Entropy

Information data coding

Information data coding   coded representation of information

$$\text{Message} \xrightarrow{\textit{Injective correspondence}} \{b_n\}$$

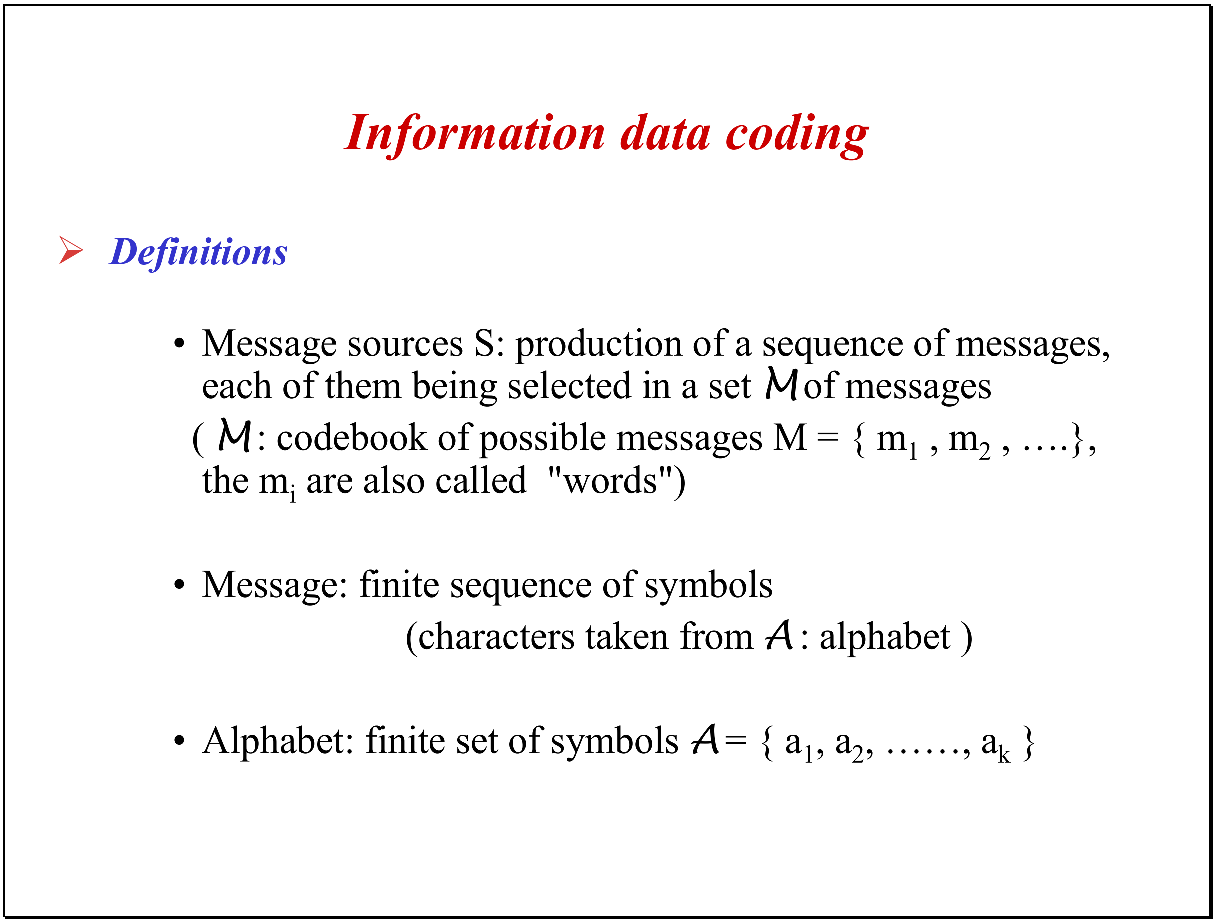
➢ *Multiples roles of coding*

- Preparing the transformation   message => transmitted signal
- Adapting the source bit rate - channel capacity (compression)
- Protective encoding against transmission errors (error detection / correction)
- Encrypting (secretive communications)
- Tattooing (ownership markers)
- Transcoding (alphabet changes, transmission constraints)

The goal of a communication system is to transport messages from a sender (the information source) towards a recipient (information user). The signal supporting the information being transmitted has to be compatible with the characteristics of the transmission channel. Information coding must establish an injective correspondence between the message produced by the source and the sequence of information $\{b_n\}$ sent to the transmitter.
This information coding plays numerous roles:

- It prepares the transformation of a message into a signal (carried out in the transmitter part: signal information encoding);
- It adapts the source of information to the capacity of the transmitting channel;
- It can be used to protect information against transmission errors (within certain limits), so as to detect and/or correct them (due to channel disturbances);
- It can also be used in certain cases for secret communications (encrypting) and watermarking to protect ownership.

A given transmission channel must be able to transport messages of various types that is why transcoding is necessary: this transforms the message representation from a codebook M_k using an alphabet A_k into a representation of the same message from a codebook M_0 using alphabet A_0. This particular alphabet is often the binary set $\{0, 1\}$, but this is not the only one.



Information data coding

➢ Definitions

- Message sources S: production of a sequence of messages, each of them being selected in a set $\mathcal{M}$ of messages
  ($\mathcal{M}$: codebook of possible messages $M = \{ m_1 , m_2 , ...\}$, the m_i are also called "words")

- Message: finite sequence of symbols
  (characters taken from $\mathcal{A}$: alphabet)

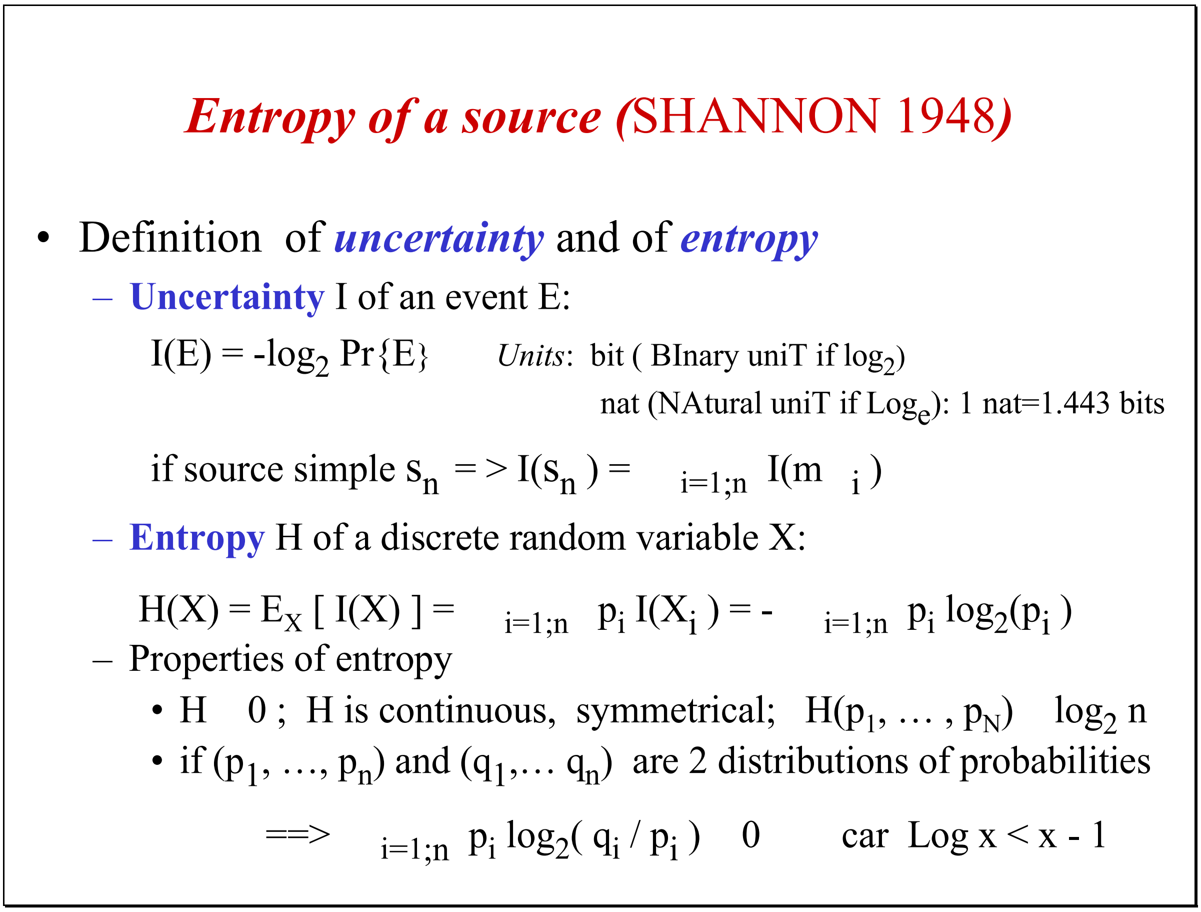
- Alphabet: finite set of symbols $\mathcal{A} = \{ a_1, a_2, ..., a_k \}$

Definitions:

A **message** is any finite set of characters taken from an **alphabet** A: a finite set of **symbols** (for example: letters, digits etc.).

A **message source** S is the system that produces a temporal series of messages m_i, each of them taken from a set of possible messages M. M is called a message (or word) codebook. The transmitted message is in fact a text formed by the syntactic rules of elementary messages called words: M = $\{m_1, m_2, \dots\}$, each word is written by a fixed, finite set of symbols taken from the alphabet A.

Depending on the applications, the message sources S can use dictionaries of very different types: from messages written using characters of the alphabet, numbers, punctuation and tab marks, to visual messages where the messages are digital images where for example, each word is a pixel represented as a sequence of 8 binary symbols taken from the alphabet {0, 1} (bit).



*Entropy of a source (*SHANNON 1948*)*

- Definition of ***uncertainty*** and of ***entropy***
  - **Uncertainty** I of an event E:

    $I(E) = -\log_2 \Pr\{E\}$ *Units*: bit (BInary uniT if $\log_2$)

    nat (NAtural uniT if Log_e): 1 nat=1.443 bits

    if source simple s_n => $I(s_n) = \; _{i=1;n} \; I(m_i)$
  - **Entropy** H of a discrete random variable X:

    $H(X) = E_X[\, I(X)\,] = \; _{i=1;n} \; p_i I(X_i) = - \; _{i=1;n} \; p_i \log_2(p_i)$
  - Properties of entropy
    - H 0 ; H is continuous, symmetrical; $H(p_1, \dots, p_N)$ $\log_2 n$
    - if $(p_1, \dots, p_n)$ and $(q_1, \dots q_n)$ are 2 distributions of probabilities

      ==> $_{i=1;n} \; p_i \log_2(q_i / p_i)$ 0 car Log x < x - 1

The ***uncertainty*** I of an event E of probability Pr(E) is defined by:

$$I(E) \;=\; \log_2 \frac{1}{Pr(E)} \;=\; -\log_2 Pr(E)$$

Notes:

- if Pr(E) = 1/2 then I (E) = 1 (unitary uncertainty)
- if Pr(E) = 1 then I (E) = 0: uncertainty is null for a certain event.
- The uncertainty unit is ***bit*** (*Binary Unit*). It is not the same as the bit: Binary digit.
- We can use the natural logarithm instead of the base 2 logarithm, therefore the unit is the ***nat*** (Natural Unit = 1.443 bit).

We now consider that the events E are in fact realizations of a random discrete variable X. We define the ***entropy*** H as being the average uncertainty of the random variable X. If we consider in fact each event x_i, i ∈ [1, n], as a realization of a random variable X (i.e. X is a random variable with values in { $x_1, x_2, \ldots, x_n$ }) :

$$H(X) = E_X\{I(X)\} = \sum_{i=1..n} Pr\{X=x_i\}\cdot I(x_i) = \sum_{i=1..n} p_i\cdot I(x_i),\ with\ p_i = Pr\{X=x_i\}$$

The entropy depends on the probability law of X but it is not a function of the values taken by X. It is expressed in bits (or nats) and represents the average number of bits necessary to binary encode the different realizations of X.

Now let's consider an information source S defined by a set of possible m_i (codebook): S{m_1, m_2, …, m_N}, and by a mechanism such as for emitting messages:

s_n = {$m_{\alpha1}$, $m_{\alpha2}$, …, $m_{\alpha n}$} with $m_{\alpha1}$: 1st emitted message, …, $m_{\alpha n}$: nth emitted message.

Warning: the index « i » in α_i defines the temporal index in the sequence of messages emitted by the source. α_i defines the index of the ith message emitted in the codebook M of possible messages, generally: N ≠ n.

The choice of m $_i$ occurs according to a given probability law. The emission of a discrete source of information thus corresponds to a sequence of random variables X_i, i∈ [1, n]:

The probability of s_n can be expressed as a product of conditional probabilities:

$$Pr(s_n) = Pr\{X_1 = m_{\alpha1}\}\ Pr\{X_2 = m_{\alpha2} / X_1 = m_{\alpha1}\} \ldots Pr\{X_n = m_{\alpha n} / X_1 = m_{\alpha1}, \ldots, X_{n-1} = m_{\alpha n-1}\}$$

In the case of simple sources, the n random variables X_i are independent and of the same law, which gives:

$$\forall\ (i,\ j)\in[1,\ n]\times[1,\ N]\ \ , Pr\{\ X_i = m_j\ \} = p_j,\ et\ Pr\{s_n\} = p_{\alpha1}.p_{\alpha2}\ldots p_{\alpha n}$$

$$\Rightarrow I(s_n) = -\log_2 Pr\{s_n\} = -\log_2(p_{\alpha_1}\ p_{\alpha_2}\cdots p_{\alpha_n}) = \sum_{i=1..n} -\log_2 p_{\alpha_i} = \sum_{i=1..n} I(m_{\alpha_i})$$

$$I(s_n) = \sum_{i=1..n} I(m_{\alpha_i})$$

In the case of a discrete source of « n » messages m_i, where each message m_i is associated with a probability p_i, the entropy H of the source S is given by:

$$H(S) = -\sum_{i=1}^{n} p_i\cdot\log_2 p_i$$

Properties of entropy:

- As $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$, then H(X) > 0: the entropy is positive.
- Given $(p_1, p_2, \ldots, p_n)$ and $(q_1, q_2, \ldots, q_n)$ two probability laws, then $\sum_{i=1}^{n} p_i \log_2 \frac{q_i}{p_i} \leqslant 0$.

  $\forall\ x>0$, we have Ln x ≤ x – 1 so

  $\ln(\frac{q_i}{p_i}) \leq \frac{q_i}{p_i} - 1$, being $\log_2(\frac{q_i}{p_i}) \leq \frac{1}{ln2}(\frac{q_i}{p_i} - 1)$ thus:

  $$\sum_{i=1}^{n} p_i \log_2 \frac{q_i}{p_i} \leq \frac{1}{ln2} \sum_{i=1}^{n} p_i(\frac{q_i}{p_i} - 1) = \frac{1}{ln2}(\sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i) = \frac{1}{ln2}(1-1) = 0$$
- The entropy of a random variable X with n possible values is maximal and is worth $\log_2 n$ when X follows a uniform probability law. By taking $q_1 = q_2 = \ldots = q_n = \frac{1}{n}$ (uniform law), in the previous property:

$$\sum_{i=1}^{n} p_i \log_2 \frac{q_i}{p_i} \leq 0 \quad \Leftrightarrow \quad -\sum_{i=1}^{n} p_i \log_2 p_i \leq -\sum_{i=1}^{n} p_i \log_2 q_i$$

$$\Leftrightarrow \quad H(X) \leq -\sum_{i=1}^{n} p_i \log_2 \frac{1}{n}$$

$$\Leftrightarrow \quad H(X) \leq -\log_2 \frac{1}{n} \sum_{i=1}^{n} p_i = \log_2 n$$

- The entropy is continuous and symmetrical.

For the rest of this course, we will systematically use the logarithm base 2.

Simple example:

Let’s consider a source S, of uniform law, that sends messages from the 26-character French (a,b,c, …, z). To this alphabet we add the "space" character as a word separator.

The alphabet is made up of 27 characters: H(S) = $-\sum_{i=1}^{27} \frac{1}{27} \log_2 \frac{1}{27} = \log_2(27) = 4.75$ bits of information per character. Actually, the entropy is close to 4 bits of information per character on a very large amount of French text.