TD6 : Entropie et codage

1 Échauffement

Question 1 : On a une pièce biaisée de probabilité $0 < p < 1$ (on ne connaît pas p). Comment peut-on simuler une pièce parfaite en utilisant la piece biaisée ? Combien de lancés sont-ils nécessaires en moyenne ?

Question 2 : On a une pièce parfaite ($p = 1/2$).

1. Comment peut-on simuler un dé à $n \in \{3, 4, 5, \ldots\}$ faces ?
2. Quel est le temps moyen d'exécution ?
3. Peut-on réaliser cette simulation en temps borné ?

2 Entropie

Question 1 : Soit X une v.a. sur un domaine fini. Quelle est (en général) la relation d'inégalité entre $H(X)$ et $H(Y)$ si

a. $Y = 2^X$?

b. $Y = cos(X)$?

Question 2 : Quelle est la valeur minimale de $H(p_1, \ldots, p_n) = H(\vec{p})$ pour $\vec{p}$ dans l'ensemble des vecteurs de probabilité de dimension n ? Donner tous les $\vec{p}$ qui réalisent ce minimum.

Question 3 : On lance une pièce parfaite jusqu'à obtenir face. La v.a. X dénote le nombre de lancers effectués.

a. Trouver $H(X)$.

b. On tire au hasard une v.a. X selon cette distribution. Trouver une séquence efficace de requêtes oui/non de la forme « Est-ce que X est dans l'ensemble S ? » pour déterminer X. Comparer $H(X)$ avec l'espérance du nombre de requêtes nécessaires pour trouver X.

Question 4 : Dans le cadre des InterENS Culturelles 2017, une série de 7 matchs d'improvisation est organisée entre les équipes de Cachan et de Rennes. La première équipe à remporter 4 matchs est déclarée gagnante. Soit X la v.a. représentant les résultats possibles, par exemple $CCCC$, $RCRCRCR$ ou $CCCRRRR$. Soit Y le nombre de matchs joués ($4 \leq Y \leq 7$). On suppose les équipes équilibrées, et les matchs indépendants. Calculer $H(X)$, $H(Y)$, $H(Y|X)$ et $H(X|Y)$.

3 Codes uniquement déchiffrables

Question 1 : [Algorithme de Sardinas & Patterson]
Si $u, v \in \Sigma^*$, on définit $u - v = \{u' \in \Sigma^* \mid u = v \cdot u'\}$. Ainsi, $u - v$ est réduit à 0 ou un élément. (par exemple, $abc - a = \{bc\}$, $ab - ba = \emptyset$).
Étant donné $S \subseteq \Sigma^*$, on définit

$$T_0 = \bigcup_{\substack{u,v \in S \\ u \neq v}} u - v$$

et on définit T comme étant le plus petit ensemble qui contient T_0 et qui satisfait l'inégalité

$$\bigcup_{s \in S, v \in T} ((s - v) \cup (v - s)) \subseteq T$$

a. L'objectif de cette question est de montrer que S est uniquement déchiffrable ssi T ne contient pas le mot vide. Pour tout $i \geq 1$, on pose

$$T_i = T_{i-1} \cup \bigcup_{u \in S, v \in T_{i-1}} ((u - v) \cup (v - u))$$

Il est facile de voir que $T = \bigcup_i T_i$ et que les T_i ne contiennent que des suffixes de mots de S. En outre, puisque $T_i \subset T_{i+1}$ pour $i \geq 0$, il existe $n \geq 0$ tel que $T = T_n = T_{n+1}$.

Démontrer les deux lemmes suivants :

(a) **Lemma 1** *Pour tout $u \in T$ et pour tous $u_1, \ldots, u_k, v_1, \ldots, v_\ell \in S$ tels que $u_1 u_2 \ldots u_k = u v_1 v_2 \ldots v_l$, on a $\varepsilon \in T$.*

(b) **Lemma 2** *Pour tout i, s'il existe $u \in T_i$ et $u_1, \ldots, u_k, v_1, \ldots, v_\ell \in S$ tels que $u u_1 \ldots u_k = v_1 \ldots v_\ell$, alors il existe $v \in T_0$ et $u'_1, \ldots, u'_k, v'_1, \ldots, v'_\ell \in S$ tels que $v u'_1 \ldots u'_k = v'_1 \ldots v'_\ell$.*

Conclure.

b. Parmi les ensembles S suivants, lesquels sont uniquement déchiffrables ?

$$S_0 = \{0, 10, 11\} \qquad S_1 = \{0, 01, 11\} \qquad S_2 = \{0, 01, 10\}$$
$$S_3 = \{0, 01\} \qquad S_4 = \{00, 01, 10, 11\} \qquad S_5 = \{110, 11, 10\}$$
$$S_6 = \{110, 11, 100, 00, 10\}$$

c. Donner un algorithme (polynomial) pour décider si S est un code uniquement déchiffrable. Estimer la complexité de l'algorithme.