Entropía → medida de lo impredecible / incierto que es el resultado de un experimento

↳ VA. X ó Y

VA. X con distribución de probab.

$$P(X) = \{P(x_1), \ldots, P(x_m)\}$$

donde $x_1, \ldots, x_m$ son los resultados del experimento X,

y $$P(x_1) = P(X = x_1)$$

$$H(X) = \sum_{i=1}^{m} P(x_i) \log \frac{1}{P(x_i)} = E\left[\log \frac{1}{P(X)}\right] \text{ bits}$$

Si la entropía de un experimento es 0'47 bits ⇒ que podemos representar la información implícita en 1000 muestras del experimento con

$$1000 \times 0'47 = 470 \text{ 0's y 1's}$$

---

2 Experimentos X e Y

$X \sim p(X) = \{p(x_1), p(x_2)\}$

$Y \sim p(Y) = \{p(y_1), p(y_2)\}$

$Y \backslash X$	x_1	x_2	Σ
y_1	12	15	27
y_2	4	25	29
Σ	16	40	56

Y\X→	x_1	x_2	P(Y)
y_1	12/56	15/56	$P(Y=y_1) = 27/56$
y_2	4/56	25/56	$P(Y=y_2) = 29/56$
P(X)	16/56	40/56	$\boxed{1}$
	$P(X=x_1)$	$P(X=x_2)$	

$$P(X,Y) = \begin{pmatrix} P(x_1,y_1) & P(x_2,y_1) \\ P(x_1,y_2) & P(x_2,y_2) \end{pmatrix}$$

$$\boxed{p_{ij} = P(X=i, Y=j)}$$

Entropía Conjunta X, Y

$$H(X,Y) = \sum_{\forall i,j} p_{ij} \lg \frac{1}{p_{ij}} = E\left[\lg \frac{1}{P_{XY}}\right]$$

→ bits/pareja de (i,j)

$$\boxed{P_{XY} = P(X,Y)}$$

$$P(X,Y) = \begin{pmatrix} 12/56 & 15/56 \\ 4/56 & 25/56 \end{pmatrix}$$

$$H(X,Y) = \sum_{i,j} p_{ij} \log \frac{1}{p_{ij}} = -\sum_{i,j} p_{ij} \log p_{ij}$$

$$H(X,Y) = -\frac{12}{52}\log\frac{12}{52} - \frac{4}{52}\log\frac{4}{52} - \frac{15}{52}\log\frac{15}{52} - \frac{25}{52}\log\frac{25}{52} =$$

$$= \boxed{1'7982}$$

Experimento con dos monedas.

$$X \sim p(X) = \{p(X=\text{cara}), p(X=\text{cruz})\} = \{0'5, 0'5\}$$

$$Y \sim p(Y) = \{p(Y=\text{cara}), p(Y=\text{cruz})\} = \{0'5, 0'5\}$$

$$p(X,Y) = \begin{pmatrix} p(X=\text{cara}, Y=\text{cara}) & p(\text{cruz}, \text{cara}) \\ p(\text{cara}, \text{cruz}) & p(\text{cruz}, \text{cruz}) \end{pmatrix} =$$

$$= \begin{pmatrix} \boxed{p(\text{cara})} \cdot p(\text{cara}) & p(\text{cruz}) \cdot p(\text{cara}) \\ \boxed{p(\text{cara})} \cdot p(\text{cruz}) & p(\text{cruz}) \cdot p(\text{cruz}) \end{pmatrix} =$$

$$= \begin{pmatrix} 0'5 \cdot 0'5 & 0'5 \cdot 0'5 \\ 0'5 \cdot 0'5 & 0'5 \cdot 0'5 \end{pmatrix} = \begin{pmatrix} 0'25 & 0'25 \\ 0'25 & 0'25 \end{pmatrix} =$$

$$P = \begin{pmatrix} 0'25 & 0'25 \\ 0'25 & 0'25 \end{pmatrix} \begin{array}{l} \rightarrow p(Y=\text{cara}) = 0'5 \\ \rightarrow p(Y=\text{cruz}) = 0'5 \end{array} \rightarrow H(Y)$$

$$\sum \quad \begin{array}{cc} \| & \| \\ p(X=\text{cara}) & p(X=\text{cruz}) \\ \| & \| \\ =0'5 & 0'5 \end{array} \rightarrow H(X)$$

$$H(X) = -0'5 \lg 0'5 - 0'5 \lg 0'5 = 1$$

$$H(Y) = 1$$

$$\boxed{H(X,Y) \rightarrow H(X), H(Y)}$$

---

2 dados de 6 caras

$$X \sim p(X) = \{p(1), p(2), p(3), p(4), p(5), p(6)\}$$

$$Y \sim p(Y) = \{p(1) \text{ ——— } p(6)\}$$

$$p(1) = p(2) = \dots = p(6) = \frac{1}{6}$$ tanto para X como Y

$$P = \begin{pmatrix} p(1,1) & \cdots & p(6,1) \\ p(1,2) & \cdots & p(6,2) \\ \vdots & & \vdots \\ p(1,6) & \cdots & p(6,6) \end{pmatrix} \begin{array}{l} \to p(Y=1) \\ \to p(Y=2) \\ \\ \to p(Y=6) \end{array}$$

$P(X=1) \quad p(X=2)$

$$p(X,Y) = p(X) \cdot p(Y)$$

HXcondY
HYcondX

function [HXY, HX, HY] = entropiaconjunta(P)

① 2 Monedas donde X, Y independientes
② 2 Dados (x6c) donde X, Y independientes
③ , ① pero X = Y
④ ≡ ② pero X = Y
⑤ Dado 6c condicionado a Moneda
$\quad\; Y \qquad\qquad\qquad\qquad\qquad X$

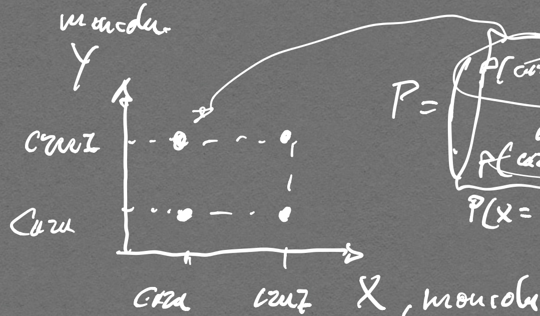
PRACTICA 3 : SESIÓN 2

	1	2	3	4	5
P	$\begin{pmatrix} 0.25 & 0'25 \\ 0'25 & 0'25 \end{pmatrix}$				
H(X,Y)					
H(X)					
H(Y)					
H(X|Y)					
H(Y|X)					
I(X,Y)					

$$H_{X|Y} = H(X|Y) = \sum_{\forall j} H(X|Y=j)\, p(Y=j)$$

$$H_{Y|X} = H(Y|X) = \sum_{\forall j} H(Y|X=j)\, p(X=j)$$

$$H(X,Y) \geq \begin{cases} H(X|Y) \\ H(Y|X) \end{cases}$$



$$P = \begin{array}{cc|l} \overset{0'25}{p(cara, cara)} & \overset{0'25}{p(cruz, cara)} & \to \overset{0'5}{p(Y=cara)} \\ \overset{0'25}{p(cara, cruz)} & \overset{0'25}{p(cruz, cruz)} & \to p(Y=cruz) \\ p(X=cara) & p(X=cruz) & \end{array}$$

$$P(A|B) = \frac{\boxed{P(A \cap B)}}{P(B)}$$

$$\boxed{X|Y=cara} \to \left\{ \frac{p(cara, cara)}{p(Y=cara)}, \frac{p(cruz, cara)}{p(Y=cara)} \right\} =$$

$$= \left\{ \frac{0'25}{0'5}, \frac{0'25}{0'5} \right\} = 1$$

$$X|Y=cruz \to \left\{ \frac{p(cara, cruz)}{p(cruz)}, \frac{p(cruz, cruz)}{p(cruz)} \right\} =$$

$$= \left\{ \frac{0'25}{0'5}, \frac{0'25}{0'5} \right\} \to 1$$

$$Y|X \begin{cases} cara \to \left\{ \frac{0'25}{0'5}, \frac{0'25}{0'5} \right\} \to 1 \\ cruz \to \left\{ \frac{0'25}{0'5}, \frac{0'25}{0'5} \right\} \to 1 \end{cases}$$

$$H(X|Y) = \sum_{j \to \substack{cara \\ cruz}} H(X|Y=j) \cdot p(Y=j) =$$

$$= H(X|Y=cara) \cdot p(Y=cara) + H(X|Y=cruz) \cdot p(Y=cruz) =$$

$$= \left[H(X=cara|Y=cara) + H(X=cruz|Y=cara) \right] \cdot P(Y=cara) +$$

$$\rightarrow + \Big[H(X=\text{cara} \mid Y=\text{cruz}) + H(X=\text{cruz} \mid Y=\text{cruz}) \Big] p(Y=\text{cruz}) =$$

$$= \left[-\frac{0'25}{0'5} \log\left(\frac{0'25}{0'5}\right) - \frac{0'25}{0'5} \log\left(\frac{0'25}{0'5}\right) \right] 0'5 \; +$$

$$\left[-\frac{0'25}{0'5} \log\left(\frac{0'25}{0'5}\right) - \frac{0'25}{0'5} \log\left(\frac{0'25}{0'5}\right) \right] 0'5 = \underline{1} \; \big(= H(Y|X) \big)$$

$$\underline{\underline{P}} = \begin{array}{c|cc|l} & x_1 & x_2 & \\ \hline Y & 0'2 & 0'1 & y_1 \to P(Y=y_1) = 0'3 \\ & 0'4 & 0 & y_2 \to P(Y=y_2) = 0'4 \\ & 0'1 & 0'2 & y_3 \to P(Y=y_3) = 0'3 \\ \hline & X & & \Sigma = 1 \end{array}$$

$$P(X=x_1) = 0'7 \qquad P(X=x_2) = 0'3 \;\rightarrow \Sigma = \underline{1}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\Big\{ \underbrace{Y \mid X = x_1}, \; \underbrace{Y \mid X = x_2} \Big\} \Rightarrow \left\{ \begin{matrix} 0'2/0'7 \\ 0'4/0'7 \\ 0'1/0'7 \end{matrix} \right. , \quad \left. \begin{matrix} 0'1/0'3 \\ \cancel{0/0'3} \to 0 \\ 0'2/0'3 \end{matrix} \right\}$$

$$\Sigma = 1$$

$$H(Y|X) = \sum_{\forall j} H(Y|X_j)\, p(X=j) = H\left(\frac{0'2}{0'7}\right) 0'7 + H\left(\frac{0'4}{0'7}\right) \cdot 0'7 \; +$$

$$+ H\left(\frac{0'1}{0'7}\right) \cdot 0'7 + H\left(\frac{0'1}{0'3}\right) \cdot 0'3 + H(0) \cdot 0'3 + H\left(\frac{0'2}{0'3}\right) \cdot 0'3$$

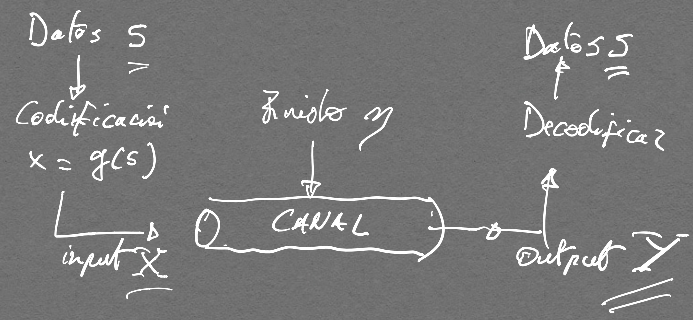
$$= \underline{1'24}$$

$$\Big\{ \underbrace{X \mid Y = y_1}, \; \overbrace{X \mid Y = y_2}, \; \underbrace{X \mid Y = y_3} \Big\} \longrightarrow \left\{ \left\{ \frac{0'2}{0'3}, \frac{0'1}{0'3} \right\}, \left\{ \frac{0'4}{0'4}, \frac{0}{0'4} \right\}, \left\{ \frac{0'1}{0'3}, \frac{0'2}{0'3} \right\} \right\}$$

$$H(X|Y) = H\left(\frac{0{,}2}{0{,}3}\right)\cdot 0{,}3 + H\left(\frac{0{,}1}{0{,}3}\right)\cdot 0{,}3 + H\left(\frac{0{,}4}{0{,}4}\right)\cdot 0{,}4 + H(0)\cdot 0{,}4 +$$

$$+ H\left(\frac{0{,}1}{0{,}3}\right)\cdot 0{,}3 + H\left(\frac{0{,}2}{0{,}3}\right)\cdot 0{,}3 = 0{,}5510$$

Información Mutua



$I(X,Y)$ nos dice la información media que ganamos ó tenemos de Y después de observar un solo valor de X

$I(X,Y)$ supone la reducción de incertidumbre sobre X tras conocer el valor de Y, y viceversa.

La incertidumbre de Y es $H(Y)$

Si X e Y están relacionados entonces tras observar un valor de X tendremos más información sobre Y y por tanto su incertidumbre se reduce a un valor que será menor de $H(Y)$

Si hacemos la media de esa reducción para todos los valores de X e Y, entonces tras observar X nuestra incertidumbre sobre Y se reduce una cantidad cuyo valor es $I(X,Y)$

La incertidumbre residual que tenemos de Y tras observar un sólo valor de X es la $H(Y|X)$

→ tb. se llama $H(\eta)$ Entropía de ruido

Entonces, la velocidad a la que comunicamos información depende de: $H(X)$, $H(Y)$ y la relación de X e Y.

y para comunicar tanto como sea posible por un canal ruidoso debemos tener $H(X)$ y $H(Y)$ alta.
y por otro lado $H(\eta)$ baja $(\equiv H(Y|X))$

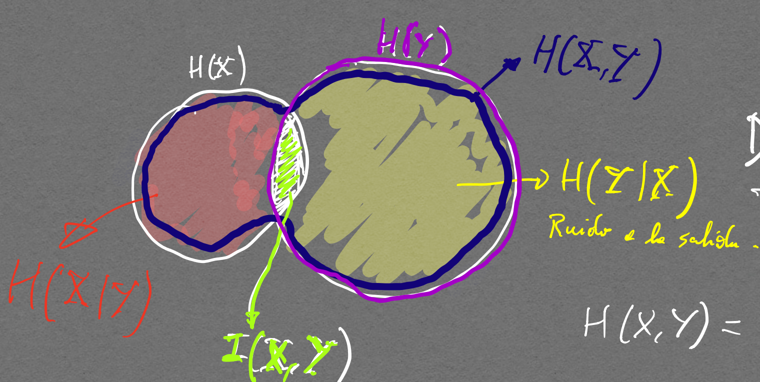


Diagrama de Venn

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

número de posibles símbolos en la entrada. $\equiv 2^{H(X)}$

por cada símbolo de entrada podríamos tener hasta $2^{H(Y|X)}$ posibles salidas en Y

$$P = \begin{pmatrix} p(x_1, y_1) & \cdots & p(x_n, y_1) \\ p(x_1, y_2) & & \\ \vdots & & \vdots \\ p(x_1, y_m) & & p(x_n, y_m) \end{pmatrix}$$

$\sum \to P(Y=y_1)$

$\sum \to P(Y=y_2)$

$\sum \to P(X=x_1)$

$$\frac{p(x_1, y_1)}{P(X=x_1) \cdot P(Y=y_1)} \quad , \quad \frac{p(x_1, y_2)}{P(X=x_1) \cdot P(Y=y_2)} \quad , \quad \cdots$$

Nota: VER VÍDEO ADICIONAL DE ESTA SESIÓN

De cada uno de estos términos vamos a hacer el logaritmo

p. ej. $\log\left(\frac{p(x_1, y_1)}{p(x_1)\, p(y_1)}\right)$, y lo multiplico por el valor del numerador, $p(x_1, y_1) \cdot \log\left(\frac{p(x_1, y_1)}{p(x_1)\, p(y_1)}\right) + \cdots +$ lo hacemos y sumamos para todos los elementos de P → El resultado de la suma es $I(X,Y)$

3ª Sesión P3

$I(X,Y) \Rightarrow$ bits.

3.3 CAPACIDAD DE UN CANAL

No Deseable

1ª Situación

$$\begin{bmatrix} H(X) & \text{baja} \\ H(Y) & \text{alta.} \end{bmatrix}$$

No Deseable

2ª Situación

$H(X)$ alta

$H(Y)$ baja