

Séries Rationnelles

Dans la suite L est un langage rationnel et $\mathcal{A}$ un automate le reconnaissant.

1. Décrivez en OCaml une structure de données pour représenter $\mathcal{A}$.
2. Donnez une fonction calculant à partir de $\mathcal{A}$ le nombre $|L \cap \Sigma^n|$.
3. En utilisant le fait que : $\sum A^n x^n = (\text{Id} - xA)^{-1}$, en déduire que la fonction définie par $S_L(x) = \sum_n |L \cap \Sigma^n| x^n$ est rationnelle. On pourra utiliser, en le justifiant, le fait que $U(\sum x_i) = \sum Ux_i$ pour U une application linéaire.
4. En déduire une fonction en OCaml qui calcule cette série. On pourra utiliser des fonctions sans les détailler lorsque leur sémantique est évidente.

Génération Aléatoire de Mots

On veut démontrer le théorème suivant :

Théorème 0.1 — Loi de Zipf-Mandelbrot La probabilité d'un mot dans une langue est équivalente à $\frac{C}{(r(X)+B)^\alpha}$ pour certaines constantes α, B, C .

Ici $r(X)$ désigne le rang du mot X , rangés par fréquence décroissante.

On considère donc un alphabet fini Σ de taille $S + 1$. On suppose que Σ contient un certain symbole **EOW** qui marque la fin d'un mot. On va tirer uniformément au hasard des lettres dans Σ pour former des mots, ces mots étant séparés par **EOW**. Soit X une variable aléatoire représentant la distribution des mots dans un texte aléatoire.

1. Donner $\mathbb{P}(X = w)$ pour $w \in \Sigma^*$. En déduire la probabilité que w soit de longueur n .
2. Calculer la fonction $n \mapsto r(n)$ qui a un mot de longueur n associe ses rangs possibles.
3. En déduire le théorème.

On change de paradigme, on considère un texte généré aléatoirement dont on ne connaît pas la distribution.

4. Donner une fonction qui permette de mesurer l'information portée par un mot de probabilité p dans un texte. On vérifiera que la fonction I est décroissante, vaut 0 en 1 et vérifie $I(p_1 p_2) = I(p_1) + I(p_2)$.
5. Donnez une mesure de la longueur moyenne d'un mot pour un code optimal associé à la distribution mesurée.

Solution: Génération Aléatoire de Mots

Consider a finite alphabet Σ of size $S + 1$. We will suppose that Σ contains a symbol **EOW** that marks the end of a word. Suppose we take letters uniformly at random in Σ to make words, and that we say that words are separated by **EOW**. Let X be a random variable representing the distribution of words in a random text : Then the probability of a given word w in a text is stated by (as it needs to start with **EOW** and end with **EOW**) :

$$\mathbb{P}(X = w) = \frac{1}{S(S+1)^{|w|}}$$

and thus the probability of a word being of length n is :

$$\mathbb{P}(|X| = n) = \sum_{w \in \Sigma^*, |w|=n} \mathbb{P}(X = w) = \frac{S^{n-1}}{(S+1)^n}$$

If we represent the rank of a word, we see that since its probability only depends on its length, its rank cannot be fully determined, just as a function $r(n)$ of its length :

$$\sum_{i=1}^{n-1} S^i < r(n) \leq \sum_{i=1}^n S^i$$

that is :

$$\frac{S}{S-1} (S^{n-1} - 1) < r(n) \leq \frac{M}{M-1} (M^n - 1)$$

Taking the log :

$$n - 1 < \log_S \left(\frac{S-1}{S} r(n) + 1 \right) \leq n$$

Raising each $1/S + 1$ to the power of this :

$$\frac{1}{(S+1)^{n-1}} > \left(\frac{1}{S+1} \right)^{\log_S \left(\frac{S-1}{S} r(n) + 1 \right)} \geq \frac{1}{(S+1)^n}$$

or :

$$\mathbb{P}(|X| = n) < \frac{C}{(r(n) + B)^\alpha} \leq \mathbb{P}(|X| = n - 1)$$

where :

$$\alpha = \frac{\log(S+1)}{\log S}, B = \frac{S}{S-1} \text{ and } C = \frac{M^{\alpha-1}}{(M-1)^\alpha}$$

Question de Cours

Soit L un langage. Montrer que les propositions suivantes sont équivalentes :

1. $\varepsilon \in L$
2. $\forall i, \varepsilon \in L^i$
3. $\forall i, L^i \subseteq L^{i+1}$

Knaster-Tarski

Soit E un ensemble. Une fonction $f : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ est croissante si pour toute partie X, Y de E , $X \subseteq Y \Rightarrow f(X) \subseteq f(Y)$.

1. Montrer que l'identité est croissante.
2. Montrer que toute fonctorisation f de $\bar{f} : E \rightarrow E$ définie par $f(Y) = \{\bar{f}(x) \mid x \in Y\}$ est croissante.
3. Donner une fonction non croissante.
4. Montrez le théorème de Knaster-Tarski. Si f est croissante, il existe $S \subseteq E$ tel que :
 - (a) $f(X) \subseteq X \Rightarrow S \subseteq X$
 - (b) $S = f(S)$.
5. Montrer que si A, B sont deux fonctions croissantes sur Σ^* , alors les équations ci-dessous ont la même plus petite solution.

$$X = A(X).X + B(X) \text{ et } X = A(X)^*.B(X)$$

Solution: Knaster-Tarski

<https://chamilo.grenoble-inp.fr/courses/ENSIMAG3MRTL1/document/TD/recueil-correction.pdf>