Prof. Kamal GHOUMID Année universitaire 2021 – 2022

4$^{\text{ème}}$ année "**Ingénierie Data Sciences & Cloud Computing**"

Cours "**Détection, Estimation & Information pour les Data Sciences**"

Série de TD $N^o.1$

Ex-1- : Digitalisation, Compression et Transmission de Data.

On cherche dans un premier temps à digitaliser un répertoire contenant 2500 photos au format bitmap en noir et blanc, de taille 256×256 blocs de 8 pixels chacun. Un pixel correspond à 256 niveaux de gris possibles. Ensuite à transmettre ces photos digitalisées depuis un ordinateur vers un téléphone mobile. Le débit de transmission sur le canal radio est $R = 10\,Mbit/s$.
Pour cela, on utilise, pour chaque photo :

- Un codage JPEG ayant un taux de compression $8,53$;
- Un codage canal de rendement $4/7$.

1. Quelles sont les quantités des données digitalisées en bit (taille du fichier 'data') avant le codage JPEG, puis après le codage canal ?
2. Calculer le temps nécessaire pour transmettre ces données.

Ex-2- : Digitalisation, Quantité des données.

Le disque compact (CD Compact Disc) est utilisé pour stocker des données sous forme numérique, il échantillonne chacune des deux voies stéréos à la fréquence $44,1\,kHz$ où chaque échantillon est codé sur 16 bits.

1. Calculer le débit binaire.
2. Évaluer la quantité de données (en bits) stokées sur un CD contenant une heure et demi de musique.
3. Pour un dictionnaire contenant 1850 pages, à raisons de 2 colonnes par page, 85 lignes par colonne, 9 mots par ligne, 6 lettres en moyenne par mot et 7 bits par lettre, déterminer le nombre de bits nécessaires pour coder un tel dictionnaire, puis évaluer le nombre d'ouvrages équivalent que l'on peut stocker sur un CD.

Ex-3- : Digitalisation, Débit, Bande passante.

Un signal analogique de largeur de bande $B = 5\,kHz$ est échantillonné à $1,2$ fois la fréquence de Nyquist, chaque échantillon étant quantifié sur 256 niveaux équiprobables. On suppose que les échantillons sont statistiquement indépendants.

1. Quel est le débit d'information de la source ?
2. Peut-on transmettre sans erreur les signaux de cette source sur un canal AWGN de $12\,kHz$ de bande passante et présentant un rapport signal sur bruit de $18\,dB$.
3. Calculer le rapport $\frac{S}{N}$ requis pour assurer une transmission sans erreur dans les conditions énoncées ci-haut.
4. Calculer la bande passante requise pour acheminer sans erreur les signaux de la source considérée sur un canal AWGN avec un rapport $\frac{S}{N} = 18\,dB$.

Ex-4- : Information, Entropie, Entropie maximale.

Soit X une source discrète sans mémoire contenant m symboles différents $\{x_1, x_2, ..., x_m\}$ de probabilités respectives $\{p_1, p_2, ..., p_m\}$.

1. Montrer que l'entropie est maximale dans le cas d'équiprobabilité des symboles.
2. Étudier le cas particulier de deux symboles x_1 et x_2 de probabilités p et $q = (1-p)$ respectivement, puis tracer la courbe de l'entropie $H(X)$ en fonction de p.

Ex-5- : Information, Codage entropique, Débit binaire, Débit d'information.

Soit une source discrète sans mémoire S qui délivre à un rythme de $T = 1\,ms$ des symboles différents selon la loi de probabilité $\{p_1 = 0,23, p_2 = 0,06, p_3 = 0,15, p_4 = 0,34, p_5 = 0.04, p_6 = 0,18\}$ sur l'alphabet à 6 symboles différents $\{S_1, S_2, S_3, S_4, S_5, S_6\}$.

1. Calculer l'entropie $H(S)$ de la source, puis en déduire la valeur D_i du débit d'information.
2. Coder les symboles $S_{i_{\{i=1,2,...,6\}}}$ en utilisant le codage entropique de Huffmann.
3. Calculer la longueur moyenne, l'efficacité et la redondance du code. Quelle est la valeur du débit binaire après ce codage ?
4. On choisit maintenant d'utiliser un code à longueur fixe. Quelles sont les valeurs de son efficacité et de son débit binaire ? Conclure.
5. Est-il possible de transmettre cette source via un canal de transmission ayant une capacité de $2,4\,kbit/s$?

Ex-6- : Codage de Huffmann, Codage de Shannnon-Fano, Redondance.

Une source S émet les symboles $\{A, B, C, D, E, F\}$ avec les probabilités respectives suivantes : $P(A) = 0,17$, $P(B) = 0,03$, $P(C) = 0,40$, $P(D) = 0,06$, $P(E) = 0,20$ et $P(F) = 0,14$.

1. Coder ces symboles en utilisant la méthode de Huffmann.
2. Coder ces symboles en utilisant la méthode de Shannnon-Fano.
3. Calculer les valeurs de la redondance et de la longueur moyenne dans chaque cas ?