Entropía

Módulo 4 : Técnicas computacionales avanzadas para modelar fenómenos sociales
Concentración en Economía Aplicada y Ciencia de Datos
ITESM

5 de noviembre de 2022

Entropía de Shannon

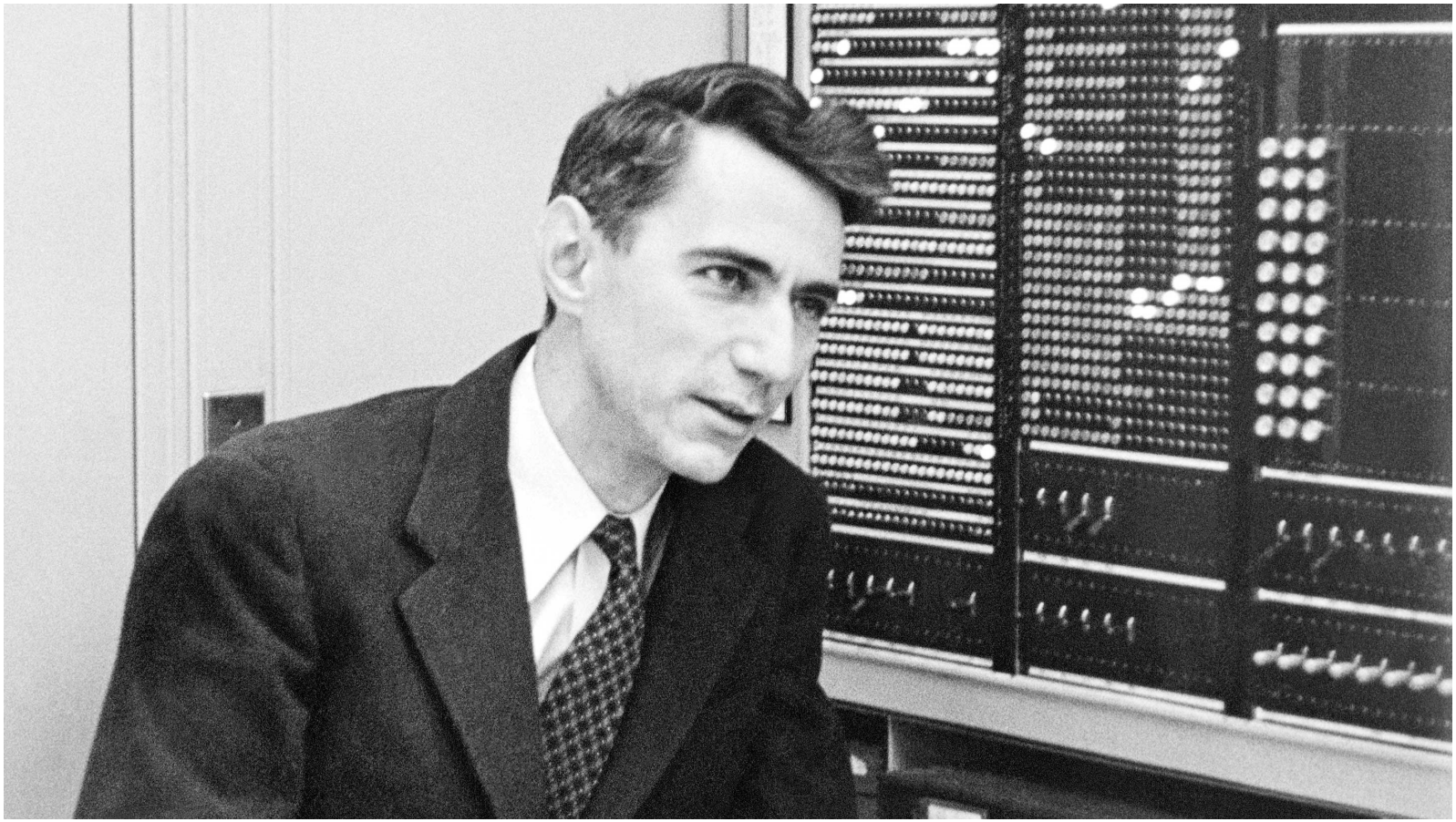
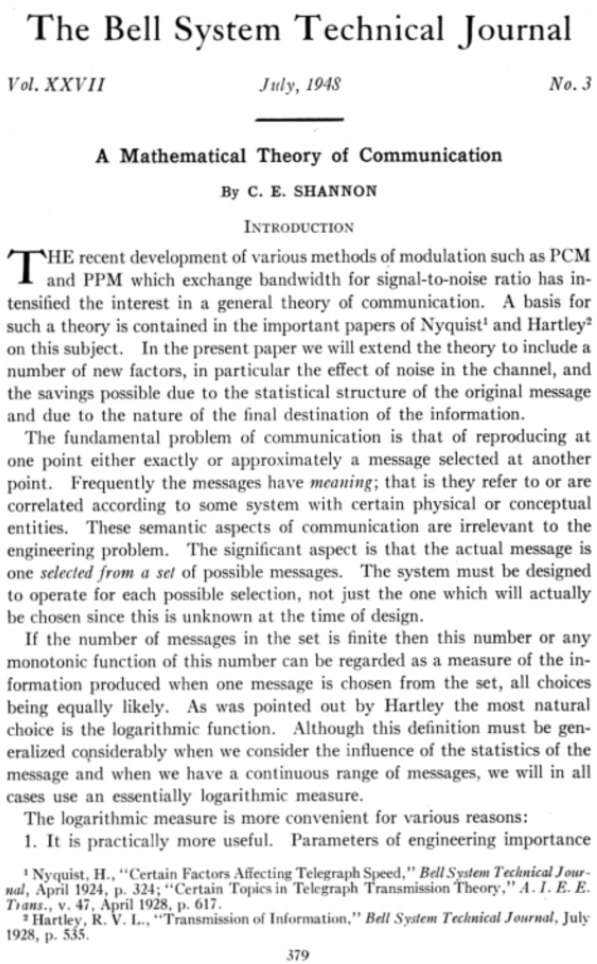


Figura: Claude Shannon

Entropía de Shannon



The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.

[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

379

Figura: A Mathematical Theory of Communication

Intuición

De las siguientes tres declaraciones:

- El cumpleaños de mi hermano es en un día particular del año.
- El cumpleaños de mi hermano es en la segunda mitad del año.
- El cumpleaños de mi hermano es el dia 25 de algún mes.

¿Cómo cuantificamos la utilidad de esta información?

Intuición

Las probabilidades de los eventos anteriores son los siguientes:

- $P_1 = 1$
- $P_2 = \frac{1}{2}$
- $P_3 = \frac{12}{365}$

Sí las variables aleatorias X y Y son independientes, la probabilidad conjunta es igual a $P(x, y) = P(x)P(y)$.
De manera que la probabilidad conjunta de los eventos 2 y 3 es igual a:

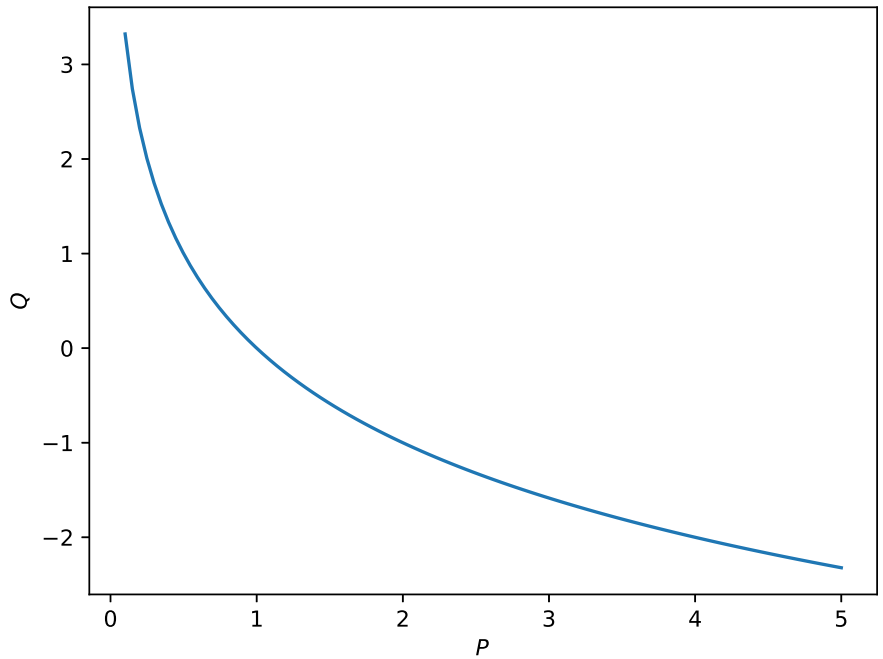
$$P_{2,3} = \frac{1}{2} \times \frac{12}{365} = \frac{6}{365}$$

Relación entre probabilidad e información

La **información** Q contenida en una declaración está definida por:

$$Q = -\log_2 P, \tag{1}$$

donde P es la probabilidad que esa declaración sea cierta.
Si P aumenta, Q disminuye. Si P disminuye, Q aumenta.
Se usa $\log_2$ dado que la información será proporcionada en **bits**.



Relación para varias declaraciones

Si tenemos un conjunto de declaraciones con probabilidad P_i con correspondiente información $Q_i = -\log P_i$, entonces el promedio ponderado de la información contenida está dada por

$$H = \sum_{i=1}^{n} Q_i P_i = -\sum_{i=1}^{n} P_i \log P_i \tag{2}$$

- $\sum_i P_i = 1$
- $P_i \geq 0$ con $1 \leq i \leq n$

H es la **entropía de Shannon**.

Ejemplo

- Un dado normal produce las siguientes salidas : $1, 2, 3, 4, 5$ y 6.
- La probabilidad de cada una de las salidas es : $\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$ y $\frac{1}{6}$.
- La información asociada a cada salida es $Q = -\log_2 \frac{1}{6} = \log_2 6$.
- La entropía de Shannon del sistema es:

$$H = \sum_{i=1}^{6} \frac{1}{6} \log_2 6 = 2.58 \tag{3}$$

Ejemplo

- Un dado cargado produce las siguientes salidas : $1, 2, 3, 4, 5$ y 6.
- La probabilidad de cada una de las salidas es : $\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}$ y $\frac{1}{2}$.
- La información asociada a cada salida es

$$\log_2 10, \log_2 10, \log_2 10, \log_2 10, \log_2 10, \log_2 2$$

- La entropía de Shannon del sistema es:

$$H = 5 \times \frac{1}{10} \log_2 10 + \frac{1}{2} \log_2 2 = \log_2 \sqrt[2]{20} = 1.16 \qquad (4)$$

La entropía disminuye cuando las probabilidades no son iguales en el dado.

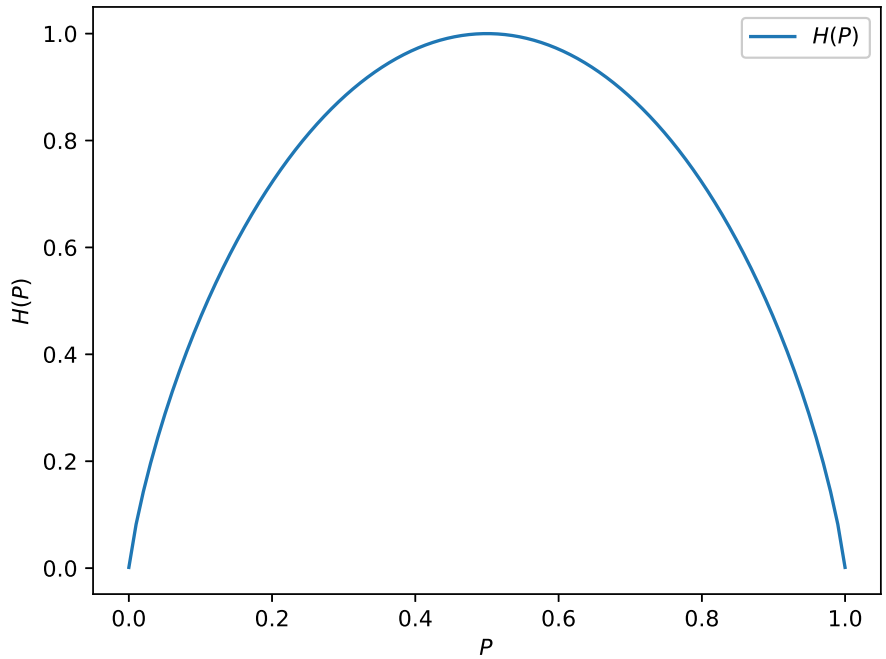
La entropía de Shannon cuantifica la cantidad de información que obtenemos, en promedio ponderado, siguiendo una medición de una cantidad particular.

La entropía de Shannon cuantifica la cantidad de incertidumbre que tenemos sobre una cantidad antes de medirla.

Ejemplo

Consideremos la tirada de una moneda. Una cara con probabilidad $P_1 = P$ y la otra con probabilidad $P_2 = 1 - P$.
La entropía del sistema es

$$H(P) = -\sum_i P_i \log_2 P_i = -P \log_2 P - (1-P)\log_2(1-P) \qquad (5)$$



- Máximo cuando $P = \frac{1}{2}$.
  Mucha incertidumbre sobre la salida, mayor información ganada.
- Mínimo cuando $P = 0$ o $P = 1$.
  Poca incertidumbre acerca de la salida, menor información ganada.

Definition

Supongamos que X es una variable aleatoria discreta aleatoria que toma valores en un conjunto finito X. Entonces, la entropía de la variable aleatoria X está definida por la cantidad

$$H(\mathrm{X}) = -\sum_{x \in X} \Pr[x] \log_2 \Pr[x] \tag{6}$$

Algunos conjuntos pueden ser:

- $X_1 = \{\text{aguila}, \text{sol}\}$.
- $X_2 = 1, 2, 3, 4, 5, 6$

Entropía Condicional

Supón que X y Y son dos variables aleatorias. Entonces, para cualquier realización y de Y, tenemos que la probabilidad (condicional) X$|y$ es

$$H(\mathrm{X}|y) = -\sum_{x} \Pr[x|y] \log_2 \Pr[x|y] \tag{7}$$

Entropía Condicional

Definimos la entropía condicional, denotada como $H(\mathrm{X}|\mathrm{Y})$, como el peso promedio (respecto a las probabilidades $\Pr[y]$) de la entropía $H(\mathrm{X}|y)$ sobre todos los posibles valores y. Este cálculo es igual a la expresión

$$H(\mathrm{X}|\mathrm{Y}) = -\sum_{y}\sum_{x} \Pr[y]\Pr[x|y]\log_2 \Pr[x|y] \tag{8}$$

La entropía condicional mide el promedio de la información acumulada acerca de X tal que no es revelada por Y.

References I

Stinson, D. R. (2005). *Cryptography: theory and practice*. Chapman and Hall/CRC.