TD4 : CDC.

Q 2.2 Quelle est la particularité des arbres de codages binaires préfixes optimaux ?

Q 2.3 Quels sont les arbres des codages préfixes binaires optimaux d'une source de trois symboles ? de quatre symboles ?

Pour qu'un code soit optimal, il faut que $K(c) = 1$.

Supposons que $c_1, c_2, \dots, c_n$ sont de longueur $l_1 \leq l_2 \leq \dots \leq l_n$ avec $l_{n-3} < l_{n-2} = l_{n-1} = l_n$

$$K(C) = \sum_{i=1}^{m} \frac{1}{2^{|c|}} = \sum_{i=1}^{m} \frac{1}{2^{l_i}} = \sum_{i=1}^{m-3} \frac{1}{2^{l_i}} + \sum_{i=m-2}^{m} \frac{1}{2^{l_i}}$$

$$= \sum_{i=1}^{m-3} \frac{2^{l_m - l_i}}{2^{l_m}} + \frac{3}{2^{l_m}}$$

$$= \underbrace{\frac{2^{l_m - l_1} + 2^{l_m - l_2} + \dots + 2^{l_m - l_{m-3}} + 3}{2^{l_m}}}_{\neq 1.}$$

pair

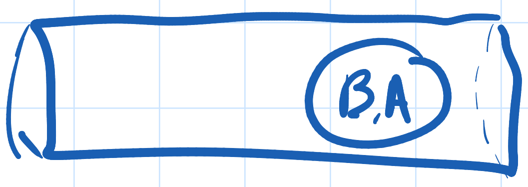
impair

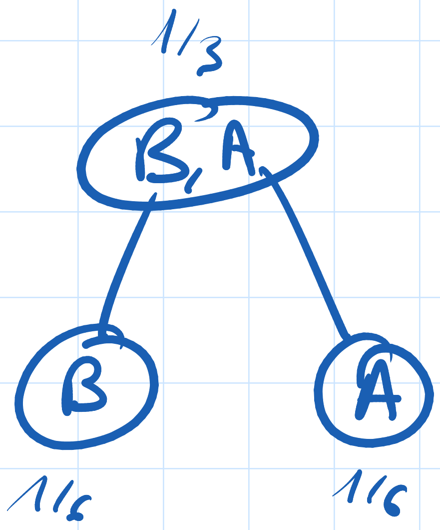
pair

Q 2.5 La distribution des longueurs des mots d'un codage préfixe binaire optimal est-elle unique ? non, contre exemple :

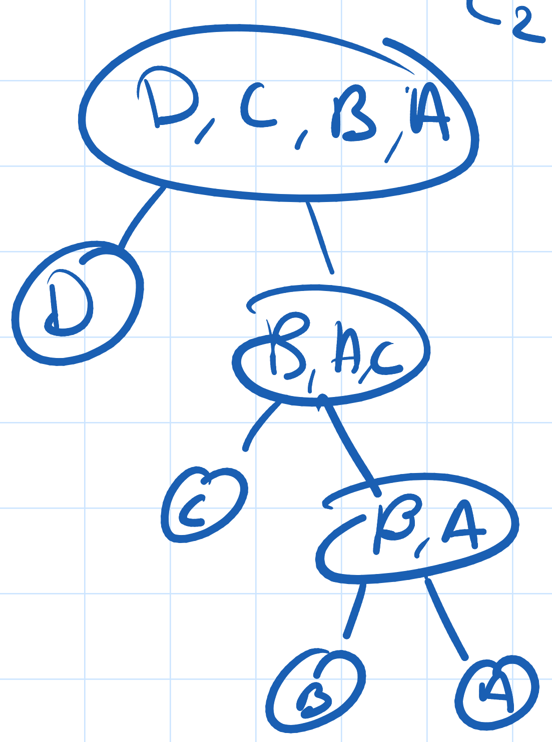
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
S	A	B	C	D

D, C, B, A

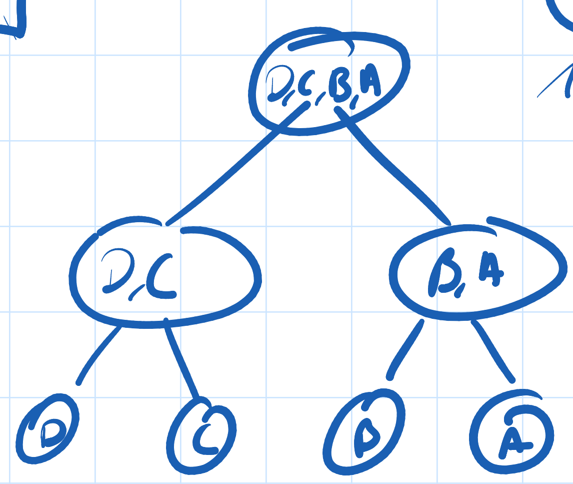


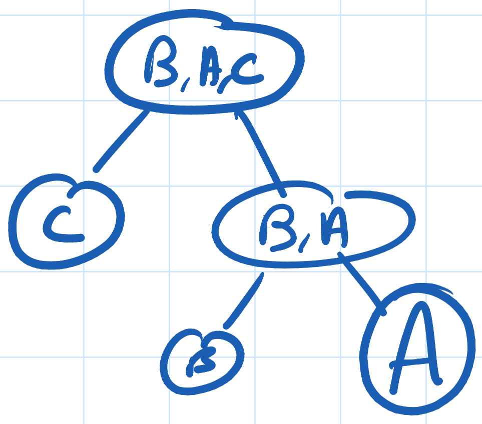


C_2



C_1





$C_1(D) = 00$
$C_1(C) = 01$
$C_1(B) = 10$
$C_1(A) = 11$

$C_2(D) = 0$
$C_2(C) = 10$
$C_2(B) = 110$
$C_2(A) = 111$

$\overline{n_{c_2}} \quad 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + \frac{3}{6} + \frac{3}{6}$

$= \frac{3}{3} + \frac{6}{6} = 2.$

$\overline{n_{c_1}} = 2$

Exercice 1.3 : ***Entropie d'un message***

On considère un message M écrit avec les symboles d'un alphabet $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$. On désigne par n_i le nombre d'occurrences du symbole s_i dans le message M. La longueur du message est désignée par N. On a évidemment

$$N = \sum_{i=1}^{m} n_i.$$

Exprimez l'entropie de M en fonction des entiers N et n_i, $i = 1, \ldots, m$, de manière à ce que N n'apparaisse pas dans la somme.

Rappel : Entropie de Shannon

$$H(S) = \sum_{s \in S} \mathbb{P}(S=s) \times \underbrace{I(s)}_{\text{Quantité d'information [bit]}}$$

Avec $I(s) = -\log_2(\mathbb{P}(S=s))$

On suppose que $\mathbb{P}(S=s_i) \approx$ fréquence de s_i dans M

$$\begin{cases} n_i = \text{compte de } s_i \text{ dans } M \\ N = |M| \end{cases}$$

Donc $f_i = \frac{n_i}{N}$.

Il faut montrer que :

$$H(M) = \log_2(N) - \frac{\sum_{i=1}^{m} n_i \log_2(n_i)}{N}$$

$$H(M) = -\sum_{i=1}^{m} \frac{n_i}{N} \times \log_2\left(\frac{n_i}{N}\right)$$

$$= -\sum_{i=1}^{m} \frac{n_i}{N} \times (\log_2(n_i) - \log_2(N))$$

$$= \log_2(N) \underbrace{\sum_{i=1}^{n} \frac{n_i}{N}}_{=1} - \frac{1}{N} \sum_{i=1}^{m} n_i \log_2(n_i) \quad \text{ok.}$$

Exercice 1.4 :

Q 4.2 Quelle contrainte a-t-on sur $I(s)$ et donc sur $p(s)$ pour qu'un tel codage existe ?

Imaginons un codage c pour lequel la longueur du mot associé à chaque symbole soit égale à la quantité d'information du symbole. Autrement dit pour chaque symbole $s \in \mathcal{S}$, $|c(s)| = I(s)$.

Q 4.1 Pourquoi un tel codage serait-il optimal ?

Si $\overline{n_c} = H(S)$ => Le code C est nécessairement optimal (cond nécessaire suffisant).

Thm du codage sans bruit :

$H(S) \leq \overline{n_c} < H(S)+1.$

$\sum_{s \in S} \mathbb{P}(S=s) I(s) = \sum_{s \in S} \mathbb{P}(S=s) . |c(s)|.$

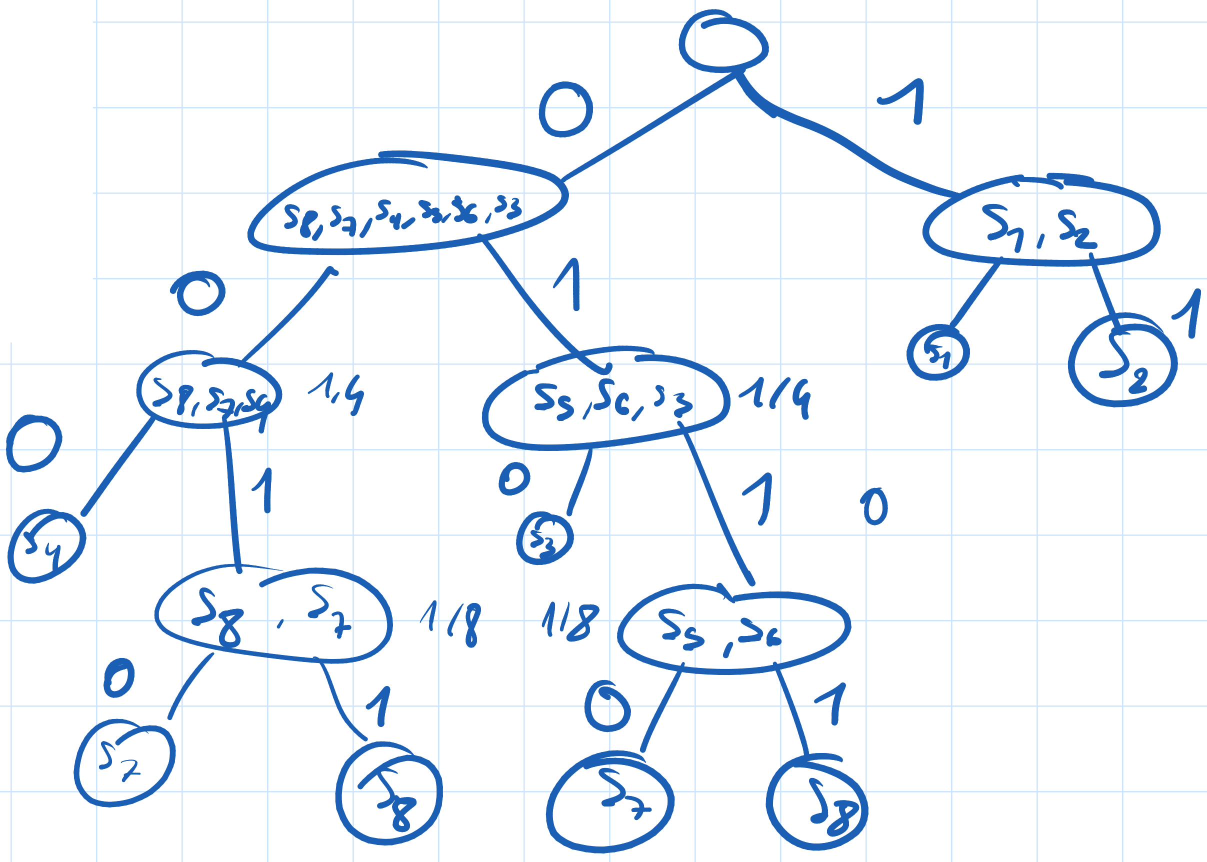
Donc $I(s) = |c(s)|$. $I(s)$ doit donc être entière > 0 :>.

Comme $I(s) = -\log_2(\mathbb{P}(S=s))$, il faut que $\mathbb{P}(S=s) = \frac{1}{2^k}$

On considère une source de 8 symboles décrites par le tableau ci-dessous.

s	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
$p(s)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Q 4.3 Déterminez un codage optimal pour cette source.



$4 \times \frac{1}{4} + \frac{3\times 2}{8} + 4 \times \frac{4}{16}$

$= 2 + \frac{3}{4} = 2.75.$

Q 4.4 Combien y a-t-il de codages binaires préfixes optimaux pour cette source ?

↳ feuille de taille 2 : $\binom{4}{2} = 6$

↳ feuilles de profondeur 3 : $\binom{4}{2} = 6$

↳ feuilles de profondeur 4 : $\binom{4}{4} = 1$

=> $6 \times 6 = 36$ arbres possibles.

Il y a $2 \times 2 \times 4! \times 36 = 2 \times 2 \times 2 \times 3 \times 4 \times 36$
$= 16 \times 4 \times 3 \times 36$
$= 16 \times 3 \times 36 = 3456$