Théorie de l'information - TD3

Exercices à rendre pour le 24/11 : 2 et 5.

Exercice 1 (Propriétés des processus stochastiques)**.** *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires à valeurs dans $\{1\dots 50\}$ indépendantes et de même loi uniforme. Parmi les processus suivants, lesquels sont sans mémoire ? Markoviens ? Stationnaires ?*

1. $Y_n = \sum_{i=0}^{n} X_i^2$.
2. $Z_n = \max_{0\leqslant i\leqslant n}(X_i + X_n)$.
3. $T_n = X_n - X_{n-1}$ *pour* $n \geqslant 1$.

Exercice 2 (Une chaîne de Markov)**.** *Une université compte trois bâtiments : A, B et C. Une personne est déposée au hasard dans l'un des trois bâtiments selon une loi de probabilité donnée. La personne se déplace ensuite aléatoirement sur le campus en suivant les règles suivantes.*

- *Quand il est dans le bâtiment A, il va dans le bâtiment B avec probabilité 1/4, dans le bâtiment C avec probabilité 1/2, ou dans une autre aile du bâtiment A avec probabilité 1/4.*
- *Quand il est dans le bâtiment B, il va dans le bâtiment A avec probabilité 1/2, ou dans le bâtiment C avec probabilité 1/2.*
- *Quand il est dans le bâtiment C, il va dans le bâtiment A avec probabilité 1/4, dans le bâtiment B avec probabilité 1/4, ou dans une autre aile du bâtiment C avec probabilité 1/2.*

1. *Modéliser la situation par une chaîne de Markov $X = (X_n)_{n\geqslant 0}$ dont on donnera le graphe de transition.*
2. *Donner la matrice de transition de cette chaîne.*
3. *Montrer que cette chaîne admet une unique loi invariante et la calculer.*
4. *Si X_0 suit cette loi invariante, calculer le taux d'entropie de X.*

Exercice 3 (Séquences typiques)**.** *On considère une suite de variables de Bernoulli indépendantes $X = (X_n)_{n\geqslant 0}$ de même loi valant 1 avec probabilité* $1/4$ *et* 0 *avec probabilité* $3/4$.

1. *Le processus X est-il sans mémoire ? Markovien ? Stationnaire ?*
2. *Donner l'entropie de X_0 et le taux d'entropie de X.*
3. *Donner une séquence* 0.5*-typique de longueur 2 de X.*
4. *Pour quelles valeurs de ε est-ce que* 0011 *est une séquence ε-typique de X ?*

Exercice 4 (Codage de Lempel-Ziv-Welch)**.** *Une source produit des messages sur l'alphabet $\{a, b\}$. On considère un message m produit par cette source.*

1. *Le début de m est ababababaabaaabbbb : le coder grâce à l'algorithme de Lempel-Ziv-Welch vu en cours.*
2. *La fin du message ainsi codé est* 4b6a7a9b *: en déduire la fin de m.*

Exercice 5 (Codage des sources vérifiant l'AEP)**.** *Soit $X = (X_n)_{n\geqslant 0}$ un processus stochastique discret à valeurs dans un ensemble S. Pour tous $n \in \mathbb{N}$ et $\varepsilon > 0$, on note :*

- *U_n un code régulier sur A^n de longueur fixe minimale ;*
- *$V_{n,\varepsilon}$ un code régulier sur $T_\varepsilon^{(n)}(X)$ de longueur fixe minimale ;*
- *$D_{n,\varepsilon} \colon A^n \to \{0,1\}^\star$ le code sur A^n défini par :*

$$D_{n,\varepsilon}(a_1 \dots a_n) = \begin{cases} 0\|U_n(x) & si \quad x \notin T_\varepsilon^{(n)}(X) \\ 1\|V_{n,\varepsilon}(x) & si \quad x \in T_\varepsilon^{(n)}(X). \end{cases}$$

1. *Soient $n \in \mathbb{N}$ et $\varepsilon > 0$. Quelles sont les longueurs moyennes de U_n et $V_{n,\varepsilon}$?*
2. *Soient $n \in \mathbb{N}$ et $\varepsilon > 0$.*
   (a) *Montrer que $D_{n,\varepsilon}$ est un code régulier sur A^n.*
   (b) *Exprimer sa longueur moyenne relativement à $X_{<n} \coloneqq (X_0, \dots, X_{n-1})$ en fonction de $p(T_\varepsilon^{(n)}(X))$ et $|T_\varepsilon^{(n)}(X)|$.*
3. *On suppose désormais que X vérifie l'AEP. Soit $\varepsilon < \bar{H}(X)$. Montrer qu'il existe $\delta > 0$ tel que pour n assez grand,*
$$\frac{\ell_{X_{<n}}(D_{n,\delta})}{n} \leqslant \bar{H}(X) + \varepsilon.$$
4. *On suppose toujours que X vérifie l'AEP. On définit pour tout $\varepsilon > 0$ un codage $C_\varepsilon \colon A^\star \to \{0,1\}^\star$ de la façon suivante. Soient $n \in \mathbb{N}$ et $x \in A^n$. Notons $l_n = \lfloor \log_2(n) \rfloor$. Écrivons $n = \sum_{i=0}^{l_n} \alpha_i 2^i$ avec $\alpha_i \in \{0,1\}$. Alors*
$$C_\varepsilon(x) = 0^{l_n}\|\alpha_{l_n} \cdots \alpha_0\|D_{n,\varepsilon}(x)$$
où 0^{l_n} est le mot de longueur l_n dont toutes les lettres sont 0.
   (a) *Soit $\varepsilon > 0$. Montrer que C_ε est uniquement décodable.*
   (b) *Soit $\varepsilon \in]0, \bar{H}(X)[$. Montrer qu'il existe $\delta > 0$ tel que pour n assez grand,*
$$\frac{1}{n} \sum_{x \in A^n} p(X_{<n} = x)\ell(C_\delta(x)) \leqslant \bar{H}(X) + \varepsilon.$$

Exercice 6 (Pire cas de LZW)**.** *On considère le codage de Lempel-Ziv-Welch C d'une source à valeurs dans l'alphabet $A = \{a, b\}$. Les mots codés sont de la forme $u||v$ où u est un mot binaire de longueur fixe (correspondant à l'écriture binaire du numéro du sous-mot codé) et $v \in A^\star$. On s'intéresse, pour chaque entier positif k, au mot $w_k = a|b|a^2|ab|ba|b^2|\dots|b^k$ contenant tous les mots de longueur $\leq k$ sur A : ces mots sont ceux produisant les mots de code les plus longs.*

1. *Quelle est la longueur de w_k ?*
2. *En combien de mots w_k est-il subdivisé pendant le codage ?*
3. *Quelle est la longueur du mot codé $C(w_k)$?*
4. *Montrer que*
$$\frac{\ell(C(w_k))}{\ell(w_k)} = 1 + \mathcal{O}_{k\to\infty}\left(\frac{1}{k}\right).$$
5. *Interpréter ce résultat.*