

TD N°6 : Entropie d'un source, entropie conditionnelle et information mutuelle

- 1- Trois lampes colorées sont utilisées dans un système de signalisation.
- a- Combien de symboles pourraient être transmis par cette source sachant qu'une lampe est soit éteinte, soit allumée.
 - b- En déduire l'entropie permise par la taille de cet alphabet
 - c- En réalité, seulement trois symboles équiprobables sont utilisés. Calculer l'entropie réelle de la source et en déduire sa redondance.

Solution :

- a- Puisque ce système possède 3 lampes qui peuvent être soit éteintes soit allumées il y a $2^3 = 8$ combinaisons possibles. 8 symboles différents peuvent donc être émis.

- b- L'entropie de cette source est maximale lorsque ces symboles sont équiprobables :

$$S = \log_2(2^3) = \boxed{3 \text{ bits/symbole}}$$

- c- Si seulement 3 symboles sont utilisés, et ils sont équiprobables, la quantité moyenne d'information qu'ils peuvent apporter sera de :

$$S = \log_2(3) = \boxed{1.585 \text{ bits/symbole}}$$

Et ainsi la redondance pour cette source :

$$\mathcal{R} = 1 - \frac{S(3)}{S_{\max}} = 1 - \frac{1.585}{3} = \boxed{47.17 \%}$$

- 2- Calculer l'entropie d'une image de résolution 320×200 avec 4 niveaux de gris

- a- Dans le cas où les niveaux de gris sont supposés équiprobables
- b- Dans les cas où leurs probabilités sont $4/8$, $1/8$, $1/8$ et $2/8$. Quelle hypothèse il faut effectuer pour faire ces calculs ?

Solution :

- a- Si on suppose les pixels de cette image indépendants les uns des autres on a 64000 réalisations indépendantes de la même variable aléatoire, donc l'entropie de l'image sera :

$$S(\text{Image}) = S(P^N) = \log_2(P^{64000}) = \boxed{64000 \log_2(P)}$$

Donc, si les 4 niveaux de gris sont supposés équiprobables

$$S(P_{\text{niveau}}) = \log_2(4) = 2$$

Ainsi

$$S(\text{Image}) = \boxed{2 \times 64000 = 128000 = 1.28010^5 \text{ bits/symbole}}$$

- b- Dans ce cas-là on doit pondérer par la probabilité de chaque niveau de gris :

$$S(P) = \sum_{k=1}^4 p_i \log_2 \left(\frac{1}{p_i} \right)$$

$$S(P) = 0.5 \log_2 \left(\frac{1}{0.5} \right) + 0.125 \log_2 \left(\frac{1}{0.125} \right) + 0.125 \log_2 \left(\frac{1}{0.125} \right) + 0.25 \log_2 \left(\frac{1}{0.25} \right)$$

$$S(P) = 0.5 + 2 * 0.375 + 0.5 = 1.75 \text{ bits/symbole}$$

$$S(\text{Image}) = \boxed{1.75 \times 64000 = 112000 = 1.12010^5 \text{ bits/symbole}}$$

- 3- Cet exercice permet essentiellement d'apprécier l'ordre de grandeur de l'entropie de diverses sources qui correspondent à 3 possibilités différentes pour un être humain d'acquérir l'information.
- a- **Lecture** : en supposant que l'être humain peut lire en 1 seconde 10 mots appartenant à un vocabulaire de 1024 mots équiprobables. Calculer l'entropie d'une seconde de lecture.
 - b- **Conversation téléphonique** : une conversation téléphonique est échantillonnée à une fréquence de 8 kHz avec une quantification non uniforme des signaux appelée log MIC sur 256 niveaux. Calculer l'entropie acquise en écoutant cette conversation pendant 1 seconde
 - c- **Télévision** : Une image de télévision en noir et blanc possède une résolution verticale de 625 lignes et une résolution horizontale de 500 points, et est émise toutes les $1/24^{\text{ème}}$ de seconde. On suppose que l'œil peut distinguer 10 degrés de luminosité différents. Calculer l'entropie acquise en regardant une seconde de télévision.

Solution :

- a- On assimile cette activité à la réception de 10 fois par seconde d'un symbole choisi parmi 1024 possibles et indépendants : l'entropie reçue est alors de

$$S(\text{lecture}) = 10 \log_2(1024) = 10 \log_2(2^{10}) = 100 \log_2(2) = \boxed{100 \text{ bits/symbole}}$$
- b- Si on assimile à la réception à une fréquence de 8 kHz = 8000 symb/s choisis parmi 256 valeurs possibles et indépendants, l'entropie reçue est alors de

$$S(\text{téléphone}) = 8000 \log_2(256) = 8000 \log_2(2^8) = 8000 * 8 \log_2(2) = \boxed{64000 \text{ bits/symbole}}$$
- c- On assimile la réception correspondant à des images chacune constituée de $625 \times 500 \text{ pixels}$ prenant 10 valeurs possibles chaque $1/24^{\text{ème}}$ de seconde : l'entropie reçue est alors de

$$S(\text{vidéo}) = 24 \times 625 \times 500 \log_2(10) = 7.5 \cdot 10^6 \log_2(2^3) = 7.5 \cdot 10^6 * 3 \log_2(2) = \boxed{22.5 \cdot 10^6 \text{ bits/symbole}}$$

- 4- Une source binaire produit des symboles 0 et 1 avec des probabilités $2/5$ et $3/5$. La durée d'émission d'un 0 est de 1 ms et celle d'un 1 est de 2 ms. Calculer
- a- L'entropie de cette source
 - b- Le débit moyen
 - c- Le taux d'émission.

Solution :

- a- L'entropie de cette d'une source binaire vaut en générale :

$$S = \sum_{i=1}^2 p_i \log_2 \left(\frac{1}{p_i} \right) = p \log_2 \left(\frac{1}{p} \right) + (1-p) \log_2 \left(\frac{1}{1-p} \right)$$

$$S = \frac{2}{5} \log_2 \left(\frac{5}{2} \right) + \frac{3}{5} \log_2 \left(\frac{5}{3} \right) = 0.528 + 0.4421 = \boxed{0.9701 \frac{\text{bits}}{\text{symb}}}$$

- b- Le débit moyen de la source D_s se calcule :

$$D_s = \frac{1}{\sum_{i=1}^2 p_i \Delta t_i}$$

$$D_s = \frac{1}{\frac{2}{5} 10^{-3} + \frac{3}{5} 10^{-3}} = \boxed{625 \text{ symb/s}}$$

- c- Le taux d'émission est défini comme :

$$T = D_s S = 625 \frac{\text{symb}}{\text{s}} * 0.9701 \frac{\text{bits}}{\text{symb}} = \boxed{606 \text{ bits/s}}$$

- 5- Des symboles e_1, e_2, e_3 sont émis à l'entrée d'un canal bruité avec des probabilités 0.6, 0.3 et 0.1. Les symboles e_1 émis sont reçus pour 3/5 comme des symboles r_1 , pour 1/5 comme des symboles r_2 et 1/5 comme des symboles r_3 . Les symboles e_2 émis sont reçus pour 2/3 comme des symboles r_2 , et pour 1/3 comme des symboles r_3 . Le symbole e_3 est reçu intégralement comme r_3 .
- Calculer les probabilités pour chaque symbole r en sortie du canal.
 - Calculer l'information moyenne (entropie) apportée par les symboles à la sortie du canal
 - Calculer l'entropie (information) apportée par chaque symbole reçu (lorsque les symboles émis sont connus)
 - Calculer l'information transmise par le canal.

Solution :

- a- L'information transmise par le canal correspond à l'information mutuelle (information commune entre l'entrée et la sortie). Nous allons calculer les probabilités des symboles reçus avec le théorème de Bayes :

$$P(y) = P(y|x)P(x)$$

$$P(r_1) = P(r_1|e_1)P(e_1)$$

$$P(r_2) = P(r_2|e_1)P(e_1) + P(r_2|e_2)P(e_2)$$

$$P(r_3) = P(r_3|e_1)P(e_1) + P(r_3|e_2)P(e_2) + P(r_3|e_3)P(e_3)$$

$$P(r_1) = \frac{3}{5} \cdot 0.6 = 0.36$$

$$P(r_2) = \frac{1}{5} \cdot 0.6 + \frac{2}{3} \cdot 0.3 = 0.32$$

$$P(r_3) = \frac{1}{5} \cdot 0.6 + \frac{1}{3} \cdot 0.3 + 0.1 = 0.32$$

- b- Nous pouvons alors calculer l'entropie apportée par les symboles reçus :

$$S(r_1, r_2, r_3) = \sum_{i=1}^3 p_i \log_2 \left(\frac{1}{p_i} \right) = 0.36 * \log_2 \left(\frac{1}{0.36} \right) + 2 * 0.32 * \log_2 \left(\frac{1}{0.32} \right) = 0.53 + 1.05 = \boxed{1.58 \text{ bits/symb}}$$

- c- Entropie apportée par chaque symbole reçu (avec le symbole émis connu):

$$S(y_i|x_k) = \sum_{i=1}^3 p(y_i|x_k) \log_2 \left(\frac{1}{p(y_i|x_k)} \right) \text{ pour un } k \text{ fixe.}$$

$$S(r_1, r_2, r_3|e_1) = -\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - 2 * \frac{1}{5} \log_2 \left(\frac{1}{5} \right) = 1.37 \text{ bits/symb}$$

$$S(r_1, r_2, r_3|e_2) = -\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) = 0.918 \text{ bits/symb}$$

$$S(r_1, r_2, r_3|e_3) = -1 \log_2(1) = 0$$

Maintenant chaque entropie doit être pondérée avec la probabilité d'émission de chaque symbole e_i :

$$S(r_1, r_2, r_3|e_1, e_2, e_3) = P(e_1)S(r_1, r_2, r_3|e_1) + P(e_2)S(r_1, r_2, r_3|e_2) + P(e_3)S(r_1, r_2, r_3|e_3)$$

$$S(r_1, r_2, r_3|e_1, e_2, e_3) = 0.6 * 1.37 + 0.3 * 0.918 = \boxed{1.097 \text{ bits/symb}}$$

- d- L'information mutuelle représente l'information réellement transmise par le canal :

$$I(e_i, r_k) = S(r_1, r_2, r_3) - S(r_1, r_2, r_3|e_1, e_2, e_3)$$

$$I(e_i, r_k) = 1.58 - 1.097 = \boxed{0.48 \frac{\text{bits}}{\text{symb}}}$$

- 6- Dans un laboratoire se trouve une cage avec 100 souris présentant deux caractères : **sexe** (mâle ou femelle) et **couleur** (blanche ou noire) : 80 sont mâles, 60 sont blanches et 50 sont mâles et blanches
- a- Compléter le tableau suivant avec les effectifs et donner les vecteurs $P(S)$ et $P(C) \in \mathcal{R}^2$ avec les probabilités marginales de la variable sexe et la variable couleur

Effectif	Mâle (M)	Femelle (F)	Total
Blanche (B)	50	10	60
Noir (N)	30	10	40
Total	80	20	100
Probabilités marginales			
$P(S) = (0.8, 0.2)$			
$P(C) = (0.6, 0.4)$			

- b- Donner les valeurs de probabilité des événements conjoints : sont les intersections de chaque critère par exemple (blanche $\cap$ mâle = 0.5)

Evènements conjoints	Probabilités conjoints
$P(B, M)$	0.5
$P(B, F)$	0.1
$P(N, M)$	0.3
$P(N, F)$	0.1

- c- Rappeler le **Théorème de Bayes** pour un couple des variables aléatoires (X, Y) et donner la matrice (2×2) des probabilités conditionnelles $P(B|M)$ et $P(N|M)$ de la variable couleur pour le cas de la souris mâle.

Théorème de bayes :

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

$P(B M)$	$P(N M)$
$\frac{P(B, M)}{P(M)} = \frac{0.5}{0.8} = \frac{5/10}{8/10} = \frac{5}{8}$	$\frac{P(N, M)}{P(M)} = \frac{0.30}{0.80} = \frac{3/10}{8/10} = \frac{3}{8}$

- d- Une assistante doit prendre 6 souris au hasard. Calculer la probabilité d'extraire 6 souris blanches dans 6 extractions indépendantes (avec remise de la souris dans la cage)

La probabilité d'extraire une souris blanche est $P(B) = 0.6 = \frac{6}{10}$

Ainsi pour 6 extractions indépendantes :

$$P(6B) = 0.6^6 = \left(\frac{6}{10}\right)^6$$

- 7- Dans les jeux de rôle, on utilise des dés à plusieurs faces. Si les dés sont parfaitement équilibrés les faces sont toutes équiprobables. S'il y a des défauts, les faces auront des probabilités distinctes dans les tirages et pourront conditionner le résultat du tirage.
- a- Donner la notion d'information mutuelle de deux variables aléatoires

L'information mutuelle mesure la quantité d'information apportée par la réalisation de la variable X sur les probabilités de réalisation de la variable Y , elle mesure l'information partagée entre ce couple de variables.

- b- Ecrire l'expression mathématique de l'information mutuelle Les 3 expressions sont correctes :

$$I(x, y) = H(x) - H(x|y) = H(y) - H(y|x) = H(x) + H(y) - H(x, y)$$

Calculer l'information mutuelle dans le cas où le couple de variables aléatoires est indépendant.

$$I(x, y) = 0$$

- c- Quel sera le rapport entre les informations mutuelles d'un dé à 8 faces et un dé à 16 faces, si on suppose que les deux dés sont parfaitement équilibrés ? Quelle hypothèse faisons-nous sur l'entropie conditionnelle dans ce cas ?

Une fois observée la face X , la valeur de Y est déterminée avec certitude ainsi le terme d'entropie conditionnelle (réalisation de la face X sachant Y) devient $H(x|y) = 0$

$$I_8(x, y) = H(x) = \log_2(8) = 3$$

$$I_{16}(x, y) = H(x) = \log_2(16) = 4$$

$$\frac{I_8(x, y)}{I_{16}(x, y)} = \frac{3}{4} = 0.75$$

- 8- Dans un sachet « prêt à l'emploi » acheté dans la section jardinière d'un supermarché, on trouve 100 graines présentant deux caractères : **couleur** (rouge ou jaune) et **type** (bégonia ou dahlia) : 80 graines sont rouges, 60 sont bégonias et 50 sont bégonias rouges
- a- Compléter le tableau suivant avec les pourcentages pour chaque type de graine et donner les vecteurs $P(C)$ et $P(T) \in \mathcal{R}^2$ avec les probabilités marginales de la variable couleur et la variable type :

	Rouge (R)	Jaune (J)	Total
Bégonia (B)	50	10	60
Dahlia (D)	30	10	40
Total	80	20	100

Probabilités marginales
$P(C) = (0.8 R, 0.2 J)$
$P(T) = (0.6 B, 0.4 D)$

- b- Donner les valeurs de probabilité des événements conjoints, ou les deux caractères s'expriment à la fois :

On a les intersections de chaque critère par exemple (bégonia $\cap$ jaune = 0.1)

Evènements conjoints	Probabilités conjointes
$P(B, R)$	0.5
$P(B, J)$	0.1
$P(D, R)$	0.3
$P(D, J)$	0.1

- c- Rappeler le **Théorème de Bayes** pour un couple des variables aléatoires (X, Y)

Théorème de bayes :

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

ou toutes les autres combinaisons

- d- On prend 10 graines au hasard. Calculer la probabilité d'extraire 10 dahlias dans 4 extractions indépendantes (avec remise des graines dans le sachet). Utiliser l'élément des probabilités marginales qui réfère aux dahlias.

La probabilité d'extraire une graine de dahlia est $P(D) = 0.4$ (c'est un élément des probabilités marginales de la variable type)

Ainsi pour 4 extractions indépendantes (on multiplie les probabilités marginales de cette variable) :

$$P(4 D) = 0.4 * 0.4 * 0.4 * 0.4 = 0.4^4 = \left(\frac{4}{10}\right)^4 = 0.0256$$