Corrigé du TD04 – D'après un article de M. Minsky et S. Papert (1965)

Soit $A \subseteq \mathbb{N}^*$. On s'intéresse au langage représentant A en base 2 et on cherche un critère pour dire que ce langage n'est pas rationnel. On dira que A est rationnel si le langage associé l'est. Les automates finis considérés lisent d'abord les bits de poids fort et refusent les mots commençant par le chiffre zéro. Le critère recherché fait appel à une notion simple de densité. On note $\pi_A(n)$ le cardinal de $[\![1, n]\!] \cap A$.

1. Soit $L \subseteq \{0,1\}^*$ un langage rationnel. Montrer qu'il existe un entier N tel que pour tout mot x vérifiant $x^{-1}L \neq \emptyset$, il existe un mot y tel que

$$|y| < N \text{ et } xy \in L$$

R. Puisque L est rationnel il existe un automate fini déterministe $\mathcal{A}$ qui le reconnaît, soit N le nombre d'états de cet automate.
Étant donné un mot x tel que $x^{-1}L \neq \emptyset$, en notant q_x l'état de $\mathcal{A}$ que l'on atteint en lisant x, il existe un chemin de q_x vers un état final (puisque le résiduel de x est non vide), il existe un tel chemin qui ne fait pas de boucle et est donc de longueur au plus N ce qui nous donne le mot y recherché.

2. Soit $A \subseteq \mathbb{N}^*$ un ensemble rationnel infini. Montrer qu'il existe une constante $K > 0$ telle que

$$\forall n, \pi_A(n) > K \log(n) - 1$$

R. Pour tout entier k, il existe un mot x de longueur k qui est préfixe d'un mot de A (car A est infini), et donc d'après la question précédente il existe un mot y de longueur inférieure à N tel que $xy \in A$.
Finalement on a montré que pour tout entier k il existait un mot de A de longueur entre k et $(k+N)$. Ainsi, il existe un mot de A de longueur entre kN et $(k+1)N$ pour tout k.
L'ensemble des entiers inférieurs à $(2^k - 1)$ est exactement l'ensemble des mots de longueur au plus k. Ainsi on a montré que pour tout k,

$$\pi_A(2^k - 1) \geq k/N$$

On conclut en encadrant n entre deux puissances de deux et en prenant $K = 1/N$.

3. Soit $A \subseteq \mathbb{N}^*$. Soit α un entier qui n'est préfixe d'aucun élément de A. Montrer que $\forall m \in \mathbb{N}$

$$A \cap [\![2^m\alpha, 2^m(\alpha+1) - 1]\!] = \emptyset$$

R. les entiers de $[\![2^m\alpha, 2^m(\alpha+1) - 1]\!]$ sont exactement les nombres dont l'écriture en base 2 est le mot α suivi d'un mot de m lettres (le premier s'écrit $\alpha.00\ldots0$ et le dernier $\alpha.11\ldots1$).
Par hypothèse, aucun de ces nombres n'appartient à A.

4. Soit $A \subseteq \mathbb{N}^*$, $\alpha \in \mathbb{N}$ tel que $\alpha^{-1}A = \emptyset$. On suppose que la suite

$$\left(\frac{\pi_A(n)}{\pi_A(\frac{\alpha+1}{\alpha}n)} \right)_{n \in \mathbb{N}}$$

converge. Montrer que sa limite est 1.

R. Puisque l'on suppose que la suite converge, il nous suffit de montrer que l'une de ses sous-suites extraites tend vers 1.
La question précédente nous indique que $\pi_A(2^n\alpha)$ et $\pi_A(2^n(\alpha+1))$ diffèrent d'au plus 1 pour tout n puisqu'aucun entier entre $2^n\alpha$ et $(2^n(\alpha+1)-1)$ n'appartient à A.
Ainsi, en prenant $u_n = 2^n\alpha$ on a

$$\frac{\pi_A(u_n)}{\pi_A(\frac{\alpha+1}{\alpha}u_n)} \to 1$$

5. Soit $A \subseteq \mathbb{N}^*$ un ensemble infini dont un des résiduels est vide. Soit a_r le r-ième élément de A selon l'ordre usuel. Montrer que la suite

$$\left(\frac{a_{r+1}-a_r}{a_r}\right)_{r\in\mathbb{N}^*}$$

ne converge pas vers 0.

R. Soit α un entier dont le résiduel selon A est vide. Pour tout n on définit r_n comme étant l'indice du plus grand élément de A plus petit que $2^n\alpha$.
On a alors (par définition de r_n et d'après la question 3)

$$\begin{array}{rcl} a_{r_n} & \leq & 2^n \\ a_{r_n+1} & \geq & 2^n(\alpha+1) \end{array}$$

et donc $a_{r_n+1} - a_r \geq 2^m$ et

$$\frac{a_{r_n+1}-a_{r_n}}{a_{r_n}} \geq \frac{1}{\alpha}$$

ce qui prouve que la suite ne tend pas vers 0.

6. Soit $A \subseteq \mathbb{N}^*$ un ensemble rationnel dont aucun résiduel n'est vide, montrer qu'il existe $N \in \mathbb{N}$ tel que pour tout n à partir d'un certain rang,

$$\frac{\pi_A(n)}{n} \geq 2^{-N}$$

R. D'après la question 1, pour tout entier n, tout mot x de longueur n est préfixe d'un mot x' de A de longueur au plus $(n+N)$.
Ainsi, il y a au moins autant de mots dans A de longueur $(n+N)$ que de mots dans $\{0,1\}^*$ de longueur n (deux mots distincts x et y de même longueur ne peuvent pas être préfixes d'un même mot) d'où

$$\begin{array}{rcl} \pi_A(2^{n+N}-1) & = & \text{nbre de mots de } A \text{ de longueur inférieure à } (n+N) \\ & \geq & \text{nbre de mots de longueur } n \\ & \geq & 2^n \end{array}$$

On déduit donc

$$\frac{\pi_A(2^{n+N}-1)}{2^{n+N}} \geq \frac{2^n}{2^{n+N}} = 2^{-N}$$

Et l'on conclut pour tout $x \in \mathbb{N}$ en l'encadrant entre deux puissances de 2 (on obtiendra alors le résultat annoncé pour $(N+1)$ au lieu de N).

Remarque : ce résultat n'est vrai qu'à partir d'un certain rang, mais c'est le comportement à l'infini qui nous intéressera par la suite.

7. Déduire des questions précédentes le critère suivant.
Un ensemble $A \subseteq \mathbb{N}^*$ infini n'est pas rationnel si il vérifie la condition 1 et l'une des conditions 2 et $2'$ suivantes :

1 - $\pi_A(n)/n \to 0$

2 - $\pi_A(n)/\pi_A(\lambda n)$ converge pour tout $\lambda > 0$ et la limite est différente de 1 pour $\lambda \neq 1$

$2'$ - $(a_{n+1} - a_n)/a_n \to 0$

R. Soit A un langage rationnel infini. On va montrer que A ne vérifie pas la condition (1) ou que A ne vérifie ni la condition (2) ni la condition (2') ce qui prouvera bien qu'un ensemble infini qui vérifie (1) et l'une parmi (2) et (2') n'est pas rationnel.
Il y a deux possibilités :
– Si A a n'a aucun résiduel vide on sait par la question 6 que $\pi_A(n)/n$ ne tend pas vers 0 et donc que A ne vérifie pas (1).
– Si A a un résiduel vide (celui d'un mot α) alors la suite

$$\frac{\pi_A(n)}{\pi_A(\frac{\alpha+1}{\alpha}n)}$$

ne converge pas vers 1 d'après la question 4 et la suite $(a_{n+1} - a_n)/a_n$ ne tend pas vers 0 d'après la question 5. Ainsi A ne vérifie ni (2) ni (2').

8. En utilisant le critère de la question 7, montrer que les ensembles suivants ne sont pas rationnels :

– $A_k = \{m^k | m \in \mathbb{N}^*\}$ $(k \geq 2)$;
– $\mathcal{P}$ l'ensemble des nombres premiers.
Indication : on pourra utiliser l'équivalence bien connue $\pi_{\mathcal{P}}(n) \sim n/log(n)$.

R. Il est immédiat que la densité de ces deux ensembles est nulle donc ils vérifient la condition (1). Pour les nombres premiers l'équivalent de $\pi_{\mathcal{P}}(n)$ montre rapidement que $\mathcal{P}$ vérifie la condition (2) tandis que les puissances n-ièmes vérifient la condition (2') puisque la différence $((n+1)^k - n^k)$ est de l'ordre de n^{k-1}.