INFOF-303 : Réseaux, information et communication

R. Petit

année académique 2016 - 2017

Table des matières

1 Introduction et inégalités de Kraft/Mc Millan

1.1 Définitions

Définition 1.1. Soit Σ un ensemble de cardinalité finie. Si on appelle les éléments $\sigma \in \Sigma$ des *symboles*, on dit que Σ est un *alphabet*.
Définition 1.2. Soit Σ un alphabet. On pose $\ell \in \mathbb{N}$, un naturel. Toute séquence de ℓ symboles de Σ concaténés est appelée *mot* de l'alphabet Σ. Si m est un mot, on peut écrire $m \in \Sigma^\ell$.
Définition 1.3. Soit Σ un alphabet. On définit l'ensemble :

$$\Sigma^* \coloneqq \bigcup_{\ell \in \mathbb{N}} \Sigma^\ell.$$

Remarque. L'ensemble Σ^* contient donc tous les mots de cardinalité naturelle qui peuvent être faits à l'aide de l'alphabet Σ. On peut également noter que l'ensemble Σ^* est toujours de cardinalité infinie alors que, par définition, l'alphabet Σ est de cardinalité finie.
Définition 1.4. Soit Σ un alphabet et $m \in \Sigma^*$ un mot sur Σ. On définit la fonction :

$$\ell_\Sigma : \Sigma^* \to \mathbb{N} : m \mapsto n \in \mathbb{N} \text{ t.q. } m \in \Sigma^n \subsetneq \Sigma^*.$$

On appelle cette fonction la fonction *longueur* des mots sur Σ.
Remarque. Lorsque l'alphabet du mot n'est pas ambigu, on note simplement cette fonction ℓ.
Définition 1.5. Soient deux alphabets S et C. Soit $K : S \to C^* : s \mapsto (c_i)_{i \leqslant n}$. On dit que la fonction K est une fonction de codage si K est injective. Dans ce contexte, le terme *univoque* est préféré à *injectif*.
Remarque.

- Par une fonction de codage, chaque symbole de l'alphabet de départ S est codé par une *suite* de symboles de l'alphabet d'arrivée C ;
- l'ensemble $K(S) \neq C^*$ car K est injective et donc $\left|K(S)\right| = |S|$. Or $|S| \in \mathbb{N}$ et $|C^*| = +\infty$. La fonction K ne peut donc pas être bijective ;
- il est usuel de noter les cardinaux des ensembles S et C respectivement par q et r.

Définition 1.6. On étend la fonction de codage $K : S \to C^*$ codant un unique symbole de S en la fonction :

$$K : S^k \to C^* : (s_i)_{1 \leqslant i \leqslant k} \mapsto \left(K(s_i)\right)_{1 \leqslant i \leqslant k} = \left(\left(c_{ij}\right)_{1 \leqslant j \leqslant n}\right)_{1 \leqslant i \leqslant k} = \left(c_{ij}\right)_{\substack{1 \leqslant j \leqslant n \\ 1 \leqslant i \leqslant k}}.$$

Remarque. En théorie de l'information, la fonction K de codage est totalement déterministe afin de permettre le décodage.

1.2 Familles de codes

1.2.1 Codes blocs

Définition 1.7. Soient S et C deux alphabets et $K : S \to C^*$ une fonction de codage. S'il existe $n \in \mathbb{N}$ tel que $K(S) \subseteq C^n$, on dit que K est une fonction de *code bloc*.
Remarque. Une fonction de code bloc code donc tous les symboles de S par une suite d'un nombre fixé de symboles de C.
Remarque. La majorité des codes correcteurs d'erreurs (CCE) sont des codes blocs. En effet, si un canal de transmission est bruyant, il faut connaitre au préalable la longueur des blocs à lire afin de les décoder et de les corriger.

1.2.2 Codes préfixes

Définition 1.8. Soient $S = \{s_1, s_2, \ldots, s_n\}$ et C deux alphabets et $K : S \to C^*$ une fonction de codage. On dit que K est une fonction de *code préfixe* si :

$$\forall 1 \leqslant i \leqslant n : \nexists 1 \leqslant j \leqslant n \text{ t.q. } (i \neq j) \wedge \left(\forall 1 \leqslant k \leqslant \min \left\{ \left|K(s_i)\right|, \left|K(s_j)\right| \right\} : \left(K(s_i)\right)_k = \left(K(s_j)\right)_k \right).$$

Remarque.

- Autrement dit, une fonction de code est dite *préfixe* lorsqu'aucun mot du code n'est préfixe d'un autre mot du code ;
- les codes préfixes présentent l'avantage d'être déchiffrables à la volée à l'aide d'un automate fini (ou d'un arbre de décision n-aire où $n = |C|$ [1]) ;
- les codes blocs sont un cas particulier de code préfixe : en effet, aucun mot du code n'est préfixe d'un autre étant donné qu'ils ont tous la même longueur et que la fonction de code est injective.

1.3 Théorèmes de Kraft et Mc Millan

Théorème 1.9 (Inégalité de Kraft)**.** *Soient* $S = \{s_1, \ldots, s_q\}$ *et* $C = \{c_1, \ldots, c_r\}$*. On pose* $\ell_i \coloneqq \ell(K(s_i))$*. Alors, il existe un code préfixe* $K : S \to C^*$ *si et seulement si :*

$$\sum_{i=1}^{q} r^{-\ell_i} \leqslant 1.$$

Démonstration. Réorganisons les s_i de manière à ce que $\forall 1 \leqslant i \lneqq q : \ell_i \leqslant \ell_{i+1}$.

Montrons d'abord que s'il existe un code préfixe, alors l'inégalité est vérifiée.

Soit $\mathcal{A}$, l'arbre r-aire complet de hauteur l_q. Pour $1 \leqslant i \leqslant q$, on pose $a_i \coloneqq K(s_i)$. Notons $\mathcal{A}_i$ le sous-arbre r-aire ayant a_i pour racine. Par définition de code préfixe, la famille $\{\mathcal{A}_i\}_{1 \leqslant i \leqslant q}$ est distincte deux à deux.

On observe aisément que $\mathcal{A}_i$ est un arbre r-aire de hauteur $\ell_q - \ell_i$. On a donc :

$$|\mathcal{A}_i| = r^{\ell_q - \ell_i}.$$

On peut donc écrire :

$$r^{\ell_q} = |\mathcal{A}| \geqslant \left| \bigcup_{k=0}^{q} \mathcal{A}_k \right| = \sum_{k=0}^{q} |\mathcal{A}_k| = \sum_{k=0}^{q} r^{\ell_q - \ell_k}.$$

En divisant de part et d'autre par r^{ℓ_q}, on obtient :

$$1 \geqslant \sum_{k=0}^{q} r^{-\ell_k}.$$

Supposons maintenant que l'inégalité est vérifiée et montrons qu'il existe un code préfixe.

Si $(\ell_i)_{1 \leqslant i \leqslant q}$ est un vecteur de naturels satisfaisant l'inégalité de Kraft et tels que :

$$\forall 1 \leqslant i \lneqq q : \ell_i < \ell_{i+1},$$

1. Les codes blocs peuvent également être représentés par un arbre de décision n-aire. On peut alors dire qu'un code est un code bloc si et seulement si l'arbre de décision associé est complet.

alors on construit $\mathcal{A}$ un arbre r-aire de hauteur ℓ_q. Pour tout $i \leqslant q$, on élague l'arbre $\mathcal{A}_i$ en supprimant un nœud de hauteur ℓ_i. Cela supprime à l'itération i, $r^{\ell_q - \ell_i}$ nœuds de l'arbre. Donc à la qème itération, sont supprimés au total :

$$\sum_{k=0}^{q} r^{\ell_q - \ell_k} = r^{\ell_q} \sum_{k=0}^{q} r^{-\ell_k} \leqslant r^{\ell_q}$$

nœuds de l'arbre. Il est donc possible de placer les q mots afin de former un code préfixe dans l'arbre car si ce n'était pas possible, l'arbre devrait contenir strictement moins de nœuds que r^{ℓ_q}, ce qui n'est pas le cas. □

Théorème 1.10 (Théorème de Mc Millan)**.** *Tout code univoque satisfait l'inégalité de Kraft.*

Démonstration. Soient $i_1, \ldots, i_n \in \mathbb{N}$, C et S deux alphabets. Soit $K : S \to C^*$ une fonction de code. On pose :

$$j \coloneqq \left| K(s_{i_1} s_{i_2} \ldots s_{i_n}) \right|.^{2}$$

Si x_j est le nombre de mots de longueur j, on sait que $x_j \leqslant r^j$. Par définition de j, on peut écrire :

$$j = \sum_{k=1}^{n} \ell_{i_k}.$$

On pose :

$$\alpha \coloneqq \sum_{k=1}^{q} r^{-\ell_k}.^{3}$$

Dès lors, on a :

$$\alpha^n = \left(\sum_{k=1}^{q} r^{-\ell_k} \right)^n = \sum_{\gamma_1, \ldots, \gamma_n = 1}^{n} r^{-\sum_{k=1}^{n} \ell_{\gamma_k}}.$$

On pose :

$$\mu \coloneqq \max_{\gamma_1, \ldots, \gamma_n} \left\{ \sum_{i=1}^{n} \ell_{\gamma_i} \right\}.$$

On peut alors exprimer

$$\alpha^n \leqslant \sum_{j=1}^{\mu} x_j r^{-j} \leqslant \sum_{j=1}^{\mu} r^{-1} r^j = \sum_{j=1}^{\mu} 1 = \mu.$$

Or, $\mu = n \cdot \max_i\{\ell_i\}$. Notons $L \coloneqq \max_i\{\ell_i\}$. On a alors :

$$\alpha^n \leqslant nL.$$

En divisant par n de part et d'autre, on obtient :

$$\frac{\alpha^n}{n} \leqslant L,$$

2. Ici, j est une fonction de n paramètres i_1 jusque i_n, mais on peut considérer la valeur constante car les valeurs i_k sont fixées.
3. Idem.

où L est une constante naturelle. La suite $\left(\frac{\alpha^n}{n}\right)_n$ est donc bornée par L. On peut alors déduire que la limite de cette suite existe également. [4]

Dès lors, $|\alpha| = \alpha \leqslant 1$. □

4. Il faut pour cela que la suite $\left(\left|\frac{\alpha^n}{n}\right|\right)_n$ soit bornée, mais la suite $\left(\frac{\alpha^n}{n}\right)_n$ est définie positive, donc la suite valeur absolue est la même, et est donc bornée.