Alphabet $A = \{x_1 \to x_K\}$.

Proba d'apparition $p_k = \mathbb{P}(X = x_k)$ (typiqt estimé empiriqt).

Code $C_k = c(x_k)$, longueur $|C_k|$

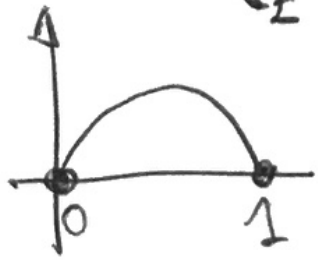
Longueur moyenne : $L(C) = \mathbb{E}_X(|c(X)|)$

$= \sum_k p_k |C_k|$ (= longueur texte codé / # symboles)

Code préfixe : $C = \text{Feuilles}(T)$

<u>Pbm</u> : $\boxed{\min_{C=\text{Feuilles}(T)} L(C)}$

<u>Déf</u> : $H(p) \triangleq \sum_k -p_k \log_2(p_k)$



<u>Propo</u> : $0 \leq H(p) \leq \log_2(N)$

<u>Preuve</u> : $\min_{(p_k)} \underbrace{\sum p_k \log p_k}_{f(p)}$ st. $(p_k \geq 0)$. $\underbrace{\sum_k p_k = 1}_{g(p) = 1}$

p solution $\Rightarrow \exists \lambda$ tq. $\nabla f(p) = \lambda \nabla g(p)$.

$\left[\frac{d}{dx}[x \log x] = \log x + 1\right] \Rightarrow \forall k \; \log p_k + 1 = \lambda \Rightarrow p_k = 2^{\lambda - 1} = \text{cste}$

$\sum p_k = 1 \Rightarrow p_k = 1/N$ (unique solution)

$H((1/N)) = \sum_k -\frac{1}{N} \log_2(1/N) = \log_2(N)$

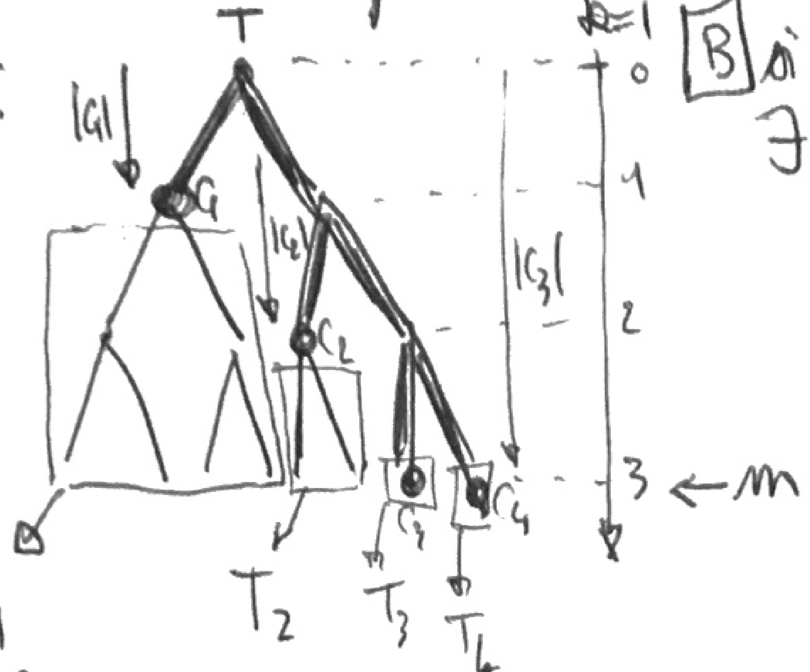
Thm : $\boxed{1}$ $\forall$ c = Feuille(T), $L(c) \geq H(p)$.

$\boxed{2}$ $\exists$ c = Feuille(T), $L(c) \leq H(p) + 1$

Lemme : [ineg de Kraft] $\boxed{A}$ $\sum_{k=1}^{m} 2^{-|c(x_k)|} \leq 1$ $\forall$ c = Feuille(T)

$\boxed{B}$ si $\sum 2^{-l_k} \leq 1$, alors $\exists$ c tq. $|c(x_k)| = l_k$

Dem $\boxed{A}$:



$\text{Prof}(T_k) = m - |c_k|$, $|\text{Feuille}(T_k)| = 2^{m-|c_k|}$

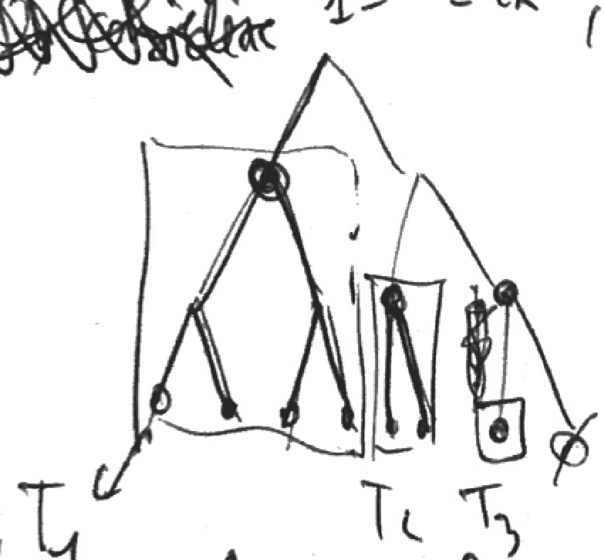
$\hat{C}$ les T_k sont disjoints, $\sum_k |\text{Feuilles}(T_k)| \leq |\text{Feuilles}(T)|$, où $|\text{Feuilles}(T_k)| = 2^{m-|c_k|}$ et $|\text{Feuilles}(T)| = 2^m$

$\Rightarrow \sum 2^{m-c_k} \leq 2^m$ QED

$\boxed{B}$ OPS que $l_1 \leq \dots \leq l_k$, $m \triangleq \max\{l_1 - l_k\}$.

Algo :
- T_1 le 1er arbre de prof. $m - l_1$
- T_2 le 2e arbre de profondeur $m - l_2$
- T_3 etc...

ex : $l_1 = 1$, $l_2 = 2$, $l_3 = 3$, $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} \leq 1$.



- $\hat{C}$ $\sum_k 2^{m-l_k} \leq 2^m$ on peut terminer cette construction.
- Les mot c_k sont associés au sommets des arbres

Preuve (thm entropie shannon) :

Préliminaire

On considère : $\min_{(l_k)} \underbrace{\sum p_k l_k}_{f(l)}$ st. $\underbrace{\sum 2^{-l_k} \leq 1}_{g(l)}$.

!! $\circledast$ Montrer que néces^t $\sum 2^{-l_k^*} = 1$

$\exists \lambda$ tq. $\nabla f(l^*) = \lambda \nabla g(l^*)$. $\quad \left(\frac{d}{dx} 2^{-x}\right) = -\ln(2) 2^{-x}$

$\Rightarrow \overset{\shortparallel}{(p_k)_k} = -\lambda \ln(2) 2^{-l_k^*}$.

$\Rightarrow \underbrace{\sum p_k}_{=1} = -\lambda \ln(2) \underbrace{\sum 2^{-l_k^*}}_{=1} \Rightarrow -\lambda \ln 2 = 1$

$\Rightarrow$ sol$^\circ$: $l_k^* = -\log(p_k)$.

On définit (pour avoir des nbr entiers) $l_k \triangleq \lceil -\log(p_k) \rceil$

~~Alors comme~~ **Preuve** $\boxed{1}$: si c = feuille(T) avec Kraft $\boxed{A}$

$\sum 2^{-|c_k|} \leq 1$ donc néc^t $(l_k^*)_k$ est meilleur que $(|c_k|)_k$ donc $\underbrace{\sum_k p_k l_k^*}_{= H(p)} \leq \underbrace{\sum_k p_k |c_k|}_{= L(c)}$

Preuve $\boxed{2}$: on définit $l_k \triangleq \lceil -\log(p_k) \rceil \in \mathbb{N}^*$ alors

$\sum_k 2^{-l_k} \leq \sum_k 2^{-\log(p_k)} = \sum_k p_k = 1$

donc avec Kraft $\boxed{B}$ $\exists$ code c = feuille(T) $|c_k| = l_k$

$L(c) = \sum_k p_k \lceil -\log(p_k) \rceil \leq \sum_k p_k(-\log p_k + 1) = H(p) + 1$ $\blacksquare$

$\circledast$ Si $\sum 2^{-l_k}$ (optimal l_k) $= 2^{-u} < 1 \Rightarrow \sum 2^{-\overbrace{(l_k - u)}^{l'_k}} = 1$

et $\sum p_k l'_k = \sum p_k l_k - u < \sum p_k \ldots \Rightarrow$ contradiction