

TD N°2 : Information, symboles et entropie

CORRIGE

- 1- En admettant que le macro-état décrivant 1 mg d'eau à température ordinaire correspond à un certain nombre de micro-états quantiques également probables, évaluer ce nombre, sachant que l'entropie de l'eau est $70 \frac{J}{mol \cdot K}$. ($M_{eau} = 18 \frac{g}{mol}$, $k = 1.38 \cdot 10^{-23} \frac{J}{K}$). Discuter le résultat.

Solution :

Nous allons chercher l'entropie de 1 mg d'eau :

$$\boxed{n = \frac{m}{M}} \quad \frac{S}{mol} = 70 \frac{J}{K} \rightarrow \frac{S}{g} = \frac{70}{18} \frac{J}{K} = 4 \frac{J}{K}$$

Ainsi pour 1 mg d'eau $S = 4 \cdot 10^{-3} \frac{J}{K}$

L'entropie statistique est donnée par la formule de Boltzmann

$$S = k \ln(\Omega) \rightarrow \Omega = e^{\frac{S}{k}}$$

$$\Omega = e^{\frac{4 \cdot 10^{-3}}{1.38 \cdot 10^{-23}}} = \boxed{e^{3 \cdot 10^{20}}}$$

Ce résultat montre la répartition statistique des molécules élémentaires de la matière. Comment un grand nombre de configuration microscopiques peut amener au même état macroscopique.

- 2- Évaluer (en unités thermodynamiques) l'ordre de grandeur de la quantité d'information en moyenne contenue dans la Bibliothèque Nationale de France. On supposera qu'elle contient de l'ordre de $7 \cdot 10^6$ volumes de 500 pages en moyenne avec 1000 signes par page, soit 26 symboles (alphabet latin), sans compter les signes de ponctuation. ($k = 1.38 \cdot 10^{-23} \frac{J}{K}$)

Comparer le résultat précédent avec l'ordre de grandeur de l'information génétique par individu sachant que l'ADN humain contient $3 \cdot 10^9$ paires de bases, pour chaque paire de base il y a 4 choix possibles (A, C, T et G)

Solution :

Dans les volumes de la Bibliothèque Nationale de France il y a environ $7 \cdot 10^6 \times 500 \times 1000 = 3,5 \cdot 10^{12}$ signes qui sont les lettres de l'alphabet (soit 26 choix, sans compter les signes de ponctuation). Si on ne tient pas compte des redondances, on trouve :

$$S = \langle I \rangle = N_{\text{symboles}} k \ln(\Omega) = 3,5 \cdot 10^{12} (1,38 \cdot 10^{-23} \ln(26)) = 15,7 \cdot 10^{-11} = \boxed{1,6 \cdot 10^{-10} \frac{J}{K}}$$

Pour l'ADN :

$$S = \langle I \rangle = N_{\text{base}} k \ln(\Omega) = 3 \cdot 10^9 (1,38 \cdot 10^{-23} \ln(4)) = \boxed{5,7 \cdot 10^{-14} \frac{J}{K}}$$

$$\frac{\langle I \rangle_{\text{bibliotheque}}}{\langle I \rangle_{\text{ADN}}} = \frac{1,6 \cdot 10^{-10}}{5,7 \cdot 10^{-14}} = 2800$$

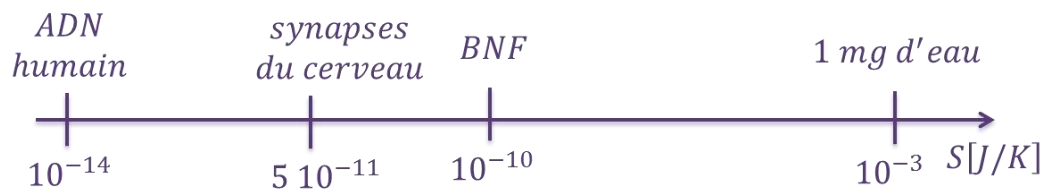
- 3- Le cerveau contient 10^{10} neurones, chacun établit des synapses avec 10^3 voisins. La liaison entre deux neurones est une synapse, qui a deux états possibles : elle est inhibitrice ou excitatrice. Quelle quantité d'information décrit l'état de toutes synapses du cerveau ?

Rassembler sur un axe, en échelle logarithmique, les quantités d'information obtenues dans les questions précédentes. Commenter. Que pensez-vous de la possibilité que l'ensemble du câblage du cerveau soit contenu dans le patrimoine génétique ? (Noter que ce câblage évolue avec l'éducation, l'âge. . .).

Solution :

Le nombre de micro états pour les synapses du cerveau est $10^3 10^{10} = 10^{13}$ et l'information correspondante

$$S = \langle I \rangle = N_{\text{synapses}} k \ln(\Omega) = 10^{13} (1.38 \cdot 10^{-23} \ln(2)) = \boxed{4.8 \cdot 10^{-11} \frac{\text{J}}{\text{K}}}$$



4- Pour l'information « 2013 » spécifier :

- a- L'ensemble des variables aléatoires (symboles)
- b- La probabilité d'occurrence de chaque symbole
- c- L'information *en bits* apportée par ce chiffre

Solution :

a- $S = \{0,1,2,3\}$

b- $S = \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}$

c- $I = \sum_{i=1}^N P_i \log_2 \left(\frac{1}{P_i}\right) = \log_2(4) = \boxed{2 \text{ bits}}$

5- Calculer l'entropie du mot **information**.

Données : $S = \{s_1 = a; s_2 = f; s_3 = i; s_4 = m; s_5 = n; s_6 = o; s_7 = r; s_8 = t\}$

$$P = \left\{p_1 = \frac{1}{11}; p_2 = \frac{1}{11}; p_3 = \frac{2}{11}; p_4 = \frac{1}{11}; p_5 = \frac{2}{11}; p_6 = \frac{2}{11}; p_7 = \frac{1}{11}; p_8 = \frac{1}{11}\right\}$$

Solution :

$$\begin{aligned} H &= \sum_{i=1}^n p_i \log\left(\frac{1}{p_i}\right) = \frac{1}{11} \log(11) + \frac{1}{11} \log(11) + \frac{2}{11} \log(5,5) + \frac{1}{11} \log(11) + \frac{2}{11} \log(5,5) \\ &+ \frac{2}{11} \log(5,5) + \frac{1}{11} \log(11) + \frac{1}{11} \log(11) \\ &= \frac{1}{11} (5 \log(11) + 6 \log(5,5)) = \log(11) - \frac{6}{11} \log(2) \quad \text{car : } \log\left(\frac{a}{b}\right) = \log(a) - \log(b) \\ &= 0,877 \end{aligned}$$

6- Quelle est l'entropie d'une source d'informations associée au message

Solution :

Pour cette source d'information nous avons

$$\left(S = \{0;1;a;b\}; P = \left\{ p_1 = \frac{1}{4}; p_2 = \frac{1}{4}; p_3 = \frac{1}{4}; p_4 = \frac{1}{4} \right\} \right).$$

Cette source d'information possède tous les symboles équiprobables donc :

$$S = \langle I \rangle = \sum_{i=1}^N P_i \log_2 \left(\frac{1}{P_i} \right) = \log_2(4) = \boxed{2 \text{ bits}}$$

7- Dans l'alphabet imaginaire des atlantes les symboles ont les probabilités suivantes :

Symbole						
P_i	0,083	0,041	0,5	0,125	0,083	0,16

Calculer l'entropie de cet alphabet et la comparer à l'alphabet phénicien ou tous ses symboles sont équiprobables



Alphabet phénicien de 14 symboles

Solution :

$$S = \sum_{i=1}^N P_i \log_2 \left(\frac{1}{P_i} \right)$$

$$S_{atlante} = -2 * 0.083 \log_2 0.083 - 0.041 \log_2 0.041 - 0.5 \log_2 0.5 - 0.125 \log_2 0.125 - 0.16 \log_2 0.16$$

$$S_{atlante} = 2.083$$

$$S_{phénicien} = - \sum_{i=1}^N P_i \log_2(P_i) = - \log_2 \left(\frac{1}{14} \right) = 3.80$$

8- On dispose pour écrire un message de n symboles différents

- Rappeler la définition d'un bit d'information
- On considère les symboles tous équiprobables. Donner l'expression littérale de l'information en bits apporté par chacun des symboles. Calculer sa valeur pour 4 symboles.
- Un problème plus réaliste suppose que chaque symbole n'a pas la même probabilité d'être utilisés à priori. P_i va être la probabilité du $i^{ème}$ symbole ($i = 1, 2 \dots 4$). Donner l'expression littérale de l'information moyenne H par symbole. Calculer sa valeur.

P_i des différents symboles:

- Symbole n°1 : 50 %
- Symbole n°2 : 20 %
- Symbole n°3 : 20 %
- Symbole n°4 : 10 %

Solution :

- a- Un *bit*, est la quantité d'information nécessaire pour connaître le symbole qui va apparaître entre deux symboles équiprobables. Par exemple $\{0,1\}$ pour une source binaire

- b- Expression littérale de I (information) :

$$I = - \sum_{i=1}^N P_i \log_2(P_i)$$

Or, $P_i = \frac{1}{N} = P$ si tous les états sont équiprobables :

$$I = \sum_{i=1}^N P_i \log_2(N) = \log_2(N) \left(\sum_{i=1}^N P_i \right)$$

$$I = \log_2(N) \rightarrow \text{si } N \text{ équiprobables}$$

Pour $N = 4$ nous avons :

$$H = \log_2(4) = \log_2(2^2) = 2 \log_2(2) = \boxed{2 \text{ bits}}$$

- c- Expression littérale de I (information) : moyenne pondérée par les probabilités

$$I = - \sum_{i=1}^N P_i \log_2(P_i)$$

Nous allons calculer I pour chaque probabilité :

$$I = -0.5 \log_2(0.5) - 2 \times 0.2 \log_2(0.2) - 0.1 \log_2(0.1)$$

$$I = -0.5 \log_2(2^{-1}) - 2 \times 0.2 \left(\frac{\ln(0.2)}{\ln(2)} \right) - 0.1 \left(\frac{\ln(0.1)}{\ln(2)} \right)$$

$$I = 0.5 - 2 \times 0.2(-2.32) - 0.1(-3.32)$$

$$I = 0.5 + 0.928 + 0.332 = \boxed{1.76 \text{ bits}}$$

- 9- On cherche à calculer l'information contenue dans une phrase écrite. On considère les lettres comme des symboles choisis pour construire une phrase. Un alphabet comporte 27 symboles (26 lettres + les espaces en blanc)

- a- En considérant les symboles équiprobables, calculer l'information contenue dans une phrase écrite de G lettres
- b- En réalité les différents symboles n'ont pas la même probabilité d'être utilisés à priori. P_i va être la probabilité du $i^{\text{ème}}$ symbole ($i = 1, 2 \dots 27$), écrire l'information moyenne H par symbole. Calculer H pour la langue anglaise.

P_i des différents lettres:

Blanc : 0.2 E : 0.105 T : 0.072 O : 0.065 A : 0.063 N : 0.059 I : 0.055

R : 0.054 S : 0.052 H : 0.047 D : 0.035 L : 0.029 C : 0.023 FU : 0.0225 M : 0.021

P : 0.0175 YW : 0.012 G : 0.011 B : 0.0105 V : 0.008 K : 0.003 X : 0.002 JQZ : 0.001

Solution :

- a- L'information contenue dans un message de G lettres est :

$$I_{tot} = G H = G \log_2(N)$$

Avec $N = 27$ lettres

$$I = \log_2(N) = \log_2(27) = \frac{\ln(27)}{\ln(2)} = \boxed{4.75 \text{ bits/lettre}}$$

- b- Nous allons calculer I pour chaque probabilité :

$$H = - \sum_{i=1}^N P_i \log_2(P_i)$$

$$\begin{aligned} I = & -0.2 \log_2(0.2) - 0.105 \log_2(0.105) - 0.072 \log_2(0.072) - 0.065 \log_2(0.065) - 0.063 \log_2(0.063) - \\ & 0.059 \log_2(0.059) - 0.055 \log_2(0.055) - 0.054 \log_2(0.054) - 0.052 \log_2(0.052) - 0.047 \log_2(0.047) - \\ & 0.035 \log_2(0.035) - 0.029 \log_2(0.029) - 0.023 \log_2(0.023) - 0.0225 \log_2(0.0225) - 0.021 \log_2(0.021) - \\ & 0.0175 \log_2(0.0175) - 2 * 0.012 \log_2(0.012) - \\ & 0.011 \log_2(0.011) - 0.0105 \log_2(0.0105) - 0.008 \log_2(0.008) - 0.003 \log_2(0.003) - 0.002 \log_2(0.002) - \\ & 3 * 0.001 \log_2(0.001) \end{aligned}$$

$$I = 0.464 + 0.341 + 0.273 + 0.256 + 0.251 + 0.241 + 0.23 + 0.227 + 0.221 + 0.207 + 0.169 + 0.148 + 0.125 + 0.123 + 0.117 + 0.102 + 0.153 + 0.0715 + 0.069 + 0.0557 + 0.0251 + 0.0179 + 0.0298$$

$$\boxed{I = 3.9 \text{ bits/lettre}}$$