Exercices du cours de Théorie de l'Information et Codage
cours 2 du 24 février 2009.

1. Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires i.i.d. tirées selon une loi de probabilité $p(x)$. Trouver
$$\lim_{n\to\infty} [p(X_1, \ldots, X_n)]^{1/n}.$$
Dans le cas particulier où $p(2) = 1/2$, $p(3) = 1/3$ et $p(4) = 1/6$, trouver la limite du produit
$$(X_1 X_2 \ldots X_n)^{1/n}.$$

2. Soient (X_i, Y_i) une suite de v.a. i.i.d. tirées selon une loi de probabilité $p(x, y)$. On note $X^n = (X_1, \ldots X_n)$ et $Y^n = (Y_1, \ldots Y_n)$. Quelle est la limite de
$$\frac{1}{n} \log \frac{p(X^n)p(Y^n)}{p(X^n, Y^n)}?$$

3. Montrer que pour un ensemble dénombrable de mots code qui forment un code instantané, les longueurs des mots code satisfont l'inégalité de Krat étendue:
$$\sum_{i=1}^{\infty} D^{-\ell_i} \leq 1.$$
Inversement, étant donné une suite $\ell_1, \ell_2 \ldots$ vérifiant cette inégalité, montrer que l'on peut construire un code instantané avec ces longueurs de mots code. (Pour le i-ème mot code $y_1 y_2 \ldots y_{\ell_i}$, on pourra considére l'intervalle $[t(i), t(i) + D^{-\ell_i})$ où $t(i) = \sum_j y_j D^{-j}$).

4. Montrer que le Théorème de McMillan (Th 2.3 du cours) est valable quand l'alphabet de la source est infini.

5. **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 $p(x_6) = 0.02$.