ITDM - LEZIONE 02

DEL 26/09/2019

Given r.v. X with pmf $p_1, p_2, \dots, p_{N_x}$ we know its entropy is

$$H(X) = \sum_{i=1}^{N_x} p_i \cdot \log_2 \frac{1}{p_i}$$

OSS: The entropy of X does not depend on the particular values X can take. It only depends on the pmf of X, that is it depends on $p_1, \dots, p_{N_x}$.

PROPERTIES OF ENTROPY

$H(X)$ has the following properties:

- P_1: CONTINUITY

$H(X)$ is a combination (SUM) of continuous functions of the form $p \cdot \log_2 \frac{1}{p}$, and therefore it is continuous with respect to $p_1, \dots, p_{N_x}$

- P_2 : NON-NEGATIVITY

$$H(X) > 0 \quad \forall \text{ values } p_1, \dots, p_{N_x}$$

This property allows us to consider $H(X)$ as a MEASURE.

NOTATION :

Sometimes instead of writing $H(X)$ we write the probabilities in an explicit way

$$H_{N_x}\underbrace{(p_1, p_2, \dots, p_{N_x})}_{\text{PROBABILITY DENSITY}}$$

OF POSSIBLE OUTCOMES

For example, in the typical COIN TOSSING experiment we have

$$H_2(p_0, p_1) = H_2(p_0, 1-p_0)$$

$$= H_2(p_0) \quad \left(\begin{array}{l}\text{BINARY ENTROPY} \\ \text{FUNCTION}\end{array}\right)$$

$$= p_0 \cdot \log_2 \frac{1}{p_0} + (1-p_0) \cdot \log_2 \frac{1}{1-p_0}$$

- P_3: EXPANDIBILITY

$$H_{N_x}(p_1, \dots, p_{N_x}) = H_{N_x+1}(p_1, \dots, p_{N_x}, 0)$$

This property means that the entropy does not change if we add an impossible event.

It can be proved by noting that

$$\lim_{x \to 0^+} x \cdot \ln \frac{1}{x} = 0$$

and therefore, if $p = 0$, we define $p \cdot \ln \frac{1}{p} := 0$, which means that the contribution of the impossible event is 0.

- P_4: MAXIMALITY

$$H_{N_x}(p_1, \dots, p_{N_x}) \leq \log_2 N_x, \quad \forall p_1, \dots, p_{N_x}$$

Also, $H_{N_x}(p_1, \dots, p_{N_x}) = \log_2 N_x$ only when $p_K = \frac{1}{N_x}$
This means that THE ENTROPY IS MAXIMIZED WITH UNIFORM DISTRIBUTIONS.

This property can be proved in the following way:

$$H_{N_x}(r_1, \dots, r_{N_x}) \leq \log N_x \iff H_{N_x}(r_1, \dots, r_{N_x}) - \log N_x = 0$$

Since $(r_1, \dots, r_{N_x})$ is a density, we have

$$H_{N_x}(x) - 1 \cdot \log N_x = \sum_{k=1}^{N_x} r_k \cdot \log_2 \frac{1}{r_k} - \sum_{k=1}^{N_x} r_k \cdot \log N_x$$

$$= \sum_{k=1}^{N_x} r_k \cdot \log_2 \frac{1}{r_k \cdot N_x}$$

Note now that $\begin{cases} \log x \leq x - 1, & x > 0 \\ \log x = x - 1, & x = 1 \end{cases}$

We thus get,

$$\sum_{k=1}^{N_x} r_k \cdot \log \frac{1}{r_k \cdot N_x} \leq \sum_{k=1}^{N_x} r_k \cdot \left(\frac{1}{r_k \cdot N_x} - 1 \right)$$

$$= \sum_{k=1}^{N_x} r_k \left(\frac{1 - r_k \cdot N_x}{r_k \cdot N_x} \right)$$

$$= \sum_{k=1}^{N_x} \frac{1}{N_x} - r_k$$

$$= \frac{1}{N_x} \cdot \sum_{k=1}^{N_x} (1 - r_k \cdot N_x)$$

$$= \frac{1}{N_x} \cdot \left(N_x - N_x \cdot \sum_{k=1}^{N_x} r_k \right)$$

$$= 1 - 1 = 0 \qquad \blacksquare$$

DIFFERENTIAL ENTROPY

Consider now a continuous r.v. X with $p_X(x)$ as p.d.f. (PROBABILITY DENSITY FUNCTION) and $S = [a, b] \subseteq \mathbb{R}$ as the SUPPORT SET

DOMAIN S SUCH THAT $\forall s \in S : p(s) > 0$

We will now extend the concept of entropy for the case of a continuous r.v. X by introducing DIFFERENTIAL ENTROPY, which is defined as follows

$$h(X) := \int_a^b p_X(x) \cdot \log \frac{1}{p_X(x)} \, dx$$

$$= - \int_a^b p_X(x) \cdot \log p_X(x) \, dx$$

We have to make the following important distinction between entropy and differential entropy:

ENTROPY → ABSOLUTE measure of information that has a meaning in and of itself.

DIFFERENTIAL ENTROPY → RELATIVE measure of information whose meaning has to be interpreted with other data.

Depends on the particular possible values of the r.v., and not only on their probabilities.

EXAMPLE ($X \sim U[a,b]$):

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{else} \end{cases}$$

$$h(X) = \int_a^b \frac{1}{b-a} \cdot \ln(b-a)\, dx$$

$$= \ln(b-a)$$

$h(X)$ depends on a and b, which define X. Thus, if we increase the interval, we also increase the amount of uncertainty.

EXAMPLE ($X \sim \mathcal{N}(\mu, \sigma^2)$):

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \quad S = (-\infty, +\infty)$$

$$h(X) = -\int_{-\infty}^{+\infty} f_X(x) \cdot \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right) dx$$

$$= -\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \cdot \int_{-\infty}^{+\infty} f(x)\, dx + \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{2\sigma^2} f(x)\, dx$$

$$= \frac{1}{2} \cdot \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \cdot \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\, dx$$

$$= \frac{1}{2} \cdot \log(2 \cdot \pi \cdot \sigma^2) + \frac{1}{2\cancel{\sigma^2}} \cdot \cancel{\sigma^2}$$

$$= \frac{1}{2}\left(\log(2 \cdot \pi \cdot \sigma^2) + 1\right)$$

Notice that $h(X) \propto \sigma^2$, which means that the more "SPREAD" the p.d.f., the more the differential entropy is HIGH.