ENTROPIA

Supongamos que un evento probabilístico resulta de la observación, a un intervalo regular de tiempo, de la salida de una fuente de información discreta. La salida de esta fuente puede ser modelada como una variable discreta aleatoria *S* escrita de la siguiente manera:

1) $$S = \{s_0, s_1, \ldots, s_{K-1}\}$$

Cada elemento de esta fuente tiene una probabilidad de ocurrencia:

2) $$P(S = s_k) = p_k \qquad k = 0, 1, \ldots, K - 1$$

donde obviamente se cumple la condición:

3) $$\sum_{k=0}^{K-1} p_k = 1$$

definimos a la cantidad de información recibida o como:

4) $$I(s_k) = \log\left(\frac{1}{p_k}\right)$$

donde I es la cantidad de información recibida luego de observar el evento s_k que tiene probabilidad de ocurrencia p_k.

Esta definición presenta las siguientes propiedades que se pueden corroborar intuitivamente:

1. $I(s_k) = 0$ para $p_k = 1$ 5)

2. $I(s_k) \geq 0$ para $0 \leq p_k \leq 1$ 6)

Esto significa que la ocurrencia de un evento implica alguna ganancia de información, o al menos nada, pero nunca una pérdida de información.

3. $I(s_k) > I(s_i)$ para $p_k < p_i$ 7)

Menos probable es el evento, entonces más ganancia de información se produce.

4. $I(s_k \cdot s_i) = I(s_k) + I(s_i)$ si s_k y s_i son estadísticamente independientes. 8)

Si bien en la ecuación (4) la base del logaritmo puede ser arbitraria se usa el logaritmo en base 2. Bajo esta condición la unidad de información resulta ser el *bit* (contracción de **bi**nary uni**t**). No hay que confundir este término con su análogo *bit* pero que significa **bi**nary digi**t** y que se refiere a uno de los dos estados posibles de transmisión dentro de una secuencia de unos y ceros. Podemos escribir entonces:

Para el caso en que p_k = ½ entonces la cantidad de información vale 1. Cuando se transmite un símbolo entre dos posibles y ambos son igualmente probables, se gana 1 bit de información.

Ejemplo. *Una fuente produce los símbolos A, B, C y D con probabilidades ½, ¼, 1/8 y 1/8, respectivamente. Los símbolos se generan en forma independiente (fuente sin memoria). Calcular la información producida al transmitirse la secuencia "BACA"*

9) $$I(BACA) = I(B) + I(A) + I(C) + I(A)$$

$$I(B) = \log_2\left(\frac{1}{1/4}\right) = 2$$

$$I(A) = \log_2\left(\frac{1}{1/2}\right) = 1$$

$$I(C) = \log_2\left(\frac{1}{1/8}\right) = 3$$

$$I(BACA) = 2 + 1 + 3 + 1 = 7\text{ bits}$$

$I(s_k)$ es la cantidad de información producida cuando un cierto elemento s_k es transmitido desde la fuente de información.

Dado que la fuente de información tiene varios elementos o símbolos, cada uno con su probabilidad de ocurrencia p_k, es más útil describir cuál es la cantidad de información promedio que produce la fuente a medida que transmite sus símbolos.

Tal cantidad la podemos escribir de la siguiente manera:

10) $$H(S) = E[I(s_k)]$$

La cantidad H se llama entropía de una fuente discreta sin memoria. Es una medida de la cantidad promedio de información generada por una fuente de información. Nótese que la entropía depende sólo de las probabilidades de cada uno de los símbolos de la fuente de información.

Algunas propiedades de la entropía

Consideremos una fuente discreta sin memoria que está caracterizada por las ecuaciones (1) y (2). La entropía de dicha fuente está limitada por los siguientes valores:

$$0 \le H(S) \le \log_2 K$$

donde *K* es la raíz o número de símbolos de la fuente *S*. Se puede establecer que:

1. $H(S) = 0$ si y sólo si $p_k = 1$ para un cierto valor de k y el resto de las probabilidad son todas cero. En este caso no hay incertidumbre (es decir, con toda seguridad se sabe qu será transmitido el símbolo s_k).

2. $H(S) = \log_2 K$ si y sólo si $p_k = 1/K$ para todo k (es decir, todos los símbolos del alfabeto S son equiprobables). En este caso se tiene la máxima incertidumbre.

Ejemplo. *Consideremos una fuente binaria (cuyos símbolos pueden representarse, por ejemplo, con un 0 y un 1) y supongamos que la probabilidad de que ocurra un 0 es* p_0 *y la probabilidad de que ocurra un 1 es* $p_1 = 1-p_0$*. Supongamos que la fuente es discreta sin memoria, es decir la emisión de los símbolos es estadísticamente independiente. La entropía de una fuente de estas características es:*

$$\begin{aligned} H(S) &= -p_0 \log_2 p_0 - p_1 \log_2 p_1 \\ &= -p_0 \log_2 p_0 - (1-p_0)\log_2(1-p_0) \quad \text{bits} \end{aligned}$$

La gráfica de la entropía para una fuente binaria se muestra en la Figura 1.

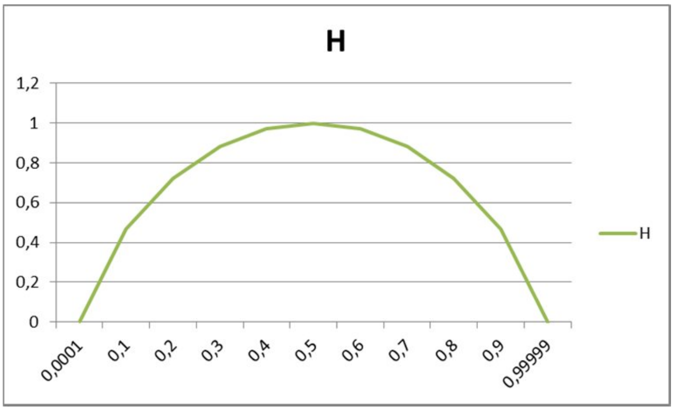


Figura 1. Entropía de una fuente discreta binaria sin memoria.

Entropía (Incertidumbre)

- Contenido de información de Shannon: medido en bits, es una medida de la cantidad de información del evento $x=a_i$.

$$h(x) = \log_2 \frac{1}{P(x)}$$

- Entropía de un espacio de probabilidad: se define como el promedio o valor esperado del contenido de información de Shannon

$$H(X) \equiv \sum_{x \in A_x} P(x) \log \frac{1}{P(x)}$$

Algunas propiedades

- H(X) = 0 si y solo si p_i=1
- H(X) > 0
- H(X)< log($|A_x|$), donde $|A_x|$ es la cardinalidad del conjunto A_x
- H(X)= log($|A_x|$), la entropía es máxima si p es uniforme, p_i=1/ $|A_x|$
- Por convención, para p=0, se usa

$$\lim_{\theta\to 0} \theta \log \frac{1}{\theta} = 0$$

EJEMPLO: Entropia del alfabeto en español

i	a_i	p_i	i	a_i	p_i
1	a	0.0575	14	n	0.0596
2	b	0.0128	15	o	0.0689
3	c	0.0263	16	p	0.0192
4	d	0.0285	17	q	0.0008
5	e	0.0913	18	r	0.0508
6	f	0.0173	19	s	0.0567
7	g	0.0133	20	t	0.0706
8	h	0.0313	21	u	0.0334
9	i	0.0599	22	v	0.0069
10	j	0.0006	23	w	0.0119
11	k	0.0084	24	x	0.0073
12	l	0.0335	25	y	0.0164
13	m	0.0235	26	z	0.0007
			27	-	0.1928

- En el ejemplo anterior tenemos que

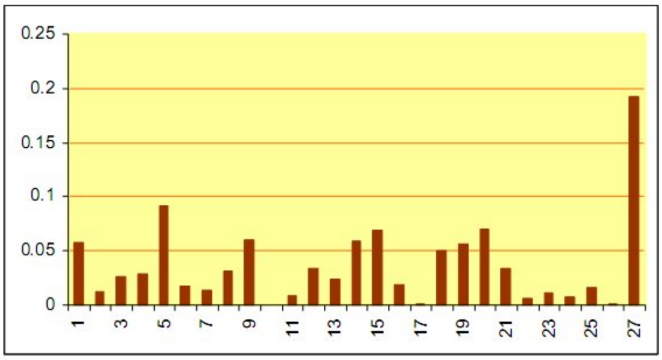
$$H(p) = \sum_i p_i \times h(p_i) = 4.11$$

- Además sabemos que la cota superior de H(p) es

$$H(p) < \log(27) = 4.7548875$$

- Como podemos interpretar este resultado?

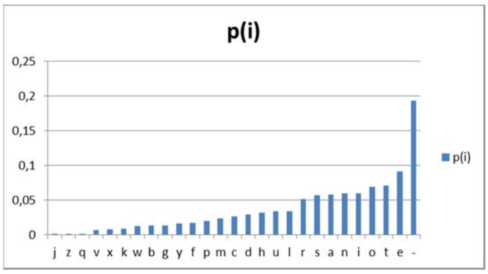
letra	p(i)	I(i)	I(i)*pi)
a	0,0575	4,12	0,23691692
b	0,0128	6,29	0,08048272
c	0,0263	5,25	0,13804327
d	0,0285	5,13	0,14628749
e	0,0913	3,45	0,31528093
f	0,0173	5,85	0,10125836
g	0,0133	6,23	0,08289132
h	0,0313	5,00	0,15642781
i	0,0599	4,06	0,24327188
j	0,0006	10,70	0,00642165
k	0,0084	6,90	0,05792132
l	0,0335	4,90	0,16413979
m	0,0235	5,41	0,12716309
n	0,0596	4,07	0,24248521
o	0,0689	3,86	0,26590937
p	0,0192	5,70	0,1094928
q	0,0008	10,29	0,00823017
r	0,0508	4,30	0,21839061
s	0,0567	4,14	0,23476677
t	0,0706	3,82	0,26998767
u	0,0334	4,90	0,16379387
v	0,0069	7,18	0,0495364
w	0,0119	6,39	0,07607545
x	0,0073	7,10	0,05181458
y	0,0164	5,93	0,09725463
z	0,0007	10,48	0,00733625
-	0,1928	2,37	0,45786588
		Entropia	4,1094

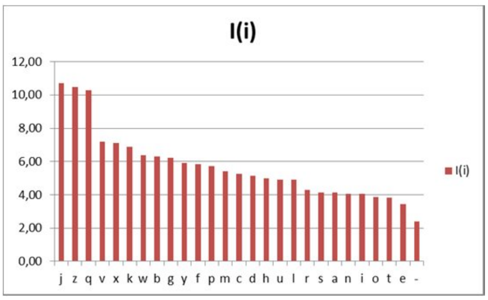


$$H(p) = \sum_i p_i \times h(p_i) = 4.11$$

$$H(p) < \log(27) = 4.7548875$$

letra	p(i)	I(i)	I(i)*pi)
j	0,0006	10,70	0,00642165
z	0,0007	10,48	0,00733625
q	0,0008	10,29	0,00823017
v	0,0069	7,18	0,0495364
x	0,0073	7,10	0,05181458
k	0,0084	6,90	0,05792132
w	0,0119	6,39	0,07607545
b	0,0128	6,29	0,08048272
g	0,0133	6,23	0,08289132
y	0,0164	5,93	0,09725463
f	0,0173	5,85	0,10125836
p	0,0192	5,70	0,1094928
m	0,0235	5,41	0,12716309
c	0,0263	5,25	0,13804327
d	0,0285	5,13	0,14628749
h	0,0313	5,00	0,15642781
u	0,0334	4,90	0,16379387
l	0,0335	4,90	0,16413979
r	0,0508	4,30	0,21839061
s	0,0567	4,14	0,23476677
a	0,0575	4,12	0,23691692
n	0,0596	4,07	0,24248521
i	0,0599	4,06	0,24327188
o	0,0689	3,86	0,26590937
t	0,0706	3,82	0,26998767
e	0,0913	3,45	0,31528093
-	0,1928	2,37	0,45786588
		Entropia	4,1094





Tasa de información

Dos fuentes de información pueden tener la misma entropía pero, por ejemplo, una de ellas puede estar generando más mensajes por unidad de tiempo que la otra fuente. O también podríamos pensar en una cierta fuente de información que tiene una cierta entropía *H(S)* cuyo comportamiento indudablemente no será igual según a qué velocidad transmita sus mensajes. Cuanto más rápida sea la fuente, más información por unidad de tiempo estará transmitiendo.

Decimos entonces que si una fuente genera mensajes a una tasa o velocidad de r mensajes por segundo, la tasa de información es

$$R = r \cdot H(S) \text{ = número promedio de bits de información por segundo}$$

Si la duración promedio del símbolo es $\bar{t}$ segundos, entonces podemos escribir la tasa de información de la siguiente manera:

$$R = \frac{H(S)}{\bar{t}}$$

Es decir que $1/\bar{t}$ es el número promedio de símbolos por unidad de tiempo (o sea, r). Más precisamente, el tiempo promedio queda definido como:

$$\bar{t} = \sum_{k=0}^{K-1} p_k t_k$$

donde t_k es el tiempo de duración de cada símbolo s_k con probabilidad de ocurrencia p_k.

Ejemplo. *Calcular la tasa de información de una fuente que utiliza el código Morse. Suponer que:*

$p_{punto} = 2/3$; $p_{raya} = 1/3$; $t_{punto} = 0{,}2$ seg; $t_{raya} = 0{,}4$ seg.

Solución.

$$H(S) = \frac{2}{3}\log_2\left(\frac{3}{2}\right) + \frac{1}{3}\log_2(3) = 0{,}92\,\textit{bits de información/símbolo}$$

$$\bar{t} = p_{punto} \cdot t_{punto} + p_{raya} \cdot t_{raya}$$
$$= \frac{2}{3} \cdot 0{,}2 + \frac{1}{3} \cdot 0{,}4 = 0{,}267 \text{ seg}$$

Finalmente

$$R = \frac{0{,}92}{0{,}267} = 3{,}44 \text{ bits de información / segundo}$$