TD n° 1 : Probabilités, entropie et information

1.
Montrer que la *quantité d'information*, exprimée en *shannons*, associée à une source S, dont les N signes de son alphabet sont équiprobables, peut se mettre sous la forme :

$$H(S)=\frac{\log_B N}{\log_B 2}[sh]$$

La *quantité d'information* est parfois exprimée en « *dits* », cette unité représentant l'information d'une expérience présentant N=10 issues équiprobables. Préciser la relation qui existe entre le « *dit* » et le « *shannon* ».

Calculer la *quantité d'information* associée à une expérience dont le résultat est certain. Choisissons, par exemple, la une source qui émet la phrase suivante : « Demain il fera jour ».

Analyser la fonction H(p) qui définit l'entropie d'une expérience à deux issues (dont les probabilités sont égales à p et 1−p). Les propriétés de continuité, valeur maximum, intersection avec l'axe devront être abordées. Représenter graphiquement cette fonction lorsque p varie de 0 à 1.

2.
Calculer la *quantité d'information* associée à une expérience ayant 8 résultats équiprobables. Montrer que l'on peut décomposer cette expérience en 3 expériences successives indépendantes ayant chacune 2 résultats équiprobables. Montrer que la quantité d'information totale est la somme des quantités d'information associées à chacune des 3 expériences partielles.

3.
Quelle est la *quantité d'information* associée à l'expérience qui consiste à déterminer si « une carte extraite d'un jeu de 52 cartes est un honneur » ou « n'est pas un honneur ». C'est à dire de calculer la quantité moyenne d'information contenue dans le message créé par l'expérience : « la carte tirée est un honneur ». Comparer cette valeur avec celle qui serait obtenue pour le message « la carte tirée est le roi du coeur ».

4.
Considérons un texte écrit en francais. Ce texte a été modifié de manière à ne plus présenter majuscules, accents et ponctuation. Calculer $H_{max}(S)$ − la valeur maximum de la *quantité d'information moyenne* par lettre. Quand est−ce que ce maximum est atteint?

Une étude probabiliste effectuée sur document d'environ 150 pages écrit en français conduit à une modélisation du texte par une source sans mémoire (cf. tableau ci−dessous). Calculer $H_0(S)$ − la *quantité d'information moyenne* par lettre. Si on disposait d'une modélisation du texte par une source avec une mémoire d'ordre 1, quelle serait la valeur de $H_1(S)$ par rapport à $H_0(S)$.

lettre	*%*	*Lettre*	*%*	*Lettre*	*%*	*Lettre*	*%*	*Lettre*	*%*	*Lettre*	*%*
A	6,4	B	0,64	C	2,59	D	2,6	E	14,86	G	0,83
H	0,61	I	5,91	J	0,23	K	0,01	L	4,65	M	2,45
N	6,23	O	4,59	P	2,56	Q	0,81	R	5,55	S	6,97
T	5,72	U	5,06	W	0	X	0,31	Y	0,21	Z	0,08
Espace	18,35										

5.
Supposons que nous avons une course de chevaux dont les probabilités de gagner pour chacun de 8 chevaux participants à la course sont : $(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{64},\frac{1}{64},\frac{1}{64},\frac{1}{64})$. Calculer *l'entropie* de cette course. Deduire la *logueur moyenne du message* contenant le résultat de cette course. Proposer une *méthode de codage du message* contenant le résultat de cette course .

6.
Une source émet selon un code décimal codé binaire (DCB). Déterminer la *quantité moyenne d'information* par symbole binaire. Calculer cette quantité par symbole décimale.

7.
Soit deux sources discrètes X et Y liées par la loi de couple $P(y_i x_i)$ donnée ci−dessous :

y_j \ x_i	x_1	x_2	x_3	P_j
y_1	0.1	0.2	0	0.3
y_2	0	0.3	0	0.3
y_3	0	0.2	0.2	0.4
P_i	0.1	0.7	0.2	

Déterminer :
- l'entropie H(XY) de la source composée (XY),
- les entropies H(X) et H(Y) des sources X et Y,
- les entropies conditionnelles moyennes $H(Y|X)$ et $H(X|Y)$,
- l'information mutuelle moyenne $I(X;Y)$ des sources X et Y. Monter que l'on peut calculer cette quantité de 3 manières différentes.