Notes de

v.a. X $\mathcal{A}=\{x_1,\dots,x_N\}$
v.a Y $\mathcal{B}=\{y_1,\dots,y_M\}$

Entropie conditionnelle : $H(Y|X)$

Mesure l'entropie restante (i.e. l'incertitude) d'une v.a Y sachant que la valeur d'une autre v.a X est connue.

$H(Y|X)$ est la moyenne de $H(Y|X=x_i)$ sur toutes les valeurs possibles x_i que peut prendre la v.a X

$$H(Y|X) \triangleq \sum_{i=1}^{N} p(X=x_i).\, H(Y|X=x_i)$$

$$= \sum_{i=1}^{N} p(X=x_i) \left\{ - \sum_{j=1}^{M} p(Y=y_j|X=x_i).\log p(Y=y_j|X=x_i) \right\}$$

$$= - \sum_{i=1}^{N} \sum_{j=1}^{M} \underbrace{p(X=x_i).\,p(Y=y_j|X=x_i)}_{\downarrow}.\log p(Y=y_j|X=x_i) \Big\}$$

$$= - \sum_{i=1}^{N} \sum_{j=1}^{M} p(X=x_i, Y=y_j).\log \frac{p(Y=y_j, X=x_i)}{p(X=x_i)}$$

$$H(Y|X) = \sum_{i=1}^{N} \sum_{j=1}^{M} p(X=x_i, Y=y_j).\log \frac{p(X=x_i)}{p(Y=y_j, X=x_i)}$$

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

$$H(X,Y) \triangleq - \sum_{j=1}^{M} \sum_{i=1}^{N} p(X=x_i, Y=y_j) \log p(X=x_i, Y=y_j) \quad \text{Bayes}$$

$$= - \sum_j \sum_i p(X=x_i|Y=y_j).\,p(Y=y_j) \log \{ p(X=x_i|Y=y_j).\,p(Y=y_j) \}$$

$$= - \sum_j \sum_i p(X=x_i|Y=y_j).\,p(Y=y_j) \log p(X=x_i|Y=y_j) \qquad \log ab = \log a + \log b$$
$$- \sum_j \sum_i p(X=x_i|Y=y_j).\,p(Y=y_j) \log p(Y=y_j)$$

$$= - \sum_j p(Y=y_j) \sum_i p(X=x_i|Y=y_j) \log p(X=x_i|Y=y_j)$$
$$- \sum_j p(Y=y_j) \log p(Y=y_j) \sum_i p(X=x_i|Y=y_j)$$

1ère ligne = entropie conditionnelle $H(X|Y=y_j) \Rightarrow H(X|Y)$

2ème ligne $\sum_i p(X=x_i|Y=y_j)$ loi conditionnelle $P(X|Y=y_j)$ $=1$

$$= - \sum_j p(Y=y_j).\,H(X|Y=y_j) - \sum_j p(Y=y_j).\log p(Y=y_j)$$

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

De la m̂ façon $H(X,Y) = H(Y|X) + H(X)$

Information mutuelle. Montrer que $I(X;Y) = H(Y) - H(Y|X)$.

$$I(X;Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} p(X=x_i, Y=y_j) \cdot \log \frac{p(X=x_i, Y=y_j)}{p(X=x_i)\, p(Y=y_j)}$$

Bayes

$$I(X;Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} p(Y=y_j \mid X=x_i) \cdot p(X=x_i) \cdot \log \frac{p(Y=y_j \mid X=x_i) \cdot p(X=x_i)}{p(X=x_i) \cdot p(Y=y_j)}$$

$\log \frac{a}{b} = \log a - \log b$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} p(Y=y_j \mid X=x_i) \cdot p(X=x_i) \left\{ \log p(Y=y_j \mid X=x_i) - \log p(Y=y_j) \right\}$$

$$= \cancel{\sum_{i=1}^{N} p(X=x_i) \sum_{j=1}^{M} p(Y=y_j) \{ -\log p(Y=y_j) \}}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} p(Y=y_j \mid X=x_i) \cdot p(X=x_i) \log p(Y=y_j \mid X=x_i)$$

$$- \sum_{i=1}^{N} \sum_{j=1}^{M} p(Y=y_j \mid X=x_i) \cdot p(X=x_i) \log p(Y=y_j)$$

$$= \sum_{i=1}^{N} p(X=x_i) \underbrace{\sum_{j=1}^{M} p(Y=y_j \mid X=x_i) \cdot \log p(Y=y_j \mid X=x_i)}_{-H(Y|X=x_i)}$$

d'après Bayes.

$$- \sum_{i=1}^{N} \sum_{j=1}^{M} p(X=x_i \mid Y=y_j) \cdot p(Y=y_j) \log p(Y=y_j)$$

$$= -H(Y|X) - \sum_{j=1}^{M} p(Y=y_j) \cdot \log p(Y=y_j) \underbrace{\sum_{i=1}^{N} p(X=x_i \mid Y=y_j)}_{\text{loi conditionnelle}}$$

$$\sum_i p(X \mid Y=y_j) = 1$$

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

De la m̂ façon on peut montrer que $I(X;Y) = H(X) - H(X|Y)$

Montrer que $H(X)$ est une auto information $I(X;X)$.

1ère façon : utiliser la définition de l'information mutuelle et la démonstration est directe.

2ième façon :
$$H(X) = -\sum_{i=1}^{N} p(X=x_i) \log p(X=x_i)$$
$$= -\sum_{i=1}^{N} p(X=x_i) \log \frac{p(X=x_i)^2}{p(X=x_i)}$$
$$H(X) = I(X;X)$$

Source sans mémoire : chaque symbole émis est indépendant du précédent

Source S, avec un alphabet $\mathcal{A} = \{a_1, a_2 \dots\}$ émet des symboles $x_1, x_2 \dots$

La probabilité d'avoir en sortie de la source S $x_1, x_2, \dots x_n$ est donnée par $p(x_1, x_2, \dots, x_n) = \prod_i^n p(x_i)$

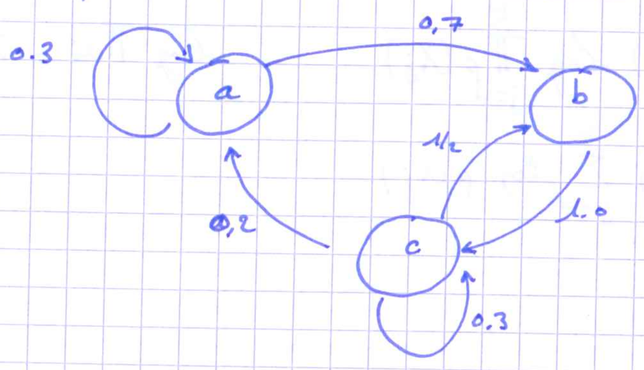
Source avec mémoire ou source de Markov.

Les symboles émis par la source ne sont pas indépendants.

Par exemple pour une source de Markov d'ordre 1 :

$$p(x_1, x_2, \dots, x_n) = p(x_1) . p(x_2 | x_1) . p(x_3 | x_2) \dots p(x_n | x_{n-1})$$

Exemple d'une telle source : source ternaire $\mathcal{A} = \{a, b, c\}$



On suppose être dans l'état a

$\iff x_k = a$

Donner les probas pour les symboles qui peuvent suivre :

$p(x_{k+1} = a \,|\, x_k = a) = 0,3$

$p(x_{k+1} = b \,|\, x_k = a) = 0,7$

$p(x_{k+1} = c \,|\, x_k = a) = 0$

Noter que si le symbole suivant ~~est~~ b $(x_{k+1} = b)$ alors $x_{k+2} = c$!

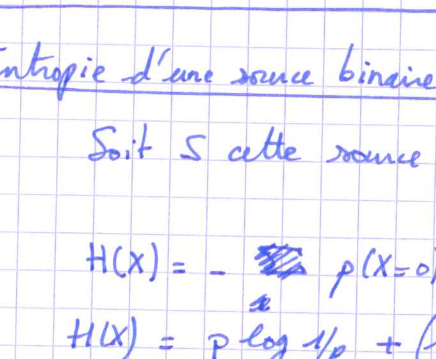
Entropie d'une source binaire sans mémoire :

Soit S cette source, définie sur $\mathcal{A} = \{0, 1\}$ avec $p(X=0) = p$

$p(X=1) = 1-p$

$H(X) = - p(X=0) . \log p(X=0) - p(X=1) \log p(X=1)$

$H(X) = p \log 1/p + (1-p) \log \frac{1}{1-p}$



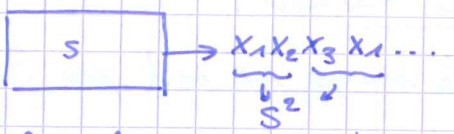
l'incertitude est maximale pour $p = 1/2$!

Entropie d'une source sans mémoire étendue à du codage par bloc.

Soit une source S avec un alphabet de taille N

La sortie de la source S (sans mémoire) est groupée par bloc de K symboles.

$\Rightarrow$ définition d'une nouvelle source S^k



Quelle est la valeur de l'entropie de cette nouvelle source ?

$$H(S^k) = -\sum_{i=1}^{N^k} p(X_1, X_2 \dots X_k) . \log p(X_1, X_2 \dots X_k)$$

Or, la source est sans mémoire :

$$H(S^k) = -\sum_{i=1}^{N^k} \prod_{j=1}^{k} p(X_j) \log \prod_{j=1}^{k} p(X_j) \qquad \log \prod a_i = \sum \log a_i$$

$$H(S^k) = -\sum_{i=1}^{N^k} \prod_{j=1}^{k} p(X_j) \sum_{j=1}^{k} \log p(X_j)$$

⋮

Soit la source S $\quad x_1, x_2 \quad N = 2$

Soit la source S^2 $\quad x_1x_1 \quad x_1x_2 \quad x_2x_1 \quad x_2x_2 \quad N = 4$

<u>Montrer que $H(S^2) = 2\, H(S)$</u>

$$H(S^2) = -\big(p(x_1, x_1) \log p(x_1, x_1) + p(x_2, x_2) \log p(x_2, x_2) + 2\, p(x_1, x_2) \log p(x_1, x_2) \big)$$

$$= -\big(2 p^2(x_1) \log p(x_1) + 2 p^2(x_2) \log p(x_2) + 2\, p(x_1) . p(x_2) (\log p(x_1) + \log p(x_2)) \big)$$

$$= -2 \underbrace{\big(p(x_1) + p(x_2) \big)}_{= 1} \big(p(x_1) \log p(x_1) + p(x_2) \log p(x_2) \big)$$

$$= 2\, H(S)$$

Binary Erasure Channel   CAPACITY

Inputs X vad 0, 1

Outputs Y vad 0, 1 and an "erasure" symbol written as "?"

For a given distribution $p_0 = p(X=0)$

+ Erasure probability f

The mutual information between channel input X and channel output Y can be written as:

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

$H(X) = ?$

$$H(X) = -\sum_{i=\{0,1\}} p(X=x_i) \log p(X=x_i)$$

$$H(X) = p_0 \log 1/p_0 + (1-p_0) \log \frac{1}{1-p_0}$$

$H(X|Y) = ?$

$$H(X|Y) = \sum_{i=0}^{2} p(Y=y_i)\, H(X|Y=y_i)$$

with $y_0 = 0$, $y_1 = 1$ and $y_2 =$ "?"

① $H(X|Y=y_0) = 0$

$H(X|Y=y_1) = 0$

② $$H(X|Y=y_2) = -\sum_{i=\{0,1\}} p(X=x_i|Y=y_2) \log p(X=x_i|Y=y_2)$$

$$p(X=x_0|Y=y_2) = \frac{p(X=x_0, Y=y_2)}{p(Y=y_2)} = \frac{p(Y=y_2|X=x_0)}{p(Y=y_2)}\, p(X=x_0)$$

$$= \frac{f}{\underbrace{f\cdot p_0 + f(1-p_0)}_{p(X=y_2) = p(X=x_0, Y=y_2) + p(X=x_1, Y=y_2)}} \cdot p_0$$

$$p(X=x_0|Y=y_2) = p_0$$

Idem $p(X=x_1|Y=y_2) = 1-p_0$

In conclusion $H(X|Y) = \left\{p_0 \log 1/p_0 + (1-p_0)\log\frac{1}{1-p_0}\right\} \times f$

$I(X;Y) = (1-f)\left\{p_0 \log 1/p_0 + (1-p_0)\log\frac{1}{1-p_0}\right\}$

The capacity of the channel is obtained by maximizing $I(X;Y)$

$\Rightarrow p_0 = 1/2 \Rightarrow C = 1-f$

$H(X^N) = N \cdot H(X)$, Extensions of a source

Original source alphabet, $\mathcal{A}_X$, has I symbols.
Its N^{th} extension, $\mathcal{A}_X^N$ will have I^N symbols

$$H(X^N) = \sum_{i_1=1}^{I} \cdots \sum_{i_N=1}^{I} p_{i_1} \cdots p_{i_N} \times \log\left(\frac{1}{p_{i_1} \cdots p_{i_N}}\right)$$

$$= \sum_{i_1=1}^{I} \cdots \sum_{i_N=1}^{I} p_{i_1} \cdots p_{i_N} \cdot \sum_{j=1}^{N} \log\left(\frac{1}{p_{i_j}}\right)$$

$$= \sum_{j=1}^{N} \sum_{i_1=1}^{I} \cdots \sum_{i_N=1}^{I} p_{i_1} \cdots p_{i_N} \cdot \log\left(\frac{1}{p_{i_j}}\right)$$

$$= \sum_{j=1}^{N} \sum_{i_j=1}^{I} \sum_{i_k \text{ for } k \neq j} p_{i_1} \cdots p_{i_N} \cdot \log\left(\frac{1}{p_{i_j}}\right)$$

$$= \sum_{j=1}^{N} \sum_{i_j=1}^{I} p_{i_j} \cdot \log\left(\frac{1}{p_{i_j}}\right) \times \sum_{i_k \text{ for } k \neq j} p_{i_1} \cdots p_{i_{j-1}} p_{i_{j+1}} \cdots p_{i_N}$$

$$H(X^N) = N \times H(X)$$