Exercices du cours de Théorie de l'Information et Codage
cours 3 du 23 février 2010.

1. **Codage de Huffman pour $D > 2$.** Le codage de Huffman dans le cas binaire, $D = 2$ a été vu en cours et nous en reprennons les notations. Pour $D > 2$, le Lemme 1 reste valide (vérifier!) mais le Lemme 2 ne s'applique plus. On définit un arbre de codage complet comme un arbre de codage pour lequel tous les noeuds intermédiaires ont D enfants. Ainsi le plus petit arbre de codage complet a D feuilles. Montrer le Lemme 3: le nombre de feuilles d'un arbre de codage complet est de la forme $D + (m-1)D$ pour $m \in \mathbb{N}^*$. Nous regardons maintenant un arbre de codage en le complétant, de telle sorte que B feuilles sont rajoutées et non utilisées par le code. Pour un code optimal, toutes les feuilles inutilisées doivent être au même niveau que le mot code le plus long et ne diffèrent que par le dernier digit. En déduire qu'un code optimal ne peut avoir qu'au plus $D-2$ feuilles inutilisées. En utilisant le Lemme 3, montrer que le nombre de feuilles inutilisées B est donc donné par $D - 2 - R(K-2|D-1)$ où $R(K-2|D-1)$ est le reste dans la division euclidienne de $K-2$ par $D-1$. Proposer un algorithme de codage qui généralise le codage vu en cours pour $D \geq 2$. Construire l'arbre de codage dans le cas suivant: $p(x_1) = 0.4$, $p(x_2) = 0.3$, $p(x_3) = 0.2$, $p(x_4) = 0.05$, $p(x_5) = 0.03$ et $p(x_6) = 0.02$.

2. Faire l'exercice 4 du TD sur le codage de Shannon.

3. Montrer le lemme du cours: $c(u_1^n) = O\left(\frac{n}{\log n}\right)$.