Information Theory: lecture notes

Riccardo Lo Iacono

June 12, 2025

- 1 - Basis of Information theory.

Information theory, alongside data compression, have a key role in modern CS. The former defines a set of theoretical tools, usefull to understand the limitations of what is computable; the latter allows to reduce the amount of space required (in terms of bits) without losing information.

- 1.1 - Shannon's Entropy.

When talking about information we all have a general notion of what it represents, but can we define it formally? And also, is there a way to measure information?

The first to do so was *Claude Shannon*, considered by many the father of modern Information theory. In is paper [ ], Shannon analizes various comunication systems, from the discrete noiceless one to the continous noisy one. A general structure of these systems is shown in *Figure 1*

TODO: Add reference to the paper.

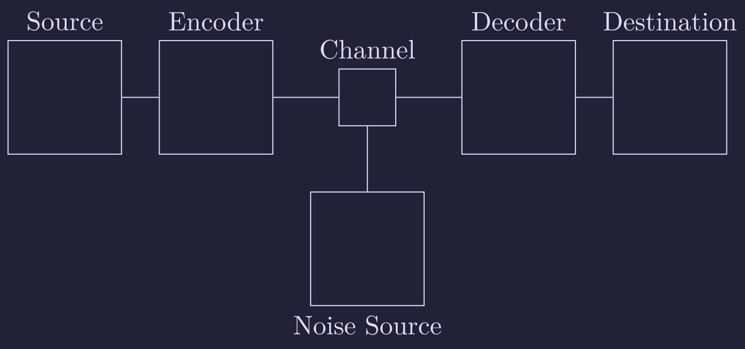

Figure 1: Diagram of a general communication system.

- The *Source* or, more precisely, the *Information Source* referes to some entity (a human, a computer, etc) that produces messages.
- The *Encoder* encodes the messages coming from the source, and transmits them trough the channel.
- The *channel* is the physical mean trough which the messages are transmitted.
- The *Decoder* has the opposite role of the *Encoder*.
- The *Destination* is the entity to which the messages are meant for.

Remark. What follows in this section, and the those following, refers just to the discrete noiceless case.

Let S be a source of information. Let Σ be the alphabet of symbols used by the source, and for each symbols let p_i be $\Pr(S = \sigma_i)$. We define the entropy of S as follow:

$$H = -K \sum_{i=1}^{n} p_i \ln p_i \tag{1}$$

The above definition of *entropy* comes out of it's axiomatic definition, which will be proved (more like a sketch of the proof) in the following.

From *Equation* (1) it's easy to understand that, the entropy measures the average information in each symbols of the source. Thus, we can conclude that information is just the uncertainty lost when the outcome of the event is know.

Note: In this context, by event we mean the next symbol transmitted/recived.

- 1.1.1 - Entropy: an axiomatic definition.

Even though we could define the entropy using a more general approach, we now proceed to define it axiomatically.

Let $H(S) = H_n(p_1, \ldots, p_n)$ be a mesure of entropy, then:

1. it must be a continuos in the p_i in the range $[0, 1]$
2. $H_2(\frac{1}{2}, \frac{1}{2}) = 1$
3. any permutation of the p_i does not change the value of H_n
4. if $A(n) = H_n(\frac{1}{n}, \ldots, \frac{1}{n})$, then H_n is monotonically increasing
5. if any choice can be broke down into K consecutive ones, then the value of H_n must be the averaged sum of the new H. In symbols

$$\begin{aligned} H_n(p_1, \ldots, p_n) =& H_{n-k+1}(p_1 + \ldots + p_k, p_{k+1}, \ldots, p_n) \\ &+ (p_1 + \ldots + p_k) H_k \left(\frac{p_1}{p_1 + \ldots + p_k}, \ldots, \frac{p_k}{p_1 + \ldots + p_k} \right) \end{aligned}$$

Theorem. *The only function H that satisfies the above axioms, has the form of* Equation (1).

Proof Let's observe that:

1. From the 5th axiom we can derive that $A(nm) = A(n)A(m)$
2. From its definition $A(n^k) = kA(n)$, for any positive n and k
3. $A(n) = K \ln n$ for any value of n, with K a positive constant
4. $H_{p,1-p} = -C[p \ln p + (1-p) \ln(1-p)]$, for $p \in [0, 1]$
5. $H_n(p_1, \ldots, p_n) = K \sum_{i=1}^{n} p_i \ln(\frac{1}{p_i})$

■