math202 : mathématiques pour le numérique
TD 2 : codes

Pierre Hyvernat
Laboratoire de mathématiques de l'université de Savoie
bâtiment Chablais, bureau 17, poste : 94 22
email : `Pierre.Hyvernat@univ-savoie.fr`
www : `http://www.lama.univ-savoie.fr/~hyvernat/`

Partie 1 : Codes de largeur fixe

Question 1. Donnez un code à taille fixe permettant de coder les trois couleurs primaires rouge, vert, bleu sur un nombre minimal de bits.

Question 2. Sans l'aide d'une calculatrice, estimez le nombre de bits nécessaire pour construire un code à taille fixe qui encode :
- les 24 heures de la journée,
- les 197 pays reconnus officiellement par l'ONU.
- les 67 millions de français.
- les 250 000 000 000 000 000 000 transistors produits au cours de l'année 2014.

Question 3. De manière générale, combien de bits sont nécessaire pour construire un code à taille fixe encodant n objets ?

Question 4. Utiliser une calculatrice pour calculer le nombre exact de bits nécessaire pour les encodages demandés dans la question précédente.

Partie 2 : codes ambigus et non-ambigus

Question 1. Montrez que les codes suivants sont ambigus en fournissant deux décodages pour la suite de bits fournie.
- code : $[a \mapsto \mathtt{0}, b \mapsto \mathtt{10}, c \mapsto \mathtt{11011}, d \mapsto \mathtt{01010}]$, suite : $\mathtt{110110101011011}$.
- code : $[a \mapsto \mathtt{1}, b \mapsto \mathtt{011}, c \mapsto \mathtt{01110}, d \mapsto \mathtt{1110}, e \mapsto \mathtt{10011}]$, suite : $\mathtt{011101110011}$.

Question 2. Quelle condition doit satisfaire un code pour être un code préfixe ?

Considérez le code :

$$C = \{\mathtt{0}, \mathtt{01}, \mathtt{011}, \mathtt{0111}, \mathtt{01111}, \mathtt{011111}, \mathtt{0111111}, \mathtt{01111111}, \mathtt{011111111}, \mathtt{0111111111}\}$$

Question 3. Le code C est-il un code préfixe ? (Justifiez...)

Question 4. Donnez tous les décodages de la suite de bits $\mathtt{0111010010110010111111}$ avec le code C

Question 5. Ce code est-il ambigu ? (Justifiez...)

Question 6. On définit un code suffixe comme étant un code dont aucun mot n'est suffixe d'un autre. Un code suffixe peut-il être ambigu ?

Question 7. Transformez le code $\{\mathtt{1}, \mathtt{010}, \mathtt{100}, \mathtt{000}\}$ en code préfixe de même taille. Qu'en pensez-vous ?

Question 8. Supposons qu'on ne dispose que du début d'un mot, peut-on décoder le (début du) mot avec les codes non-ambigus suivants :
- $\{\texttt{1}, \texttt{0}\}$,
- $\{\texttt{01}, \texttt{10}\}$,
- $\{\texttt{0110}, \texttt{0100}\}$,
- $\{\texttt{0}, \texttt{01}, \texttt{011}\}$,
- $\{\texttt{1}, \texttt{10}, \texttt{00}\}$.

Question 9. En reprenant les codes de la question précédente, que se passe-t'il si on ne dispose que de la fin d'un mot ?

Idem s'il manque le début et la fin d'un mot : peut-on encore décoder la partie du milieu ?

Partie 3 : Inégalité de Kraft

Question 1.
- Peut-on construire un code non-ambigu avec des mots de longueur 1, 2, 3, 3, 3 ?
- Peut-on construire un code non-ambigu avec des mots de longueur 1, 3, 3, 3, 3 ?
- Peut-on construire un code non-ambigu avec des mots de longueur 1, 3, 3, 3, 3, 3 ?

Quant cela est possible, construisez un tels code. Quand cela est impossible, justifiez.

Partie 4 : Code Huffman

Soit l_1 la longueur de l'encodage d'une suite de symboles par un code C_1 et l_2 la longueur de l'encodage de la même suite par un code C_2. On dit le code C_1 offre un gain en espace de $100 \times (1 - l_1/l_2)\%$ par rapport à C_2.

Question 1. Quel est le pourcentage de gain en espace si $l_1 = l_2$?

Question 2. Quel est le pourcentage de gain en espace si $l_1 = l_2/2$?

Question 3. Quel est le pourcentage de gain en espace si $l_1 = l_2/4$?

Question 4. Pour chacun des cas suivants, proposez un code à taille fixe de taille minimum. Donnez ensuite l'arbre obtenu par l'algorithme de Huffman ainsi que le code correspondant.

Finalement, calculez le gain en espace du code de Huffman par rapport au code à taille fixe.
- nombre d'occurrence des symboles : `A` : 140, `B` : 65, `C` : 30, `D` : 22, `E` : 12.
- nombre d'occurrence des symboles : `a` : 216, `e` : 541, `i` : 274, `o` : 233, `u` : 224, `y` : 10.
- nombre d'occurrence des symboles : "$\bullet$" : 126, "$\triangleright$" : 129, "$\diamond$" : 127, "$\ominus$" : 130.

Question 5. Que peut-on dire d'un code de Huffman pour 2^n éléments ayant tous des fréquences semblables ?