Information Theory and Channel Coding

Prof. Rodrigo C. de Lamare
CETUC, PUC-Rio, Brazil
delamare@cetuc.puc-rio.br

Textbooks and Assessment

- Textbooks:
  - Thomas Cover and Joy Thomas, Elements of Information Theory, 2nd Edition, 2006.
  - Jorge Moreira and Patrick Farrell, Essentials of Error-Control Coding, 2006, Wiley.

- Assessment:
  - 1 Exam papers (E) on Information Theory.
  - 1 Project (P) on a topic on Channel Coding chosen by the student in agreement with the lecturer.
  - 8 Lists of exercises (LE) on all topics.
  - Final grade (FG) = (E + P + LE)/3

Syllabus

Part I: Information Theory

I. Introduction
   - Fundamental limits
   - Uncertainty, information and entropy

II. Source coding
   - Prefix coding
   - Huffman coding
   - Lempel-Ziv coding
   - Quantisation

III. Channel capacity
   - Continuous and discrete channels
   - Mutual information
   - Channel capacity theorem

IV. Channel coding theorem

V. Rate distortion theory

I. Introduction

- Information theory deals with the mathematical analysis and modelling of communications systems.

- It is a branch of probability theory that focuses on abstract models and the derivation of bounds and inequalities.

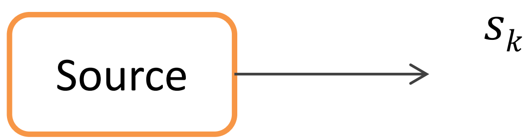
- In this course, we focus on the following fundamental limits:
  - Compression
  - Transmission
  - Reliability

A. Fundamental limits

- Compression → Entropy (H) of a source
- Transmission → Channel capacity (C)
- Reliability → If $H \leq C$ then a system can communicate without errors

B. Uncertainty, information and entropy

- Consider a discrete source illustrated by



- Let us define the alphabet of this source as

$\xi = \{s_0, s_1, \dots, s_{K-1}\}$ with probabilities $P(s = s_k) = p_k, k = 0,1,\dots,K-1$

- This set of probabilities must satisfy the following

$$\sum_{k=0}^{K-1} p_k = 1$$

discrete memoryless source

Information

- Given an event $s = s_k$ which occurs with probability p_k, the amount of information is given by

$$I(s_k) = \log\left(\frac{1}{p_k}\right)$$

Information (continued)

- The information using a base-2 logarithm is defined by

$$
\begin{aligned}
I(s_k) &= \log_2\left(\frac{1}{p_k}\right) \text{ bits} \\
&= -\log_2(p_k), \quad \text{for } k = 0, 1, \cdots, K-1
\end{aligned}
$$

Example 1

Compute the amount of information when $p_k = ½$

$$I(s_k) = \log_2\left(\frac{1}{p_k}\right) = -\log_2(p_k)$$

$$= -\log_2(½) = 1 \text{ bit}$$

Properties

i) $I(s_k) = 0$ when $p_k = 1$

ii) $I(s_k) \geq 0$ for $0 \leq p_k \leq 1$

iii) $I(s_k) > I(si)$ for $p_k < p_i$

iv) $I(s_k s_i) = I(s_k) + I(si)$ if s_k and s_i are statistically independent

Entropy

- The entropy is defined as the average information content measure per symbol of a source as described by

$$H(\xi) = E[I(s_k)] = \sum_{k=0}^{K-1} p_k I(s_k)$$

$$= \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) \qquad \text{bits,}$$

where ξ is the alphabet

$I(s_k)$ is the information content and

p_k is the symbol probability for $k = 0,1,\ldots,K-1$

and the entropy depends only on the probabilities of the symbols in ξ.

C. Properties of Entropy

- Consider the discrete memoryless source previously defined.
- The entropy $H(\xi)$ of this source is bounded as follows:

$$0 \leq \mathrm{H}(\xi) \leq \log_2 \mathrm{K}$$

- Properties:

i) $H(\xi) = 0$ if and only if $p_k = 1$ (no uncertainty)

ii) $H(\xi) = \log_2 K$ if and only if $p_k = \frac{1}{K}$ for all k (maximum uncertainty)

Proof

- Because $p_k \leq 1$ each term of $p_k \log_2(1/p_k)$ in $H(\xi)$ is always non negative. Therefore, we have

$$H(\xi) \geq 0$$

- Noting that $p_k \log_2(1/p_k) = 0$ if $p_k = 0$ or $p_k = 1$ we have that $H(\xi) = 0$ if $p_k = 1$ for any k and the remaining probabilities $p_j = 0$, for $j \neq k$.

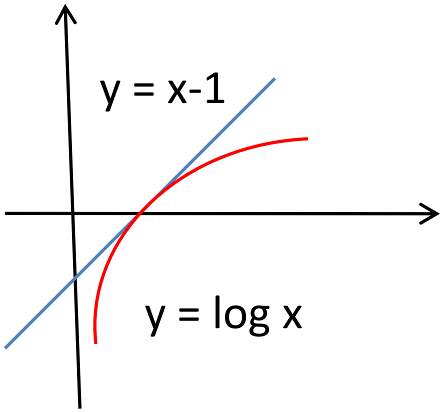
- This completes the first part of the proof that

$$H(\xi) \geq 0$$

Proof (continued)

- In order to obtain the upper bound $H(\xi) \leq \log_2 K$ we adopt the following strategy:

- Consider the inequality $\log x \leq x - 1, \; x \geq 0$



- Consider 2 probability distributions given by $\{p_0, p_1, \dots, p_{K\,-\,1}\}$ and $\{q_0, q_1, \dots, q_{K\,-\,1}\}$ and the alphabet $\xi = \{s_0, s_1, \dots, s_{K\,-\,1}\}$ of a discrete memoryless source.

Proof (continued)

- We can then write the following expression using the natural logarithm:

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) = \frac{1}{\log_2} \sum_{k=0}^{K-1} p_k \log\left(\frac{q_k}{p_k}\right)$$

- Using the inequality $\log x \leq x - 1$, we have

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) \leq \frac{1}{\log_2} \sum_{k=0}^{K-1} p_k \left(\frac{q_k}{p_k} - 1\right)$$

$$\leq \frac{1}{\log_2} \sum_{k=0}^{K-1} q_k - p_k$$

$$\leq \frac{1}{\log_2} \sum_{k=0}^{K-1} q_k - \sum_{k=0}^{K-1} p_k = 0$$

Proof (continued)

- Therefore, we have

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) \le 0$$

- For $q_k = p_k = 0$ we have

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) = 0$$

- For $q_k = 1/K$, k =0,1, ..., K-1 (equiprobable symbols) we have

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) \le \log_2 K$$

Proof (continued)

- Since the entropy of a discrete memoryless source with equiprobable symbols is given by

$$H(\xi)=\sum_{k=0}^{K-1} q_k \log_2\left(\frac{1}{q_k}\right)=\sum_{k=0}^{K-1}\frac{1}{K}\log_2 K=\log_2 K$$

- We conclude the second part of the proof which states that

$$H(\xi) \leq \log_2 K$$

- The equality is only obtained for equiprobable symbols.

Example 2: Entropy of a binary memoryless source

Consider a binary memoryless source with symbols s_k, $k = 0{,}1$ and symbol probabilities p_0 and $p_1 = 1 - p_0$

a) Compute the entropy of the source

$$H(\xi) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) = p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right)$$

$$= -p_0 \log_2 p_0 - p_1 \log_2 p_1 = -p_0 \log_2 p_0 - (1-p_0)\log_2(1-p_0) \quad \text{bits}$$

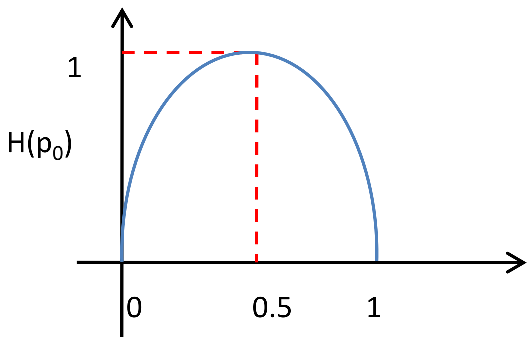
$$H(p_0) = -p_0 \log_2 p_0 - (1-p_0)\log_2(1-p_0)$$

b) Plot $H(p_0)$

when $p_0 = 0 \rightarrow H(\xi) = 0$ ($x \log x \rightarrow 0$ when $x \rightarrow 0$)

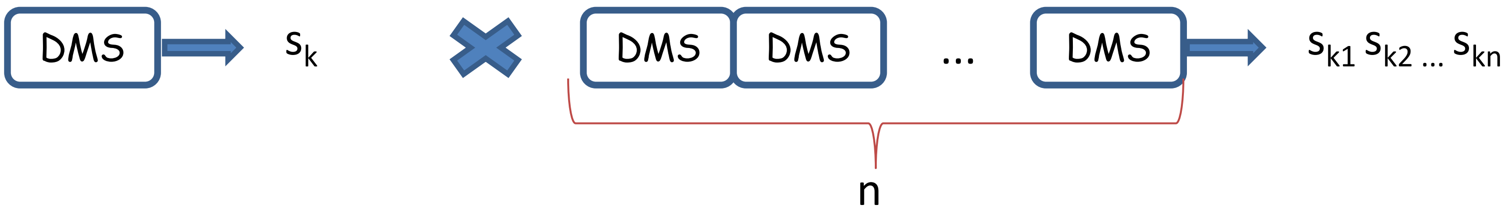
when $p_0 = 1 \rightarrow H(\xi) = 0$

$H(\xi)$ is maximum ($H(\xi) = 1$) when $p_0 = p_1 = 1/2$



D. Extension of a Discrete Memoryless Source

- In information theory, it is useful to consider blocks rather individual symbols:



- Consider a block with n successive symbols and K the number of distinct symbols in the source alphabet ξ of the original source.

- We can view each block as being produced by an extended source with a source alphabet ξ^n that has K^n distinct blocks.

- The entropy of the extended source is given by

$$H(\xi^n) = n\,H(\xi)$$

Joint entropy

- Consider a block of symbols organized in the vector

$$\boldsymbol{s} = [s_1 \quad s_2 \quad \cdots \quad s_n]^T$$

- The joint entropy is given by

$$H(\boldsymbol{s}) = \sum p(\boldsymbol{s}) \log \frac{1}{p(\boldsymbol{s})}$$

$$= \sum_{i_1 \quad i_2 \quad \dots \quad i_n} p(s_1 \quad s_2 \quad \cdots \quad s_n) \log \frac{1}{p(s_1 \quad s_2 \quad \cdots \quad s_n)}$$

- Since we are dealing with a DMS the source symbols are statistically independent.

- Hence, the probabilities involved in $p(\boldsymbol{s})$ are decoupled and $p(\boldsymbol{s})$ is equal to the product of the probabilities

$$p(\boldsymbol{s}) = p(s_1)p(s_2) \dots p(s_n)$$

- Therefore, the joint entropy can be rewritten as

$$H(\boldsymbol{s}) = \sum p(\boldsymbol{s}) \log \frac{1}{p(\boldsymbol{s})} = H(\xi^n)$$

$$= \sum_{i_1 \quad i_2 \quad \dots \quad i_n} p(s_1 \quad s_2 \quad \cdots \quad s_n) \log \frac{1}{p(s_1 \quad s_2 \quad \cdots \quad s_n)}$$

$$= \sum_{i_1 \quad i_2 \quad \dots \quad i_n} p(s_1)p(s_2) \dots p(s_n) \log \left(\frac{1}{p(s_1)p(s_2)\dots p(s_n)}\right)$$

$$= \sum_{i_1} p(s_{i_1}) \log \left(\frac{1}{p(s_{i_1})}\right) \sum_{i_2} p(s_{i_2}) \log \left(\frac{1}{p(s_{i_2})}\right) \dots \sum_{i_n} p(s_{i_n}) \log \left(\frac{1}{p(s_{i_n})}\right)$$

$$= n \sum_{i_1} p(s_{i_1}) \log \left(\frac{1}{p(s_{i_1})}\right) = \mathrm{nH}(\xi)$$

- Thus, the entropy of the extended source is given by

$$H(\xi^n) = n\, H(\xi)$$

Example 3: Entropy of an extended source

Consider a discrete memoryless source with alphabet $\xi = \{s_0, s_1, s_2\}$ with symbol probabilities $p_0 = 1/4, p_1 = 1/4$ and $p_2 = 1/2$.

a) Compute the entropy

$$H(\xi) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) = p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right)$$

$$= \frac{1}{4}\log_2(4) + \frac{1}{4}\log_2(4) + \frac{1}{2}\log_2(2) = \frac{3}{2} \text{ bits}$$

Consider now an extension of the previous source with 2 symbols in a block as

The alphabet $\xi^2 = \{\sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8\}$ corresponds to the sequences of ξ

$$s_0s_0, s_0s_1, s_0s_2, s_1s_0, s_1s_1, s_1s_2, s_2s_0, s_2s_1, s_2s_2$$

with probabilities

$p_0 = 1/16, p_1 = 1/16, p_2 = 1/8\,, p_3 = 1/16\,, p_4 = 1/16\,, p_5 = 1/8\,, p_6 = 1/8\,, p_7 = 1/8$ and $p_8 = 1/4$.

b) Compute the entropy

$$H(\xi^2) = \sum_{k=0}^{8} p_k \log_2\left(\frac{1}{p_k}\right) = \frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8)$$

$$+\frac{1}{16}\log_2(16) + \frac{1}{16}\log_2(16) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8) + \frac{1}{4}\log_2(4) = 3\,\text{bits}$$

or

$$H(\xi^2) = nH(\xi) = 2\frac{3}{2} = 3\text{bits}$$