



Universidad de Granada

decsai.ugr.es

Teoría de la Información y la Codificación

Grado en Ingeniería Informática

Tema 2.- La entropía como medida de información.



DECSAI

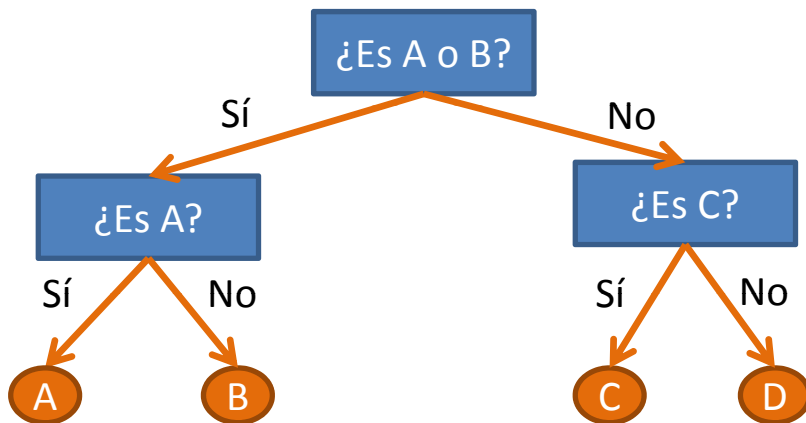
**Departamento de Ciencias de la
Computación e Inteligencia Artificial**

- 1. ¿Qué es la entropía?**
- 2. Entropía de más de una variable.**
- 3. Información mutua.**
- 4. Ejemplo adicional de aplicación de la Entropía: Sistemas Expertos.**

1. **¿Qué es la entropía?**
2. **Entropía de más de una variable.**
3. **Información mutua.**
4. **Ejemplo adicional de aplicación de la Entropía: Sistemas Expertos.**

- La cuantificación de la **Información** nos indica, cuantitativamente, cuánta información nos aporta un mensaje producido aleatoriamente por una fuente.
- En el tema anterior, nos planteábamos el concepto de cantidad de información de forma intuitiva.
- Nos planteábamos qué mínimo número de preguntas tendríamos que hacer, en media, a lo largo del tiempo para conocer qué símbolo proporcionaba una fuente.

$$p(A) = p(B) = p(C) = p(D) = 0,25$$



Fuente 1

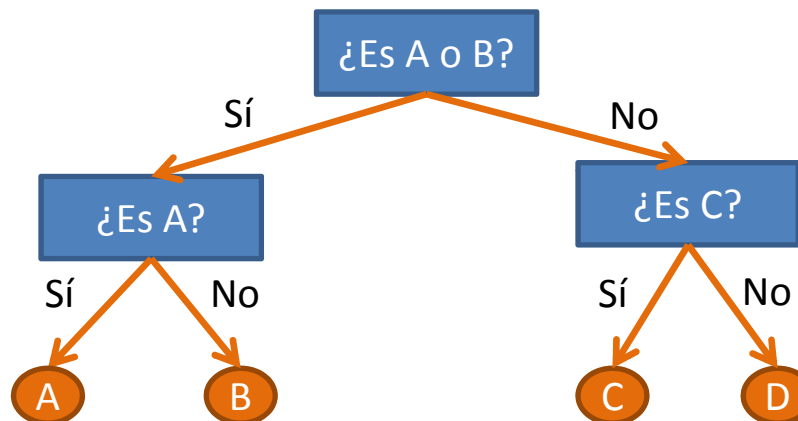


¿A? ¿B? ¿C? ¿D?

$$N.Preguntas = 0,25 * 2 + 0,25 * 2 + 0,25 * 2 + 0,25 * 2 = 2 \text{ preguntas}$$

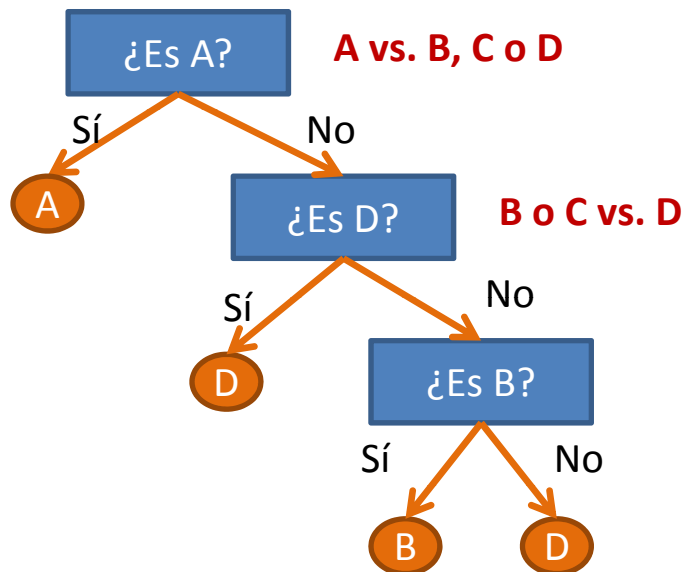
- Llamaremos, de momento, **cantidad de información de la fuente** como el valor calculado para el promedio del 'mínimo número de preguntas.
- Vamos a abstraer la fórmula un poco:

$$\begin{aligned} \text{N.Preguntas} = & p(A) * \text{Nivel}(A) + \\ & p(B) * \text{Nivel}(B) + \\ & p(C) * \text{Nivel}(C) + \\ & p(D) * \text{Nivel}(D) = \text{Cantidad de información de la fuente} \end{aligned}$$



- Llamaremos, de momento, **cantidad de información de la fuente** como el valor calculado para el promedio del mínimo número de preguntas.
- Vamos a abstraer la fórmula un poco:

$$p(A) = 0,5; p(B) = p(C) = 0,125; p(D) = 0,25$$



Fuente 2



¿A? ¿B? ¿C? ¿D?

$$C.\text{Inf Fuente 2} = 0,5 * 1 + 0,125 * 3 + 0,125 * 3 + 0,25 * 2 = 1,75 \text{ preguntas}$$

- Podemos comprobar que la cantidad de información de las fuentes depende directamente del **Nivel del árbol** en el que se decide cada símbolo.
- Este nivel, a su vez, **depende a su vez de la probabilidad de que ocurra cada símbolo!**
- **Shannon** descubrió esta relación y decidió proponer una medida de información, a la que denominó **entropía ($H(s)$)**:

$$H(s) = -p(s) \cdot \log(p(s))$$

- **La entropía de un evento es igual a la probabilidad de ese evento por su logaritmo, cambiado de signo.**
- **La entropía de una fuente es, por tanto, la suma de las entropías de los mensajes que puede transmitir. Equivale a definirla como la esperanza de la información de una fuente:**

$$H(S) = E\{I(S)\} = \sum_{i=1}^n -p(S = s_i) \cdot \log(p(S = s_i))$$

- En términos generales, escribiremos la entropía de una fuente S como:

$$H(S) = - \sum_{i=1}^n p(S = s_i) \cdot \log_2(p(S = s_i))$$

- **Entropía normalizada:**

- En ocasiones, nos interesará comparar la información que proporcionan 2 fuentes capaces de generar conjuntos de símbolos diferentes. Supongamos que una **fente 1 ($S1$)** tiene un alfabeto de n símbolos, y que otra **fente 2 ($S2$)** tiene m símbolos. Podemos normalizar la entropía de ambas fuentes dividiendo por el logaritmo del número de símbolos:

$$\tilde{H}(S1) = \frac{H(S1)}{\log_2(n)}; \tilde{H}(S2) = \frac{H(S2)}{\log_2(m)}$$

– En el ejemplo anterior:

- Como la **Fuente 1** tenía 4 símbolos A, B, C, D, con

$$p(A) = p(B) = p(C) = p(D) = 0,25$$

- Para la **Fuente 2**: $p(A) = 0,5$; $p(B) = p(C) = 0,125$; $p(D) = 0,25$

– ¿Cuál es la entropía de la fuente 1?

$$H(\text{fuente1}) = \sum_{i=1}^n -p(s_i) \cdot \log(p(s_i)) = -p(A) * \log(p(A)) - p(B) * \log(p(B)) \\ - p(C) * \log(p(C)) - p(D) * \log(p(D)) = \\ 0,25 * \log(0,25) * 4 = 0,6021$$

– ¿Cuál es la entropía de la fuente 2?

$$H(\text{fuente2}) = \sum_{i=1}^n -p(s_i) \cdot \log(p(s_i)) = -0,5 * \log(0,5) - 0,125 * \log(0,125) \\ - 0,125 * \log(0,125) - 0,25 * \log(0,25) = \\ 0,5268$$

- Podemos observar que **la entropía es una medida del desorden, del caos.**
 - Un evento con probabilidad 1, tendrá una entropía $H=0$
 - Un evento con baja probabilidad tendrá una entropía muy alta.
- ¿Qué ocurre con eventos **s** cuya probabilidad de ocurrencia es 0?

$$H(s) = -p(s) \cdot \log(p(s)) = 0 \cdot \log(0)$$

- Como el término $\log(0)$ no existe, **establecemos por definición que $0 \cdot \log(0) = 0$** . Esto significa que:
 - Un evento que tiene probabilidad de ocurrir igual a 0 no proporciona información alguna. No posee incertidumbre.
- La medida de la información es el **BIT (Binary digIT)**. Proviene de la naturaleza del mínimo número de preguntas a realizar con respuesta “Sí/No” para descubrir el símbolo.
- **En estos casos el logaritmo debe ser en Base 2.**

- La entropía nos puede servir para calcular el número de bits necesarios para codificar un símbolo:
 - ¿Cuántas preguntas de respuesta “Sí”/”No” son necesarias para discriminar entre 2 símbolos equiprobables?: **1 pregunta**. Se necesita **1 BIT de información** para averiguar el símbolo.
 - ¿Cuántas preguntas de respuesta “Sí”/”No” son necesarias para discriminar entre 4 símbolos equiprobables?: **2 preguntas**. Se necesitan **2 BITS de información** para averiguar el símbolo.

- La entropía nos puede servir para calcular el número de bits necesarios para codificar un símbolo:
- Supongamos que tenemos que transmitir mensajes desde una fuente **S**, compuestos por símbolos del siguiente alfabeto:
 - Las 26 letras del abecedario (sin la ñ),
 - junto con 5 símbolos de puntuación,
 - y un espacio en blanco.
- **Total: 32 símbolos**
- Por simplificar, vamos a asumir que todos los símbolos del alfabeto son equiprobables (máxima incertidumbre).
Entonces:

$$H(S) = - \sum_{i=1}^{n=32} p(S = s_i) \cdot \log_2(p(S = s_i)) = 5$$

- **Se necesitarán 5 bits de información para codificar cada símbolo.**

– **Propiedades de la entropía de una fuente $H(S)$:**

- **La entropía está acotada** (n =número de símbolos diferente de la fuente):

$$0 \leq H(S) \leq \log_2(n)$$

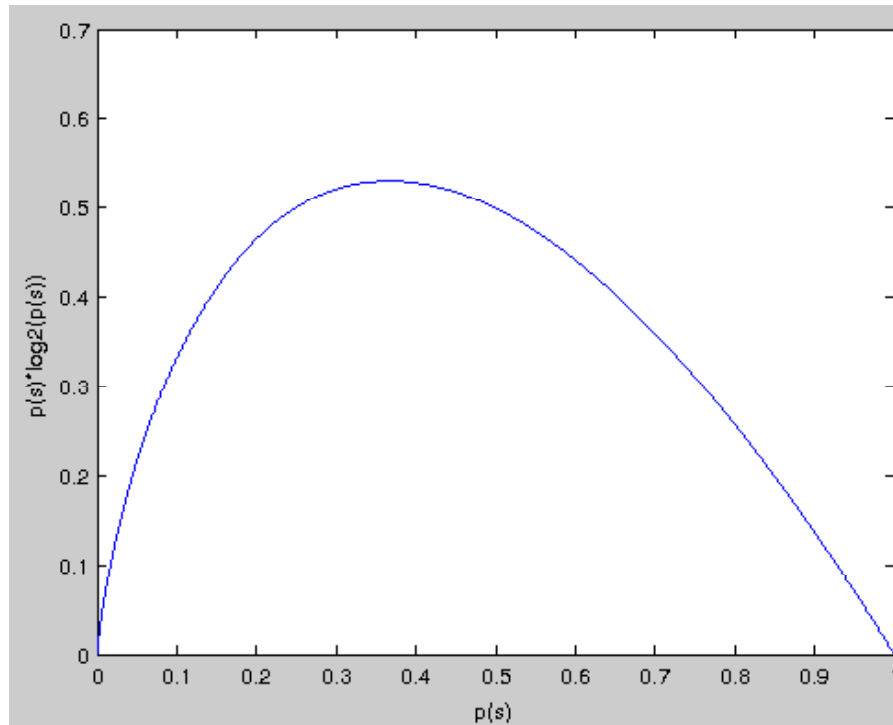
- **Si se conoce al 100% qué se va a recibir, no se aporta información:**

$$H(S) = 0 \leftrightarrow \exists s_i \in \{s_1, s_2, \dots, s_n\} : p(S = s_i) = 1$$

- **La entropía es máxima cuando todos los símbolos de la fuente son equiprobables:**

$$H(S) = \log_2(n) \leftrightarrow \forall s_i \in \{s_1, s_2, \dots, s_n\} : p(S = s_i) = 1/n$$

- Propiedades de la entropía de una fuente $H(S)$:
 - La entropía tiende a 0 cuando la probabilidad de ocurrencia de un evento tiende a 1 o a 0:



- Cuanto más seguros estemos sobre un evento, menos incertidumbre tendremos sobre él.

- **Ejemplo:** supongamos que tenemos un sistema S que genera 2 símbolos “0” y “1”, con $p(S=\text{“0”})=0,2$ y $p(S=\text{“1”})=0,8$. ¿Cuál es la información que genera el símbolo “1”? ¿Y el símbolo “0”? ¿Cuál es la entropía del sistema?

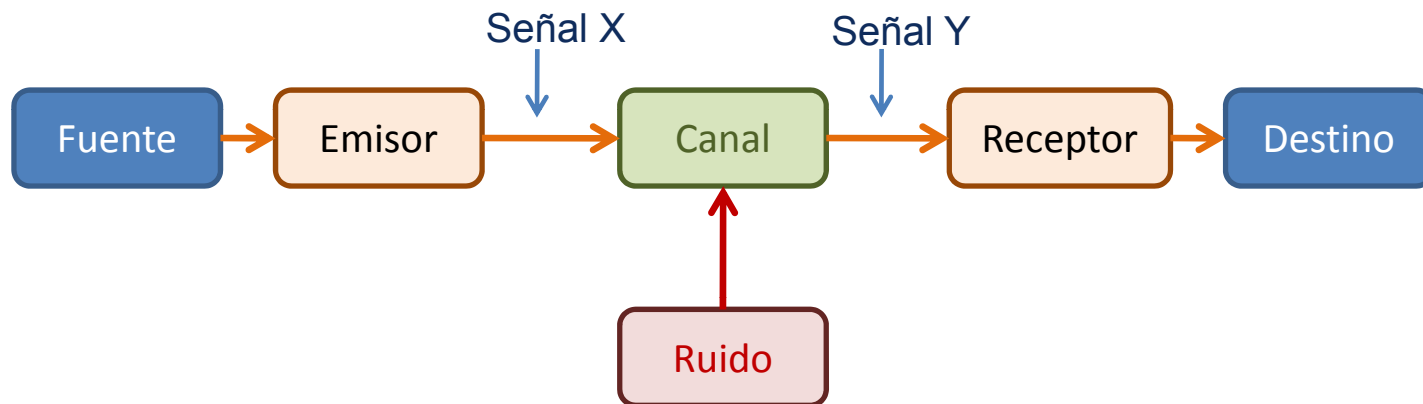
$$I(S = \text{“0”}) = -\log(P(S = \text{“0”})) = -\log(0,2) = 2,3219$$

$$I(S = \text{“1”}) = -\log(P(S = \text{“1”})) = -\log(0,8) = 0,3219$$

$$\begin{aligned} H(S) &= -\sum_{i=1}^{n=2} p(S = s_i) \cdot \log_2(p(S = s_i)) = \\ &\sum_{i=1}^{n=2} p(S = s_i) \cdot I(S = s_i) = 0,2 * I(S = \text{“0”}) + 0,8 * I(S = \text{“1”}) = \\ &0,2 * 2,3219 + 0,8 * 0,3219 = 0,7219 \end{aligned}$$

1. **¿Qué es la entropía?**
2. **Entropía de más de una variable.**
3. **Información mutua.**
4. **Ejemplo adicional de aplicación de la Entropía: Sistemas Expertos.**

- En un sistema de comunicaciones como el siguiente:



- Las entropías que nos podría interesar estudiar son:
 - $H(X)$: Entropía en el emisor.
 - $H(Y)$: Entropía en el receptor.
 - $H(X,Y)$: Entropía por par de símbolos emitido y recibido.
 - $H(Y|X)$: Entropía del símbolo recibido Y supuesto emitido X.
 - $H(X|Y)$: Entropía del símbolo emitido X supuesto recibido Y.

- En estos casos tendremos que conocer la matriz de probabilidades de que se reciba un símbolo Y dado un símbolo X emitido.
- Por ejemplo, suponiendo que tenemos 4 símbolos (A, B, C, D) y la siguiente matriz de probabilidades P:

		Y				
		A	B	C	D	
X	A	0,25	0,1	0	0	0,35
	B	0,025	0,1	0	0	0,125
	C	0	0	0,4	0	0,4
	D	0,025	0	0	0,1	0,125
		0,3	0,2	0,4	0,1	

La probabilidad de que Y reciba A, B, C y D (independientemente de qué haya enviado la fuente) es, respectivamente, $p(Y=A)= 0,3$; $p(Y=B)= 0,2$; $p(Y=C)= 0,4$; $p(Y=D)= 0,1$;

- En estos casos tendremos que conocer la matriz de probabilidades de que se reciba un símbolo Y dado un símbolo X emitido.
- Por ejemplo, suponiendo que tenemos 4 símbolos (A, B, C, D) y la siguiente matriz de probabilidades P:

		Y				
		A	B	C	D	
X	A	0,25	0,1	0	0	0,35
	B	0,025	0,1	0	0	0,125
	C	0	0	0,4	0	0,4
	D	0,025	0	0	0,1	0,125
		0,3	0,2	0,4	0,1	

La probabilidad de que X envíe A, B, C y D (independientemente de qué reciba Y) es, respectivamente, $p(X=A)= 0,35$; $p(X=B)= 0,125$; $p(X=C)= 0,4$; $p(X=D)= 0,125$;

		Y				
		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	
X	<i>A</i>	0,25	0,1	0	0	0,35
	<i>B</i>	0,025	0,1	0	0	0,125
	<i>C</i>	0	0	0,4	0	0,4
	<i>D</i>	0,025	0	0	0,1	0,125
		0,3	0,2	0,4	0,1	

- **$p(X=D, Y=A)$** vale 0,025, es la probabilidad de que X envíe D e Y reciba A.
- **$P(Y=A \mid X=B)$** vale $0,025/0,125=0,2$, es la probabilidad de que Y detecte A, sabiendo que X ha emitido B.

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

- La entropía de dos variables discretas X e Y $H(X, Y)$ se define como:

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(X = x_i, Y = y_j) \cdot \log_2(p(X = x_i, Y = y_j))$$

- Donde n es el número de posibles valores de X y m el número de posibles valores de Y . En nuestro ejemplo:

$$\begin{aligned} H(X, Y) = & p(X = A, Y = A) \cdot \log_2(p(X = A, Y = A)) + \\ & p(X = A, Y = B) \cdot \log_2(p(X = A, Y = B)) + \\ & p(X = A, Y = C) \cdot \log_2(p(X = A, Y = C)) + \\ & p(X = A, Y = D) \cdot \log_2(p(X = A, Y = D)) + \\ & p(X = B, Y = A) \cdot \log_2(p(X = B, Y = A)) + \\ & \dots + \end{aligned}$$

- Si el canal de transmisión contiene ruido, entonces $H(X, Y) > 0$ y viceversa. A mayor valor, mayor ruido.

- **Ejemplo:** Supongamos que tenemos un canal de transmisión binario que transmite con ruido atendiendo a las siguientes probabilidades conjuntas ($P(X,Y)$):

		Y	
		0	1
X	0	0,3	0,3
	1	0,3	0,1

- ¿Cuál es la entropía conjunta del receptor y del emisor?
- **CUIDADO:** lo que se muestra en el ejercicio debe interpretarse como probabilidades **CONJUNTAS** $p(X=0, Y=0)$, $p(X=0, Y=1)$, y $p(X=1, Y=0)$, $p(X=1, Y=1)$, dado que se representa como *“probabilidad de que X envía 0/1 e Y recibe 0/1”*.

- ¿Cuál es la entropía conjunta del receptor y del emisor?

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(X = x_i, Y = y_j) \cdot \log_2(p(X = x_i, Y = y_j)) =$$

$$p(X = 0, Y = 0) \cdot \log_2(p(X = 0, Y = 0)) +$$

$$p(X = 0, Y = 1) \cdot \log_2(p(X = 0, Y = 1)) +$$

$$p(X = 1, Y = 0) \cdot \log_2(p(X = 1, Y = 0)) +$$

$$p(X = 1, Y = 1) \cdot \log_2(p(X = 1, Y = 1)) =$$

$$-3 * 0,3 * \log_2(0,3) - 0,1 * \log_2(0,1) = 1,8955$$

- ¿Y la entropía del emisor? ¿Y la entropía del receptor?

– ¿Y la entropía del emisor? ¿Y la entropía del receptor?

$$p(X = 0) = 0,3 + 0,3 = 0,6; p(X = 1) = 0,3 + 0,1 = 0,4$$

$$\begin{aligned} H(X) &= - \sum_{i=1}^{n=2} p(X = x_i) \cdot \log_2(p(X = x_i)) = \\ &= -0,6 * \log_2(0,6) - 0,4 * \log_2(0,4) = 0,971 \end{aligned}$$

$$P(Y = 0) = 0,6; P(Y = 1) = 0,4$$

$$\begin{aligned} H(Y) &= - \sum_{i=1}^{m=2} p(Y = y_i) \cdot \log_2(p(Y = y_i)) = \\ &= -0,6 * \log_2(0,6) - 0,4 * \log_2(0,4) = 0,971 \end{aligned}$$

- **Dato curioso:** Observar que la incertidumbre dada por dos variables conjuntamente no supera la incertidumbre de ambas variables consideradas por separado:

- En nuestro caso:

$$H(X) = H(Y) = 0,971$$

$$H(X, Y) = 1,8955$$

$$H(X, Y) = 1,8955 \leq H(X) + H(Y) = 2 * 0,971 = 1,942$$

- La entropía o incertidumbre aportada por dos variables conjuntamente siempre será menor (o igual, en el caso de independencia) que la entropía o incertidumbre de ambas variables consideradas por separado.

- La entropía condicionada de una variable Y con respecto a otra X , $H(Y | X)$ se define como:

$$H(Y | X) = - \sum_{i=1}^n p(X = x_i) \cdot \sum_{j=1}^m p(Y = y_j | X = x_i) \cdot \log_2(p(Y = y_j | X = x_i))$$

- Se puede reescribir, desarrollando la probabilidad condicionada, como:

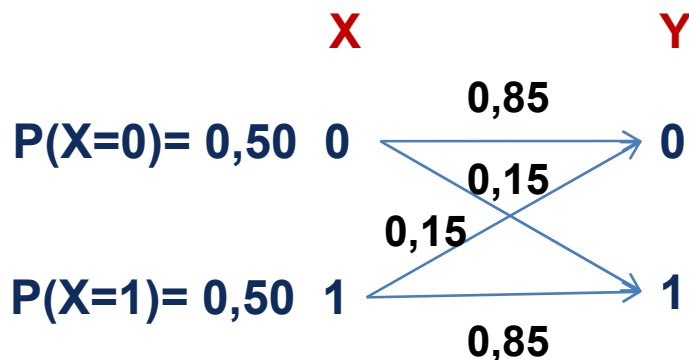
$$H(Y | X) = - \sum_{i=1}^n P(X = x_i) H(Y | X = x_i)$$

- Con $H(Y | X=x_i)$ como:

$$H(Y | X = x_i) = - \sum_{j=1}^m P(Y = y_j | X = x_i) \log_2(P(Y = y_j | X = x_i))$$

- $P(Y=y_j | X=x_i)$ se interpreta como: ¿Cuál es la probabilidad de que Y tome el valor y_j , sabiendo que X toma el valor x_i ?

- **Ejemplo:** Supongamos que tenemos un canal de transmisión binario que transmite con ruido atendiendo a las siguientes probabilidades:



- ¿Cuál es la entropía condicionada del receptor con respecto al emisor?
- **CUIDADO:** lo que se muestra en el diagrama debe interpretarse como probabilidades condicionadas $p(Y=0|X=0)$, $p(Y=0|X=1)$, y $p(Y=1|X=0)$, $p(Y=1|X=1)$, dado que se sabe qué probabilidad hay de que Y reciba 0 (o 1) supuesto X enviando 0 (o 1).
- Además, el diagrama indica que X envía un 0 con probabilidad 0,5 y que envía un 1 con probabilidad 0,5

- ¿Cuál es la entropía condicionada del receptor con respecto al emisor?

$$H(Y | X) = - \sum_{i=1}^n p(X = x_i) \cdot \sum_{j=1}^m p(Y = y_j | X = x_i) \cdot \log_2(p(Y = y_j | X = x_i))$$

$$\begin{aligned} & p(X = 0) \cdot (-p(Y = 0 | X = 0) * \log_2(p(Y = 0 | X = 0)) \\ & - p(Y = 1 | X = 0) * \log_2(p(Y = 1 | X = 0))) + \\ & p(X = 1) \cdot (-p(Y = 0 | X = 1) * \log_2(p(Y = 0 | X = 1)) \\ & - p(Y = 1 | X = 1) * \log_2(p(Y = 1 | X = 1))) = \\ & - 2 * 0,5 * 0,85 * \log_2(0,85) - 2 * 0,5 * 0,15 * \log_2(0,15) = 0,6098 \end{aligned}$$

- ¿Y cuál sería la entropía conjunta?

- ¿Cuál es la entropía conjunta?

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(X = x_i, Y = y_j) \cdot \log_2(p(X = x_i, Y = y_j))$$

- ¿Cómo calculamos las probabilidades conjuntas? Fórmula del cálculo de probabilidades condicionadas.

$$P(Y | X) = \frac{P(X \cap Y)}{P(X)} \Rightarrow P(X \cap Y) = P(Y | X) * P(X)$$

$$P(X = 0, Y = 0) = P(Y = 0 | X = 0) * P(X = 0) = 0,85 * 0,5 = 0,425$$

$$P(X = 0, Y = 1) = P(Y = 1 | X = 0) * P(X = 0) = 0,15 * 0,5 = 0,075$$

$$P(X = 1, Y = 0) = P(Y = 0 | X = 1) * P(X = 1) = 0,15 * 0,5 = 0,075$$

$$P(X = 1, Y = 1) = P(Y = 1 | X = 1) * P(X = 1) = 0,85 * 0,5 = 0,425$$

- Ya podemos calcular $H(X, Y)$

– ¿Cuál es la entropía conjunta?

$$\begin{aligned}
 H(X, Y) &= - \sum_{i=1}^n \sum_{j=1}^m p(X = x_i, Y = y_j) \cdot \log_2(p(X = x_i, Y = y_j)) = \\
 &\quad - p(X = 0, Y = 0) \cdot \log_2(p(X = 0, Y = 0)) \\
 &\quad - p(X = 0, Y = 1) \cdot \log_2(p(X = 0, Y = 1)) \\
 &\quad - p(X = 1, Y = 0) \cdot \log_2(p(X = 1, Y = 0)) \\
 &\quad - p(X = 1, Y = 1) \cdot \log_2(p(X = 1, Y = 1)) = \\
 &\quad - 2 \cdot 0,425 \cdot \log_2(0,425) - 2 \cdot 0,075 \cdot \log_2(0,075) = \\
 &\quad 1,0493 + 0,5605 = 1,6098
 \end{aligned}$$

- ¿Y la entropía del emisor? ¿Y la entropía del receptor?

$$P(X = 0) = 0,5; \quad P(X = 1) = 0,5$$

$$\begin{aligned} H(X) &= -\sum_{i=1}^{n=2} p(X = x_i) \cdot \log_2(p(X = x_i)) = \\ &= -0,5 * \log_2(0,5) - 0,5 * \log_2(0,5) = 1 \end{aligned}$$

$$P(Y = 0) = 0,425 + 0,075 = 0,5; \quad P(Y = 1) = 0,075 + 0,425 = 0,5;$$

$$\begin{aligned} H(Y) &= -\sum_{i=1}^{m=2} p(Y = y_i) \cdot \log_2(p(Y = y_i)) = \\ &= -0,5 * \log_2(0,5) - 0,5 * \log_2(0,5) = 1 \end{aligned}$$

- Seguimos observando que $H(X,Y) < H(X)+H(Y)$

- Propiedades de la entropía de más de una variable:
 - La entropía o incertidumbre aportada por dos variables conjuntamente siempre será menor (o igual, en el caso de independencia) que la entropía o incertidumbre de ambas variables consideradas por separado.

$$H(X, Y) \leq H(X) + H(Y)$$

- La entropía de dos variables es simétrica:

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

- La entropía de dos variables se puede calcular a partir de la entropía condicionada y de las entropías individuales (Ley de Entropías totales):

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

1. ¿Qué es la entropía?
2. Entropía de más de una variable.
3. Información mutua.
4. Ejemplo adicional de aplicación de la Entropía: Sistemas Expertos.

- La **información mutua entre dos variables**, que notamos como $I(X;Y)$, es una medida para cuantificar la dependencia entre el input X y el output Y de un canal.

- La información mutua calcula la entropía de una variable, a la que se le quita la entropía condicionada:

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

- Es decir, la entropía de X , quitándole la entropía de X supuesto que supiésemos Y .

- La **información mutua entre dos variables es simétrica**, por lo que:

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

- La información mutua es una forma de expresar la dependencia entre lo que se recibe (Y) y lo que se envía (X).

- **Ejemplo:** Supongamos que tenemos un canal de transmisión binario que transmite con ruido atendiendo a las siguientes probabilidades conjuntas ($P(X,Y)$):

		Y	
		0	1
X	0	0,3	0,3
	1	0,3	0,1

- ¿Cuál es la información mutua entre X e Y?

- ¿Cuál es la información mutua entre X e Y?

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

- Por la Ley de entropías totales, podemos calcular $H(X | Y)$

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

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

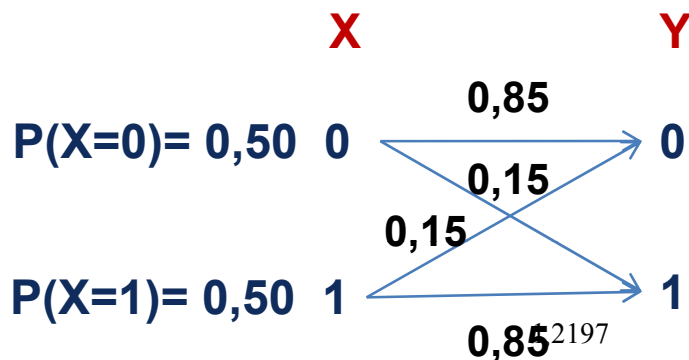
$$1,8955 - 0,971 = 0,9245$$

- Y ahora calcular la información mutua:

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

$$0,971 - 0,9245 = 0,0465$$

- **Ejemplo:** Supongamos que tenemos un canal de transmisión binario que transmite con ruido atendiendo a las siguientes probabilidades:

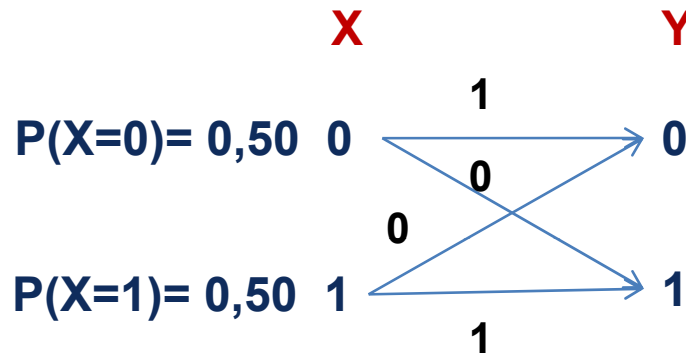


- ¿Cuál es la información mutua entre X e Y ?

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

$$1 - 0,6098 = 0,3902$$

- **Ejemplo:** Supongamos que tenemos un canal de transmisión binario que transmite con ruido atendiendo a las siguientes probabilidades:



- ¿Cuál es la información mutua entre X e Y ?
 - **Primero calculamos $H(X)$**

$$H(X) = -2 * 0,5 * \log_2(0,5) = 1$$

- ¿Cuál es la información mutua entre X e Y?
 - Ahora calculamos $H(X|Y)$. Para ello, antes tenemos que saber $P(X,Y)$ para calcular $P(Y)$ y $H(Y)$.

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \Rightarrow P(X \cap Y) = P(Y|X) * P(X)$$

$$P(X=0, Y=0) = P(Y=0|X=0) * P(X=0) = 1 * 0,5 = 0,5$$

$$P(X=0, Y=1) = P(Y=1|X=0) * P(X=0) = 0 * 0,5 = 0$$

$$P(X=1, Y=0) = P(Y=0|X=1) * P(X=1) = 0 * 0,5 = 0$$

$$P(X=1, Y=1) = P(Y=1|X=1) * P(X=1) = 1 * 0,5 = 0,5$$

– Por tanto:

$$P(Y=0) = 0,5 + 0 = 0,5; \quad P(Y=1) = 0 + 0,5 = 0,5$$

$$H(Y) = -2 * 0,5 * \log_2(0,5) = 1$$

– ¿Cuál es la información mutua entre X e Y?

– **Ahora calculamos $H(X,Y)$:**

$$\begin{aligned}
 H(X,Y) &= - \sum_{i=1}^n \sum_{j=1}^m p(X = x_i, Y = y_j) \cdot \log_2(p(X = x_i, Y = y_j)) = \\
 &\quad - p(X = 0, Y = 0) \cdot \log_2(p(X = 0, Y = 0)) \\
 &\quad - p(X = 0, Y = 1) \cdot \log_2(p(X = 0, Y = 1)) \\
 &\quad - p(X = 1, Y = 0) \cdot \log_2(p(X = 1, Y = 0)) \\
 &\quad - p(X = 1, Y = 1) \cdot \log_2(p(X = 1, Y = 1)) = \\
 &\quad - 2 * 0,5 \cdot \log_2(0,5) - 2 * 0 \cdot \log_2(0) = 1
 \end{aligned}$$

– **Y aplicamos Ley de Entropías Totales para $H(X|Y)$:**

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

$$H(X | Y) = H(X,Y) - H(Y) = 1 - 1 = 0$$

- ¿Cuál es la información mutua entre X e Y?
- Ya podemos calcular $I(X;Y)$:

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

$$I(X;Y) = 1 - 0 = 1$$

- Conclusiones:
 - En canales sin ruido, $H(X|Y) = H(Y|X) = 0$
 - Esto indica que se puede conocer unívocamente el símbolo recibido en el receptor sabiendo el símbolo del emisor
 - Además, $I(X; Y) = \text{máximo}$
 - La información mutua transmitida de X a Y es máxima.

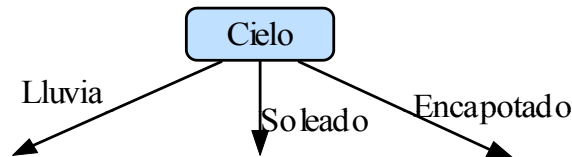
1. ¿Qué es la entropía?
2. Entropía de más de una variable.
3. Información mutua.
4. **Ejemplo adicional de aplicación de la Entropía: Sistemas Expertos.**

- La Entropía es una medida muy útil no sólo en comunicaciones, sino en gran parte de la IA moderna.
- **Un ejemplo son los Sistemas Expertos (algoritmo ID3).**

Día	Cielo	Temp.	Humedad	Viento	Jugar
1	Soleado	Calor	Alta	Débil	No
2	Soleado	Calor	Alta	Fuerte	No
3	Encapotado	Calor	Alta	Débil	Si
4	Lluvia	Suave	Alta	Débil	Si
5	Lluvia	Frío	Normal	Débil	Si
6	Lluvia	Frío	Normal	Fuerte	No
7	Encapotado	Frío	Normal	Fuerte	Si
8	Soleado	Suave	Alta	Débil	No
9	Soleado	Frío	Normal	Débil	Si
10	Lluvia	Suave	Normal	Débil	Si
11	Soleado	Suave	Normal	Fuerte	Si
12	Encapotado	Suave	Alta	Fuerte	Si
13	Encapotado	Calor	Normal	Débil	Si
14	Lluvia	Suave	Alta	Fuerte	No

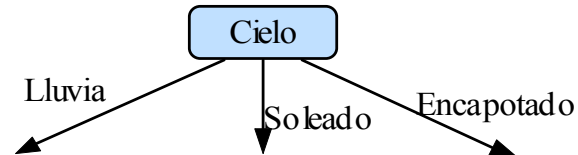
- En un sistema experto, intentamos reducir la incertidumbre de un atributo clase u objetivo con respecto a los demás
 - ¿Qué atributo tiene menos incertidumbre para jugar?
 - $H(\text{Jugar} \mid \text{Cielo})$?
 - $H(\text{Jugar} \mid \text{Temperatura})$?
 - $H(\text{Jugar} \mid \text{Humedad})$?
 - $H(\text{Jugar} \mid \text{Viento})$?

- **Ejemplo:** Tras calcular la medida de entropía (**mínima entropía= mínima incertidumbre**), el atributo que menor entropía tiene es **Cielo**. Generamos el árbol de Decisión con este atributo como raíz.



- A cada nodo hijo le asignamos una subtabla, eliminando el atributo **Cielo** y quedándonos sólo con aquellos patrones donde Cielo tenga el valor de la rama del nodo hijo.

Ejemplo:



Cielo=Lluvia

Temp	Humd.	Viento	Jugar
Suave	Alta	Débil	Si
Frío	Normal	Débil	Si
Frío	Normal	Fuerte	No
Suave	Normal	Débil	Si
Suave	Alta	Fuerte	No

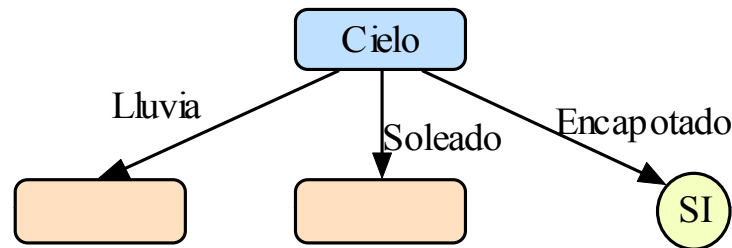
Cielo=Encapotado

Temp	Humd.	Viento	Jugar
Calor	Alta	Débil	Si
Frío	Normal	Fuerte	Si
Suave	Alta	Fuerte	Si
Calor	Normal	Débil	Si

Cielo=Soleado

Temp	Humd.	Viento	Jugar
Calor	Alta	Débil	No
Calor	Alta	Fuerte	No
Suave	Alta	Débil	No
Frío	Normal	Débil	Sí
Suave	Normal	Fuerte	Sí

- Ejemplo:** Para el valor **Cielo=Encapotado**, todos los patrones tienen el mismo valor de clase $Jugar=SI$. Criterio de parada de nodo puro: Se asigna el valor de la clase del nodo.



- De aquí podremos extraer la regla “***Si** cielo=Encapotado **entonces** Jugar=Si*”

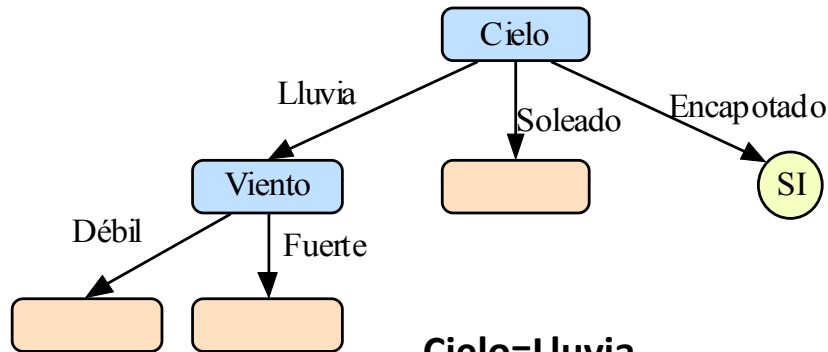
- **Ejemplo:** Pasemos ahora, por ejemplo, a analizar la tabla que queda cuando Cielo=Lluvia.

Cielo=Lluvia

Temp	Humd.	Viento	Jugar
Suave	Alta	Débil	Si
Frío	Normal	Débil	Si
Frío	Normal	Fuerte	No
Suave	Normal	Débil	Si
Suave	Alta	Fuerte	No

- El atributo que menos incertidumbre proporciona es “Viento”, con valor mínimo de entropía **$H(\text{Jugar} \mid \text{Viento}^{\wedge} \text{Cielo}=\text{“Lluvia”})$** .

- Ejemplo:** Volvemos a dividir la tabla según los valores de Viento, y asignamos las subtablas a los nodos creados.



- En ambos nodos se da el criterio de parada de nodo puro: Se asigna el valor de la clase del nodo.

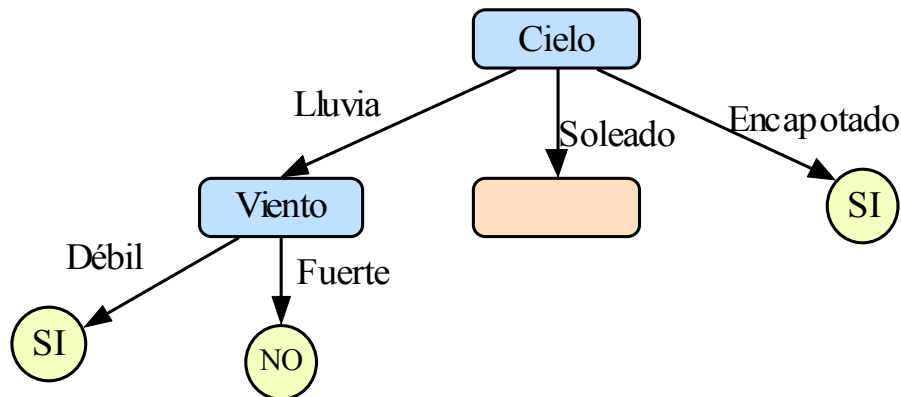
Cielo=Lluvia
Viento= Débil

Temp	Humd.	Jugar
Suave	Alta	Si
Frío	Normal	Si
Suave	Normal	Si

Cielo=Lluvia
Viento= Fuerte

Temp	Humd.	Jugar
Frío	Normal	No
Suave	Alta	No

- Ejemplo:** Volvemos al atributo “Cielo”, para expandir el nodo que nos quedaba de antes:

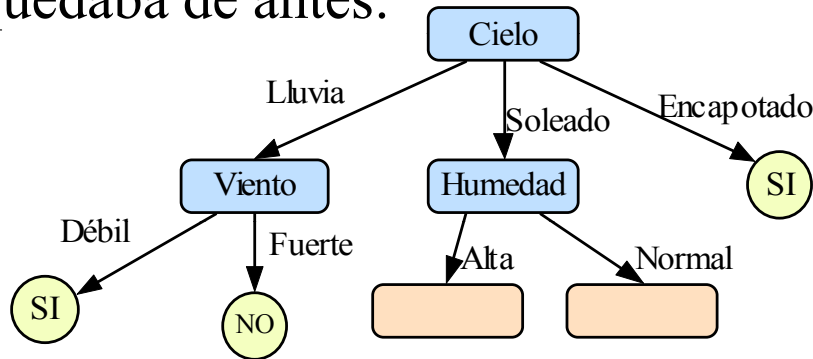


Cielo=Soleado

Temp	Humd.	Viento	Jugar
Calor	Alta	Débil	No
Calor	Alta	Fuerte	No
Suave	Alta	Débil	No
Frío	Normal	Débil	Sí
Suave	Normal	Fuerte	Sí

- El atributo de mínimo valor de entropía es “Humedad” ($H(\text{Jugar}|\text{Humedad}^{\wedge}\text{Cielo}=\text{Soleado})$). Volvemos a hacer la operación para generar hijos.

- **Ejemplo:** Volvemos al atributo “Cielo”, para expandir el nodo que nos quedaba de antes:



**Cielo=Soleado
Humedad=Alta**

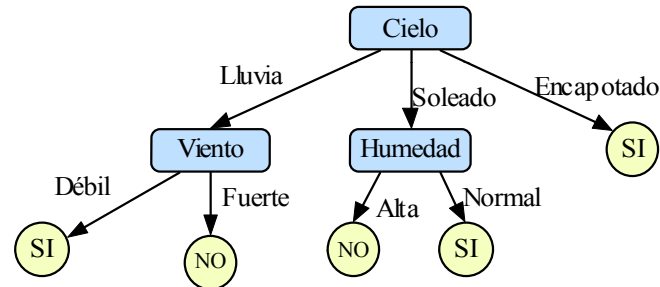
Temp	Viento	Jugar
Calor	Débil	No
Calor	Fuerte	No
Suave	Débil	No

**Cielo=Soleado
Humedad=Normal**

Temp	Viento	Jugar
Frío	Débil	Sí
Suave	Fuerte	Sí

- En ambos nodos se da el criterio de parada de nodo puro: Se asigna el valor de la clase del nodo.

- **Ejemplo:** El algoritmo para porque ya no hay más nodos que expandir.



- Las reglas se extraen del árbol de decisión recorriéndolo de la raíz a las hojas:
 - **Si** Cielo=Encapotado **entonces** Jugar=SI
 - **Si** Cielo=Soleado **Y** Humedad=Alta **entonces** Jugar=NO
 - **Si** Cielo=Soleado **Y** Humedad=Normal **entonces** Jugar=SI
 - Etc.



Universidad de Granada

decsai.ugr.es

Teoría de la Información y la Codificación

Grado en Ingeniería Informática

Tema 2.- La entropía como medida de información.



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**