

# Práctica 1. Fuente de Información Asociada a un Texto

Fuente de información

- Alfabeto  $A = \{a_1, \dots, a_m\}$
- Probabilidad de emisión de cada uno de los símbolos  
 $\{p_1, \dots, p_m\}$



$A = \{a_1, \dots, a_m\}$

$P = [p_1, \dots, p_m]$

q-ario

sin ruido

↕

C.D.U. (códigos de decodificación  
única)

↕  
codificación de la fuente

Para hacer una codificación que comprima la información va a ser conveniente trabajar con fuentes de información.

$(A, p)$

# Análisis de Frecuencias

Partimos de un texto

→ localizar los símbolos distintos que aparecen en el texto

↘ Contar el número de veces que aparece cada símbolo en el texto

texto = "SECRETO DE UNO SECRETO SEGURO"

|   |   |   |   |  |  |  |  |  |  |  |  |
|---|---|---|---|--|--|--|--|--|--|--|--|
| S | E | C | R |  |  |  |  |  |  |  |  |
| 1 | 2 | 1 | 1 |  |  |  |  |  |  |  |  |

alfabeto

frecuencias absolutas

Nota Es importante que todos usemos el mismo orden para los símbolos del alfabeto. (Práctica 3)

(nº de veces que aparece cada símbolo)

texto = "SECRETO DE UNO SECRETO SEGURO"

| S | E | C | R | T | O |   | D | U | N | L |       |
|---|---|---|---|---|---|---|---|---|---|---|-------|
| 3 | 6 | 2 | 3 | 2 | 4 | 4 | 1 | 2 | 1 | 1 | 29    |
|   |   |   |   |   |   |   |   |   |   |   | total |

alfabeto  
frecuencias

fuentes de información

| S              | E              | C   | R | T | O |  | D | U | N | L |  |
|----------------|----------------|-----|---|---|---|--|---|---|---|---|--|
| $\frac{3}{29}$ | $\frac{6}{29}$ | ... |   |   |   |  |   |   |   |   |  |

alfabeto  
probabilidades

asumimos  $\rightarrow$  probabilidad = frecuencia relativa

# Entropía de una Fuente (sin memoria)

$$p(a_i a_j) \\ \downarrow \\ p_i \cdot p_j$$

## Medidas de la información

(1) Cantidad de información que aporta la emisión de un mensaje

• Depende únicamente de la probabilidad de emisión

• A mayor probabilidad menor cantidad de información

$$I(p_i p_j) = I(p_i) + I(p_j)$$

• Unidad de medida el bit = cantidad de información que aporta cada símbolo de una fuente binaria equiprobable.

$$I(1/2) = \underline{\underline{1 \text{ bit}}}$$

$$I(p) = \log_2(1/p) = \log_2(p^{-1}) = -\log_2(p)$$

fórmula para  
la cantidad  
de información  
medida en bits

(2) Entropía de una Fuente  $R = (A, p)$

Es un promedio de la cantidad de información que aporta la emisión de un símbolo de  $R$ .

$$H(R) = \sum_{i=1}^m p_i \cdot \underbrace{\log_2 \left( \frac{1}{p_i} \right)}_{I(a_i)}$$

↓  
probabilidad de  
emisión del símbolo  $a_i$



COTAS  
PARA DIVERSOS  
PROCESOS  
DE  
CODIFICACIÓN

(1) Cálculo usando logaritmos neperianos  $\rightarrow \log_2(x) = \frac{\ln(x)}{\ln(2)}$

(2) Cálculo a partir de las frecuencias absolutas

$$\begin{aligned} (1) H(P) &= \sum_{i=1}^m p_i \log_2 \left( \frac{1}{p_i} \right) = \sum_{i=1}^m p_i \frac{\ln(1/p_i)}{\ln(2)} = \\ &= \frac{1}{\ln(2)} \sum_{i=1}^m p_i \ln(1/p_i) = - \frac{1}{\ln(2)} \sum_{i=1}^m p_i \ln(p_i) \end{aligned}$$

(2) Como  $p_i = \frac{f_i}{\text{total}}$  se tiene

$$\begin{aligned} H(F) &= \sum_{i=1}^m \frac{f_i}{\text{total}} \log_2 \left( \frac{\text{total}}{f_i} \right) = \frac{1}{\text{total}} \sum_{i=1}^m f_i (\log_2(\text{total}) - \log_2(f_i)) \\ &= \frac{1}{\text{total}} \left( \underbrace{\sum_{i=1}^m f_i}_{\text{total}} \underbrace{\log_2(\text{total})}_{\text{total}} - \sum_{i=1}^m f_i \log_2(f_i) \right) \\ &= \log_2(\text{total}) - \frac{1}{\text{total}} \sum_{i=1}^m f_i \log_2(f_i) \end{aligned}$$

\* Podemos dar una fórmula que trabaje con frecuencias y repeticiones.

la entropía de la fuente asociada al texto

texto = "SECRETO DE UNO SECRETO SEGURO"

es

$$H(F) \approx 3,236617201725714$$

15 cifras decimales

Para  
Comprobar