

Exercice 2:

① a) Pour k une clef de U , on définit :

$X_j^k | 1 \text{ si } k \text{ est hachée en } T[j] (h(k) = j)$
0 sinon

$$\text{On a: } X_j = \sum_{k \in k} X_j^k$$

Et, par linéarité de l'espérance : $\mathbb{E}[X_j] = \sum_{k \in k} \mathbb{E}[X_j^k]$

X_j^k est une variable de Bernoulli

$$\mathbb{E}[X_j^k] = 0 \times \Pr[X_j^k = 0] + 1 \Pr[X_j^k = 1]$$

$$= \Pr[X_j^k = 1] = \frac{1}{n} \text{ car } h \text{ est uniforme}$$

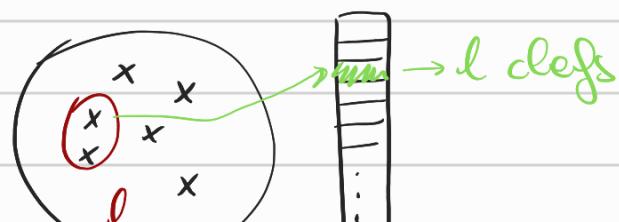
$$\text{Du coup, } \mathbb{E}[X_j] = \sum_{k \in k} \frac{1}{n} = \frac{|k|}{n} = \frac{n}{n} = 1$$

b) On cherche $\max \{X_j : j = 0 \dots n-1\}$ en moyenne et pas $\max \{\mathbb{E}[X_j] : j = 0 \dots n-1\}$ qui avait 1...

But: trouver le max par case. On pourrait dire 1, mais on fait pas le max des espérances (qui valent toutes 1).

② a) On veut majorer $\Pr[X_j \geq l]$

Compte la probabilité de clefs dans la case j



$$\Pr[X_j \geq l] \leq \Pr[\exists X \subseteq k, |X|=l, \forall k \in X, h(k)=j]$$

$$\leq \Pr\left[\bigcup_{\substack{X \subseteq k \\ |X|=l}} \{\forall k \in X, h(k)=j\}\right]$$

$$\leq \sum_{\substack{X \subseteq k \\ |X|=l}} \Pr[\forall k \in X, h(k)=j]$$

$$\leq \sum_{\substack{X \subseteq k \\ |X|=l}} \left(\frac{1}{n}\right)^l \leq \binom{n}{l} \left(\frac{1}{n}\right)^l$$

nbr de sous-ens. de taille l dans un ensemble de taille n

Si $X = \{x_1, \dots, x_n\}$ | $\Pr[h(x_i)=j] = \frac{1}{n}$

$$\Pr[\forall i=1 \dots l \ h(x_i)=j] = \left(\frac{1}{n}\right)^l$$

b) $\binom{n}{l} = \frac{n!}{l!(n-l)!} = \frac{1}{l!} \times \frac{n \times (n-1) \times \dots \cancel{x}}{(n-l) \times (n-l-1) \times \dots \times 1}$

$$= \frac{1}{l!} \frac{\overset{\leq n}{n} \times \overset{\leq n}{(n-1)} \times \dots \times \overset{\leq n}{(n-l+1)}}{n^l}$$

c) On a : $\Pr[X_j \geq l] \leq \binom{n}{l} \left(\frac{1}{n}\right)^l \leq \frac{1}{l!} \times \frac{1}{n^l} = \frac{1}{l!}$

Par ailleurs $(l!)^2 \geq l^l$ donc $l! \geq l^{l/2}$ et $\frac{1}{l!} \leq \frac{1}{l^{l/2}}$

Probas de 10 clés dans la case j: $\frac{1}{l^{l/2}}$

10⁵

③ @ On appelle N le max de clés dans une case
Si on écrit $t = \log n$

$$\frac{c \log n}{\log \log n} = \frac{c \cdot t}{\log t} = \frac{c \cdot t}{\frac{\log t}{\sqrt{t}}} = \frac{c \cdot \sqrt{t}}{\log t} \xrightarrow[t \rightarrow +\infty]{} +\infty$$

et $\sqrt{\log_1} = \sqrt{t}$

La racine
l'emporte sur le log

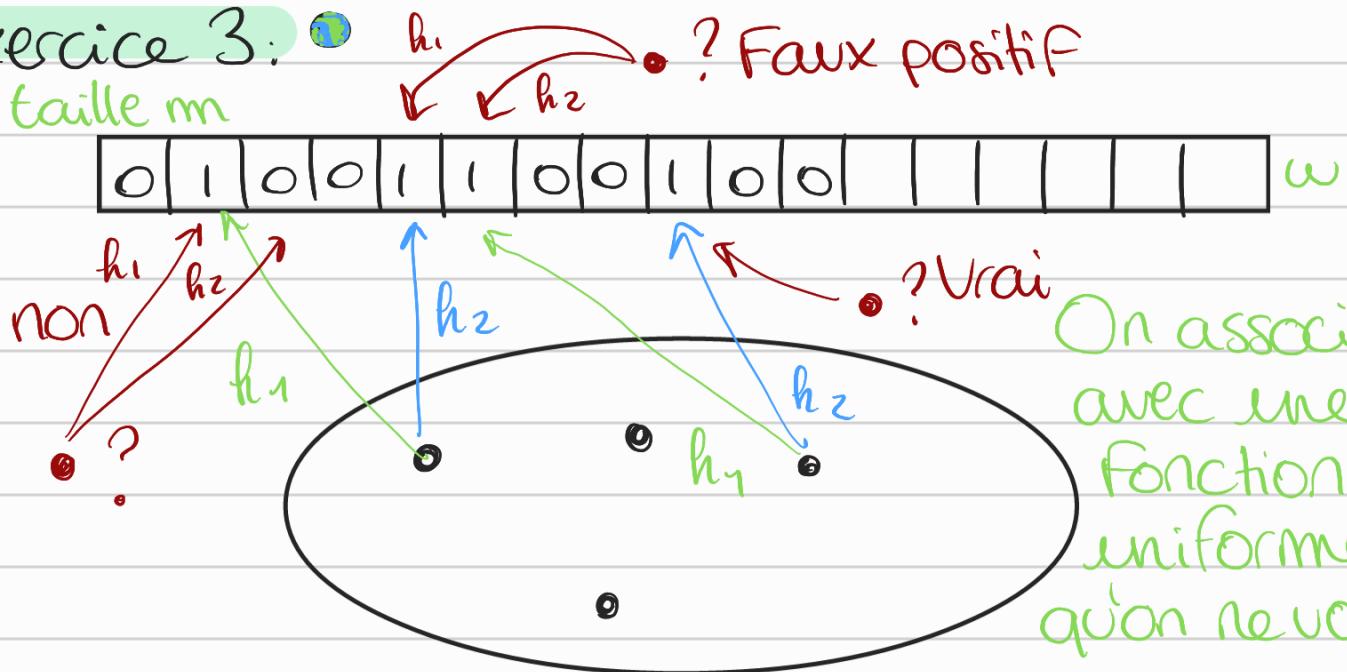
skip le reste

A la fin du ⑤ on trouve la taille
de la plus grande liste chaînée

$\frac{\log}{(\log(\log(n)))} \leftarrow$ voir ~~fa~~ Haché 1000 clé
 $\log(1000)$; j'ai pas suivi
Liste chaînée de taille 3
maximum
Donc plutôt efficace

Exercice 3:

taille m ? Faux positif



On associe
avec une
fonction
uniforme
qu'on ne voit pas

à chaque élément des indices dans w .
Pour la vérification rapide, on regarde les
indices associés à l'élément.

Ce que ça fait, c'est qu'il y a 1 élément qui a

Si tous sont déjà à i, c'est possible que l'élément ait déjà été mappé, ou faux positifs d'autres éléments.

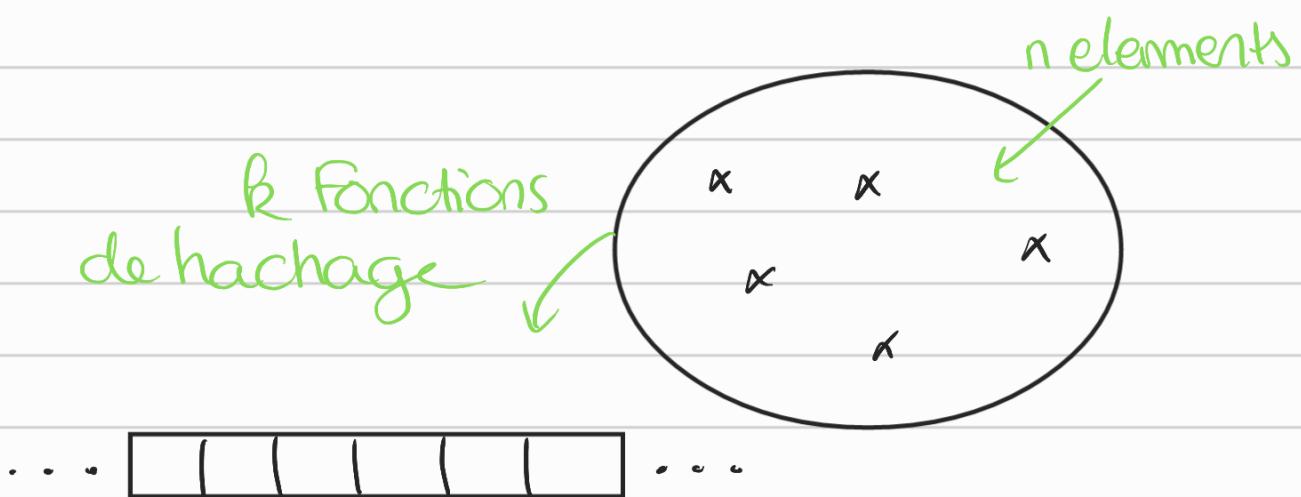
Il y a k fonctions de hachage, chacune d'elles map une case à un élément.

① Si le filtre répond non sur x , on est sûr que $x \notin w$

② C'est la définition de l'algorithme.

Si $w_i = 1$, alors à la construction de w , un élément $x \in X$.

③



Si on fixe $x \in X$ et $j \in \{1 \dots k\}$

$$\Pr[h_j(x) = i] = \frac{1}{n} \text{ car } h_j \text{ est uniforme}$$

chance de tomber sur 1 case avec chaque fonction de hachage

$$\Pr[w_i = 0] = \Pr[\forall x \in X, \forall j \in \{1 \dots k\}, h_j(x) \neq i]$$

$$= P_r \left[\bigcap_{x \in X} \bigcap_{j \in \{1..k\}} h_j(x) \neq i \right]$$

$$= \prod_{x \in X} \prod_{j \in \{1..k\}} P_r [h_j(x) \neq i] \text{ car les } h_j \text{ sont indépendants.}$$

$$= \prod_{x \in X} \prod_{j \in \{1..k\}} \left(1 - \frac{1}{n}\right) = \prod_{x \in X} \left(1 - \frac{1}{m}\right)^k$$

$$\text{Donc } P_r [\omega_i = 0] = \left(1 - \frac{1}{m}\right)^{kn} =: p$$

Proba case $i=0$? C'est la proba qu'après avoir fait toutes les fonctions de hachage, aucune n'aït touché la case i
 $h_j(x) \neq i \rightarrow$ la fonction ne touche pas la case i

④ L'espérance du nombre de bits à 0 est $p \times m$, mais on est pas complètement sûr d'avoir ce nombre de bits à 0.

C'est comme lancer 10 fois une pièce, on a pas forcément 5 piles 5 faces.

⑤ On se sert de la proba trouvée avant qu'une case soit à 0.

$P_r [y \text{ faux positif}]$

$= P_r [\forall j=1..k, h_j(y) \text{ ième bit de } \omega \text{ soit égal à 1}]$

$= P_r [\forall j=1..k, \omega_{h_j(y)} = 1]$

$$\begin{aligned}
 &= P_r \left[\bigcap_{j=1}^k \{\omega_{h_j}(y) = 1\} \right] \\
 &= \prod_{j=1}^k P_r [\omega_{h_j}(y) = 1] \text{ car } h_j \text{ indépendants} \\
 &= \prod_{j=1}^k (1-p) = (1-p)^k
 \end{aligned}$$

log népérien
 ↓

⑥ Montrer qu'en prenant $k = \frac{m \cdot \log 2}{n}$, la proba de faux positifs est au plus: $\left(\frac{3}{4}\right) \frac{m \log 2}{n}$ (l'énoncé avait une erreur)

Exemple:

$$\log 2 \approx 0.7 \quad m=20\,000$$

$$n=1\,000 \quad k=1h$$

$$\text{Probas Faux positifs} \leq \left(\frac{3}{4}\right)^{1h} \approx 0,3\%.$$

Par ⑤ la proba de faux positifs est

$$\begin{aligned}
 (1-p)^k &= \left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^k \text{ par ③} \\
 &= \left(1 - \left(1 - \frac{1}{m}\right)^{\frac{m \log 2}{n} \times n}\right)^{\frac{m \log 2}{n}} \\
 &= \left(1 - \left(1 - \frac{1}{m}\right)^{m \log 2}\right)^{\frac{m \log 2}{n}}
 \end{aligned}$$

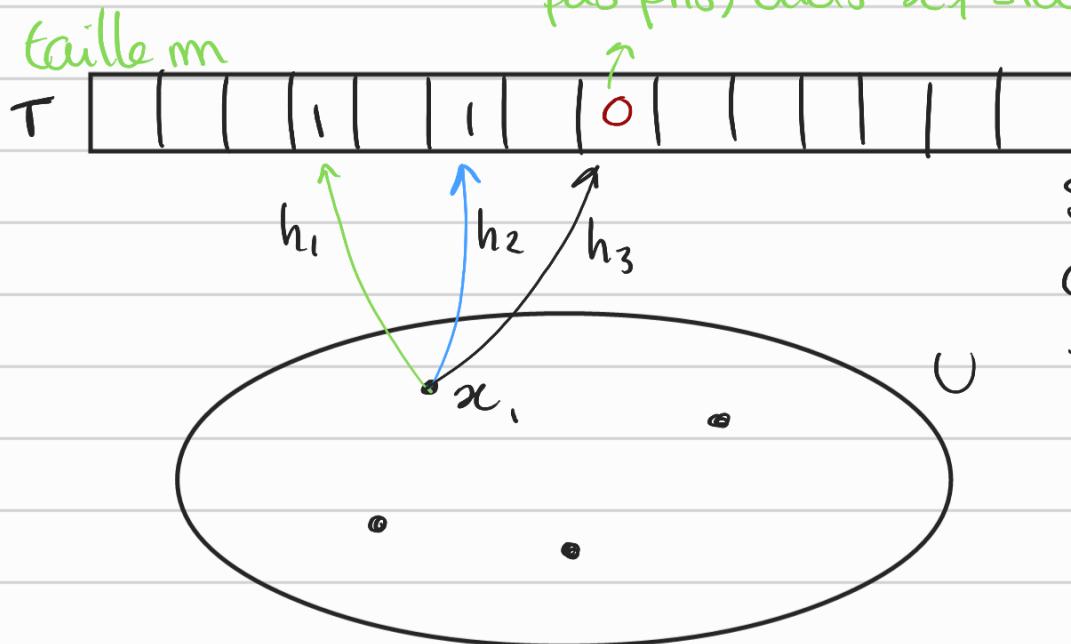
Par l'indication $\left(1 - \frac{1}{m}\right)^m \geq e^{-\frac{2}{m}}$ (on a bien $\frac{1}{m} \leq \frac{1}{2}$)

$$\begin{aligned}
 \text{donc } 1 - \left(1 - \frac{1}{m}\right)^{m \log 2} &\leq 1 - e^{\frac{-2m \log 2}{m}} = 1 - e^{-2 \log 2} \\
 &= 1 - (e^{\log 2})^{-2} = 1 - 2^{-2} = 1 - \frac{1}{4} = \frac{3}{4}
 \end{aligned}$$

$$\text{Et: } (1-p)^k \leq \left(\frac{3}{4}\right)^{\frac{m \log 2}{n}}$$

Exercice 4 : On sort du filtre de Bloom, c'est une table de hachage qui stocke un couple dans chaque case pour chaque élément.

pas pris, alors x_1 stocké avec h_3



Si $h_1(x_1)$ est déjà pris, on essaie $h_2(x_1)$

① Combien de cases occupées dans T ? n cases choisies aléatoirement et uniformément.

$$\Pr [T_{h_0}(x) \text{ occupé}] = \frac{1}{m}$$

$$\Pr [T_{h_0}(x) \text{ libre}] = 1 - \frac{1}{m}$$

② $E_{m,n} = 1 + \frac{n}{m} E_{m-1,n-1}$ utiliser l'espérance conditionnelle

③ Par récurrence

