Durée 2h, Notes de cours et de TD autorisées. Calculatrice autorisée.

Exercice 1. On considère une source à trois éléments $\mathcal{S} = (S = \{a, b, c\}, P = \{p_a = 0.1, p_b = 0.2, p_c = 0.7\})$.

1. On considère le code $C = \{0, 0010, 0001100\}$ pour $\mathcal{S}$ où $a \mapsto 0, b \mapsto 0010$ et $c \mapsto 0001100$, donner sa longueur moyenne.
2. Le code $C = \{0, 0010, 0001100\}$ est-il préfixe ?
3. Décomposer la séquence 0001000000110000100 en une séquence de mots de $C = \{0, 0010, 0001100\}$.
4. Le code $C = \{0, 0010, 0001100\}$ est-il uniquement décodable ? Rappel : un code C pour une source S est uniquement décodable si toute séquence de bit est décomposable en *au plus* une séquence de mot de C.
5. Montrez que le code $C' = \{0, 0010, 000100\}$ n'est pas uniquement décodable : vous donnerez une séquence qui peut être décomposée de deux manières différentes.

Correction.

1. *On calcule la longueur moyenne (somme pondérée des mots du code) :*

$$L(C) = 1 \times 0.1 + 4 \times 0.2 + 7 \times 0.7 = 5.8$$

2. *Le code C n'est pas préfixe car le mot* 0 *apparaît au début du mot* 0010 *(par exemple).*
3. *On décompose* 0001000000110000100 *en* $0, 0010, 0, 0, 0001100, 0010, 0$.
4. *Le code C est uniquement décodable : les séquences de bits correspondant à une séquence de mot de C n'ont que des séquences de* 1 *de longueur 1 ou 2. Les* 1 *isolés (encadrés par deux* 0*) doivent correspondre au mot* 0010 *et les séquences* 11 *encadrés par des zéros doivent correspondre au mot* 0001100*. Le reste des bits ne peuvent être que des zéros et correspondent à une séquence du mot "*0*".*
5. *C' n'est pas uniquement décodable car la séquence* 0001000 *peut être décomposée en* $0, 0010, 0, 0$ *ou en* $000100, 0$.

Exercice 2. On considère une source à deux symboles $S = \{a, b\}$ avec pour probabilité $p_a = 0.1$ et $p_b = 0.9$.

1. Calculer l'entropie de la source et trouver un code optimal dont vous donnerez la longueur moyenne pour cette source.
2. On considère maintenant la source $S' = S \times S \times S$ formé de triplet symbole de S. Les probabilités sont donnés par $p_{\lambda\mu\rho} = p_\lambda p_\mu p_\rho$ pour $\lambda, \mu, \rho \in S$. Calculer les probabilités de chaque triplet et en déduire l'entropie de la source.
3. Appliquer l'algorithme de Huffman pour déterminer un code optimal pour S'. Vous déterminerez sa longueur moyenne ainsi que la longueur moyenne en bit pour encoder un symbole de S.

Correction.

1. *L'entropie de S est calculée ci-dessous :*

$$H(S) = 0.1 \times \log_2(0.1) + 0.9 \times \log_2(0.9) = 0.47(logons)$$

Pour cette source on peut pas faire mieux que choisir un mot de longueur 1, le mot $c_a = 1$ pour a et un mot de longueur 1, le mot $c_b = 0$ pour b. La longueur moyenne de ce code est

$$L(C) = 0.3 + 0.7 = 1$$

.

2. *Les probabilités sont*

$$\begin{array}{rcl} p(aaa) & = & 0.001 \\ p(aab) & = & 0.009 \\ p(aba) & = & 0.009 \\ p(abb) & = & 0.081 \\ p(baa) & = & 0.009 \\ p(bab) & = & 0.081 \\ p(bba) & = & 0.081 \\ p(bbb) & = & 0.7290 \end{array}$$

L'entropie est la suivante :

$$H(S') = 0.001 \cdot \log_2(0.001) + 3 \cdot 0.009 \cdot \log_2(0.009) + 3 \cdot 0.081 \cdot \log_2(0.081) + 0.7290 \cdot \log_2(0.7290)$$

qui vaut $H(S') = 1.416$.

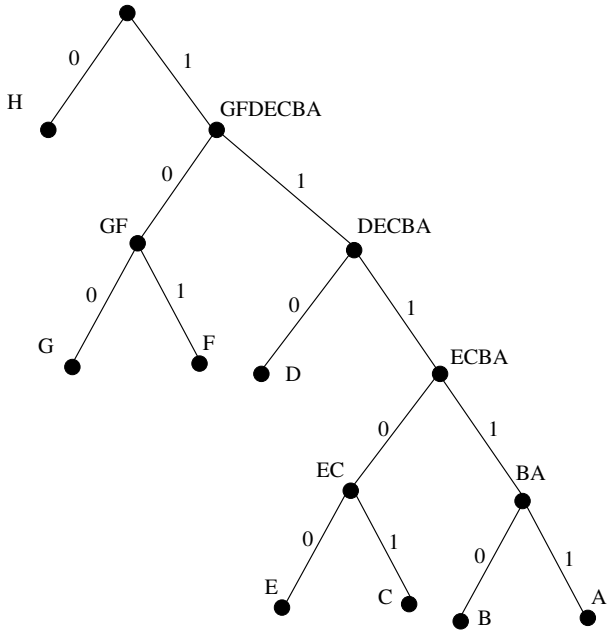
3. *On applique Huffman, mais avant de procéder, pour plus de lisibilité on renomme les 8 symboles de S'*

$$\begin{array}{rcl} A & = & aaa, \\ B & = & aab, \\ C & = & aba, \\ D & = & abb, \\ E & = & baa, \\ F & = & bab, \\ G & = & bba, \\ H & = & bbb, \end{array}$$

L'application de la phase de réduction de l'algorithme de Huffman produit :

$H, 0.7290$	$G, 0.081$	$F, 0.081$	$D, 0.081$	$E, 0.009$	$C, 0.009$	$B, 0.009$	$A, 0.001$
$H, 0.7290$	$G, 0.081$	$F, 0.081$	$D, 0.081$	$BA, 0.01$	$E, 0.009$	$C, 0.009$	
$H, 0.7290$	$G, 0.081$	$F, 0.081$	$D, 0.081$	$EC, 0.018$	$BA, 0.01$		
$H, 0.7290$	$G, 0.081$	$F, 0.081$	$D, 0.081$	$ECBA, 0.028$			
$H, 0.7290$	$DECBA, 0.109$	$G, 0.081$	$F, 0.081$				
$H, 0.7290$	$GF, 0.162$	$DECBA, 0.109$					
$H, 0.7290$	$GFDECBA, 0.271$						

On obtient l'arbre de Huffman ci-dessous :



qui fournit le codage

$$\begin{array}{lcl} A & \mapsto & 11111, \\ B & \mapsto & 11110, \\ C & \mapsto & 11101, \\ D & \mapsto & 110, \\ E & \mapsto & 11100, \\ F & \mapsto & 101, \\ G & \mapsto & 100, \\ H & \mapsto & 0, \end{array}$$

qui a la longueur moyenne :

$$\begin{array}{rcl} L(C') & = & 4 \times (0.001 + 0.009 + 0.009 + 0.009) + 3 \times (0.081 + 0.081 + 0.081) + 1 \times 0.7290 \\ & = & 1.57 \end{array}$$

ce qui donne une longueur moyenne de $1.57/3 = 0.523$ *bits par symbole de* S.

Exercice 3. Soit le code binaire qui à chaque (x_1, x_2, x_3) de $\mathbb{F}_2^3$ associe le vecteur $(x_1, x_2, x_3, x_4, x_5, x_6)$ défini par :

$$\left\{ \begin{array}{rcl} x_4 & = & x_1 + x_2 + x_3 \\ x_5 & = & x_1 + x_3 \\ x_6 & = & x_2 + x_3 \end{array} \right.$$

1. Donner une matrice génératrice G de ce code.
2. Encoder le mot source 011 à l'aide de G.
3. Donner la distance minimale de ce code. Quelle est la capacité de correction de ce dernier ?
4. Déterminer la matrice de contrôle H du code C.
5. À quel syndrome correspond l'erreur à la position j ?
6. Exprimer $H \cdot (100110)^t$ en fonction des colonnes h_i de H.
7. Décoder les messages reçus suivants avec la méthode de décodage par syndrome :

$a.$ 111100 $\qquad\qquad$ $b.$ 010011

Correction.

1. *La matrice génératrice G se déduit de*

$$\begin{aligned}(x_1, x_2, x_3, x_4, x_5, x_6) &= (x_1, x_2, x_3, x_1 + x_2 + x_3, x_1 + x_3, x_2 + x_3) \\ &= x_1 \cdot (1, 0, 0, 1, 1, 0) + x_2 \cdot (0, 1, 0, 1, 0, 1) + x_3 \cdot (0, 0, 1, 1, 1, 1)\end{aligned}$$

   Donc

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

2. *On encode $x = 011$ comme suit :*

$$[0, 1, 1] \cdot G = 0 \cdot (1, 0, 0, 1, 1, 0) + 1 \cdot (0, 1, 0, 1, 0, 1) + 1 \cdot (0, 0, 1, 1, 1, 1) = (0, 1, 1, 0, 1, 0)$$

3. *Les mots non nuls du code et leur poids sont*

mot	*poids*
100110	3
010101	3
001111	4
011010	3
110011	4
101001	3
111100	4

   Le poids minimzl du code est donc 3 et donc $d(C) = 3$. Sa capacité de correction est donc de $e = (d - 1)/2 = 1$ erreur.

4. *On décompose G en $[Id_3|P]$ donc*

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

   et donc (d'après le cours) la matrice de contrôle H de C est

$$H = [P^t|Id_3] = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

5. *L'erreur à la position j correspond au syndrome de $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ ou j est à la j-eme place. Or le syndrome $\sigma(e_j)$ est égal à la j-eme colonne de H.*

6. *$H \cdot (101100)^t$ est égal à $h_1 + h_3 + h_4$.*

7. *On calcule le syndrome des mots de poids 1*

vecteurs d'erreur	*syndrome*
100000	$\begin{matrix}1\\1\\0\end{matrix}$
010000	$\begin{matrix}1\\0\\1\end{matrix}$
001000	$\begin{matrix}1\\1\\1\end{matrix}$
000100	$\begin{matrix}1\\0\\0\end{matrix}$
000010	$\begin{matrix}0\\1\\1\end{matrix}$
000001	$\begin{matrix}0\\0\\1\end{matrix}$

On calcule le syndrome de 111100 qui vaut $H \cdot [111100]^t = 0$ *donc c'est un mot du code (cf. cours), on a pas à le décoder.*
On calcule le syndrome de 010011 qui vaut $H \cdot [010011]^t = [110]^t$*, le chef de classe associé est 001000 donc on décode en*

$$010011 + 100000 = 110011$$

Exercice 4. Soit C un code de distance minimale 5 et soit H une matrice de parité. Est-il possible d'avoir $H \cdot (1110 \cdots 0)^t = H \cdot (001110 \cdots 0)^t$?

Correction. *On réécrit* $H \cdot (1110 \cdots 0)^t = H \cdot (001110 \cdots 0)^t$ *en faisant tout passer à gauche*

$$H \cdot (1110 \cdots 0)^t + H \cdot (001110 \cdots 0)^t = 0$$

puis en factorisant H

$$H \cdot (1110 \cdots 0 + 001110 \cdots 0)^t = 0$$

qui donne

$$H \cdot (110110 \cdots 0)^t = 0$$

ce qui signifie que le mot $110110 \cdots 0$ *est dans le code* C*, or ce mot est de poids 4 et le poids minimum de* C *est 5, donc ce n'est pas possible.*