ITDM - LEZIONE 01

DEL 24/09/2019

INFORMATION is an abstract term which can be interpreted as additional KNOWLEDGE. Information is RECEIVED when a QUESTION is answered. We can then use the information obtained to modify our BEHAVIOURS.

OSS: The modern Information Theory is based on the following articles:

- "CERTAIN FACTORS AFFECTING TELEGRAPH SPEED",
  HARRY NYQUIST, 1924.
- "TRANSMISSION OF INFORMATION",
  RALPH V. HARTLEY, 1928.
- "A MATHEMATICAL THEORY OF COMMUNICATION",
  C.E. SHANNON, 1948

Shannon was the first to use a probability model for a communication system. Information is therefore linked to probability and randomness.

PROBABILITY THEORY

Probability theory studies the properties and patterns of RANDOM EXPERIMENTS.

DEF (RANDOM VARIABLE):

A RANDOM VARIABLE is a real-valued number $X(\omega)$ assigned to every OUTCOME of a STOCHASTIC EXPERIMENT, such that the set

$$\{ \omega : X(\omega) \leq x \}$$

is an EVENT, $\forall x \in \mathbb{R}$.

OSS: The function defining the r.v. $X : \Omega \to \mathbb{R}$ IS NOT RANDOM. What is random is actually in the outcome $\omega \in \Omega$ (SAMPLE SPACE), and not in $X(\omega)$.

Remember that...

i) The OUTCOME of an experiment is denoted as ω and it is uniquely determined by carrying out the experiment.

ii) The set of all possible outcomes is denoted by Ω and is called the SAMPLE SPACE.

iii) An $x \in \Omega$ is called a SAMPLE POINT, while an $A \subseteq \Omega$ is called an EVENT.

We have 2 different types of r.v., depending on the CODOMAIN:

- If it is discrete, we have a DISCRETE R.V., and we define

$$F_X(x) := P(X \leq x) \quad \text{(CUMULATIVE DISTRIBUTION FUNCTION)}$$

$$p_X(x_i) := P(X = x_i) \quad \text{(PROBABILITY MASS FUNCTION)}$$

- If it is continuous, we have a CONTINUOUS R.V., and we define

$$F_X(x) := \int_{-\infty}^{x} f_X(t)\,dt$$

$$f_X(x)$$

HOW TO MEASURE INFORMATION

We start with the following rules:

- If there is no knowledge, or 0 knowledge, then we have no information.

- If I is the amount of info., then $I \geq 0$ and there is no upper bound.

We thus get $I \in [0, \infty)$.

Information and probability are intuitively linked in the following way:

+ Events with LOW PROB. give HIGH INFO if they happen, and

+ Events with LOW PROB. give HIGH INFO if they happen, and

DEF (SELF-INFORMATION):
Let x_i be the outcome of the experiment, then the SELF-INFORMATION of the event $(X = x_i)$ is defined as

$$I_a(x_i) := \log_a \frac{1}{p_X(x_i)} \quad , \quad a \in \mathbb{R}^+$$

Usually $a = 2$, and the unit of measure is called BIT (BINARY UNIT).

OSS: Initially the word bit was used to represent a UNIT OF MEASURE, and not a BINARY DIGIT. With this you can also have a REAL NUMBER OF BITS.

PROPERTIES OF SELF-INFORMATION

1) CONTINUITY
2) NON-NEGATIVITY
3) MONOTINICITY,

$I(x_i)$ is a non-decreasing function of $p(x_i)$

ENTROPY

The SELF-INFORMATION measure is defined for a single event and it can be computed only after the experiment is over and the outcome is known.

We thus introduce the concept of ENTROPY to be able to compute the amount of information of a random experiment BEFORE the experiment is actually carried out.

DEF (ENTROPY):

Given a discrete r.v. X, the **ENTROPY** of X is defined as

$$H(X) := \mathbb{E}[I(X)] = \sum_{i=1}^{N_X} p(x_i) \cdot I(x_i) = \sum_{i=1}^{N_X} p(x_i) \cdot \frac{1}{\log_2 p(x_i)}$$

PROPERTIES OF ENTROPY

1) CONTINUITY

2) NON-NEGATIVITY

3) CONCAVITY,

$\cap$ concave function

$\cup$ convex function

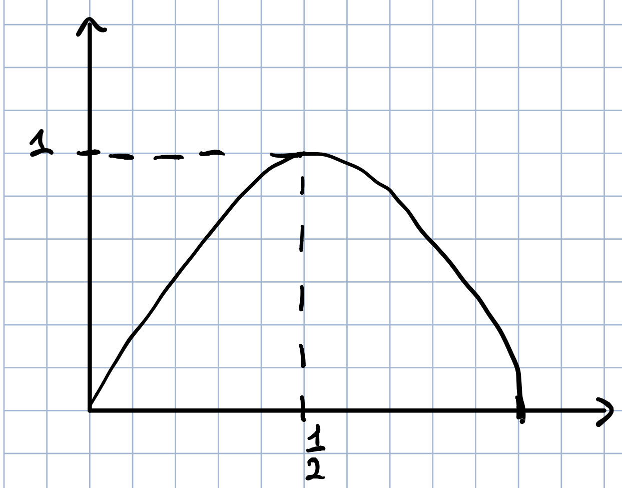
The concavity of $H(X)$ means that the entropy function has a **COMPUTABLE AND UNIQUE MAXIMUM**.

EXAMPLE: Consider the typical COIN TOSSING problem.

$P(H) = \frac{1}{2}$

$P(T) = \frac{1}{2}$

$$H(x) = \frac{1}{2} \cdot \log_2 2 + \frac{1}{2} \cdot \log_2 2 = 1$$



Therefore the binary entropy function $H(p)$ is maximized when $p = \frac{1}{2}$.