Exercices du cours de Théorie de l'Information et Codage

cours 3 du 1er mars 2011.

1. Montrer le lemme énoncé sans démonstration en cours: $c(u_1^n) = O(n/\log n)$.

2. Montrer que pour toute v.a. X à valeurs entières de moyenne $\mathbb{E}[X]$, on a:

$$H(X) \leq (\mathbb{E}[X]+1)\log(\mathbb{E}[X]+1) - \mathbb{E}[X]\log\mathbb{E}[X],$$

avec égalité quand X suit une loi géométrique: $\mathbb{P}(X=k) = q^k(1-q)$ pour $k \geq 0$.

3. Soit $\{U_i\}_{i=1}^{\infty}$ une suite de v.a. i.i.d. à valeurs dans un ensemble fini $\mathcal{U}$ et d'entropie $H(U)$. On note $C(n)$ la v.a. égale au nombre maximal de mots distincts en lequel $U^{(n)}$ peut être découpé. C'est à dire avec les notations du cours: $C(n) = c(U_1^n)$. Nous allons montrer qu'avec probabilité 1, on a:

$$\limsup_{n\to\infty} \frac{C(n)\log C(n)}{n} \leq H(U).$$

Ce résultat permet de démontrer l'optimalité de l'algorithme de Lempel-Ziv dans le cas particulier d'une source sans mémoire.

a) Pour un découpage de u_1^n en c mots distincts, on note c_ℓ le nombre de mots de longueur ℓ. Montrer que

$$\log\mathbb{P}(u_1,u_2,\ldots,u_n) \leq -\sum_\ell c_\ell \log c_\ell.$$

b) On définit la v.a. Z par $\mathbb{P}(Z=\ell) = \frac{c_\ell}{c}$. En utilisant les deux exercices précédents, montrer que

$$\lim_{n\to\infty} \frac{c}{n} H(Z) = 0$$

c) Conclure.