Feuille d'exercices n$^{\text{o}}$7

Exercice 1
On considère une variable aléatoire discrète, à valeurs dans {M, P, R, U, Y, Z}, chaque symbole ayant la probabilité suivante :

M	P	R	U	Y	Z
1/16	1/16	1/16	1/4	1/2	1/16

1. Calculer l'entropie de cette distribution.

2. a) Calculer le code de Huffman associé.
b) Pour ce code, quel est le nombre moyen de bits par lettre ?

3. Pour ce code, montrer que, quelle que soit $\epsilon_1, ..., \epsilon_n$ une suite de 0 et de 1, il existe un mot dont le code commence par $\epsilon_1 ... \epsilon_n$.

4. a) Déterminer le code du mot YUM.
b) Si on change le premier bit de ce code, le code que l'on obtient correspond-il encore à un mot ? Si oui, lequel ?

Exercice 2 : quantification non-linéaire
Soit X une variable aléatoire réelle de densité $p(x)$.
Soit $(y_k)_{0\leq k\leq K}$ une subdivision non-uniforme de $]-\infty; +\infty[$, avec $y_0 = -\infty$ et $y_K = +\infty$. Soit Q le quantificateur associé, avec :

$$Q_{|[y_{k-1},y_k[} = \frac{y_{k-1}+y_k}{2} \stackrel{\text{def}}{=} a_k \qquad (\forall k = 2, ..., K-1)$$

Soit $(\tilde{y}_k)_{0\leq k\leq K}$ une subdivision uniforme : $\tilde{y}_0 = -\infty, \tilde{y}_K = +\infty$ et

$$\tilde{y}_2 - \tilde{y}_1 = \tilde{y}_3 - \tilde{y}_2 = ... = \tilde{y}_{K-1} - \tilde{y}_{K-2}$$

Soit $\tilde{Q}$ le quantificateur associé à cette subdivision, avec $\tilde{Q}_{|[\tilde{y}_{k-1},\tilde{y}_k[} = \tilde{a}_k$. On suppose que $\tilde{a}_k = \frac{\tilde{y}_{k-1}+\tilde{y}_k}{2}$ pour tout $k = 2, ..., K-1$.
On suppose que Q et $\tilde{Q}$ sont des quantificateurs haute résolution.

1. Montrer que l'on peut définir Q par une formule du type $Q = G^{-1} \circ \tilde{Q} \circ G$, avec G croissante et affine par morceaux sur chaque $[y_{k-1}; y_k]$. Calculer $G'(a_k)$ pour $2 \leq k \leq K-1$.

2. Calculer la distorsion $D = E(|X - Q(X)|^2)$, en fonction de $\tilde{y}_1, \tilde{y}_{K-1}, K$, des $G'(a_k)$ et $p(a_k)$.

3. Donner une expression intégrale approchée de D pour K grand.

4. a) Calculer la distorsion minimale, pour le meilleur choix de G possible.
[Indication : utiliser l'inégalité de Hölder.]
b) Comparer avec la distorsion de $\tilde{Q}$.

Exercice 3 : codage arithmétique
Soit X une variable aléatoire à valeurs dans $\{1, ..., K\}$. On note $p_k = P(X = k)$ et on définit :

$$a_s = \sum_{k=1}^{s} p_k \quad \forall s = 0, ..., K$$

On code un mot $x_1...x_n \in \{1, ..., K\}^n$ de la manière suivante :
– On définit des suites $(y_k^{\text{inf}})_{k=0,...,n}, (y_k^{\text{sup}})_{k=0,...,n}$ telles que $y_0^{\text{inf}} = 0, y_0^{\text{sup}} = 1$ et :

$$y_{k+1}^{\text{inf}} = y_k^{\text{inf}} + a_{x_k-1}(y_k^{\text{sup}} - y_k^{\text{inf}}) \qquad y_{k+1}^{\text{sup}} = y_k^{\text{inf}} + a_{x_k}(y_k^{\text{sup}} - y_k^{\text{inf}})$$

– On utilise comme code le réel (ou l'un des réels) de l'intervalle $[y_n^{\text{inf}}; y_n^{\text{sup}}[$ dont le développement binaire est fini et de longueur minimale ; le code est l'écriture en binaire de ce réel.

1. Décrire un processus de décodage (en supposant n connu).

2. On suppose qu'on encode des mots dont chaque lettre est une réalisation indépendante de X. On note l_n l'espérance de la longueur du code d'un mot de n lettres. Montrer que, pour tout $\epsilon > 0$:

$$\frac{1}{n} l_n \leq H(X) + \epsilon \quad \text{pour tout } n \text{ assez grand}$$

3. [Comparaison avec Huffman] Soit X une variable aléatoire binaire valant 0 avec probabilité $1 - \epsilon$ et 1 avec probabilité ϵ.
a) Calculer l'entropie de cette distribution.
b) Calculer le code de Huffman associé à cette distribution.
c) Pour ce code, quelle est la longueur moyenne du code d'un mot de n lettres ?

Exercice 4
Soit X une variable aléatoire prenant la valeur 0 (respectivement 1) avec probabilité p (respectivement $1 - p$). Soit T_ϵ^n l'ensemble typique associé pour n tirages indépendants de X.
Estimer le nombre de « 1 » contenus dans les suites de T_ϵ^n, sous l'hypothèse $p \neq 1/2$.

Exercice 5
On veut coder une variable aléatoire X à images gand $\mathbb{N}^*$, de loi $P(X = k) = 2^{-k}$.

1. Calculer l'entropie $H(X)$.

2. Trouver un code optimal possédant la propriété du préfixe (en s'inspirant du code de Huffman).

3. On coupe les mots de plus de N bits en n'envoyant que les N premiers bits. Donner les expressions du nombre moyen de bits et de l'erreur quadratique moyenne.