entropies for BSC (Cont.)

- The BSC satisfies

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

  with equality if and only if $p = 1/2$ or the channel is totally unreliable (P = 0) or reliable (P = 1)

Transmission through the BSC generally increases uncertainty

Note in BSC, $q = pP + \bar{p}\bar{P}$

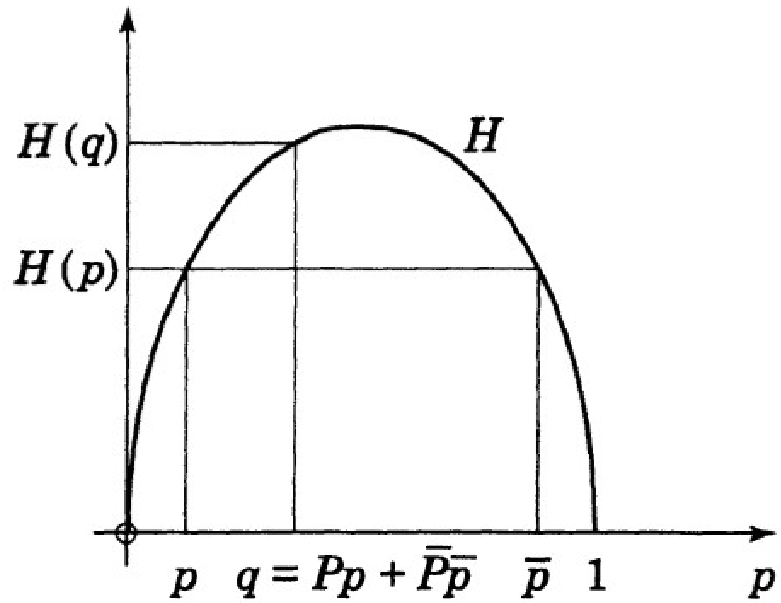


Figure 4.5

System Entropies for BSC (Cont.)

- For the BSC we have

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

- The equivocation for the BSC is

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

- The BSC satisfies

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

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

the uncertainty about B generally decreases when A is known

the uncertainty about A generally decreases when B is known

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

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 $H(\mathrm{A}|\mathrm{B})$.

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

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

Mutual Information (Cont.)

- We have

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

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

- 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.

Mutual Information (Cont.)

- 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

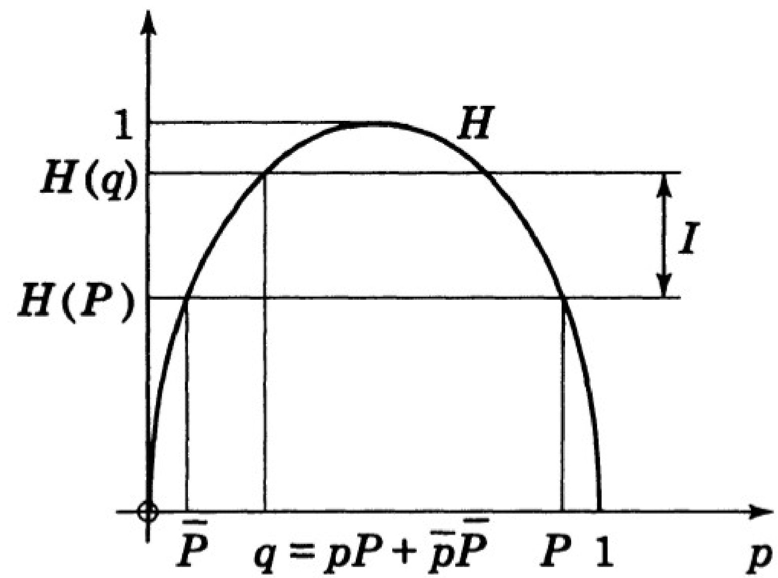
$$I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B} \mid \mathcal{A})$$

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

- 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

Channel Capacity (Cont.)

- Example 4.13
  - We saw that the BSC has channel capacity C = 1 - H(P), attained when the input satisfies p = 1/2.
  - Figure shows C as a function of P
    - C is greatest when P is 0 or 1
    - C is least when P = 1/2