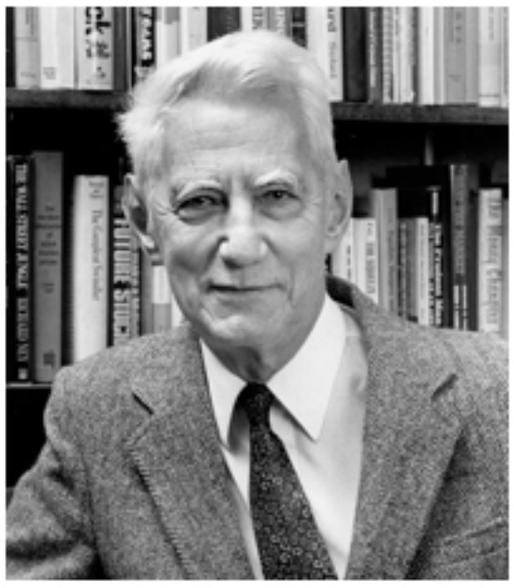
TEORÍA DE LA INFORMACIÓN

The Bell System Technical Journal

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

A Mathematical Theory of Communication

By C. E. SHANNON



CLAUDE SHANNON
(30/04/1916 - 24/02/2001)

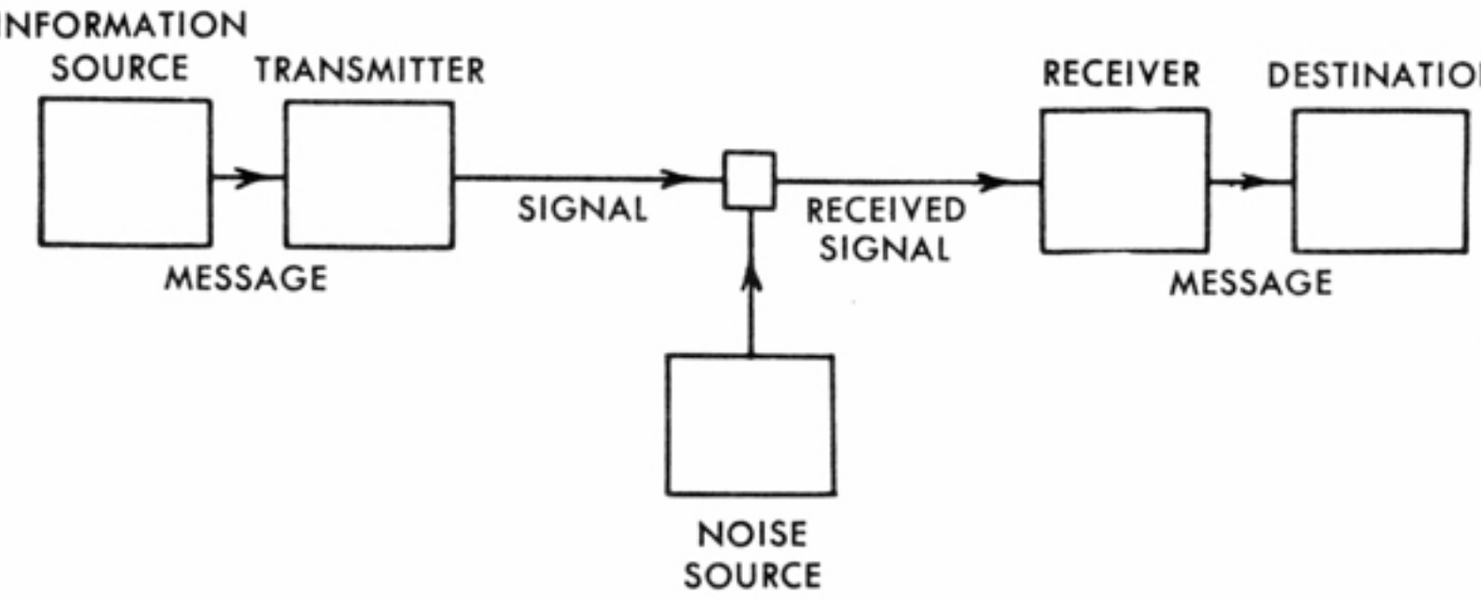
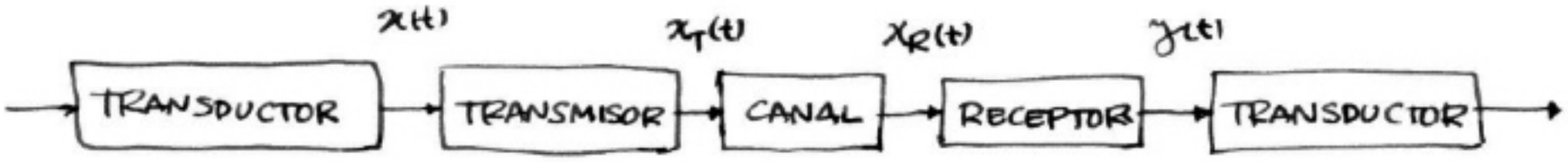


Fig. 1. — Schematic diagram of a general communication system.



$X = \{x_i\}_{i=1,\dots,m}$

1er Teorema de Shannon

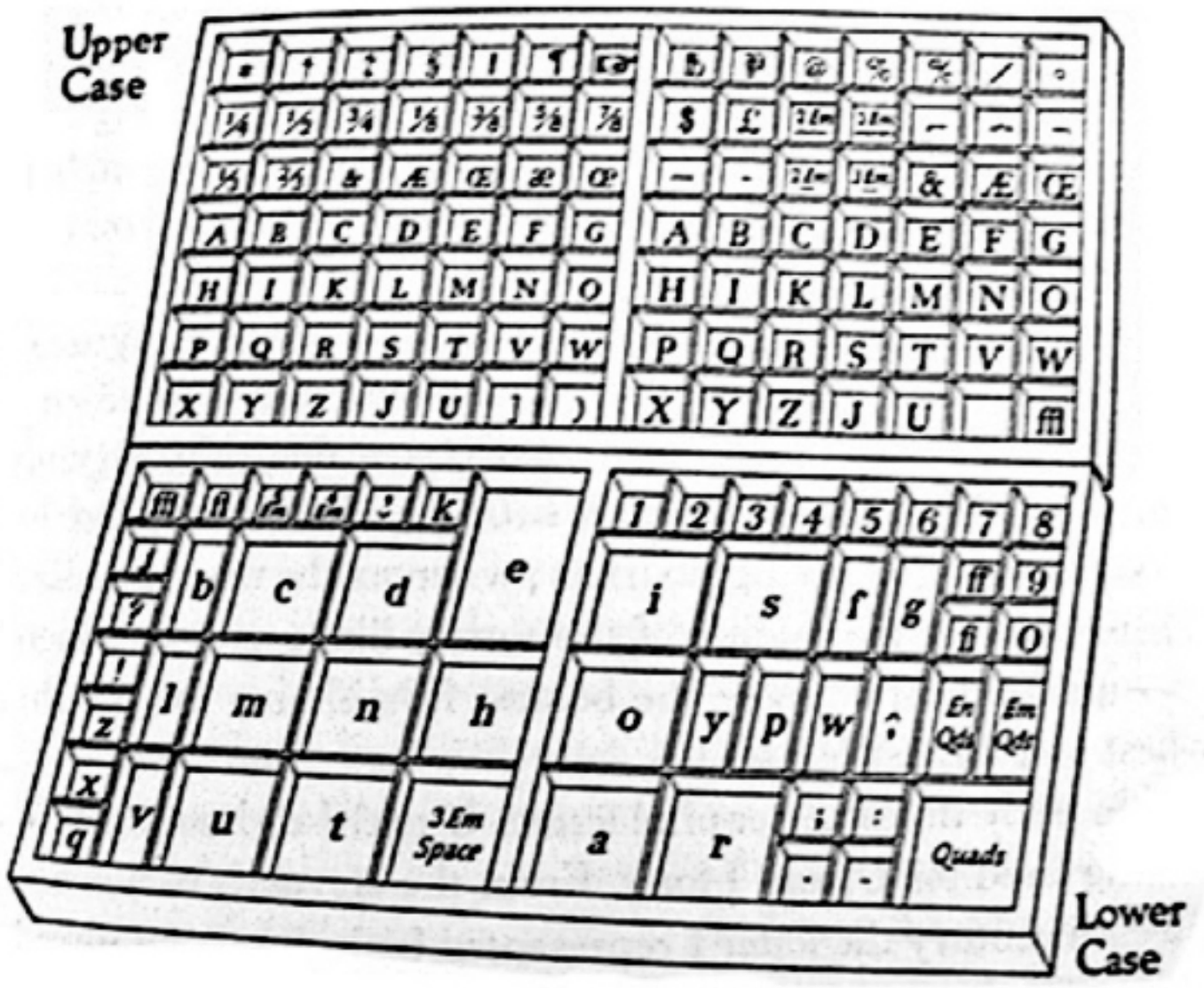
2do. Teorema de Shannon

$P(X) = (p(x_1), \dots, p(x_m))$

¿Y ANTES QUÉ?

1888 Heinrich Hertz "las ondas electromagnéticas pueden detectarse"
1861 Ecuaciones de Maxwell
1836 Código Morse
1876 Teléfono: Graham Bell
1892 William Preece: telegrafía sin hilos (5km)
1895 Marconi mejora la telegrafía sin hilos (long distance radio transmission
Guglielmo
1899
1884 Paul Gottlieb Nipkow (concibe la idea de la transmisión de imágenes y) describe un disco espiral perforado para transmitir una imagen como un ~~conjunto de~~ mosaico de líneas.
1907 Boris Rosing: concibe el CRT para la recepción de imágenes
1925 John Baird demuestra la transmisión de una silueta moviéndose (Londres)
1927 Primer broadcast de TV en UK (BBC = 1936)
1936 Edwin Howard Armstrong publica la base de la modulación FM
(V24, NS, Mayo 1936)
1946 ENIAC (Electronic Numerical Integrator and Computer) primer computador de propósito general. Digital / Programmable / Turing-complete: simula una Turing-machine
1948 Shannon
" Transistor (John Bardeen, William ~~Shot~~ Shockley and Walter Brattain)
Bell Labs.
1976 (11 de abril) se pone a la venta (Release ↑) la Apple I a USD 666.66 (Julio)
1979 Primer teléfono celular comercial (red de telefonía celular) NTT Japón
(En 1946 Bell systems tenía un dispositivo para hacer llamados telefónicos desde un auto; pesaba 36 kg.)

CÓDIGO MORSE



Definición: Información de un ~~~~ mensaje.

$I(x_i)$

Será una medida de la <u>incertidumbre</u> de la VA. que modela la fuente.

$X = \{x_i\}_{i=1,\dots,m}$

$P(X) = P(X) = (p(x_1), \dots, p(x_m)) = (p_1, \dots, p_m)$

$$I(x_i) = -\log_2 P(x_i)$$

$b=2$ BIT / $b=10$ HARTLEY / $b=e$ NATS

(i) $I_i = I(x_i) \geq 0$ con $0 \leq P_i \leq 1$

(ii) $I_i \xrightarrow{P_i \uparrow 1} 0$

(iii) $I_j > I_i$ sii $P_j < P_i$

(iv) x_i, x_j INDEPENDIENTES $\rightarrow$ $I(x_j, x_i) = I(x_j) + I(x_i)$

Ejemplos

① $P(S_1) = (1/4, 1/4, 1/4, 1/4)$ $\quad I_1 = I_2 = I_3 = I_4 = \log_2 4 = 2$ BITS $\quad H(S_1) = 2$ BITS

② $P(S_2) = (1/2, 1/4, 1/8, 1/8)$ $\quad I_1 = 1 \quad I_2 = 2 \quad I_3 = I_4 = 3 \quad H(S_2) = 1.75$ BITS

③ $P(S_3) = (1/2, 1/2)$ $\quad I_1 = I_2 = 1 \quad H(S_3) = 1$

④ $P(S_4) = (9/10, 1/10)$ $\quad I_1 = 0.15$ BITS $\quad I_2 = 3.32$ BITS

La mayoría de los resultados los vamos a ver en fuentes DMS

DMS (Discrete Memoryless Source)

- Estacionaria
- Símbolos estadísticamente independientes $P(x_i, x_j) = P(x_i)P(x_j)$
- cadencia o tasa de símbolos r_s símbolos x seg.

También mencionaremos las fuentes Markovianas o con memoria

MKS

. $P(x_i | x_j)$ ← se conocen estas probabilidades

ENTROPÍA

$$H(X) = H(X) = E\{I(x_i)\} = -\sum_{i=1}^{m} p(x_i) \log p(x_i)$$

$$= \sum_{i=1}^{m} p(x_i)\, I(x_i)$$

$$0 \leq H(x) \leq \log m$$

Regla de la cadena $H(X_1, \dots, X_n) = \sum_i H(X_i \mid X_{i-1}, \dots, X_1)$

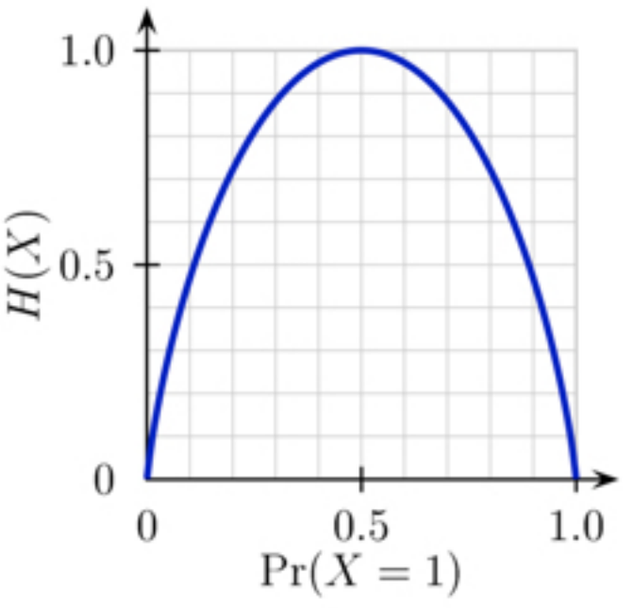
$$H(X_1, X_2) = H(X_1) + H(X_2 \mid X_1)$$

CONDICIONAR REDUCE LA ENTROPÍA $H(X \mid Y) \leq H(X)$

Markov orden k $H(X_n \mid X_{n-1}, \dots, X_1) = H(X_n \mid X_{n-1}, \dots, X_{n-k})$

En general $H_{MKS}(X) \leq H_{DMS}(X)$

ENTROPÍA BINARIA



OTRAS ENTROPÍAS

ENTROPÍA CONJUNTA

$X \in Y$ V.A. $p(x_i, y_j)$

$$H(X,Y) = \sum_{\substack{x \in X \\ y \in Y}} p(x,y) \log \frac{1}{p(x,y)} = E\{I(x,y)\} \Big/ I(x,y) = \log \frac{1}{p(x,y)}$$

ENTROPÍA CONDICIONAL

$$H(X|Y) = \sum_y p(y)\, H(X|Y=y) = E_{p(x,y)}\{-\log p(X|Y)\}$$

$$= \sum_{\substack{x \in X \\ y \in Y}} p(x,y) \log \frac{1}{p(x|y)}$$

DEFINICIÓN ENTROPÍA RELATIVA O DISTANCIA DE KULLBACK-LEIBLER

Sean p y q distribuciones de probabilidad sobre la misma fuente de m mensajes

$$D(p \| q) = \sum_{i=1}^{m} p(x_i) \log\left(\frac{p(x)}{q(x)}\right)$$

TASA DE TRANSFERENCIA DE INFORMACIÓN

Tenemos definida una información media y una cadencia de símbolo
¿cuál es la <u>tasa de transferencia de información</u>?

Observando la generación de $n+1$ símbolos de la fuente, la duración de esos n símbolos es $n r_s$ segundos. La información <u>media</u> generada es $n\,H(X)$. La tasa de transf. de info es

$$R = \frac{n \cdot H(X)}{n / r_s} = r_s H(X) \quad \frac{\text{BITS}}{\text{seg.}}$$

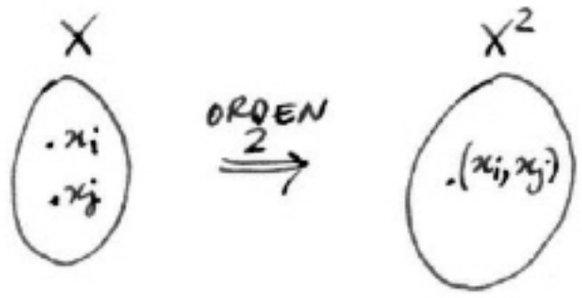
<u>Ejemplos</u>: $r_s = 2k \frac{\text{símbolos}}{\text{seg}}$

$R(S_1) = 4k \frac{\text{BITS}}{\text{seg}}$

$R(S_2) = 3.5\,k \frac{\text{BITS}}{\text{seg}}$

EXTENSIÓN DE UNA FUENTE

La extensión de una fuente es una nueva fuente cuyos símbolos son la concatenación de los símbolos de la fuente original.



En gral. $X^n = \{ (x_{i_1}, \dots, x_{i_n})_j \}$ $j = 1, \dots, m^n$

$i_k \in \{1, \dots, m\}$

Las probabilidades serán

$$p((x_{i_1}, \dots, x_{i_n}))$$

Para una DMS $H(X^n) = n\, H(X)$

¿Cuál es la nueva tasa de transferencia de info.?

Con X^2 $\quad r_{x^2} = \frac{r_x}{2}$ (los símbolos demoran el doble)

$$H(x^2) = 2\, H(X)$$

$$\rightarrow \quad R_{x^2} = \frac{2\, H(x)}{} \frac{r_x}{2} = R_x$$

Es razonable, es solamente un reordenamiento de los símbolos; no se genera ni se destruye información.

Ejemplo: 52 cartas (símbolos) Asumimos equiprobables $P(x_i) = \frac{1}{52}$

(Prob. 1b.1-2) ↑ (10 números + 3 figuras) × 4 palos

Se extrae una carta al azar.

(a.1) Cuánta información nos da saber que es de corazones?

(a.2) una figura?

(a.3) una figura de corazones?

(a.1) $I(\text{corazones}) = -\log p(\text{corazones}) = -\log 1/4 = 2 \text{ bits} = \log 4$

(a.2) $I(\text{figuras}) = -\log p(\text{figura}) = -\log(12/52) = \log \frac{52}{12} = \log \frac{13}{3}$ $\quad I(\text{fig}) + I(\text{corazones})$ " " "

(a.3) $I(\text{figura de corazones}) = I(\text{fig} \cap \text{corazon}) = -\log p(\text{fig coraz.}) = -\log \frac{3}{52} = \log \frac{52}{3} = \log \frac{52}{12} + \log 4$

(b) Si sabemos que la carta es figura roja. Cuánta info nos falta para determinar cuál es?

↳ Es una de 6.

$I(\text{Carta}) = \log 52$ bits

$I(\text{fig roja}) = \log \frac{52}{6} = I(\text{roja}) + I(\text{fig}) = (-\log \frac{1}{2}) + (-\log \frac{12}{52}) = -\log \frac{6}{52}$

$I(\text{identificar}) = \log 52 - \log \frac{52}{6} = \log 6 = -\log \left(\frac{1}{6}\right)$ → Pr. de una de las 6 posibles figuras rojas

TEOREMA DE CODIFICACIÓN DE FUENTES

CODIFICACIÓN DE FUENTE

$\mathcal{D} = \{0, 1, \cdots, D-1\}$ ALFABETO CÓDIGO $\quad \mathcal{D} = \mathcal{B} = \{0,1\}$

CÓDIGO : $\quad C : X \longrightarrow \bigcup_{k=1}^{k=\infty} d^k \qquad d_i \in \mathcal{D} = \{d_1, \cdots, d_m\} \qquad d^k = d_1 \cdots d_k$

Es una regla de los símbolos (mensajes) de la fuente en secuencias (palabras) de un código, obtenidas concatenando símbolos de $\mathcal{D}$

Clasificación de los códigos

- Los alfabetos código son chicos $\mathcal{D} = \mathcal{B} = \{0,1\}$, VARÍA la longitud de las palabras.

- código
  - SINGULAR ← dos o más símbolos se mapean en la misma palabra de código
  - NO SINGULAR
    - UNIVOCAM. DECOD. ← no hay ambigüedades al decodificar, o sea si su extensión es no singular
      - NO INSTANTANEO
      - INSTANTANEO
    - NO U.D.

↑ c/palabra se decodifica cuando termina de recibirse, no hay que esperar a otra palabra

Ejemplo

Con los códigos analizados, y la fuente $p_X(\mathcal{X}) = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$ de $H(X) = 1{,}75$bits

$\mathcal{X}$	C_1	C_2	C_3	C_4	C_5
x_1	0	0	10	0	0
x_2	0	010	00	10	10
x_3	0	01	11	110	110
x_4	0	10	110	1110	111
K	2	1.125	0.875	0.9375	1
L	1	1.75	2.25	1.875	1.75
H/L	1.75	1	0.778	0.933	1

LARGO MEDIO DE UN CÓDIGO

$$L(\mathcal{C}) = \sum_{x \in \mathcal{X}} p(x)\, l(x) = E\{l(x)\}$$

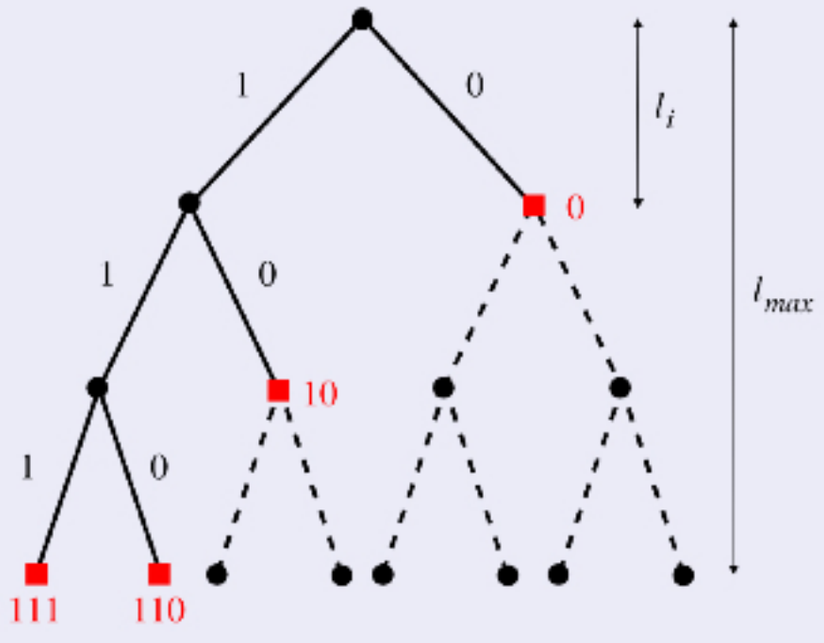
DESIGUALDAD DE KRAFT

Propiedad de Kraft: Cualquier código U.D. / INSTANTÁNEO sobre un alfabeto D-ario con largos l_i cumple $K = \sum_i D^{-l_i} \leq 1$

Recíprocamente si un set $(l_1, \dots, l_M)$ cumple ↑ $\Rightarrow$ $\exists$ un código instantáneo con esos largos

Demostración

Considerar un árbol D-ario (ramas, hojas y camino).



$$\sum_i D^{l_{\text{máx}} - l_i} \leq D^{l_{\text{máx}}}$$

□

TEOREMA DE CODIFICACIÓN DE FUENTES DE SHANNON

Teorema (Teorema de Codificación de Fuente)

El largo medio de un código C instantáneo, D-ario para una variable aleatoria X es mayor o igual a la entropía de X

$$L(C) \geq H_D(X)$$

y la igualdad se cumple sii $p_i = D^{-l_i}$

- La igualdad se da sii $p_i = q_i = D^{-l_i}$ $(K = 1)$; distribución D-ádica.
- El procedimiento para encontrar el código óptimo sería hallar la distribución D-ádica más cercana a la distribución de X (en el sentido de la entropía relativa). Pero esto no siempre es fácil.
- Es una cota para la longitud media para la descripción de una fuente, “no se puede comprimir más allá de la entropía”.

Demostración

Sea $K = \sum_i D^{-l_i} \leq 1$ y $q_i = D^{-l_i}/K$ una distribución de probabilidad

$$\begin{aligned} D(p||q) &= \sum_i p_i \log_D \frac{p_i}{q_i} = \sum_i p_i \log_D p_i - \sum_i p_i \log_D q_i \\ &= -H_D(X) - \sum_i p_i \log_D \frac{D^{-l_i}}{K} \\ &= -H_D(X) - \sum_i p_i \log_D D^{-l_i} + \sum_i p_i \log_D K \\ &= -H_D(X) + \sum_i p_i l_i + \log_D K \sum_i p_i \\ &= -H_D(X) + L(C) + \log_D K \geq 0 \end{aligned}$$

La igualdad se da si $D(p||q) = 0$ y $K = 1$. □

Qué sucede si hacemos una extensión de una fuente X^n y un nuevo código con Los largo medio por símbolo

$$H(X_1, X_2, X_n) \leq L(\overset{X^n}{(X_1, X_2, X_n)}) \leq H(X_1, X_2, X_n) + 1$$

$\longrightarrow$ DMS

$$n\,H(X) \leq L(\overset{n}{C}) \leq nH(X) + 1$$

$$\Rightarrow \quad H(X) \leq L_n \leq H(X) + \frac{1}{n}$$

Teóricamente es posible hacer una extensión de la fuente y acotar más el valor de L

$$L \xrightarrow[n\uparrow\infty]{} H(X) \qquad \frac{H(X)}{L} \longrightarrow 100\%$$

Pero no es realizable en la práctica

Ejemplo:

$P(X_4) = (9/10, 1/10)$ $\quad H(X_4) = 0.467$ $\quad$ 46,7 % de eficiencia

$\mathcal{C}_I = \{0, 1\}$ $\quad L(\mathcal{C}_I) = 1$

$P(X_4^2) = \left(\frac{81}{100}, \frac{9}{100}, \frac{9}{100}, \frac{1}{100}\right)$ $\quad H(X_4^2) = 2\, H(X_4) = 0,934$

$\mathcal{C}_I^2 = (00, 01, 10, 11)$ $\quad L(\mathcal{C}_I^2) = 2$ $\quad \frac{H(X_4^2)}{L(\mathcal{C}_I^2)} = 46.7\%$

$K = 1$

$\mathcal{C}_{II} = (0, 10, 110, 111)$ $\quad L(\mathcal{C}_{II}) = 1,29$ $\quad \frac{H(X4)}{L(\mathcal{C}_{II})} = 72\%$

$K = 1$

$P(X_2) = (1/2, 1/4, 1/8, 1/8)$ $\quad H(X_2) = 1,75$ $\quad \frac{H}{L} = 100\%$

$\mathcal{C}_{II} = (0, 10, 110, 111)$ $\quad L(\mathcal{C}_{II}) = 1,75$

código óptimo.

CÓDIGOS DE HUFFMAN

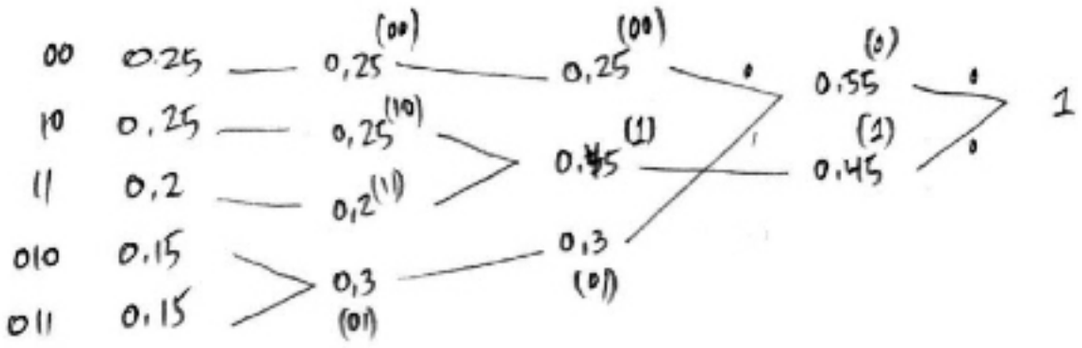
Son códigos óptimos. No hay otro codigo con $L(C)$ menor que $L(C_{HUFF})$.

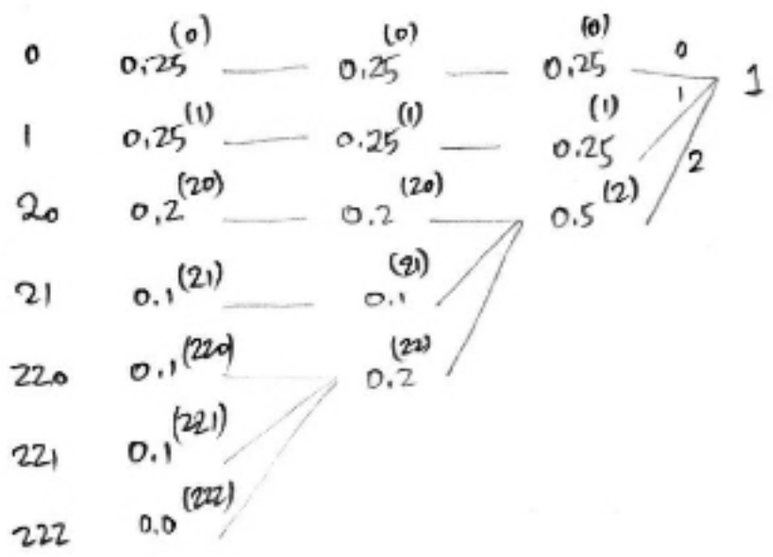
(1) Ordenar los símbolos según su prob.

(2) Combinar los D menos prob. en un nuevo símbolo con prob igual a la suma.

(3) Goto 1

(4) Se asignan dígitos del alfabeto código





$k(D-1)+1 = M$

Nos dan seis botellas de vino de las cuales una está picada. No sabemos cuál pero sí sabemos las probabilidades de que esté picada para cada botella

$$(p_1, p_2, p_3, p_4, p_5, p_6) = \left(\tfrac{8}{23}, \tfrac{6}{23}, \tfrac{4}{23}, \tfrac{2}{23}, \tfrac{2}{23}, \tfrac{1}{23}\right)$$

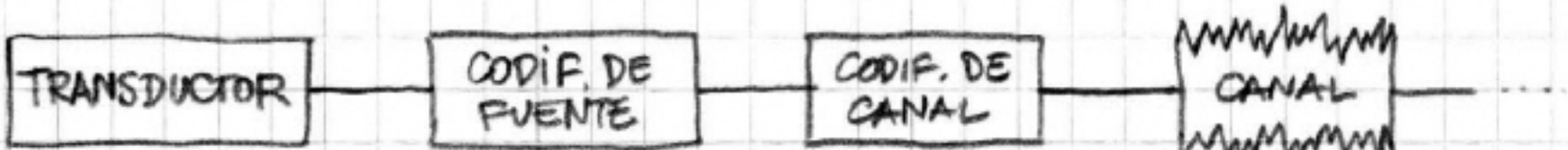
(a) Suponiendo que se puede probar de a una a la vez ¿cuál es el orden para minimizar el número medio de pruebas? ¿Cuál es el número medio de pruebas?

Suponiendo que se pueden mezclar los vinos en un vaso y probarlos

(b) ¿Cuáles deben ser las mezclas y en qué orden para minimizar el número medio de pruebas?

TEOREMA DE CODIFICACIÓN DE CANAL

CODIFICACIÓN DE CANALES CON RUIDO



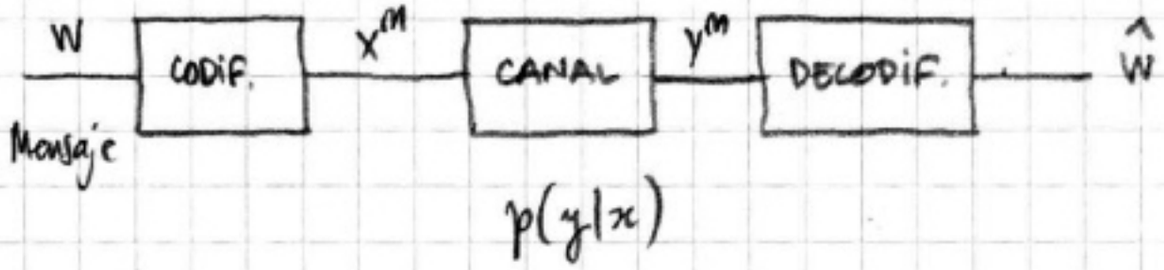
CANAL DISCRETO $(\mathcal{X}, p(y|x), \mathcal{Y})$

$\mathcal{X}, \mathcal{Y}$: dos conjuntos finitos

$p(y|x)$: probabilidades de transición

$$p(y|x) \geq 0$$

$$\sum_{y} p(y|x) = 1$$



INFORMACIÓN MUTUA

INFORMACIÓN MUTUA! X, Y variables aleatorias, con distribución de prob. conjunta $p(x,y)$ y marginales $p(x)$ y $p(y)$. La información mutua entre las dos variables es

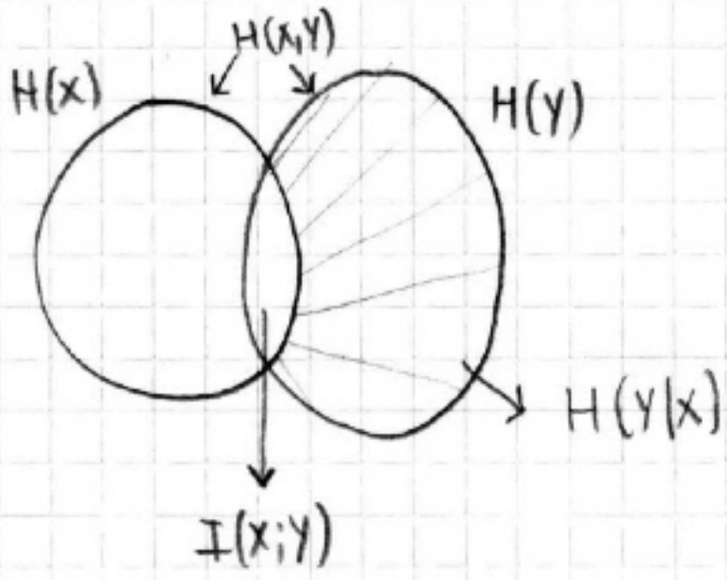
$$I(X;Y) = D\left(p(x,y) \,\|\, p(x)\,p(y)\right) =$$

$$= E_{p(x,y)}\left\{ \log \frac{p(x,y)}{p(x)\,p(y)} \right\} = \sum_{\substack{x\in X \\ y\in Y}} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

$$= H(X) - H(X|Y) \longrightarrow \text{EQUIVOCACIÓN (info. perdida en el canal)}$$

$$= H(Y) - H(Y|X) \longrightarrow \text{ENTROPÍA DEL RUIDO (entropía asociada al ruido en el canal)}$$

$$= H(X) + H(Y) - H(X,Y)$$



INFO. MUTUA: MEDIDA DE LA CANT DE INFO. QUE UNA V.A. TIENE SOBRE OTRA V.A. . ES LA REDUCCIÓN EN LA INCERTIDUMBRE SOBRE UNA V.A. DADO QUE CONOCEMOS OTRA V.A.

$$I(X;X) = H(X) - H(X|X) = H(X)$$

Algunas convenciones y nomenclatura.

$p(x_i)$: Prob. elegir x_i para transmitirlo

$p(y_j)$: Prob y_j sea recibido

$p(x_i, y_j)$: Prob. x_i transmitido y y_j recibido

$p(x_i | y_j)$: Prob. x_i transmitido dado que se recibió y_j

$p(y_j | x_i)$: Prob recibir y_j dado que se transmitió x_i

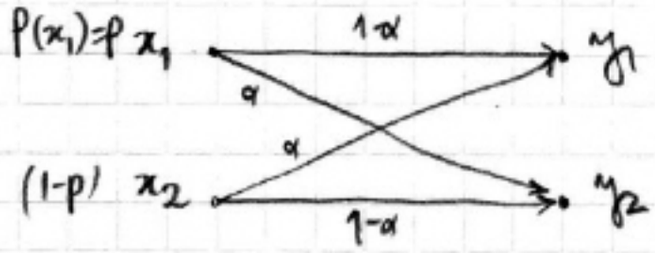
→ Channel forward transition prob.

$$I(x_i; y_j) = \log \frac{p(x_i|y_j)}{p(x_i)} = \log \frac{1}{p(x_i)} - \log \frac{1}{p(x_i|y_j)}$$

La Información Mutua representa la info media "ganada" por el destino por sobre lo transmitido

$$I(X;Y) \geq 0$$

Canal binario simétrico



$p(y_1|x_1) = 1-\alpha \qquad P(y_2|x_1) = \alpha$

$p(y_1|x_2) = \alpha \qquad P(y_2|x_2) = 1-\alpha$

$$P_{error} = P(y_2|x_1)\,P(x_1) + P(y_1|x_2)\,P(x_2) = \alpha p + \alpha(1-p) = \alpha$$

$$P(y_1) = P(y_1|x_2)\,P(x_2) + P(y_1|x_1)\,P(x_1) = \alpha(1-p) + (1-\alpha)p = \alpha + p - 2\alpha p$$

$$P(y_2) = 1 - P(y_1)$$

$$H(X) = \Omega(p)$$

$$H(Y) = \Omega(P(y_1))$$

$$H(Y|X) = \sum_{x_i} P(x_i)\,H(Y|X=x_i) = \sum_{x_i} P(x_i)\left(\sum_{y_j} P(y_j|x_i)\,\log\frac{1}{p(y_j|x_i)}\right) = \Omega(\alpha)$$

$$I(X;Y) = H(Y) - H(Y|X) = \Omega(P(y_1)) - \Omega(\alpha) = \Omega(\alpha + p - 2\alpha p) - \Omega(\alpha)$$

$\alpha \ll 1$	$I(X;Y) \simeq \Omega(p) = H(X)$
$\alpha = 1/2$	$I(X;Y) = \Omega\left(\frac{1}{2} + p - 2\frac{1}{2}p\right) - \Omega(1/2) = 0$
$\alpha = 1$	$I(X;Y) = \Omega(1 + p - 2p) - \Omega(1) = \Omega(1-p) - 0 = \Omega(p)$

↓

NO HAY ERROR

CAPACIDAD DE UN CANAL DISCRETO

CAPACIDAD DE UN CANAL DISCRETO $C(\mathcal{X}, p(y|x), \mathcal{Y})$

- Alfabetos origen y destino fijos
- Prob. de transición fijas $p(y|x)$
- Las variables son la $P(x_i)$

$\Rightarrow$ La máxima $I(X;Y)$ se da para alguna cierta $P(x_i)$

Capacidad en bits por símbolo $\quad C_S = \max_{P(x_i)} I(X;Y)$

Si s $\frac{\text{simb}}{\text{seg}}$ es la máxima tasa de transf. de símbolos del canal

$\rightarrow$ $C = s\, C_S$ $\frac{\text{bits}}{\text{seg}}$ Tasa de transf. de info del canal

Propiedades de C

$C \geq 0$

$C \leq \log |\mathcal{X}|$

$C \leq \log |\mathcal{Y}|$

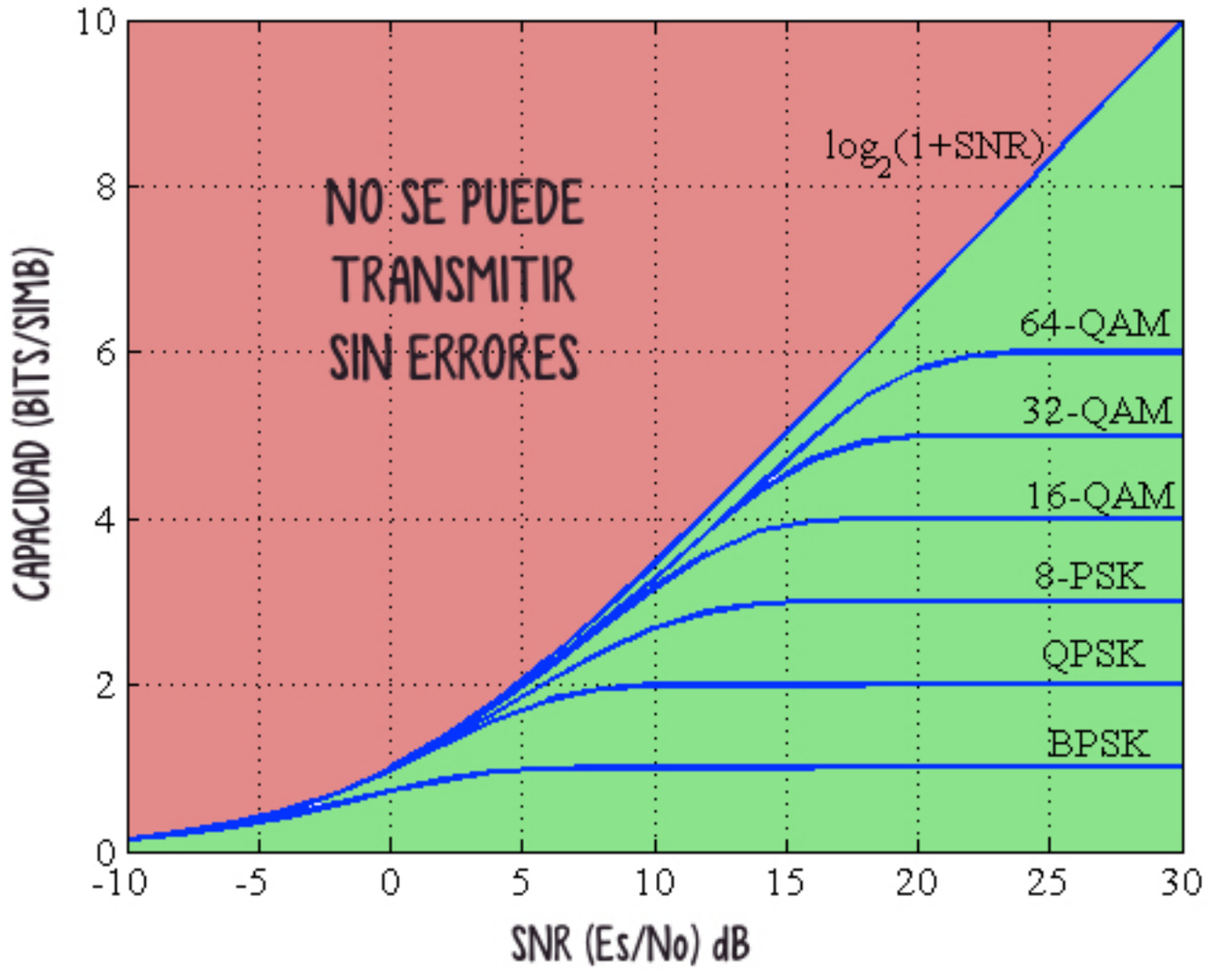
$I(X;Y)$ es continua $p(x)$ y cóncava con $p(x)$

TEOREMA DE CODIFICACIÓN DE CANAL

Todas las tasas $R \leq C$ son alcanzables. En particular, para cualquier tasa $R < C$

$\exists$ un código con máxima probabilidad de error $P_e \to 0$

Recíprocamente un código con $P_e \to 0$ tiene $R \leq C$. (No se puede transmitir sin error a $R > C$)



MODELO PARA UN CANAL CONTÍNUO

fuente: $x(t)$ $\qquad H(X) \triangleq \int_{-\infty}^{+\infty} p(x) \log \frac{1}{p(x)}\, dx$

$$I(X;Y) = \iint_{-\infty}^{+\infty} p_{XY}(x,y) \frac{\log p_X(x|y)}{p_X(x)}\, dx\, dy$$

El canal será (modelo):

(i) No distorsiona y se compensan las pérdidas de atenuación

(ii) Ancho de banda B

(iii) La señal, de potencia finita $S_x = \overline{x^2}$ y ancho de banda $W < \infty$

(iv) AWGN de potencia σ_N^2, media nula y DEP $\frac{1}{2}\eta$ $(\sigma_N^2 = \eta B)$

(v) Señal y ruido indep. $\qquad y(t) = x(t) + n(t)$

$$\overline{y^2} = \overline{x^2} + \overline{n^2} = S_x + N \qquad SNR_R = \frac{S_x}{N}$$

LEY DE HARTLEY-SHANNON

$$C = B \log_2 (1 + SNR_R)$$

$$R = W \log_2 (1 + SNR_D)$$