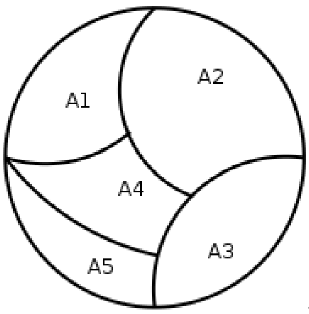
# Entropy

MUDS
Herramientas de análisis y visualización de datos
La Salle - Universitat Ramon Llull

We want to quantify the amount of uncertainty encoded into a probability distribution.

**Example** Tossing a coin and a falling toast have the same event set. They only differ in the probability distribution. If we were to bet, our options in the toast case would be much more obvious: jelly side !

If we consider a partition of a probability space, $\{A_i\}_{i=1\ldots n}$




.

we would say

- If $p(A_1) \simeq 1$ while $p(A_i) \simeq 0$, $i > 1$, there is practically no uncertainty.
- If $p(A_i) \simeq \frac{1}{n}$, we have maximal uncertainty.

A function to quantify uncertainty should have the following properties.

- Continous for $p_i = p(A_i)$.
- Refining partition should imply higher uncertainty.

$$A_n = A'_n \cup A'_{n+1}$$

- If

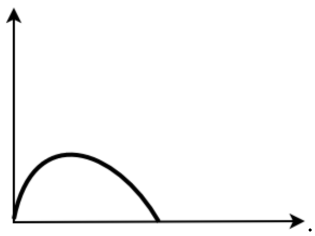
$$p(A_i) = \frac{1}{n}, \qquad i = 1, \ldots, n,$$

  and

$$p(A'_i) = \frac{1}{n+1}, \qquad i = 1, \ldots, n+1.$$

  partition $\{A'_i\}$ should have more uncertainty than partition $\{A_i\}$

The contribution of every single $p_i$ should be minimal for $p_i \simeq 1$ or $p_i \simeq 0$ and maximal in between.



Among the many functions fulfilling the above conditions we have

$$S(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \ln p_i.$$

This is

**Shannon's Entropy**

The simplest case, $n = 2$, $A_1$, $A_2 = \overline{A_1}$.

$$p(A_1) = p, \qquad p(A_2) = 1 - p$$

$$S(p) = -p \ln p - (1-p)\ln(1-p).$$

We have

$$p \to 0 \Longrightarrow S(p) \to 0,$$

$$p \to 1 \Longrightarrow S(p) \to 0.$$

The maximum:

$$\frac{d}{dp}S(p) = -\ln p + \ln(1-p)$$

so

$$p_{max} = \frac{1}{2}$$

In the general case $i = 1, \ldots, n$, the maximal entropy corresponds to

$$p_i = \frac{1}{n},$$

as expected.

We can extend the definition of Shannon's Entropy to a discrete random variable.

$$S(X) = -\sum_{i=1}^{n} P(X = x_i) \ln P(X = x_i).$$

In terms of the probability function,

$$S(X) = -\sum_{i=1}^{n} f(x_i) \ln f(x_i).$$

In other words,

$$S(X) = - < \ln f(x) >$$

For a continous random variable, this definition can be extended to the so-called relative entropy,

$$S(X) = - < \ln f(x) > = -\int_{-\infty}^{\infty} dx\ f(x) \ln f(x).$$

**Example** A Gaussian random variable is

$$S(X) = \frac{1}{2} + \frac{1}{2} \ln\left(2\pi\sigma^2\right).$$

Entropy is related to the way the result of the experiment can be codified in a binary channel.

The simplest experiment, tossing a coin, requires only one bit. This is the reason why a bit is said to be the minimal unit of information.

Shannon theorem states that the optimal code to send the result of a random experiment is the following.

- Send one bit (say 1) when the result of the experiment is the one with the largest probability.

- Send two bits (say 01) when the result of the experiment is the one with the second largest probability.

- and so on.

The average lenght of this code is given by Shannon's Entropy.

**Example** Typing an SMS on a mobile phone.