Teoría de la información y métodos de codificación

Información y Entropía

Información y entropía

Una de las características principales de la teoría de información es su enfoque en la frecuencia relativa o la ocurrencia de mensajes y no en su significado

Cada mensaje producido por una fuente de información esta formado por una combinación de símbolos que forman parte de cierto **alfabeto**

Por ejemplo, un mensaje de texto de cualquier longitud es básicamente una secuencia específica de caracteres alfanuméricos y especiales

Información y entropía

El alfabeto de una fuente se puede representar por medio del conjunto $\{X\}$ con elementos (símbolos) $x_1, x_2, \ldots, x_M$ donde M es el tamaño del alfabeto

Sobre este alfabeto es posible entonces definir las probabilidades de ocurrencia de todos los símbolos,

$$p(x_i), \quad i = 1, 2, \ldots, M$$

En donde $p(x_i) \geq 0 \quad i = 1, 2, \ldots, M$ y $\sum_{i=1}^{M} p(x_i) = 1$

Información y entropía

Una fuente de información queda completamente definida por medio de su alfabeto X y su respectivo conjunto de probabilidades de ocurrencia $\{p(x_i)\}$

Debido a eso, se dice que la teoría de la información se puede ver también como teoría de probabilidad aplicada

Información y entropía

El objetivo ahora es encontrar una **medida de la cantidad de información** que contiene el símbolo x_i basado en su probabilidad de ocurrencia $p(x_i)$

El concepto intuitivo de información sugiere que los símbolos con menor probabilidad de ocurrencia deberán representar mayor cantidad de información que aquellos que ocurren con mayor frecuencia

Por lo tanto, la medida de información deberá de ser una función decreciente de $p(x_i)$

Información y entropía

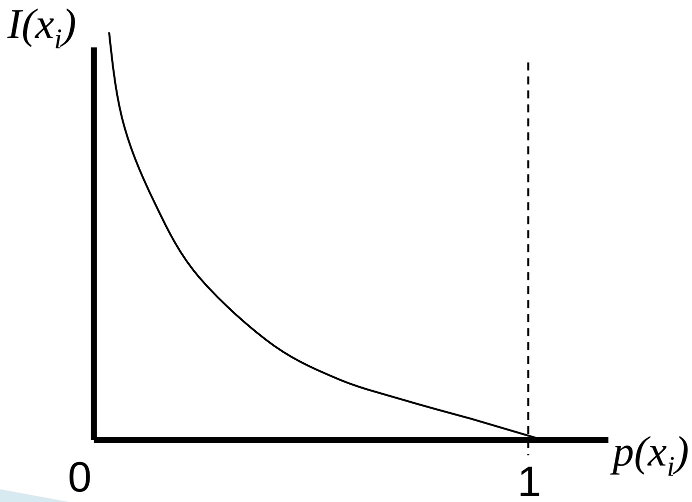
Una función con la característica antes mencionada es la función $\dfrac{1}{p(x_i)}$

Sin embargo, la función que además de satisfacer el comportamiento buscado también cumple otras propiedades que debe tener la medida de la información es la función logaritmo de $1/p(x_i)$

Información y entropía

Definición. La medida de información o también llamada la **información propia de** x_i se denota por

$$I(x_i) = \log_b \frac{1}{p(x_i)} = -\log_b p(x_i)$$



Información y entropía

La expresión anterior satisface la noción intuitiva que se tiene acerca de la información asociada a un evento dada su ocurrencia

Nótese que $I(x_i)$ es una función no-negativa y decreciente en el intervalo (0,1]

Información y entropía

Si $p(x_i) = 1$ entonces $I(x_i) = 0$ y por lo tanto la ocurrencia de dicho evento o símbolo no genera ninguna información

Por el contrario es muy pequeña o bien $p(x_i) \to 0$ entonces su ocurrencia debe generar mayor información

Información y entropía

La selección de la base del logaritmo (b) determina el tipo de unidades asociadas a la medida de información $I(x_i)$ de algún evento o símbolo

$$I(x_i) = \log_2 \frac{1}{p(x_i)} = -\log_2 p(x_i) \quad bits$$

Ejemplo. Se tiene una fuente de información la cual emite 4 símbolos A, B, C y D con probabilidades de ocurrencia 0.5, 0.25, 0.125 y 0.125, respectivamente.

Calcular la información propia del mensaje $X = BDA$ asumiendo que los símbolos son independientes

Información y entropía

Debido a que se trabaja con fuentes de información que producen secuencias de símbolos que pertenecen a un alfabeto, la información $I(x_i)$ es a su vez una variable aleatoria discreta

Una de las propiedades estadísticas mas importantes para cualquier variable aleatoria es su valor esperado

En este caso para $I(x_i)$ tenemos

$$E\left[I(x_i)\right] = H(X) = \sum_{i=1}^{M} p(x_i) I(x_i) = -\sum_{i=1}^{M} p(x_i) \log_2 p(x_i)$$

Entropía

La expresión $H(X)$ define la entropía de una fuente de información

En la teoría de información, $H(X)$ representa una medida de la incertidumbre que posee una fuente de información

Es decir, $H(X)$ calcula el número de bits que en promedio se necesitan para representar la información producida por alguna fuente

Entropía de una fuente binaria

Sea $S = \{0,1\}$ el alfabeto de una fuente binaria con probabilidades $p(0)=p$ y $p(1)=1-p$

De la expresión para calcular la entropía tenemos

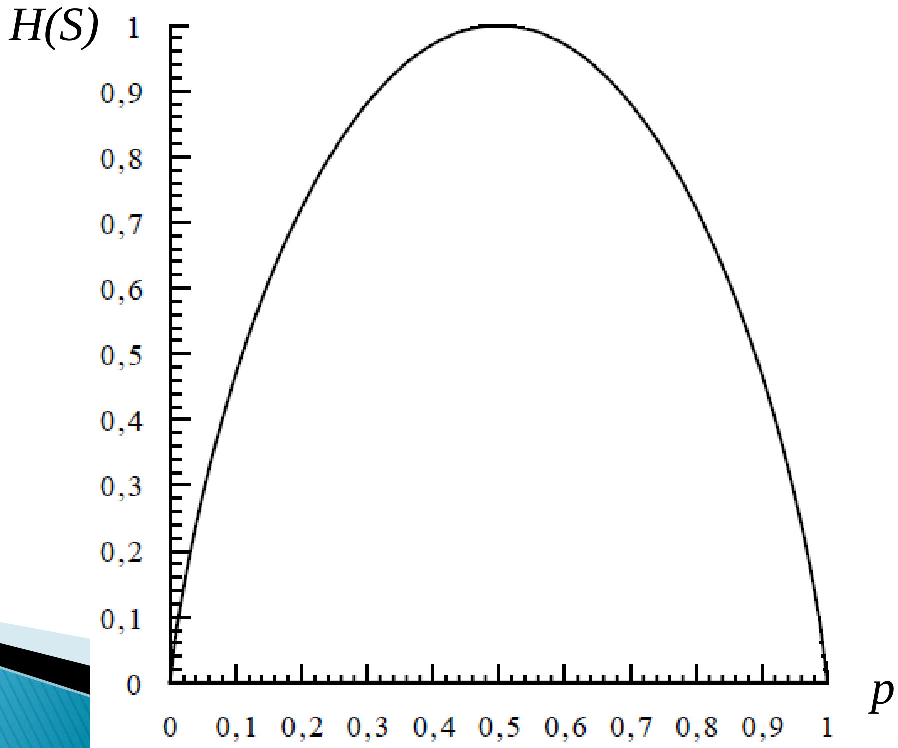
$$H(S) = -[p\log_2 p + (1-p)\log_2(1-p)]$$

La gráfica de la ecuación con respecto a p se muestra a continuación

Entropía de una fuente binaria

Propiedades.

1. *H(S)* es no negativa
2. *H(S)* es cero solo para $p = 0$ y $p = 1$
3. *H(S)* tiene un máximo para $p = 1 - p = 0.5$



Entropía de una fuente binaria

Ejemplos. Calcular la entropía de la fuente binaria para los siguientes casos

a. Los símbolos son equiprobables
b. Sea $p(0) = 0.96875$ y $p(1) = 0.03125$
c. Sea $p(0) = 1$ y $p(1) = 0$

Entropía de una fuente

Ejemplo. Calcular la entropía la fuente de información que emite 4 símbolos A, B, C y D con probabilidades de ocurrencia 0.5, 0.25, 0.125 y 0.125, respectivamente.

Repetir el cálculo pero ahora suponiendo que los símbolos son equiprobables

Entropía de una fuente

Propiedades

a. $H(S) \leq \log_2 M$

b. $H(S) = \log_2 M$

si y solo si todas las probabilidades son iguales a $p(s_i) = \frac{1}{M}$ para toda i (equiprobables)

En general, el valor de la entropía es máximo para cuando los símbolos de una fuente son equiprobables

Tipos de fuentes de información

Una fuente de información se puede definir en términos de 2 atributos principales que son su alfabeto $S=\{s_1, s_2, \ldots, s_M\}$ y la distribución de probabilidades de los símbolos que emite $p(s_1), p(s_2), \ldots, p(s_M)$

Definición. Una **fuente sin memoria** es aquella en donde los símbolos que emite son completamente independientes entre si, es decir, en todo momento no existe ningún tipo de correlación entre los símbolos

Fuentes sin memoria

Además de existir independencia entre los símbolos de salida, si las distribuciones de probabilidad son iguales entonces la fuente es además ***estacionaria***

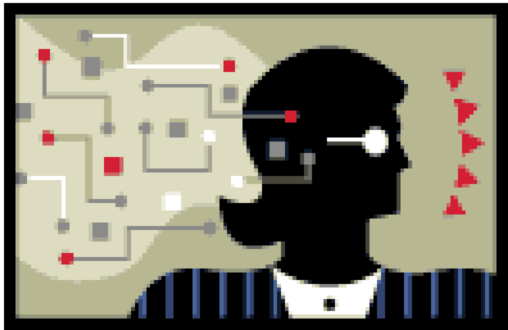
El modelo para fuentes sin memoria es el mas simple que se pueda encontrar

El comportamiento de las fuentes sin memoria es un tanto restrictivo para muchas aplicaciones reales

Fuentes con memoria

Un tipo de fuente que se emplea mas en aplicaciones reales es aquella en donde la ocurrencia de un símbolo s_i se ve afectada por un número finito de símbolos que se emitieron previamente

Debido a lo anterior se dice que este tipo de fuentes poseen ***memoria***



Fuentes con memoria

Definición. Las fuentes con memoria de orden ***n*** son aquellas en donde ***n*** representa el número de símbolos previamente emitidos y que influyen en la probabilidad de ocurrencia del siguiente símbolo a procesar

Una fuente con memoria se define en términos de su alfabeto $S = \{s_1, s_2, \ldots, s_M\}$ y el conjunto de probabilidades condicionales

$$p(s_i \mid s_{j_1}, s_{j_2}, s_{j_3}, \ldots, s_{j_n})$$

Para $i = 1,2,\ldots,M;\ j_k = 1,2,\ldots,M$

Fuentes con memoria

El conjunto de ***n*** símbolos previos representan el **estado** en el que se encuentra una fuente con memoria en un determinado momento

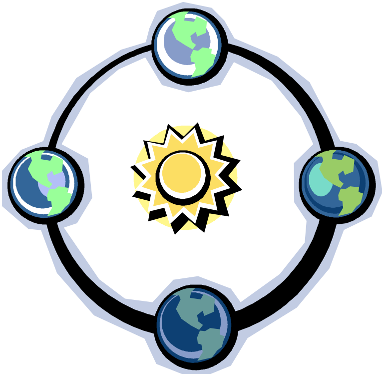
Dado que el alfabeto $S = \{s_1, s_2, \ldots, s_M\}$ contiene M símbolos, una fuente con memoria de orden ***n*** tendrá $M^n = q$ estados

La probabilidad de que la fuente se encuentre en algún estado en particular varía con el tiempo a medida que emite símbolos

Diagramas de estados

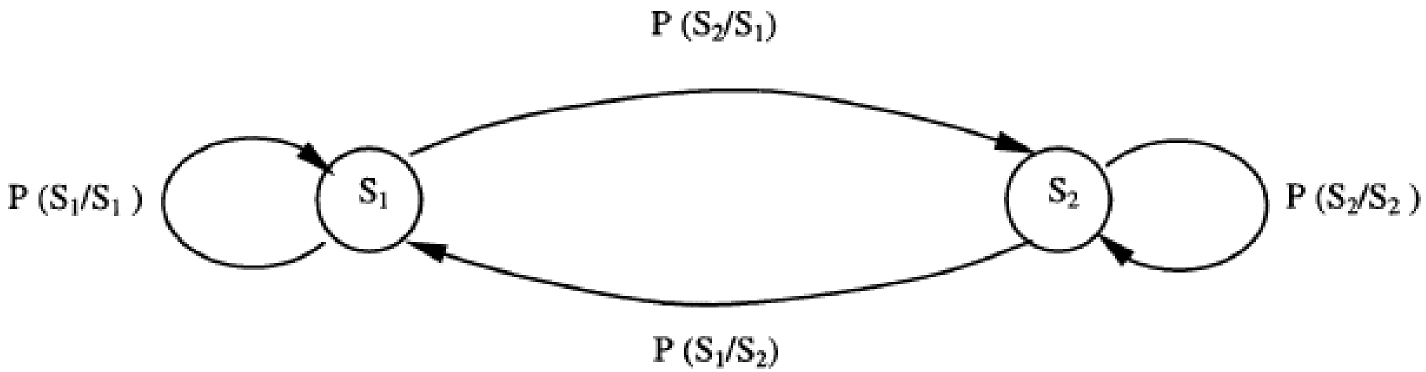
Una forma de ilustrar el comportamiento de las fuentes con memoria es por medio de diagramas de estados

En un diagrama de estados se pueden representar todos los M^n estados con vértices así como todas las posibles transiciones de estado a estado por flechas



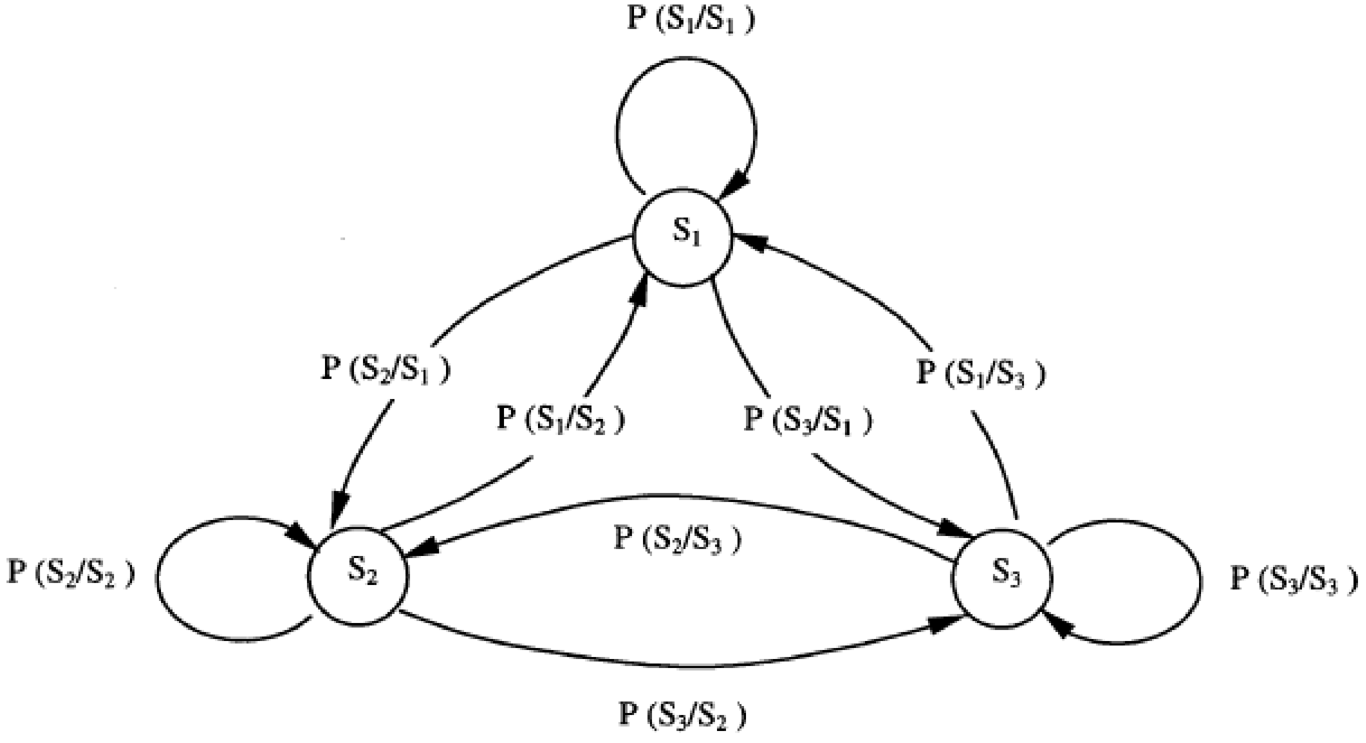
Diagramas de estados

Diagrama de estados para una fuente con memoria de primer orden con alfabeto $S = \{s_1, s_2\}$



Diagramas de estados

Diagrama de estados para una fuente con memoria de primer orden con alfabeto $S = \{s_1, s_2, s_3\}$



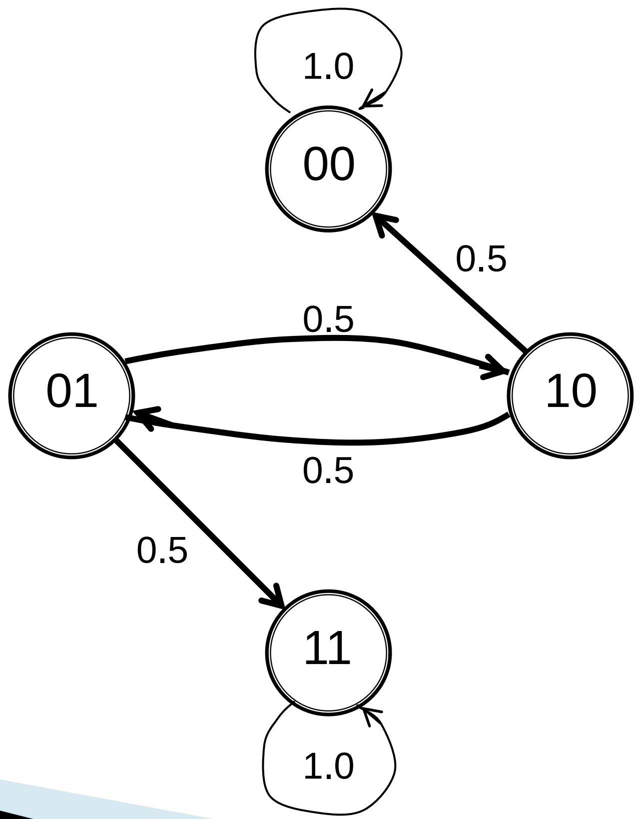
Fuentes con memoria

En nuestro estudio solo nos enfocaremos en fuentes con memoria que sean **ergódicas**

Una fuente con memoria es ergódica si es posible que se mueva de un estado a cualquier otro estado en un número finito de pasos

Fuentes con memoria no ergódica

El siguiente diagrama de estados corresponde a una fuente con memoria de segundo orden con alfabeto $S = \{0, 1\}$ y la cual es **no ergódica**



Fuentes con memoria

Definición. Una fuente con memoria la definen los siguientes atributos

a. Un alfabeto $S = \{s_1, s_2, \ldots, s_M\}$
b. Un conjunto de estados $\sum = \{\sigma_1, \sigma_2, \ldots, \sigma_q\}$
c. Dos conjuntos de probabilidades
   i. El primer conjunto define la distribución de probabilidades iniciales de todos los estados y por lo regular se representa por medio de un **vector de probabilidades iniciales** π
   ii. El segundo conjunto representa las probabilidades de transición entre estados, para cada par de estados σ_i y σ_j la probabilidad de transición de σ_i a σ_j se denota por $p(\sigma_j \mid \sigma_i)$ y todas se representan en la **matriz de probabilidades de transición**

Fuentes con memoria

El comportamiento de una fuente con memoria se representa de forma matemática utilizando los atributos anteriores en conjunto con lo que se conoce como matriz de probabilidades de transición

$$T = [p_{jk}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1q} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2q} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ p_{q1} & p_{q2} & p_{q3} & \cdots & p_{qq} \end{bmatrix}$$

Fuentes con memoria

Ejemplo 1. Encontrar la matriz de probabilidades de transición para una fuente con memoria de segundo orden y con alfabeto $S = \{0, 1\}$

Se conocen también las siguientes probabilidades

$P(00)=0.4$	$P(01)=0.3$	$P(10)=0.2$	$P(11)=0.1$

$P(000)=0.32$	$P(001)=0.08$
$P(010)=0.15$	$P(011)=0.15$
$P(100)=0.16$	$P(101)=0.04$
$P(110)=0.06$	$P(111)=0.04$

Fuentes con memoria

Ejemplo 1 (cont.) Como $M = 2$ y estamos usando una fuente con memoria de segundo orden, entonces vamos a tener un total de $4(2^2)$ estados

$$\left.\begin{aligned} \sigma_1 &= 00 \\ \sigma_2 &= 01 \\ \sigma_3 &= 10 \\ \sigma_4 &= 11 \end{aligned}\right\} \qquad \pi = \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix}$$

Fuentes con memoria

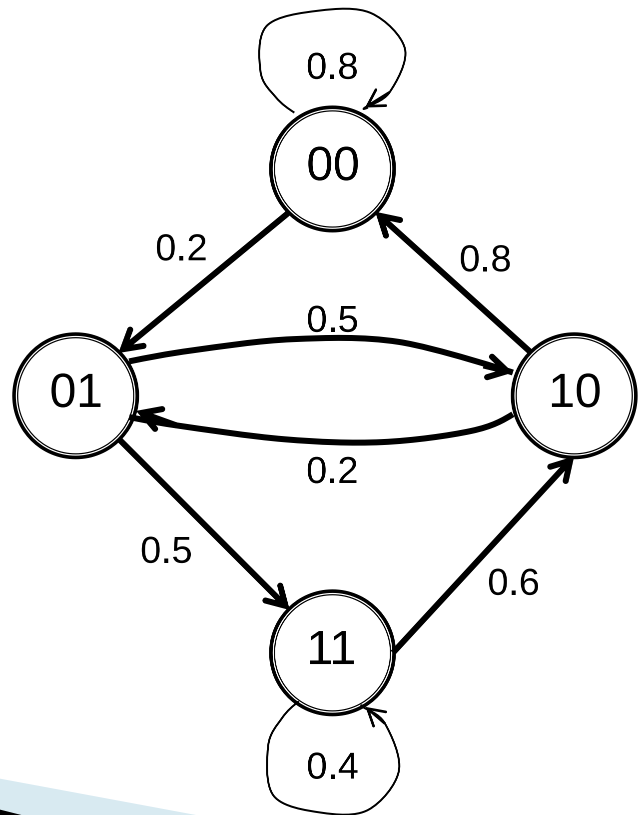
Ejemplo 1 (cont.) Dado que es un modelo con memoria de segundo orden, ahora debemos encontrar el conjunto de probabilidades condicionales de la forma

$$p(s_i \mid s_{j_1}, s_{j_2}) = p(\sigma_i \mid \sigma_j)$$

$P(0\|00)=0.8$	$P(1\|00)=0.2$
$P(0\|01)=0.5$	$P(1\|01)=0.5$
$P(0\|10)=0.8$	$P(1\|10)=0.2$
$P(0\|11)=0.6$	$P(1\|11)=0.4$

Fuentes con memoria

Ejemplo 1 (cont.) El diagrama de estados de la fuente con memoria de segundo orden queda de la siguiente forma



Fuentes con memoria

Ejemplo 1 (cont.) La matriz de probabilidades de transición en este caso es

$$T = \begin{bmatrix} p(\sigma_1 | \sigma_1) & p(\sigma_2 | \sigma_1) & p(\sigma_3 | \sigma_1) & p(\sigma_4 | \sigma_1) \\ p(\sigma_1 | \sigma_2) & p(\sigma_2 | \sigma_2) & p(\sigma_3 | \sigma_2) & p(\sigma_4 | \sigma_2) \\ p(\sigma_1 | \sigma_3) & p(\sigma_2 | \sigma_3) & p(\sigma_3 | \sigma_3) & p(\sigma_4 | \sigma_3) \\ p(\sigma_1 | \sigma_4) & p(\sigma_2 | \sigma_4) & p(\sigma_3 | \sigma_4) & p(\sigma_4 | \sigma_4) \end{bmatrix}$$

$$T = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}$$

Entropía de una fuente con memoria

Se puede probar que la entropía de una fuente con memoria de orden ***n*** esta dada por

$$E\left[I(s_i \mid s_{j_1}, s_{j_2}, \ldots, s_{j_n})\right] = H(S)$$

$$H(S) = \sum_{S^{n+1}} p(s_{j_1}, s_{j_2}, \ldots, s_{j_n}, s_i) \log_2 \frac{1}{p(s_i \mid s_{j_1}, \ldots, s_{j_n})}$$

Entropía de una fuente con memoria

Ejemplo 1 (cont.) Volviendo al ejemplo anterior de la fuente con memoria de segundo orden tenemos las siguiente probabilidades

s_{j1},s_{j2},s_i	$p(s_i \mid s_{j1},s_{j2})$	$p(s_{j1},s_{j2})$	$p(s_{j1},s_{j2},s_i)$
000	0.8	0.4	0.32
001	0.2	0.4	0.08
010	0.5	0.3	0.15
011	0.5	0.3	0.15
100	0.8	0.2	0.16
101	0.2	0.2	0.04
110	0.6	0.1	0.06
111	0.4	0.1	0.04

Entropía de una fuente con memoria

Ejemplo 1 (cont.) El valor de la entropía es

$$H(S)=\sum_{S^3} p(s_{j_1},s_{j_2},s_i)\log_2\frac{1}{p(s_i\mid s_{j_1},s_{j_2})}$$

$$=0.32\log_2\frac{1}{0.8}+0.08\log_2\frac{1}{0.2}+0.15\log_2\frac{1}{0.5}+0.15\log_2\frac{1}{0.5}+$$

$$0.16\log_2\frac{1}{0.8}+0.04\log_2\frac{1}{0.2}+0.06\log_2\frac{1}{0.6}+0.04\log_2\frac{1}{0.4}$$

Entropía de una fuente con memoria

Ejemplo 2 Sea una fuente con memoria de primer orden con alfabeto $S = \{A, B, C\}$ con un vector de probabilidades iniciales

$$\pi = \begin{bmatrix} \frac{9}{27} & \frac{16}{27} & \frac{2}{27} \end{bmatrix}$$

Una matriz de probabilidades de transición

$$T = \begin{bmatrix} 0 & 4/5 & 1/5 \\ 1/2 & 1/2 & 0 \\ 1/2 & 2/5 & 1/10 \end{bmatrix}$$

Calcular la entropía de la fuente $H(S)$

Entropía del alfabeto inglés (26 letras y el espacio)

Columna 2 =
Letras equiprobables

Columna 3 =
H(S) = 4.75 bits/sim.

□	0.0370	4.7549	0.1859	2.4274
a	0.0370	4.7549	0.0642	3.9613
b	0.0370	4.7549	0.0127	6.2990
c	0.0370	4.7549	0.0218	5.5195
d	0.0370	4.7549	0.0317	4.9794
e	0.0370	4.7549	0.1031	3.2779
f	0.0370	4.7549	0.0208	5.5873
g	0.0370	4.7549	0.0152	6.0398
h	0.0370	4.7549	0.0467	4.4204
i	0.0370	4.7549	0.0575	4.1203
j	0.0370	4.7549	0.0008	10.2877
k	0.0370	4.7549	0.0049	7.6730
l	0.0370	4.7549	0.0321	4.9613
m	0.0370	4.7549	0.0198	5.6584
n	0.0370	4.7549	0.0574	4.1228
o	0.0370	4.7549	0.0632	3.9839
p	0.0370	4.7549	0.0152	6.0398
q	0.0370	4.7549	0.0008	10.2877
r	0.0370	4.7549	0.0484	4.3688
s	0.0370	4.7549	0.0514	4.2821
t	0.0370	4.7549	0.0796	3.6511
u	0.0370	4.7549	0.0228	5.4548
v	0.0370	4.7549	0.0083	6.9127
w	0.0370	4.7549	0.0175	5.8365
x	0.0370	4.7549	0.0013	9.5873
y	0.0370	4.7549	0.0164	5.9302
z	0.0370	4.7549	0.0005	10.9658

Columna 4 =
Probabilidades de ocurrencia del alfabeto en el idioma inglés

Columna 5 =
H(S) = 4.03 bits/simb.

Entropía del idioma inglés

Empleando un modelo con memoria de primer orden y sus respectivas probabilidades condicionales se obtiene

$$H(S) = \sum_{S^2} p(s_j, s_i) \log_2 \frac{1}{p(s_i \mid s_j)} = 3.32\, bits / simbolo$$

Empleando un modelo con memoria de segundo orden se obtiene

$$H(S) = \sum_{S^3} p(s_{j_1}, s_{j_2}, s_i) \log_2 \frac{1}{p(s_i \mid s_{j_1}, s_{j_2})} = 3.1\, bits / simbolo$$

Entropía conjunta y condicional

Puede haber casos en donde sea necesario analizar el comportamiento de dos o mas fuentes de información ya sea de forma conjunta o bien de manera condicional de unas o otras

De forma similar a como se definieron los conceptos de probabilidades conjuntas y condicionales para dos o mas variables aleatorias, se pueden introducir los conceptos de entropía conjunta y condicional

Entropía conjunta y condicional

Definición. La entropía conjunta de 2 variables aleatorias discretas X y Y se define por

$$H(X,Y) = \sum_{x,y} p(x,y) \log_2 \frac{1}{p(x,y)}$$

Definición. La entropía condicional de una variable aleatoria X dada otra variable Y esta dada por

$$H(X \mid Y) = \sum_{x,y} p(x,y) \log_2 \frac{1}{p(x \mid y)}$$