2025 ~ 40° Aniversario de la Creación del Consejo Interuniversitario Nacional

PROBABILIDAD Y ESTADÍSTICA

Notas de clase - *Introducción a la Teoría de la Información*

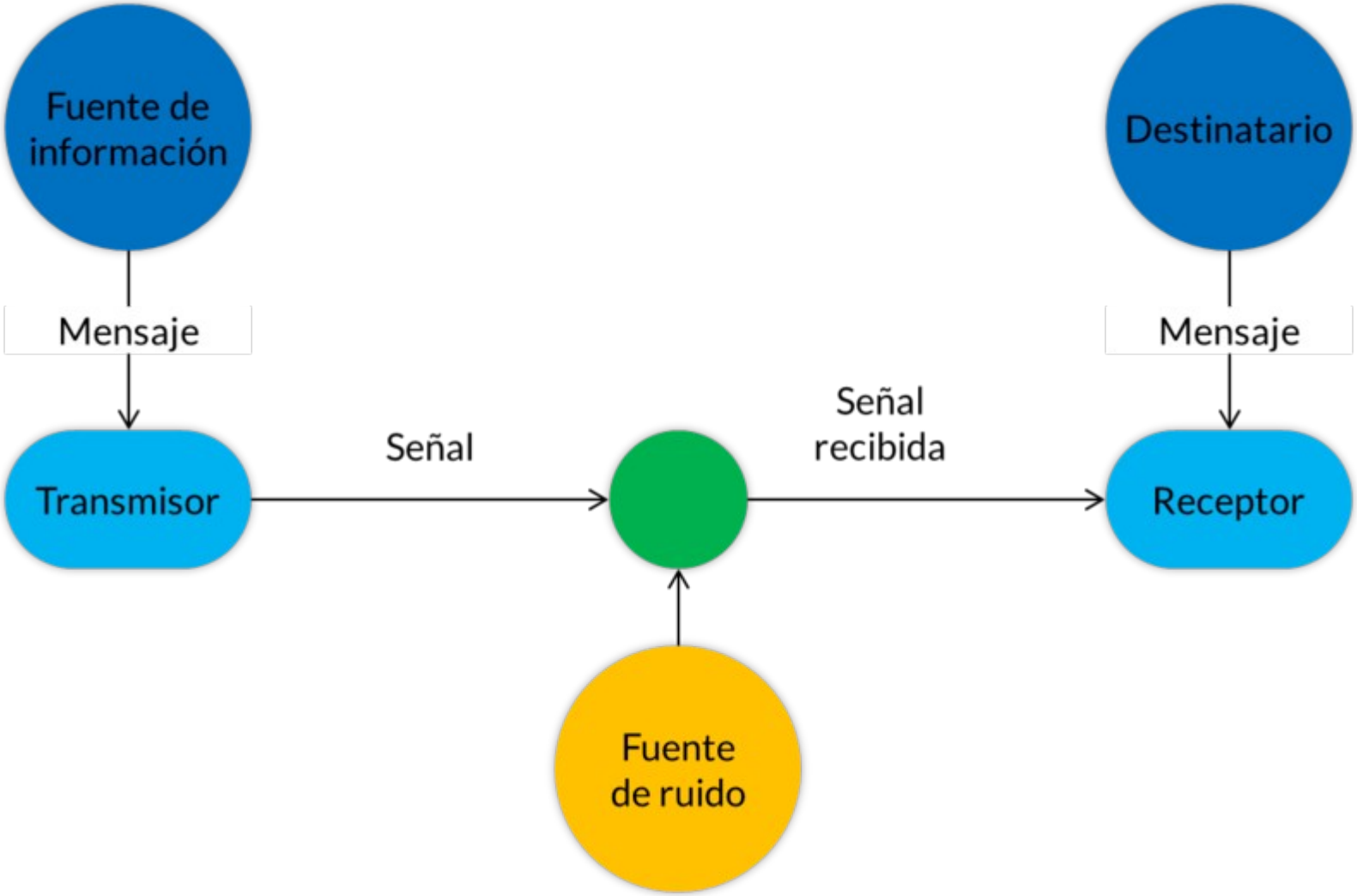
Ingeniería en Inteligencia Artificial

Año 2025

Prof. Juan Pablo Taulamet

consultas: taulamet@unl.edu.ar

MODELO DE COMUNICACIÓN DE SHANNON



1-1. Lo que no es la teoría de la información.

Teoría de la información es un nombre muy significativo para designar una disciplina científica; al aplicarse, sin embargo, al tema de que trata este libro puede resultar algo decepcionante. Los orígenes de la teoría de la información datan de la publicación, por Claude E. Shannon, de un artículo en el *Bell System Technical Journal* en 1948 (Shannon, 1948) *. Shannon, dándose quizás cuenta de las cualidades poco atractivas de la palabra *información*, tituló su artículo «Una teoría matemática de la comunicación». Si nos referimos al significado usual de la palabra información, el artículo de Shannon trata de sus soportes, los símbolos, y no de la información misma. Estudia más bien la comunicación y los medios de comunicación que el, llamémosle, producto final de ella, la información.

(Abramson,p.15)

1-2. Lo que es la teoría de la información.

El primer paso en nuestro estudio de la información consistirá en la definición de una medida de la información, investigando sus propiedades. Estas propiedades darán un sentido más práctico a la medida y ayudarán a relacionar la teoría matemática con el modelo físico que la motivó. Es importante resaltar, sin embargo, que la justificación de

(Abramson,p.16)

1-3. Codificación de la información.

Con objeto de exponer las ideas básicas de la teoría de la información, consideremos algunos ejemplos de transmisión de información. Nos limitaremos, en principio, a considerar un tipo particular pero importante de información, la información binaria. (Abramson, p17)

TABLA 1-2. CÓDIGO BINARIO

Símbolos mensaje	*Palabras código*
s_1	0
s_2	01
s_3	001
s_4	111

(Abramson,p.18)

TABLA 1-4. ESTADO DEL TIEMPO EN SAN FRANCISCO

Mensajes	*Probabilidades*
Soleado	1/4
Nublado	1/4
Lluvia	1/4
Niebla	1/4

(Abramson;p20)

Código A

Soleado	00
Nublado ...	01
Lluvia	10
Niebla	11

(soleado, niebla, niebla, nublado)

00111101

Promedio de bits por mensaje: 2

TABLA 1-5. ESTADO DEL TIEMPO EN LOS ANGELES

Mensajes	*Probabilidades*
Soleado	1/4
Nublado	1/8
Lluvia	1/8
Bruma	1/2

(Abramson;p21)

Código B

Soleado	10
Nublado	110
Lluvia	1110
Bruma	0

(soleado, bruma, bruma, nublado)

1000110

Promedio de cant. bits por mensaje:

Código B: 15/8
Código A: 2 = 16/8

consultas: taulamet@unl.edu.ar

CAPITULO 2

LA INFORMACION Y SUS FUENTES

2-1. Definición de información.

Definición. Sea E un suceso que puede presentarse con probabilidad $P(E)$. Cuando E tiene lugar, decimos que hemos recibido

$$I(E) = \log \frac{1}{P(E)} \qquad (2\text{-}1)$$

unidades de información.

La elección de la base del logaritmo que interviene en la definición equivale a elegir una determinada unidad, ya que,

$$\log_a x = \frac{1}{\log_b a} \log_b x \qquad (2\text{-}2)$$

(Abramson,p.25)

Si introducimos el logaritmo de base 2, la unidad correspondiente se denomina *bit* *

$$I(E) = \log_2 \frac{1}{P(E)} \quad \text{bits} \qquad (2\text{-}3a)$$

Empleando logaritmos naturales, la unidad de información recibe el nombre de *nat* **.

$$I(E) = \ln \frac{1}{P(E)} \quad \text{nats} \qquad (2\text{-}3b)$$

En el caso de logaritmos de base 10, la unidad de información es el Hartley. R. V. Hartley fue quien primero sugirió la medida logarítmica de la información (Hartley, 1928).

$$I(E) = \log_{10} \frac{1}{P(E)} \quad \text{Hartleys} \qquad (2\text{-}3c)$$

En general, empleando logaritmos de base r,

$$I(E) = \log_r \frac{1}{P(E)} \quad \text{unidades de orden } r \qquad (2\text{-}3d)$$

(Abramson,p.26)

$$I(E) = \log_2 \frac{1}{P(E)} \quad \text{bits} \qquad (2\text{-}3a)$$

$$\log_a x = \frac{1}{\log_b a} \log_b x \qquad (2\text{-}2)$$

De la relación (2-2), vemos que

$$1 \text{ Hartley} = 3{,}32 \text{ bits} \qquad (2\text{-}4a)$$

$$1 \text{ nat} = 1{,}44 \text{ bits} \qquad (2\text{-}4b)$$

* En adelante escribiremos el logaritmo en base 2 de x simplemente como $\log x$, omitiendo el subíndice 2 del «log». Asimismo, expresaremos el logaritmo natural como $\ln x$. En todos los demás casos indicaremos la base mediante un subíndice (p. e., $\log_{10} x$).

(Abramson,p.26)

La probabilidad de que aparezca es precisamente $P(s_i)$, de modo que la cantidad *media* de información por símbolo de la fuente es

$$\sum_S P(s_i)\, I(s_i) \quad \text{bits}$$

donde $\sum_S$ indica la suma extendida a q símbolos de la fuente S. Esta magnitud, cantidad media de información por símbolo de la fuente, recibe el nombre de *entropía* $H(S)$ de la fuente de memoria nula *.

$$H(S) \triangleq \sum_S P(s_i) \log \frac{1}{P(s_i)} \quad \text{bits} \,, \qquad (2\text{-}5a)$$

(Abramson,p.28)

$$H(S) \triangleq \sum_{S} P(s_i) \log \frac{1}{P(s_i)}$$

Ejemplo 2-1. Consideremos la fuente $S = \{s_1, s_2, s_3\}$ con $P(s_1) = 1/2$ y $P(s_2) = P(s_3) = 1/4$. Entonces

$$H(S) = 1/2 \log 2 + 1/4 \log 4 + 1/4 \log 4$$

$$= 3/2 \text{ bits}$$

(Abramson,p.28)

$$H(S) \triangleq \sum_{S} P(s_i) \log \frac{1}{P(s_i)}$$

La cantidad media de información por símbolo de la fuente, recibe el nombre de entropía H (S) de la fuente de memoria nula.

También puede ser pensada como el valor medio de la incertidumbre de un observador antes de conocer la salida de la fuente.

(Abramson,p.28)

2-3. Propiedades de la entropía.

(Abramson,p.29)

consultas: taulamet@unl.edu.ar

2-3. Propiedades de la entropía.

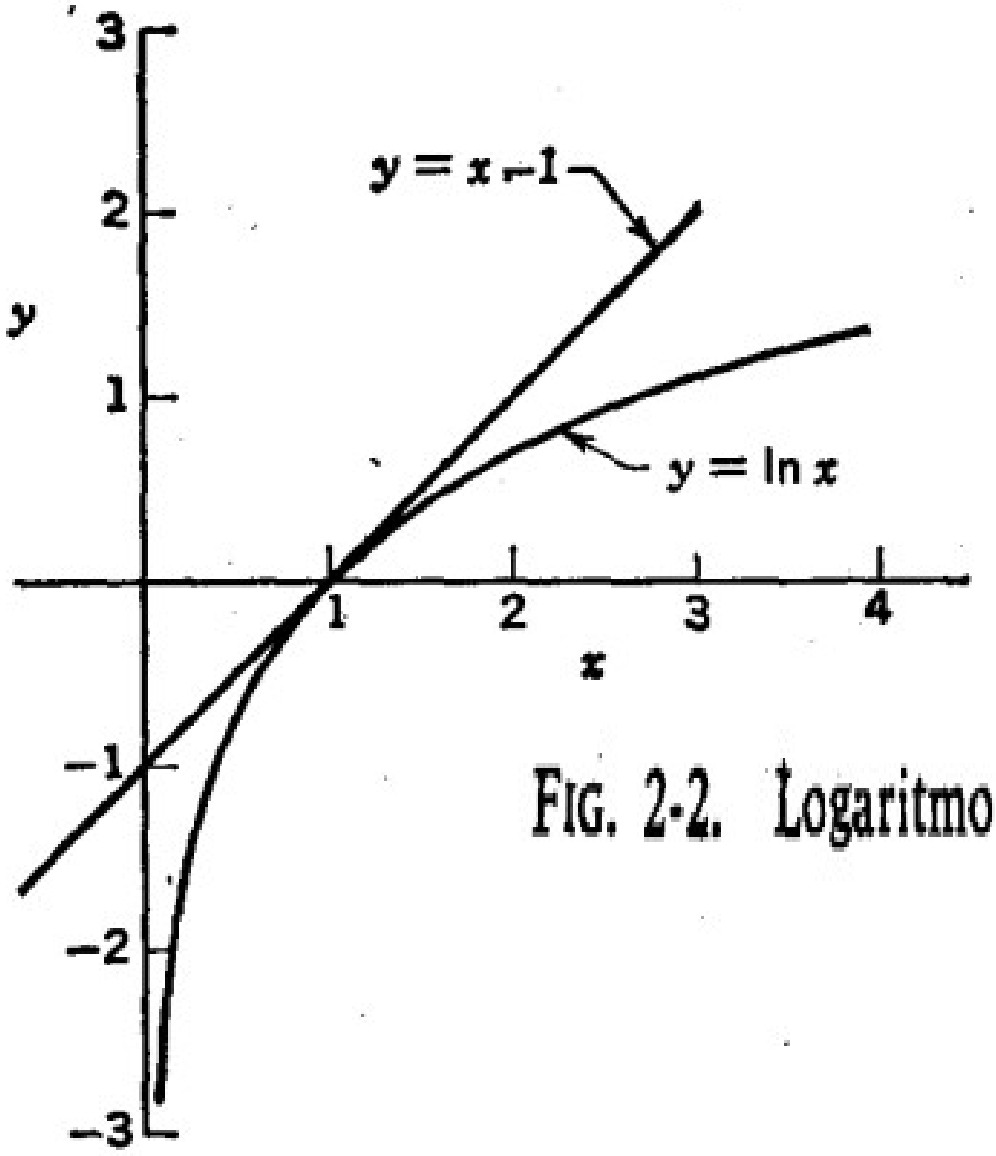


FIG. 2-2. Logaritmo natural de x y $x - 1$.

(Abramson,p.29)

2-3. Propiedades de la entropía.

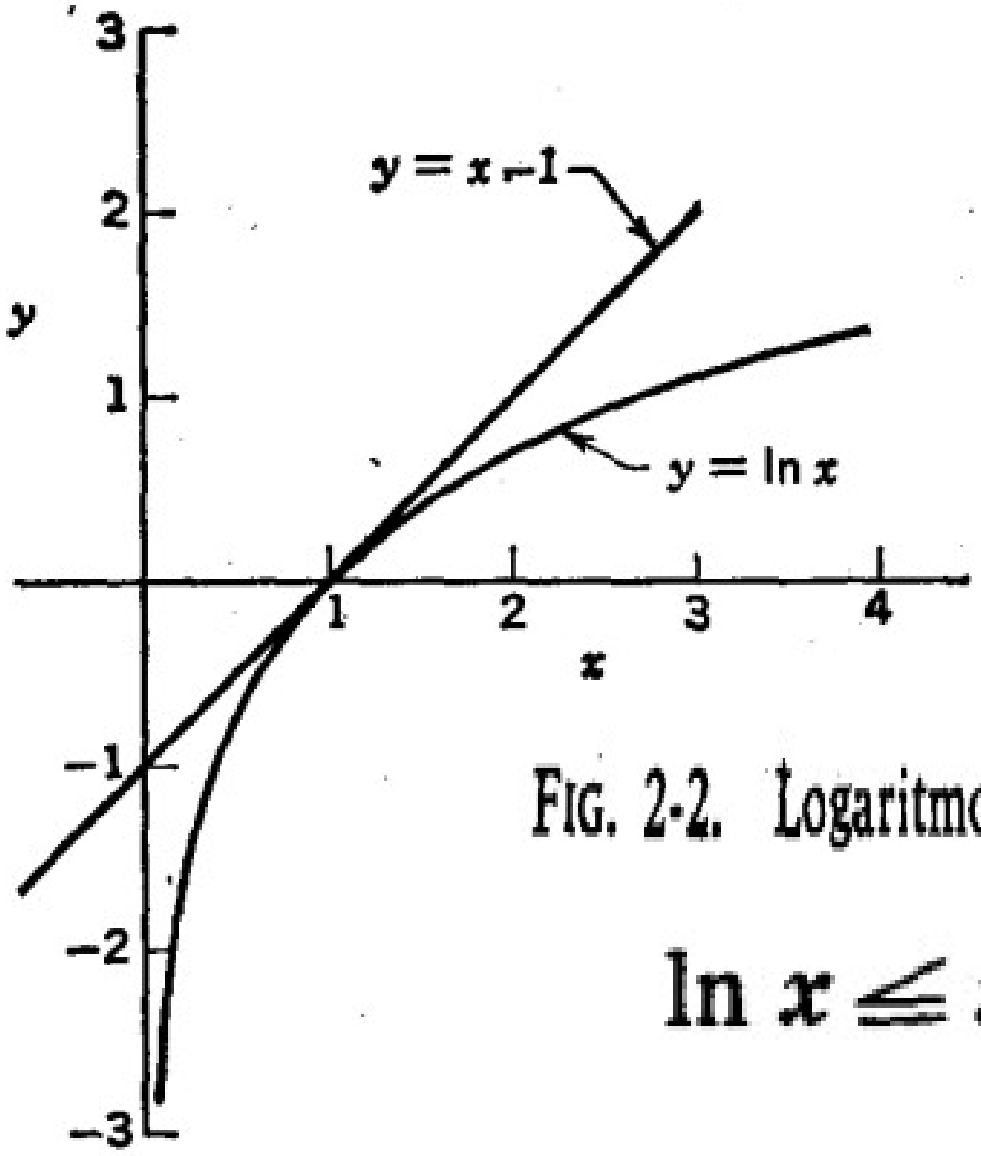


FIG. 2-2. Logaritmo natural de x y $x - 1$.

$$\ln x \leqq x - 1$$

(Abramson,p.29)

2-3. Propiedades de la entropía.

$$\ln x \leqq x - 1$$

Multiplicando por -1 a ambos lados

$$\ln \frac{1}{x} \geqq 1 - x$$

Lo cual puede trabajarse hasta llegar a:

$$\sum_{i=1}^{q} x_i \log \frac{1}{x_i} \leqq \sum_{i=1}^{q} x_i \log \frac{1}{y_i}$$

(Abramson,p.30)

ENTROPÍA DE UNA FUENTE DE MEMORIA NULA

Supongamos una fuente de memoria nula, definida por su alfabeto $S = \{s_i\}$, $i = 1, 2, \ldots, q$, y sus probabilidades $P(s_i)$, $i = 1, 2, \ldots, q$. La $H(S)$ viene dada por

$$H(S) = \sum_{i=1}^{q} P_i \log \frac{1}{P_i} \qquad (2\text{-}9)$$

Puede demostrarse que:

$$\log q \geqq H(S) = \sum_{i=1}^{q} P_i \log \frac{1}{P_i}$$

(Abramson,p.31)

ENTROPÍA DE UNA FUENTE DE MEMORIA NULA

$$\log q \geqq H(S) = \sum_{i=1}^{q} P_i \log \frac{1}{P_i}$$

una fuente de información de memoria nula con un alfabeto de q símbolos, *el valor máximo de la entropía es precisamente log q, alcanzándose solamente si todos los símbolos de la fuente son equiprobables.*

(Abramson,p.31)

consultas: taulamet@unl.edu.ar

ENTROPÍA DE UNA FUENTE BINARIA DE MEMORIA NULA

En tal fuente, el alfabeto se reduce a $\{0, 1\}$. La probabilidad de un 0 es ω y la de un 1, $1 - \omega$. Llamaremos ϖ a $1 - \omega$. Calcularemos la entropía a partir de la fórmula (2-5)

$$H(S) = \omega \log \frac{1}{\omega} + \varpi \log \frac{1}{\varpi} \quad \text{bits} \qquad (2\text{-}12)$$

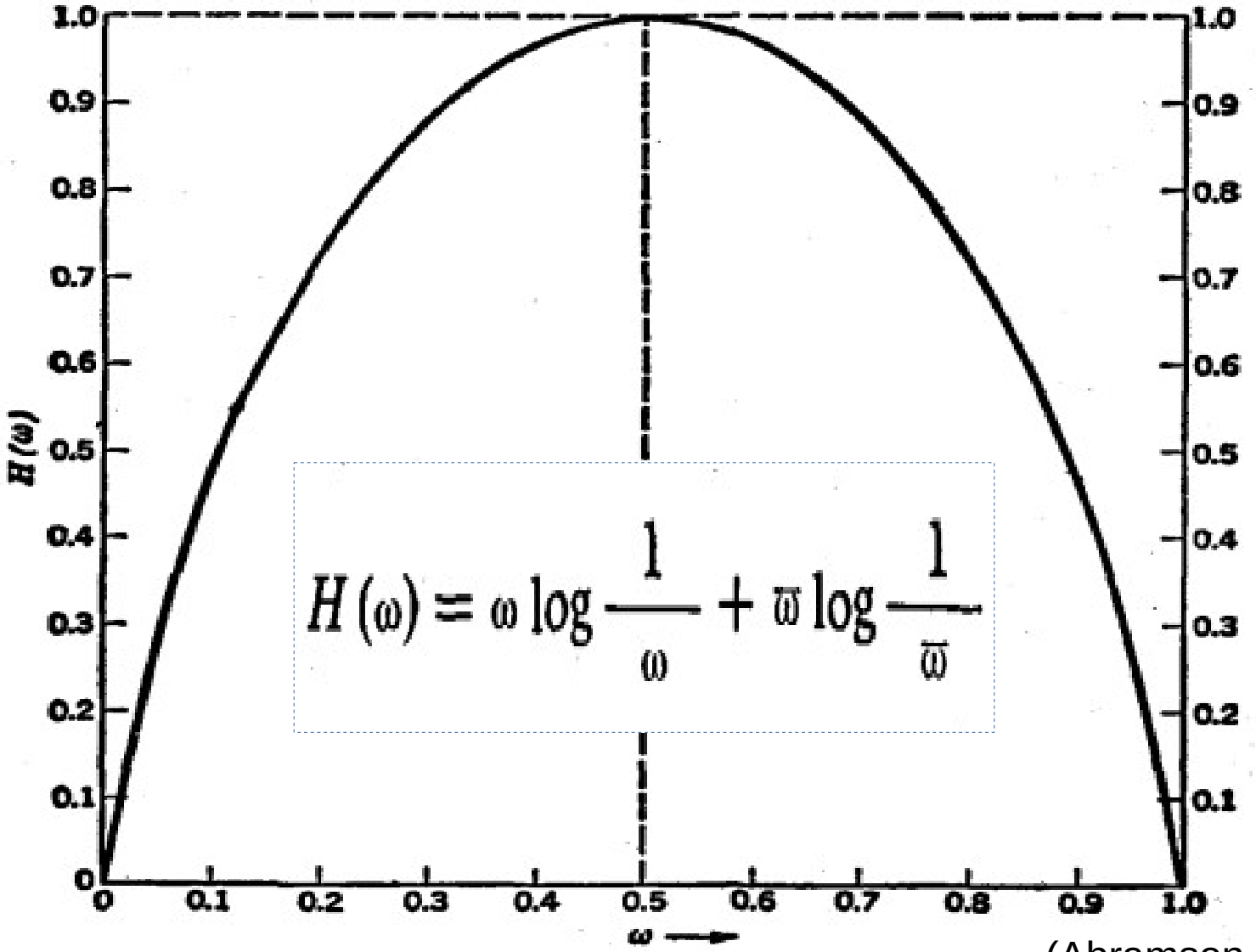
Ya que sólo depende de w, podemos definir la llamada función entropía:

$$H(\omega) = \omega \log \frac{1}{\omega} + \varpi \log \frac{1}{\varpi}$$

(Abramson,p.32)

consultas: taulamet@unl.edu.ar

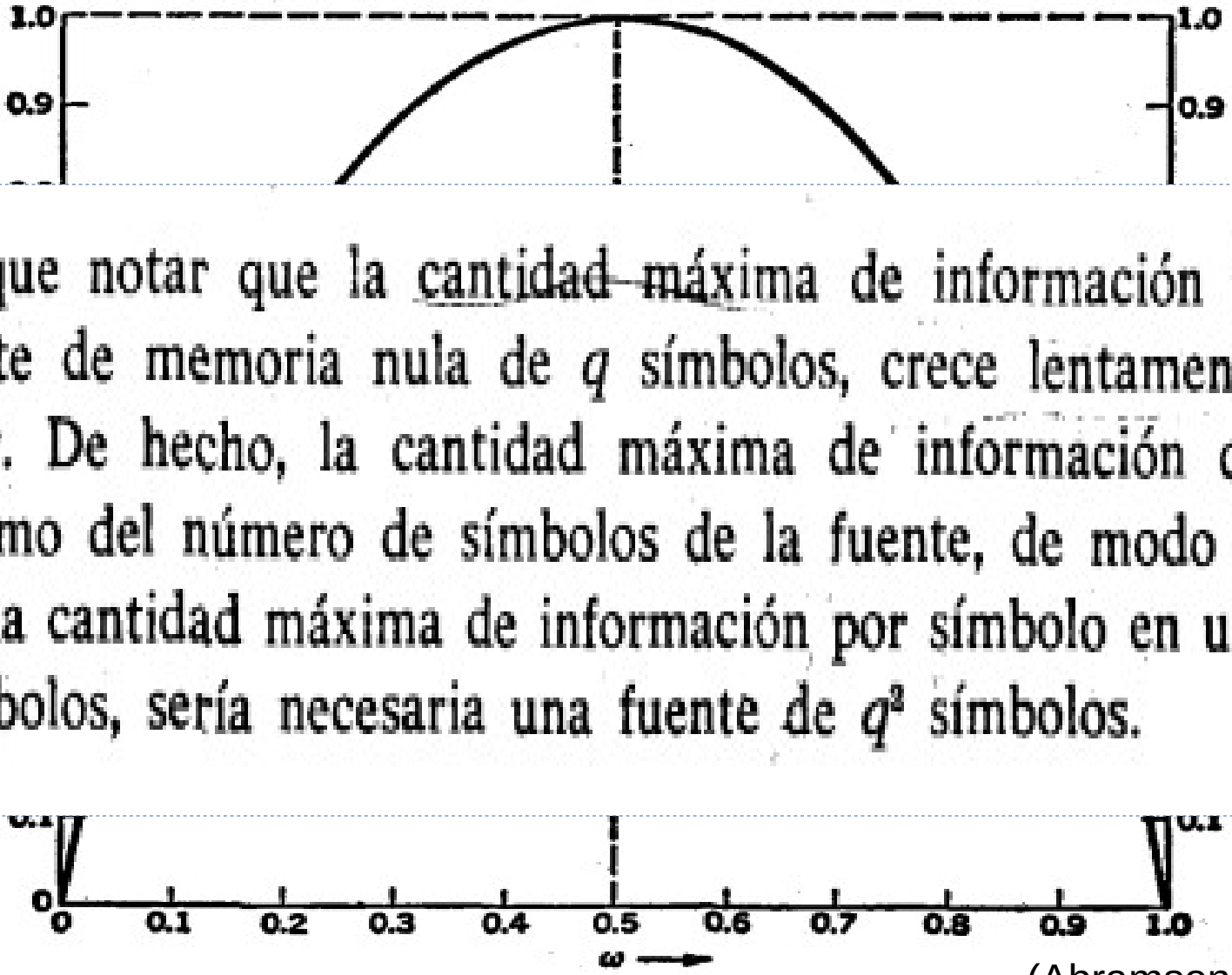
FUNCIÓN ENTROPÍA



(Abramson,p.33)

consultas: taulamet@unl.edu.ar

FUNCIÓN ENTROPÍA



Hay que notar que la cantidad máxima de información dada por una fuente de memoria nula de q símbolos, crece lentamente al aumentar q. De hecho, la cantidad máxima de información crece con el logaritmo del número de símbolos de la fuente, de modo que para duplicar la cantidad máxima de información por símbolo en una fuente de q símbolos, sería necesaria una fuente de q^2 símbolos.

(Abramson,p.33)

EXTENSIONES DE ORDEN N

Definición. Sea S una fuente de información de memoria nula, con un alfabeto $\{s_1, s_2, \ldots, s_q\}$. Sea P_i la probabilidad correspondiente a s_i. La extensión de orden n de S, S^n, es una fuente de memoria nula de q^n símbolos, $\{\sigma_1, \sigma_2, \ldots, \sigma_{q^n}\}$. El símbolo σ_i corresponde a una secuencia de n de los sq símbolos. La probabilidad de σ_i, $P(\sigma_i)$, es precisamente la probabilidad de la secuencia correspondiente. Es decir, si σ_i representa la secuencia $(s_{i_1}, s_{i_2}, \ldots s_{i_n})$, $P(\sigma_i) = P_{i_1} \cdot P_{i_2} \ldots P_{i_n}$.

(Abramson,p.34)

ENTROPÍA DE UNA EXTENSIÓN DE ORDEN N

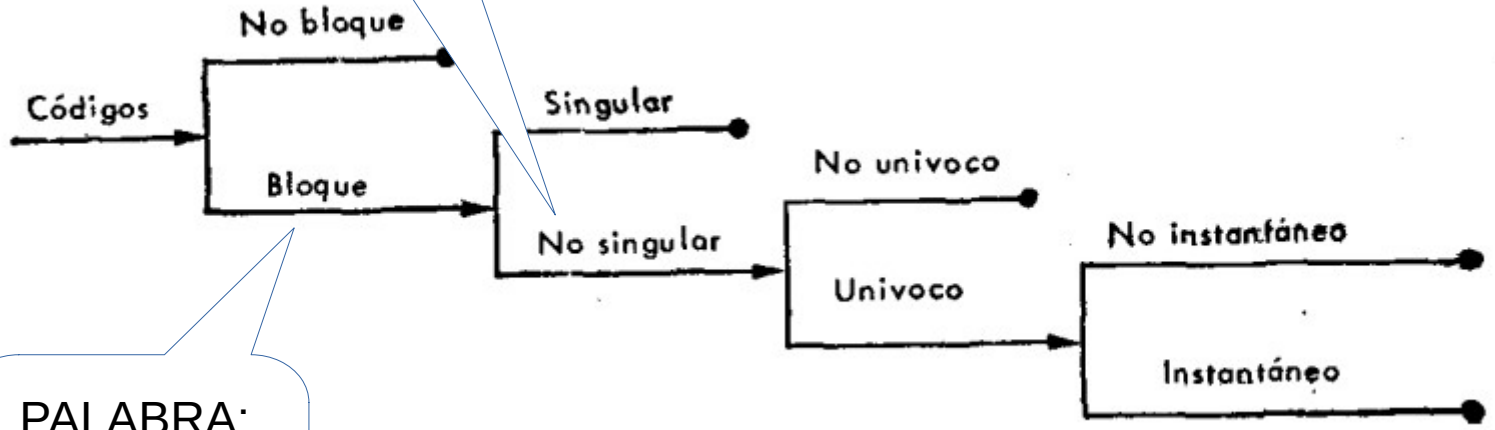
Definición. Sea S una fuente de información de memoria nula, con un alfabeto $\{s_1, s_2, \ldots, s_q\}$. Sea P_i la probabilidad correspondiente a s_i. La extensión de orden n de S, S^n, es una fuente de memoria nula de q^n símbolos, $\{\sigma_1, \sigma_2, \ldots, \sigma_{q^n}\}$. El símbolo σ_i corresponde a una secuencia de n de los sq símbolos. La probabilidad de σ_i, $P(\sigma_i)$, es precisamente la probabilidad de la secuencia correspondiente. Es decir, si σ_i representa la secuencia $(s_{i_1}, s_{i_2}, \ldots s_{i_n})$, $P(\sigma_i) = P_{i_1} \cdot P_{i_2} \ldots P_{i_n}$.

$$H(S^n) = n\,H(S)$$

(Abramson,p.35)

SUB-CLASES DE CÓDIGOS

Todas las palabras son distintas



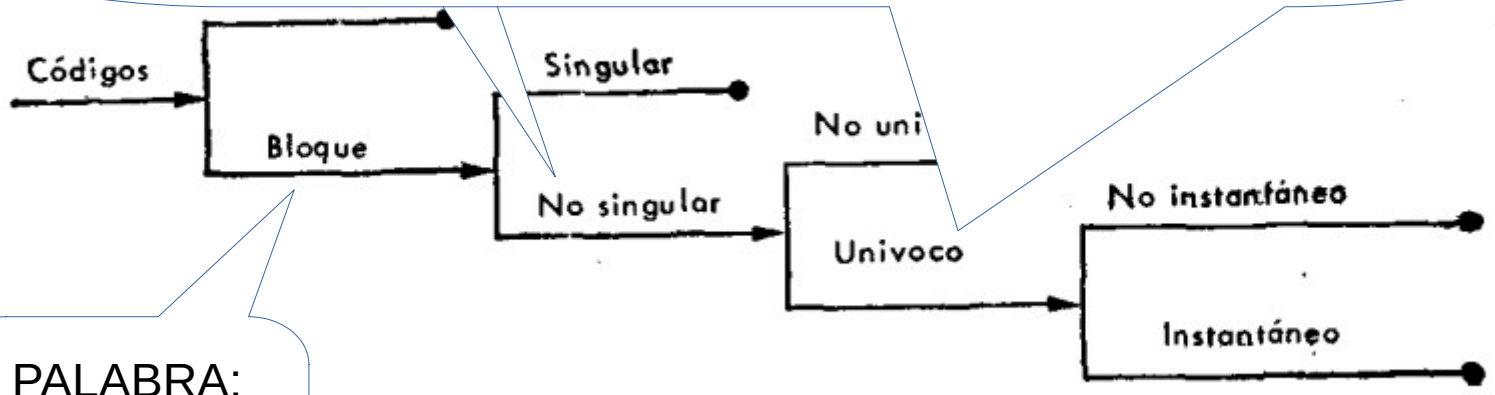
PALABRA: Secuencia símbolos (bits)

(Abramson,p.66-67)

consultas: taulamet@unl.edu.ar

SUB-CLASES DE CÓDIGOS

Definición. Un código bloque se dice *unívocamente decodificable* si, y solamente si, su extensión de orden n es no singular para cualquier valor finito de n.



PALABRA: Secuencia fija de símbolos (bits)

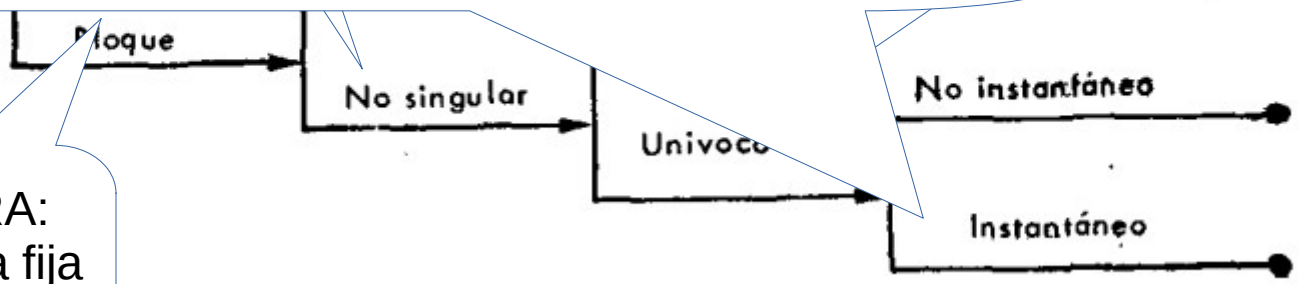
(Abramson,p.66-67)

consultas: taulamet@unl.edu.ar

SUB-CLASES DE CÓDIGOS

Definición. Un código bloque se dice *unívocamente deco-*
n es

La condición necesaria y suficiente para que un código sea instantáneo es que ninguna palabra del código coincida con el prefijo de otra.



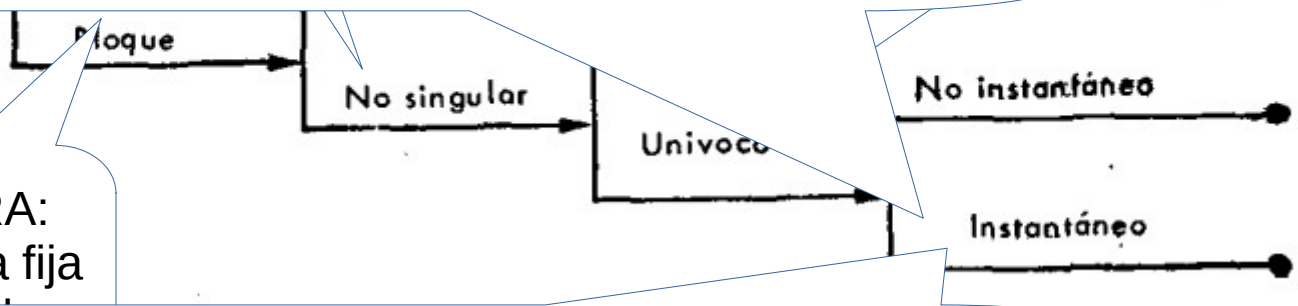
PALABRA: Secuencia fija de símbolos (bits)

(Abramson,p.66-67)

SUB-CLASES DE CÓDIGOS

Definición. Un código bloque se dice *unívocamente deco-*
n es

La condición necesaria y suficiente para que un código sea instantáneo es que ninguna palabra del código coincida con el prefijo de otra.



PALABRA:
Secuencia fija

Definición. Un código unívocamente decodificable se denomina *instantáneo* cuando es posible decodificar las palabras de una secuencia sin precisar el conocimiento de los símbolos que las suceden.

,p.66-67)

INECUACIÓN DE KRAFT (1949)

La condición necesaria y suficiente para la existencia de un código instantáneo de longitudes $l_1, l_2, \ldots, l_q$ es que

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

donde r es el número de símbolos diferentes que constituyen el alfabeto código.

(Abramson,p.69)

consultas: taulamet@unl.edu.ar

INECUACIÓN DE KRAFT (1949)

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

$$\sum_{i=1}^{b} 2^{-l_i} \leq 1$$

CASO BINARIO

(Abramson,p.69)

consultas: taulamet@unl.edu.ar

EJEMPLO DE CONSECUENCIAS

TABLA 3-7. UN CÓDIGO BINARIO PARA DÍGITOS DECIMALES

Dígitos decimales	*Codigó binario*
0	0
1	10
2	11 ...
3	11 ...
4	11 ...
5	11 ...
6	11 ...
7	11 ...
8	11 ...
9	11 ...

¿Cuántos bits necesitamos?

l?

(Abramson,p.72)

consultas: taulamet@unl.edu.ar

EJEMPLO DE CONSECUENCIAS

TABLA 3-7. UN CÓDIGO BINARIO PARA DÍGITOS DECIMALES

Dígitos decimales	*Codigó binario*
0	0
1	10

¿Cuántos bits necesitamos?

$$\sum_{i=0}^{9} 2^{-l_i} \leq 1 \qquad (3\text{-}5)$$

Por hipótesis $l_0 = 1$, $l_1 = 2$ y $l_2 = l_3 = \ldots = l_9 = l$. Introduciendo estos valores en (3-5), encontramos

$$1/2 + 1/4 + 8\,(2^{-l}) \leq 1$$

o

$$l \geq 5 \qquad (3\text{-}6)$$

(Abramson,p.72)

EJEMPLO DE CONSECUENCIAS

TABLA 3-7. UN CÓDIGO BINARIO PARA DÍGITOS DECIMALES

Dígitos decimales	*Codigó binario*
0	0
1	10
2	11000
3	11001
4	11010
5	11011
6	11100
7	11101
8	11110
9	11111

¿Cuántos bits necesitamos?

5 o más

$$l \geq 5 \qquad (3\text{-}6)$$

(Abramson,p.72)

LONGITUD MEDIA DE UN CÓDIGO

Definición. Sea un código bloque que asocia los símbolos de una fuente $s_1, s_2, \ldots, s_q$ con las palabras $X_1, X_2, \ldots, X_q$. Supongamos que las probabilidades de los símbolos de la fuente son $P_1, P_2, \ldots, P_q$ y las longitudes de las palabras $l_1, l_2, \ldots, l_q$. Definiremos la *longitud media* del código, L, por la ecuación

$$L = \sum_{i=1}^{q} P_i l_i$$

(Abramson,p.81)

consultas: taulamet@unl.edu.ar

CANALES DE INFORMACIÓN

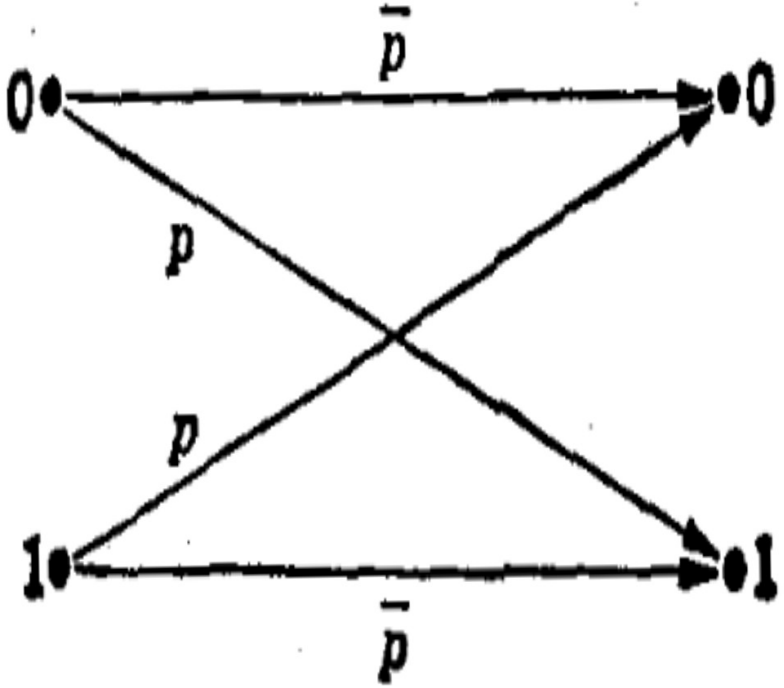
Definición. Un canal de información * viene determinado por un alfabeto de entrada $A = \{a_i\}$, $i = 1, 2, \ldots, r$; un alfabeto de salida $B = \{b_j\}$, $j = 1, 2, \ldots, s$; y un conjunto de probabilidades condicionales $P(b_j/a_i)$. $P(b_j/a_i)$ es la probabilidad de recibir a la salida el símbolo b_j cuando se envía el símbolo de entrada a_i.

(Abramson,p.35)

consultas: taulamet@unl.edu.ar

REPRESENTACIONES

$$\begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$



CANAL BINARIO SIMÉTRICO (BSC)

(Abramson,p.112-114)

consultas: taulamet@unl.edu.ar

EXTENSIONES DE UN CANAL

La extensión de orden n de un canal se obtiene meramente considerando bloques de símbolos de longitud n.

$$\Pi = \begin{bmatrix} \bar{p}^2 & \bar{p}p & p\bar{p} & p^2 \\ \bar{p}p & \bar{p}^2 & p^2 & p\bar{p} \\ p\bar{p} & p^2 & \bar{p}^2 & \bar{p}p \\ p^2 & p\bar{p} & \bar{p}p & \bar{p}^2 \end{bmatrix}$$

FIG. 5-4. Matriz del canal $(BSC)^2$.

(Abramson,p.115)

consultas: taulamet@unl.edu.ar

EXTENSIONES DE UN CANAL

$$\mathbf{\Pi} = \begin{bmatrix} \bar{p}^2 & \bar{p}p & p\bar{p} & p^2 \\ \bar{p}p & \bar{p}^2 & p^2 & p\bar{p} \\ p\bar{p} & p^2 & \bar{p}^2 & \bar{p}p \\ p^2 & p\bar{p} & \bar{p}p & \bar{p}^2 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix} \qquad \mathbf{\Pi} = \begin{bmatrix} \bar{p}\mathbf{P} & p\mathbf{P} \\ p\mathbf{P} & \bar{p}\mathbf{P} \end{bmatrix}$$

(Abramson,p.115)

consultas: taulamet@unl.edu.ar

EXTENSIONES DE UN CANAL

$$\mathbf{\Pi} = \begin{bmatrix} \bar{p}^2 & \bar{p}p & p\bar{p} & p^2 \\ \bar{p}p & \bar{p}^2 & p^2 & p\bar{p} \\ p\bar{p} & p^2 & \bar{p}^2 & \bar{p}p \\ p^2 & p\bar{p} & \bar{p}p & \bar{p}^2 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$

Puede apreciarse que la matriz del $(BSC)^2$ se expresa como una matriz de matrices. Sea **P**, igual que antes, la matriz del canal BSC. Entonces, la matriz del $(BSC)^2$ puede escribirse en la forma

$$\mathbf{\Pi} = \begin{bmatrix} \bar{p}\mathbf{P} & p\mathbf{P} \\ p\mathbf{P} & \bar{p}\mathbf{P} \end{bmatrix}$$

Esta matriz se conoce como *cuadrado de Kronecker* (Bellman, 1960) (o cuadrado tensorial) de la matriz **P**. En un caso más general, la *matriz de la extensión* de orden n de un canal es la potencia *enésima* de Kronecker de la matriz del canal original. (Abramson,p.115)

consultas: taulamet@unl.edu.ar

PROBABILIDADES HACIA ADELANTE

$$\mathrm{P}(\mathrm{b}_j / a_i)$$

Es la probabilidad de que se obtenga una una determinada salida b_j sabiendo que se ha enviado una entrada a_i. Son las probabilidades que disponemos en la matriz del canal.

$$P(b_j) = \sum_{\forall i} P(b_j / a_i) * P(a_i)$$

(Abramson,p.116)

consultas: taulamet@unl.edu.ar

PROBABILIDADES HACIA ATRÁS

$$\mathrm{P}(\mathrm{a}_i / b_j)$$

Es la probabilidad de que haya sido enviada una entrada a_i, sabiendo que se ha recibido una determinada salida b_j.

(Abramson,p.116-117)

consultas: taulamet@unl.edu.ar

PROBABILIDADES CONDICIONALES

Según la ley de Bayes, la probabilidad condicional de una entrada a_i, cuando se recibe una salida b_j, viene dada por la fórmula

$$P(a_i/b_j) = \frac{P(b_j/a_i)\,P(a_i)}{P(b_j)} \qquad (5\text{-}7a)$$

$$P(a_i/b_j) = \frac{P(b_j/a_i)\,P(a_i)}{\sum_{i=1}^{r} P(b_j/a_i)\,P(a_i)}$$

(Abramson,p.116)

consultas: taulamet@unl.edu.ar

ENTROPÍAS A PRIORI Y POSTERIORI

Denominaremos $P(a_i)$ la probabilidad *a priori* de los símbolos de entrada, es decir antes de recibir un símbolo de salida determinado. $P(a_i/b_j)$ recibirá el nombre de probabilidad *a posteriori*, probabilidad despues de la recepción de b_j. Según se explicó en el apartado 2-2 puede

$$H(A) = \sum_{A} P(a) \log \frac{1}{P(a)}$$

$$H(A/b_j) = \sum_{A} P(a/b_j) \log \frac{1}{P(a/b_j)}$$

(Abramson,p.118)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

4-1. Longitud media de un código.

Definición. Sea un código bloque que asocia los símbolos de una fuente $s_1, s_2, \ldots, s_q$ con las palabras $X_1, X_2, \ldots, X_q$. Supongamos que las probabilidades de los símbolos de la fuente son $P_1, P_2, \ldots, P_q$ y las longitudes de las palabras $l_1, l_2, \ldots, l_q$. Definiremos la *longitud media* del código, L, por la ecuación

$$L = \sum_{i=1}^{q} P_i l_i \qquad (4\text{-}1)$$

81

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

4-1. Longitud media de un código.

Definición. Sea un código bloque que asocia los símbolos de una fuente $s_1, s_2, \ldots, s_q$ con las palabras $X_1, X_2, \ldots, X_q$. Supongamos que las probabilidades de los símbolos de la fuente son $P_1, P_2, \ldots, P_q$ y las longitudes de las palabras $l_1, l_2, \ldots$ digo, L, por

Definición. Consideremos un código unívoco que asocia los símbolos de una fuente S con palabras formadas por símbolos de un alfabeto r-ario. Este código será *compacto* (respecto a S) si su longitud media es igual o menor que la longitud media de todos los códigos unívocos que pueden aplicarse a la misma fuente y el mismo alfabeto.

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

Consideremos una fuente de memoria nula, cuyos símbolos, s_1, s_2, ..., s_q tienen respectivamente las probabilidades P_1, P_2, ..., P_q. Supongamos un código bloque que codifica estos símbolos en un alfabeto de r símbolos, y definimos por l_i la longitud de la palabra correspondiente a s_i. La entropía de esta fuente de memoria nula será, entonces

$$H(S) = -\sum_{i=1}^{q} P_i \log P_i \qquad (4\text{-}2)$$

Definición. Consideremos un código unívoco que asocia los símbolos de una fuente S con palabras formadas por símbolos de un alfabeto r-ario. Este código será *compacto* (respecto a S) si su longitud media es igual o menor que la longitud media de todos los códigos unívocos que pueden aplicarse a la misma fuente y el mismo alfabeto.

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

Consideremos una fuente de memoria nula, cuyos símbolos, s_1, s_2, ..., s_q tienen respectivamente las probabilidades P_1, P_2, ..., P_q. Supongamos un código bloque que codifica estos símbolos en un alfabeto de r símbolos, y definimos por l_i la longitud de la palabra correspondiente a s_i. La entropía de esta fuente de memoria nula será, entonces

$$H(S) = -\sum_{i=1}^{q} P_i \log P_i \qquad (4\text{-}2)$$

Sean Q_1, Q_2, ..., Q_q números tales que $Q_i \geqq 0$ para cualquier valor de i, y $\sum_{i=1}^{q} Q_i = 1$. Debido a (2-8), sabemos que

$$\sum_{i=1}^{q} P_i \log \frac{1}{P_i} \leqq \sum_{i=1}^{q} P_i \log \frac{1}{Q_i} \qquad (4\text{-}3)$$

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

Consideremos una fuente de memoria nula, cuyos símbolos, s_1, s_2, ..., s_q tienen respectivamente las probabilidades P_1, P_2, ..., P_q. Supongamos un código bloque que codifica estos símbolos en un alfabeto de r símbolos, y definimos por l_i la longitud de la palabra correspondiente a s_i. La entropía de esta fuente de memoria nula será, entonces

$$H(S) = -\sum_{i=1}^{q} P_i \log P_i \qquad (4\text{-}2)$$

Sean Q_1, Q_2, ..., Q_q números tales que $Q_i \geqq 0$ para cualquier valor de i, y $\sum_{i=1}^{q} Q_i = 1$. Debido a (2-8), sabemos que

$$\sum_{i=1}^{q} P_i \log \frac{1}{P_i} \leqq \sum_{i=1}^{q} P_i \log \frac{1}{Q_i} \qquad (4\text{-}3)$$

igualdad solamente cuando $P_i = Q_i$, para todo valor de i. Por lo tanto

$$H(S) \leqq -\sum_{i=1}^{q} P_i \log Q_i \qquad (4\text{-}4)$$

con signo igual en el mismo caso.

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

La ecuación (4-4) será válida para cualquier conjunto de números positivos, Q_i, cuya suma sea la unidad. En consecuencia, se podrá elegir

$$Q_i = \frac{r^{-l_i}}{\sum_{i=1}^{q} r^{-l_i}} \qquad (4\text{-}5)$$

de donde

$$H(S) \leq -\sum_{i=1}^{q} P_i (\log r^{-l_i}) + \sum_{i=1}^{q} P_i \left(\log \sum_{j=1}^{q} r^{-l_j} \right)$$

$$\leq \log r \sum_{i=1}^{q} P_i l_i + \log \left(\sum_{j=1}^{q} r^{-l_j} \right)$$

$$\leq L \log r + \log \sum_{j=1}^{q} r^{-l_j} \qquad (4\text{-}6)$$

$$H(S) \leq -\sum_{i=1}^{q} P_i \log Q_i \qquad (4\text{-}4)$$

con signo igual en el mismo caso.

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

La ecuación (4-4) será válida para cualquier conjunto de números positivos, Q_i, cuya suma sea la unidad. En consecuencia, se podrá elegir

Si exigimos que el código sea instantáneo, la inecuación de Kraft impone que el argumento del segundo logaritmo del segundo miembro de (4-6) sea igual o menor que la unidad. Por lo tanto, su logaritmo deberá ser igual o menor que cero, y

$$H(S) \leq L \log r \qquad (4\text{-}7a)$$

o bien

$$\frac{H(S)}{\log r} \leq L \qquad (4\text{-}7b)$$

$H(S)$ viene medida en bits en la ecuación (4-7*b*). Recordemos que L es el número medio de símbolos utilizados para codificar S. Expresando la entropía asimismo en unidades r-arias, como en (2-5*c*), la relación (4-7*b*) podría escribirse en la forma

$$H_r(S) \leq L \qquad (4\text{-}7c)$$

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

La ecuación (4-4) será válid

Es importante destacar que la relación (4-7) marca un hito en el estudio de la teoría de la información. Esta ecuación constituye el primer indicio demostrativo de la relación existente entre la definición de información y una cantidad (en este caso L) *que no depende de la definición*. Con esta ecuación se comienza a desarrollar la justificación de nuestra medida de información.

$$\frac{H(S)}{\log r} \leq L \qquad (4\text{-}7b)$$

$H(S)$ viene medida en bits en la ecuación (4-7*b*). Recordemos que L es el número medio de símbolos utilizados para codificar S. Expresando la entropía asimismo en unidades r-arias, como en (2-5*c*), la relación (4-7*b*) podría escribirse en la forma

$$H_r(S) \leq L \qquad (4\text{-}7c)$$

con signo igual en el mismo caso.

(Abramson,pp.81-84)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

En el apartado anterior se ha resuelto el problema de la codificación de una fuente de memoria nula con símbolos cuyas probabilidades tienen la forma $(1/r)^{\alpha_i}$. Dedicaremos a continuación nuestra atención a las fuentes de memoria nula cuyos símbolos tienen probabilidades arbitrarias.

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

En el apartado anterior se ha resuelto el problema de la codifica-

La ecuación (4-9*b*) dice que si $\log_r (1/P_i)$ es un número entero, l_i lida- debe hacerse igual a este valor. Si no lo es, parece lógico formar un ten- código compacto eligiendo un l_i igual al número entero inmediatamen- bili- te superior a $\log_r (1/P_i)$. De hecho esta conjetura *no* es correcta, pero seleccionando l_i de acuerdo con esta regla se obtendrán algunos resultados interesantes. Por lo tanto, se hará l_i igual al número entero que satisface la relación

$$\log_r \frac{1}{P_i} \leq l_i < \log_r \frac{1}{P_i} + 1 \qquad (4\text{-}10)$$

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

En el apartado anterior se ha resuelto el problema de la codifica-

La ecuación (4-9*b*) dice que si $\log_r (1/P_i)$ es un número entero, l_i lida-debe hacerse igual a este valor. Si no lo es, parece lógico formar un ten-código compacto eligiendo un l_i igual al número entero inmediatamen-bili-te superior a $\log_r (1/P_i)$. De hecho esta conjetura *no* es correcta, pero seleccionando l_i de acuerdo con esta regla se obtendrán algunos resultados interesantes. Por lo tanto, se hará l_i igual al número entero que satisface la relación

En primer lugar, se comprobará que las longitudes definidas por este procedimiento cumplen la inecuación de Kraft y son, en consecuencia, aceptables para constituir un código instantáneo. Hallando el antilogaritmo de la primera inecuación de (4-10) se encuentra

$$\frac{1}{P_i} \leq r^{l_i}$$

o bien

$$P_i \geq r^{-l_i} \qquad (4\text{-}11)$$

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

En el apartado anterior se ha resuelto el problema de la codifica-

La ecuación (4-9*b*) dice que si $\log_r (1/P_i)$ es un número entero [illegible] lida-
debe hacerse igual a este valor. Si no lo es, pare[illegible] ten-
código compacto eligiendo un l_i igu[illegible]
te superior a log $(1/P$[illegible] ibili-
se[illegible]
s[illegible]

Sumando esta expresión, extendida a todos los valores de i, se obtiene

$$1 \geqq \sum_{i=1}^{q} r^{-l_i}$$

Que demuestra que (4-10) define un conjunto de l_i válido para un código instantáneo.

Multiplicando (4-10) por P_i y sumando para todos los valores de i se obtiene

$$H_r(S) \leqq L < H_r(S) + 1 \qquad (4\text{-}12)$$

o

$$P_i \geqq r^{-l_i} \qquad (4\text{-}11)$$

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

tal que existe entre (4-12) y el valor mínimo de L definido por (4-7). Las ecuaciones (4-7) determinan el valor mínimo de la longitud media L, independientemente del sistema de codificación empleado. El único requisito exigido es que el código sea instantáneo. La ecuación (4-12), por otra parte, se dedujo admitiendo el procedimiento de codificación definido en (4-10). En definitiva, ambas ecuaciones definen los valores máximo y mínimo de L, *válidos al utilizar el método de codificación enunciado en* (4-10).

Puesto que (4-12) puede aplicarse a cualquier fuente de memoria nula, lo haremos a la extensión de orden n de la fuente original

$$H_r(S^n) \leq L_n < H_r(S^n) + 1 \qquad (4\text{-}13)$$

L_n representa la longitud media de las palabras correspondientes a los *símbolos de la extensión de orden n* de la fuente S. Esto es, si λ_i es la longitud de la palabra correspondiente al símbolo σ_i y $P(\sigma_i)$ la probabilidad de σ_i, entonces

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

tal que existe entre (4-12) y el valor mínimo de L definido por (4-7). Las ecuaciones (4-7) determinan el valor mínimo de la longitud media L, independientemente del sistema de codificación empleado. El único requisito exigido es que el código sea instantáneo. La ecuación (4-12), por otra parte, se dedujo admitiendo el procedimiento de codificación definido en (4-10). En definitiva, ambas ecuaciones definen los valores máximo y mínimo de L, *válidos al utilizar el método de codificación enunciado en* (4-10).

Puesto que (4-12) puede aplicarse a cualquier fuente de memoria nula, lo haremos a la extensión de orden n de la fuente original

$$H_r(S^n) \leq L_n < H_r(S^n) + 1 \qquad (4\text{-}13)$$

L_n representa la longitud media de las palabras correspondientes a los *símbolos de la extensión de orden n* de la fuente S. Esto es, si λ_i es la longitud de la palabra correspondiente al símbolo σ_i y $P(\sigma_i)$ la probabilidad de σ_i, entonces

$$L_n = \sum_{i=1}^{q^n} P(\sigma_i)\,\lambda_i \qquad (4\text{-}14)$$

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

L_n/n, por lo tanto, es el número medio de símbolos * empleados en cada símbolo simple de S. Según (2-16), la entropía de S^n es igual a n veces la entropía de S. La ecuación (4-13) puede, entonces, escribirse en la forma

$$H_r(S) \leq \frac{L_n}{n} < H_r(S) + \frac{1}{n} \qquad (4\text{-}15a)$$

de modo que siempre será posible encontrar un valor de L_n/n tan próximo a $H_r(S)$ como queramos, sin más que codificar la extensión de orden n de S, en lugar de S:

$$\lim_{n \to \infty} \frac{L_n}{n} = H_r(S) \qquad (4\text{-}15b)$$

(Abramson,pp.88-89)

consultas: taulamet@unl.edu.ar

1° TEOREMA DE SHANNON

L_n/n, por lo tanto, es el número medio de símbolos * empleados en cada símbolo simple de S. Según (2-16), la entropía de S^n es igual a n veces la entropía de S. La ecuación (4-13) puede, entonces, escribirse en la forma

$$H_r(S) \leq \frac{L_n}{n} < H_r(S) + \frac{1}{n} \qquad (4\text{-}15a)$$

La ecuación (4-15*a*) se conoce como *primer teorema de Shannon* o *teorema de la codificación sin ruido*. Constituye uno de los dos teoremas fundamentales de la teoría de la información. La ecuación (4-15*a*) dice que el número medio de símbolos r-arios correspondientes a un símbolo de la fuente puede hacerse tan pequeño, pero no inferior, a la entropía de la fuente expresada en unidades de orden r. El precio que se paga por la disminución de L_n/n es un aumento en la complejidad de la codificación debido al gran número (q^n) de símbolos de la fuente que hay que manejar.

(Abramson,pp.88-90)

consultas: taulamet@unl.edu.ar

GENERALIZACIÓN

5-5. Generalización del primer teorema de Shannon.

Según el primer teorema de Shannon, la entropía de un alfabeto se interpreta como el número medio de binits necesarios para representar un símbolo de ese alfabeto. Apliquemos esta interpretación al concepto de entropía a priori y a posteriori (figura 5-7).

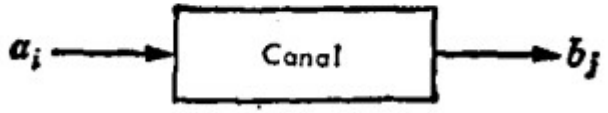


FIG. 5-7. Canal de información.

(Abramson,p.119)

consultas: taulamet@unl.edu.ar

GENERALIZACIÓN

5-5. Generalización del primer teorema de Shannon.

Según el primer teorema de Shannon, la entropía de un alfabeto se interpreta como el número medio de binits necesarios para repre-

Antes de recibir un símbolo a la salida de un canal, se asocian las probabilidades a priori $P(a_i)$ con el alfabeto de entrada A. $H(A)$ es el número medio de binits necesarios para representar un símbolo de este alfabeto. Recibido un símbolo, por ejemplo b_j, se asocian al alfabeto de entrada las probabilidades a posteriori $P(a_i/b_j)$. $H(A/b_j)$ es el número medio de binits necesarios para representar un símbolo de ese alfabeto a partir de las probabilidades a posteriori. Puesto que los símbolos se presentan a la salida con probabilidades $P(b_j)$, es de esperar que el número medio de binits necesarios (valor medio extendido también a b_j) para representar un símbolo de entrada a_i, dado un símbolo de salida determinado, sea igual a la entropía media a posteriori

$$\sum_b P(b)\, H(A/b) \qquad (5\text{-}17)$$

(Abramson,pp.119-120)

consultas: taulamet@unl.edu.ar

GENERALIZACIÓN

$$\sum_B P(b)\, H(A/b) \leq \frac{\overline{L_n}}{n} < \sum_B P(b)\, H(A/b) + \frac{1}{n} \qquad (5\text{-}24)$$

que constituye la generalización del primer teorema de Shannon. Hay que destacar la semejanza de (5-24) y (4-15*a*). Aumentando n, puede hacerse I_n/n tan próximo a

$$\sum_B P(b)\, H(A/b) \qquad (5\text{-}25)$$

(Abramson,p.122)

consultas: taulamet@unl.edu.ar

EQUIVOCACIÓN DEL CANAL

$$H(A/B) = \sum_{B} P(b)\, H(A/b)$$

$$= \sum_{B} P(b) \sum_{A} P(a/b) \log \frac{1}{P(a/b)}$$

$$= \sum_{A,B} P(a,b) \log \frac{1}{P(a/b)} \qquad (5\text{-}26)$$

$H(A/B)$ recibe el nombre de *equivocación* de A con respecto a B, o *equivocación* del canal. La ecuación (5-24) puede expresarse en función de la equivocación, en la forma siguiente

$$\lim_{n\to\infty} \frac{\overline{L_n}}{n} = H(A/B) \qquad (5\text{-}27)$$

(Abramson,p.123)

consultas: taulamet@unl.edu.ar

INFORMACIÓN MUTUA

La información mutua puede expresarse de diferentes maneras,

$$I(A\,;\,B) = H(A) - H(A/B)$$

$$= \sum_{A} P(a) \log \frac{1}{P(a)} - \sum_{A,B} P(a,b) \log \frac{1}{P(a/b)}$$

$$= \sum_{A,B} P(a,b) \log \frac{1}{P(a)} - \sum_{A,B} P(a,b) \log \frac{1}{P(a/b)}$$

$$= \sum_{A,B} P(a,b) \log \frac{P(a/b)}{P(a)} \qquad (5\text{-}31a)$$

o, puesto que

$$P(a_i, b_j) = P(a_i/b_j)\, P(b_j)$$

$$I(A\,;\,B) = \sum_{A,B} P(a,b) \log \frac{P(a,b)}{P(a)\,P(b)} \qquad (5\text{-}31b)$$

(Abramson,p.125)

consultas: taulamet@unl.edu.ar

INFORMACIÓN MUTUA

$$I(A\ ;\ B) = H(A) - H(A/B)$$

5-7. Propiedades de la información mutua.

Se ha demostrado que la información mutua es el número medio de binits necesarios para determinar un símbolo de entrada antes de conocer el símbolo de salida correspondiente, menos el número medio de binits necesarios para especificar un símbolo de entrada después de conocer el símbolo de salida. Es decir,

$$I(A;\ B) = H(A) - H(A/B) \qquad (5\text{-}33)$$

(Abramson,p.125)

consultas: taulamet@unl.edu.ar

INFORMACIÓN MUTUA

$$I(A\ ;\ B) = H(A) - H(A/B)$$

5-7. Propiedades de la información mutua.

ser negativo? Para aclarar esta pregunta se escribirá (5-31*b*) de otra forma:

$$I(A;B) = \sum_{A,B} P(a,b) \log \frac{P(a,b)}{P(a)\,P(b)} \qquad (5\text{-}31b)$$

Haciendo uso de la desigualdad (2-8*a*) encontramos

$$I(A;B) \geqq 0$$

que será una igualdad cuando

$$P(a_i, b_j) = P(a_i)\,P(b_j) \text{ para cualquier } i,\ j \qquad (5\text{-}35)$$

CANAL AMBIGUO

(Abramson,p.126)

consultas: taulamet@unl.edu.ar

INFORMACIÓN MUTUA

$$I(A\ ;\ B) = H(A) - H(A/B)$$

5-7. Propiedades de la información mutua.

$$I(A;\ B) = I(B;\ A) \qquad (5\text{-}36)$$

(Abramson,p.126)

consultas: taulamet@unl.edu.ar

ENTROPÍA AFÍN

babilidad de este suceso es $P(a_i, b_j)$, de modo que la entropía afín valdrá

$$H(A, B) = \sum_{A,B} P(a, b) \log \frac{1}{P(a, b)} \qquad (5\text{-}40)$$

$$= H(A) + H(B) - I(A; B) \qquad (5\text{-}41)$$

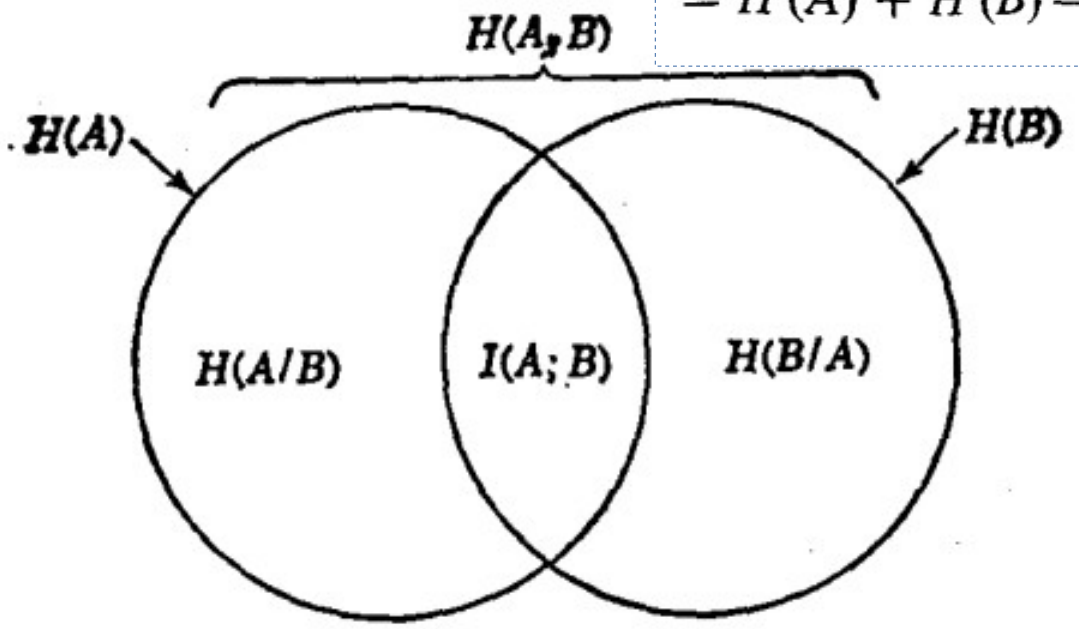


FIG. 5-9. Relaciones entre las diferentes magnitudes de un canal.

(Abramson,p.127)