

CODAGE

I) Code :

$A = \{ \text{symboles à utiliser pour coder} \}$ l'alphabet

$S = \{ \text{symboles / messages à coder} \}$ la source

un code est une application $C: S \rightarrow A^*$. Pour que C soit décodable, C doit être injectif. Alors C est régulier. Sinon, C est singulier.

NON SUFFISANT

II) Déchiffabilité :

Algo de Sardinas-Patterson : marche sur les codes réguliers.

$$X^{-1}Y = \{ w \in A^* / \exists u \in X, uw \in Y \}$$

↳ pour calculer $X^{-1}Y$: - trouver les mots de Y qui ont un préfixe dans X .
- prendre leur suffixe

Étapes : ① $X_0 = X = \{ \text{mots de code} \}$

$$② X_1 = X^{-1}X$$

$$\rightarrow ③ X_n = X^{-1}X_{n-1} \cup X_{n-1}^{-1}X$$

- ④ Arrêt lorsque : - $\varepsilon \in X_n \Rightarrow$ pas décodable
- $X_i = X_j, i < j \Rightarrow$ décodable
- $X_1 = \emptyset \Rightarrow$ décodable

↳ Exemple : $\{aaa, abu, abb, abaab\}$, $X_3 = X_5$

Comment assurer la déchiffabilité :

↳ code de longueur fixe

↳ code avec séparateurs

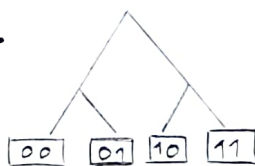
↳ code où aucun mot n'est préfixe d'un autre

↳ Alors, code "instantané" / "préfixe" / "irréductible"

Arbre d'un code :

arbre des mots de A^* dans lequel on affiche que les mots de C .

Ex:



ptio: (u préfixe de v) $\Leftrightarrow$ ($\exists$ un chemin de u à v)

(C préfixe) $\Leftrightarrow$ (les mots de C sont feuilles)

(C complet) $\Leftrightarrow$ (Arbre complet)

Inégalité de Kraft:

$$\begin{cases} |S| = N \\ |A| = q \\ l_1, \dots, l_N \text{ des entiers naturels} \end{cases}$$

Alors $\left(\sum_{i=1}^N q^{-l_i} \leq 1 \right) \Rightarrow \left(\exists \text{ un code déchiffrable de } S \text{ sur } A^* \right.$
avec les longueurs de code $l_1, \dots, l_N$

$\hookrightarrow R_q$: il est préfixe

Théorème de Mac-Millan:

$$\begin{cases} |S| = N \\ |A| = q \\ l_1, \dots, l_N \in \mathbb{N}^N \end{cases}$$

Alors $\left(\sum_{i=1}^N q^{-l_i} \leq 1 \right) \Leftrightarrow \left(\exists \text{ un code déchiffrable de } S \text{ sur } A^* \right.$
avec les longueurs de code $l_1, \dots, l_N$

III Codage Optimal:

A chaque symbole, on associe sa fréquence d'apparition. (S, f)

Déf: distribution de freq^o: Application de la forme $f: S \rightarrow [0, 1]$
 $\alpha \mapsto f(\alpha)$
 $\text{tq } \sum_{\alpha \in S} f(\alpha) = 1.$

longueur moyenne: C un code de S sur A^* . La longueur moyenne relative à (S, f) est $\bar{m}_C = \sum_{\alpha \in S} f(\alpha) |C(\alpha)|$

code optimal au regard de f : C déchiffrable tq $(\forall C' / \bar{m}_{C'} < \bar{m}_C)$

$$\underline{\text{Ppté}}: ((S, \beta) \text{ et } C \text{ optim}^\circ \text{ par } \beta) \Rightarrow \left(\begin{array}{l} \exists C' \text{ optim}^\circ \text{ par } \beta \text{ tq} \\ \forall a \in S, |C'(a)| = |C(a)| \end{array} \right)$$

$$\left((S, \beta) \text{ et } C \text{ optim}^\circ \text{ linéaire} \right. \\ \left. \text{par } \beta \right) \Rightarrow \left(\begin{array}{l} \text{Inégalité de Kraft est une} \\ \text{égalité, et } C \text{ est complet.} \end{array} \right)$$

IV) Entropie:

Déf: Quantité d'infos: $\left\{ \begin{array}{l} (S, \beta) \\ a \in S \end{array} \right\}$ Alors $I(a) = \log_2 \left(\frac{1}{\beta(a)} \right)$ en bits.

$$\text{Entropie d'une source } (S, \beta): H(S, \beta) = \sum_{a \in S} \beta(a) I(a)$$

$$\hookrightarrow \text{par } S = \{0, 1\} \text{ et } \beta(0) = \beta(1), \\ H = 1.$$

$$= - \sum_{a \in S} \beta(a) \log_2 (\beta(a))$$

THM: Limite inférieure de la longueur moyenne:

$$\left\{ \begin{array}{l} (S, \beta) \\ |A| = q \\ C \text{ déchiffrable} \end{array} \right\} \Rightarrow \left(\bar{n}_C \geq \frac{H(S, \beta)}{\log_2 q} \right)$$

Borne supérieure de la longueur moyenne:

$$\left\{ \begin{array}{l} (S, \beta) \\ |A| = q \end{array} \right\} \Rightarrow \left(\begin{array}{l} \exists C \text{ déchiffrable tq} \\ \bar{n}_C < \frac{H(S, \beta)}{\log_2 q} + 1 \end{array} \right)$$