Conditions nécessaires d'optimalité pour un code instantané

Définition : Un code déchiffrable d'une source est dit optimal s'il n'existe aucun code déchiffrable de cette source dont la longueur moyenne soit strictement inférieure

Propriété : Pour qu'un code soit optimal il est nécessaire que :

1. si $p_j > p_k$ alors $l_j \leq l_k$
2. les deux mots les plus longs ont la même longueur
3. parmi les mots de longueur maximale, deux au moins ne diffèrent que par le dernier caractère.

Code optimal de Huffmann

Le code de Hufmann

1. classer l'ensemble des réalisations possibles de la source dans l'ordre des probabilités décroissantes
2. grouper les deux évènements les moins probables en un unique évènement
3. reclasser et renouveler l'opération

Exemple "dindon dina dit-on du dos d'un dodu dindon "

1. Donner le codage de chacun des caractères par la méthode de Huffman et coder le premier mot de la phrase
2. Que vaut la compacité du code
3. commenter sa performance
4. comparer avec un code à longueur fixe

Code de Fano-Shannon

La construction du code de Fanon-Shannon reprend l'idée utilisée pour la construction de bons questionnaires.
codage de Fano-Shannon :

1. classer l'ensemble des réalisations possibles de la source dans l'ordre des probabilités décroissantes
2. diviser deux sous-ensembles de probabilités aussi voisines que possible
3. renouveler l'opération sur chacun des sous-ensembles.

Exemple "dindon dina dit-on du dos d'un dodu dindon "

1. Donner X et sa loi de probabilité,
2. calculer $H(X)$
3. Donner le codage de chacun des caractères par la méthode de Fano-Shanon et coder le premier mot de la phrase
4. Que vaut la compacité du code
5. commenter sa performance

suites typiques d'une source simple

Soit une source constituée de la suite de v.ar $X_1, \cdots X_n, \cdots$ à valeurs dans l'alphabet $\mathcal{A}$

L'entropie de la source par lettre est donnée par

$$H = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \cdots, X_n)$$

Ensemble des séquences ε-typique de longueur n

$$A_\varepsilon^{(n)} = \left\{ (x_1, \cdots, x_n) \in \mathcal{A}^n, \left| \frac{1}{n} log_2 \left(\frac{1}{p(x_1, \cdots, x_n)} \right) - H \right| \leq \varepsilon \right\}$$

suites typiques d'une source simple

Dans le cas d'une **source simple** :

$$H = H(X)$$

Notons $\mathcal{A} = \{a_1, a_2, \cdots, a_k\}$, $x = (x_1, \cdots, x_n)$ une séquence de longueur n et $f_i(x)$ la fréquence d'apparition du symbole a_i dans la suite x

Ensemble des séquences ε-typique de longueur n

$$A_\varepsilon^{(n)} = \left\{ (x_1, \cdots, x_n) \in \mathcal{A}^n, \left| \sum_{i=1}^{k} (f_i(x) - p(a_i)) \, log_2 (p(a_i)) \right| < \varepsilon \right\}$$

Ensemble des séquences ε-typique de longueur n

$$A_\varepsilon^{(n)} = \{(x_1, \cdots, x_n) \in \mathcal{A}^n, |H(f) + \mathcal{D}(f||p) - H(p)| \leq \varepsilon\}$$

Équipartition Asymptotique (AEP)

Une source vérifie l'AEP si $(\forall \varepsilon > 0)$, $\lim_{n \to \infty} P\left[A_\varepsilon^{(n)}\right] = 1$

Proposition :

- $\frac{1}{n} log_2 \left(\frac{1}{p(x_1, \cdots, x_n)} \right)$ converge presque surement vers H
- $\left| \mathcal{A}_\varepsilon^{(n)} \right| \leq 2^{n(H+\varepsilon)}$
- pour n suffisament grand $\left| \mathcal{A}_\varepsilon^{(n)} \right| \geq (1-\varepsilon) 2^{n(H-\varepsilon)}$

pour n assez grand et $\varepsilon > 0$ assez petit le nombre de suites typiques et de l'ordre de 2^{nH}, de probabilité environ 2^{-nH}

Loi faible des grand nombres

Rappel : loi faible des grand nombres : soient $Z_1, Z_2, \cdots, Z_n, \cdots$ une suite de variables aléatoires iid et d'espérance μ

$$\overline{Z_n} = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

Pour tout $\varepsilon < 0$

$$\lim_{n\to\infty} P\left\{|\overline{Z_n} - \mu| > \varepsilon\right\} = 0$$

AEP des sources sans mémoires

Une source sans mémoire vérifie l'AEP

$$-\frac{1}{n}log_2\left[p(X_1, X_2, \cdots, X_n)\right] = -\frac{1}{n}\sum_{i=1}^{n} log_2\left[p(X_i)\right]$$

d'après la loi faible des grand nombre

$$P\left\{\left|-\frac{1}{n}\sum_{i=1}^{n} log_2\left[p(X_i)\right] - E\left[-log_2(p(X))\right]\right| \leq \varepsilon\right\} \to 1$$

$$P\left\{\left|-\frac{1}{n}\sum_{i=1}^{n} log_2\left[p(X_i)\right] - H\right| \leq \varepsilon\right\} \to 1$$