

TD 9 : Applications à l'informatique

L3 Miage : Probabilités

A. Lentz

Recherche d'un élément en temps linéaire : $O(n)$.

| | | | |
|----|-----|--------|-----|
| FR | ... | manger | ... |
| EN | ... | to eat | ... |

Figure: Exemple d'utilisation : dictionnaire français/anglais.

On trie le tableau par ordre lexicographique:

→ recherche dichotomique en temps logarithmique : $O(\log n)$.

| | | | |
|----|-----|--------|-----|
| FR | ... | manger | ... |
| EN | ... | eat | ... |

Figure: Exemple d'utilisation : dictionnaire français/anglais.

Tables de hachage

On veut que la recherche d'un élément soit en temps constant : $O(1)$.

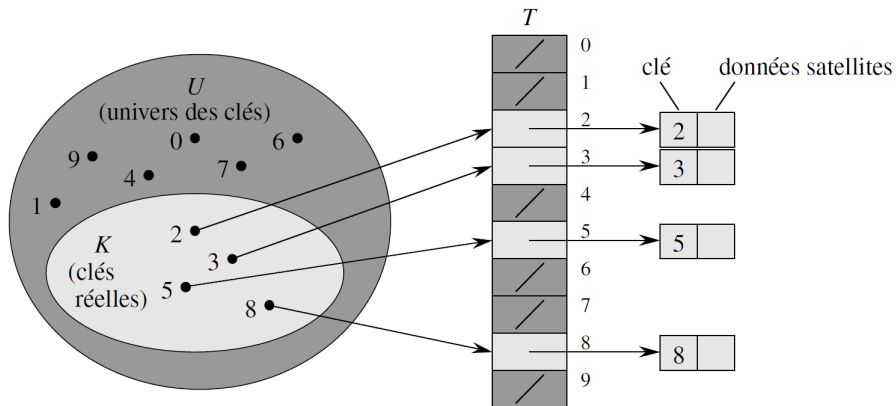


Figure: Hachage (ref : Cormen)

Tables de hachage

Problème : plusieurs éléments vont dans une même case !

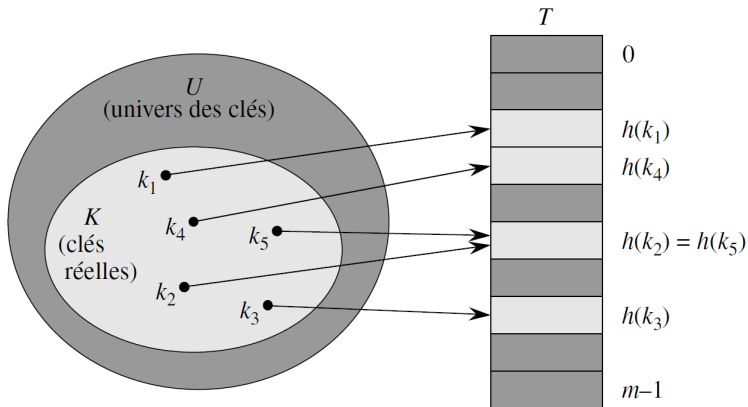


Figure: Hachage et collisions (ref : Cormen)

Tables de hachage

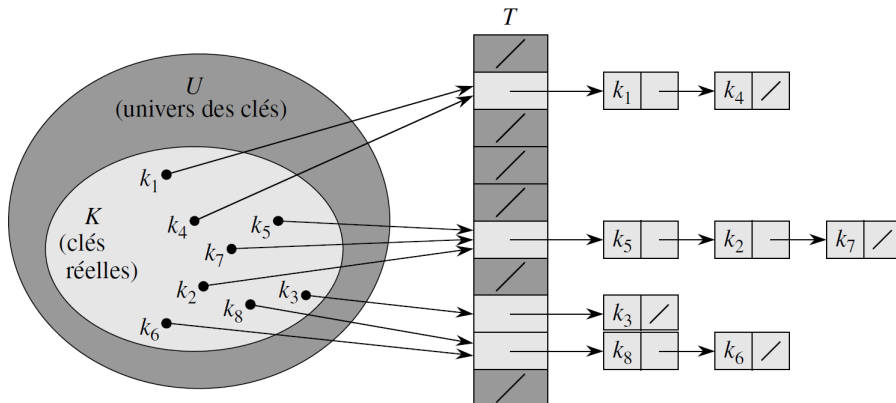


Figure: Gestion des collisions avec des listes chaînées (ref : Cormen)

Exercice 1, Question 1

Enoncé

- i la case considérée,
- n clés,
- m cases,
- X le nombre de collisions dans une case.

Exercice 1, Question 1

Énoncé

- i la case considérée,
- n clés,
- m cases,
- X le nombre de collisions dans une case.

$$\mathbb{P}[k \text{ collisions en } i] = \binom{n}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{n-k}$$

Exercice 1, Question 1

Énoncé

- i la case considérée,
- n clés,
- m cases,
- X le nombre de collisions dans une case.

$$\mathbb{P}[k \text{ collisions en } i] = \binom{n}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{n-k}$$

Donc

$$X \sim \text{Bin}\left(n, \frac{1}{m}\right)$$

Exercice 1, Question 1

Énoncé

- i la case considérée,
- n clés,
- m cases,
- X le nombre de collisions dans une case.

$$\mathbb{P}[k \text{ collisions en } i] = \binom{n}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{n-k}$$

Donc

$$X \sim \text{Bin}\left(n, \frac{1}{m}\right) = \text{Bin}\left(n, \frac{\alpha}{n}\right) \text{ avec } \alpha = \frac{n}{m}$$

Exercice 1, Question 1

Énoncé

- i la case considérée,
- n clés,
- m cases,
- X le nombre de collisions dans une case.

$$\mathbb{P}[k \text{ collisions en } i] = \binom{n}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{n-k}$$

Donc

$$X \sim \text{Bin}\left(n, \frac{1}{m}\right) = \text{Bin}\left(n, \frac{\alpha}{n}\right) \text{ avec } \alpha = \frac{n}{m}$$

Si $n > 50$ et $\alpha < 0.1$, alors on peut considérer que :

$$X \sim \text{Pois}(\alpha)$$

Recherche négative

Une recherche négative d'un élément x consiste en :

- un appel à h pour trouver la case d'indice $h(x)$,
- un parcours linéaire de la liste en $h(x)$.

Le nombre d'élément dans la liste en $h(x)$ est le nombre de collision en $h(x)$.

Recherche négative

Une recherche négative d'un élément x consiste en :

- un appel à h pour trouver la case d'indice $h(x)$,
- un parcours linéaire de la liste en $h(x)$.

Le nombre d'éléments dans la liste en $h(x)$ est le nombre de collisions en $h(x)$.

$$\begin{aligned}\mathbb{E}[\text{recherche négative}] &= 1 + \mathbb{E}[\text{nombre de collisions en } i] \\ &= \boxed{1 + \alpha}\end{aligned}$$

Exercice 1, Question 3

Notations

- x_1, x_2, \dots, x_n les n clés insérées, dans l'ordre **inverse** d'insertion.

Notations

- x_1, x_2, \dots, x_n les n clés insérées, dans l'ordre **inverse** d'insertion.
- Y_i : complexité moyenne de la recherche (positive) de x_i ,

Notations

- x_1, x_2, \dots, x_n les n clés insérées, dans l'ordre **inverse** d'insertion.
- Y_i : complexité moyenne de la recherche (positive) de x_i ,
- $X_{i,j} = \begin{cases} 1 & \text{si } h(x_i) = h(x_j) \\ 0 & \text{sinon} \end{cases}$

Exercice 1, Question 3

Notations

- x_1, x_2, \dots, x_n les n clés insérées, dans l'ordre **inverse** d'insertion.
- Y_i : complexité moyenne de la recherche (positive) de x_i ,
- $X_{i,j} = \begin{cases} 1 & \text{si } h(x_i) = h(x_j) \\ 0 & \text{sinon} \end{cases}$

On a :

$$Y_i = 1 + \sum_{j=1}^{i-1} X_{i,j}$$

Exercice 1, Question 3

Notations

- x_1, x_2, \dots, x_n les n clés insérées, dans l'ordre **inverse** d'insertion.
- Y_i : complexité moyenne de la recherche (positive) de x_i ,
- $X_{i,j} = \begin{cases} 1 & \text{si } h(x_i) = h(x_j) \\ 0 & \text{sinon} \end{cases}$

On a :

$$Y_i = 1 + \sum_{j=1}^{i-1} X_{i,j}$$

$$\begin{aligned} \mathbb{E}[\text{recherche positive}] &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n Y_i \right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \end{aligned}$$

Exercice 1, Question 3

Notations

- x_1, x_2, \dots, x_n les n clés insérées, dans l'ordre **inverse** d'insertion.
- Y_i : complexité moyenne de la recherche (positive) de x_i ,
- $X_{i,j} = \begin{cases} 1 & \text{si } h(x_i) = h(x_j) \\ 0 & \text{sinon} \end{cases}$

On a :

$$Y_i = 1 + \sum_{j=1}^{i-1} X_{i,j}$$

$$\begin{aligned} \mathbb{E}[\text{recherche positive}] &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n Y_i \right] & \mathbb{E}[Y_i] &= \mathbb{E} \left[1 + \sum_{j=1}^{i-1} X_{i,j} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] & &= 1 + \sum_{j=1}^{i-1} \mathbb{E}[X_{i,j}] \end{aligned}$$

Exercise 1, Question 3

$$\mathbb{E}[X_{i,j}] = \boxed{\frac{1}{m}}$$

Exercise 1, Question 3

$$\mathbb{E}[X_{i,j}] = \boxed{\frac{1}{m}}$$

$$\begin{aligned}\mathbb{E}[Y_i] &= \mathbb{E}\left[1 + \sum_{j=1}^{i-1} X_{i,j}\right] \\ &= 1 + \sum_{j=1}^{i-1} \mathbb{E}[X_{i,j}] \\ &= 1 + \sum_{j=1}^{i-1} \frac{1}{m} \\ &= \boxed{1 + \frac{i-1}{m}}\end{aligned}$$

Exercice 1, Question 3

$$\mathbb{E}[\text{recherche positive}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right]$$

Exercice 1, Question 3

$$\mathbb{E}[\text{recherche positive}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i]$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right)\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m}\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m} \\ &= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (i-1)\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m} \\ &= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (i-1) = 1 + \frac{1}{n \cdot m} \sum_{i=0}^{n-1} i\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m} \\ &= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (i-1) = 1 + \frac{1}{n \cdot m} \sum_{i=0}^{n-1} i \\ &= 1 + \frac{1}{n \cdot m} \frac{n(n-1)}{2}\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m} \\ &= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (i-1) = 1 + \frac{1}{n \cdot m} \sum_{i=0}^{n-1} i \\ &= 1 + \frac{1}{n \cdot m} \frac{n(n-1)}{2} = 1 + \frac{n-1}{2m}\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m} \\ &= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (i-1) = 1 + \frac{1}{n \cdot m} \sum_{i=0}^{n-1} i \\ &= 1 + \frac{1}{n \cdot m} \frac{n(n-1)}{2} = 1 + \frac{n-1}{2m} \\ &= 1 + \frac{n}{2m} - \frac{1}{2m}\end{aligned}$$

Exercice 1, Question 3

$$\begin{aligned}\mathbb{E}[\text{recherche positive}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{i-1}{m}\right) = \frac{1}{n} \sum_{i=1}^n 1 + \frac{1}{n} \sum_{i=1}^n \frac{i-1}{m} \\ &= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (i-1) = 1 + \frac{1}{n \cdot m} \sum_{i=0}^{n-1} i \\ &= 1 + \frac{1}{n \cdot m} \frac{n(n-1)}{2} = 1 + \frac{n-1}{2m} \\ &= 1 + \frac{n}{2m} - \frac{1}{2m} = \boxed{1 + \frac{\alpha}{2} - \frac{\alpha}{2n}}\end{aligned}$$

Exercice 3, Questions 1 et 2

Question 1

$$\mathbb{P}[\text{il existe une collision}] \leq \mathbb{E}[\text{nombre de collisions}]$$

Question 1

$$\begin{aligned}\mathbb{P}[\text{il existe une collision}] &\leq \mathbb{E}[\text{nombre de collisions}] \\ &\leq \binom{n}{2} \cdot \frac{1}{m}\end{aligned}$$

Avec :

- $\binom{n}{2}$ le nombre de paires de clés,
- $\frac{1}{m}$ la probabilité qu'une paire de clés donnée entre en collision.

Question 1

$$\begin{aligned}\mathbb{P}[\text{il existe une collision}] &\leq \mathbb{E}[\text{nombre de collisions}] \\ &\leq \binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2} \frac{1}{n^2}\end{aligned}$$

Avec :

- $\binom{n}{2}$ le nombre de paires de clés,
- $\frac{1}{m}$ la probabilité qu'une paire de clés donnée entre en collision.

Question 1

$$\mathbb{P}[\text{il existe une collision}] \leq \mathbb{E}[\text{nombre de collisions}]$$

$$\leq \binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2} \frac{1}{n^2}$$

$$\leq \frac{n-1}{2n}$$

Avec :

- $\binom{n}{2}$ le nombre de paires de clés,
- $\frac{1}{m}$ la probabilité qu'une paire de clés donnée entre en collision.

Question 1

$$\mathbb{P}[\text{il existe une collision}] \leq \mathbb{E}[\text{nombre de collisions}]$$

$$\leq \binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2} \frac{1}{n^2}$$

$$\leq \frac{n-1}{2n} < \frac{1}{2}$$

Avec :

- $\binom{n}{2}$ le nombre de paires de clés,
- $\frac{1}{m}$ la probabilité qu'une paire de clés donnée entre en collision.

Exercice 3, Questions 1 et 2

Question 1

$$\mathbb{P}[\text{il existe une collision}] \leq \mathbb{E}[\text{nombre de collisions}]$$

$$\leq \binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2} \frac{1}{n^2}$$

$$\leq \frac{n-1}{2n} < \frac{1}{2}$$

Avec :

- $\binom{n}{2}$ le nombre de paires de clés,
- $\frac{1}{m}$ la probabilité qu'une paire de clés donnée entre en collision.

Question 2

Pour chaque i , si n_i nombre de collisions en i , tirage aléatoire de h_i pour une table de taille n_i^2 . Si collision, on retire, etc... \rightarrow loi géométrique.

Exercice 3, Question 3

$$\mathbb{E}[S] = \mathbb{E} \left[\sum_{i=1}^n N_i^2 \right]$$

Avec :

- N_i la v.a. du nombre de collision en i ,

Exercice 3, Question 3

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E} \left[\sum_{i=1}^n N_i^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \left(N_i + 2 \cdot \binom{N_i}{2} \right) \right]\end{aligned}$$

Avec :

- N_i la v.a. du nombre de collision en i ,
- $N_i^2 = N_i + \binom{N_i}{2}$,

Exercice 3, Question 3

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E} \left[\sum_{i=1}^n N_i^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \left(N_i + 2 \cdot \binom{N_i}{2} \right) \right] \\ &= \mathbb{E}[n] + 2 \cdot \mathbb{E} \left[\sum_{i=1}^n \binom{N_i}{2} \right]\end{aligned}$$

Avec :

- N_i la v.a. du nombre de collision en i ,
- $N_i^2 = N_i + \binom{N_i}{2}$,

Exercice 3, Question 3

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}\left[\sum_{i=1}^n N_i^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n \left(N_i + 2 \cdot \binom{N_i}{2}\right)\right] \\ &= \mathbb{E}[n] + 2 \cdot \mathbb{E}\left[\sum_{i=1}^n \binom{N_i}{2}\right] \\ &= n + 2 \cdot \mathbb{E}[\text{nombre de collision}]\end{aligned}$$

Avec :

- N_i la v.a. du nombre de collision en i ,
- $N_i^2 = N_i + \binom{N_i}{2}$,

Exercice 3, Question 3

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}\left[\sum_{i=1}^n N_i^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n \left(N_i + 2 \cdot \binom{N_i}{2}\right)\right] \\ &= \mathbb{E}[n] + 2 \cdot \mathbb{E}\left[\sum_{i=1}^n \binom{N_i}{2}\right] \\ &= n + 2 \cdot \mathbb{E}[\text{nombre de collision}] \\ &= n + n - 1 < 2n\end{aligned}$$

Avec :

- N_i la v.a. du nombre de collision en i ,
- $N_i^2 = N_i + \binom{N_i}{2}$,
- $\mathbb{E}[\text{nombre de collisions}] = \frac{n-1}{2}$.

Question 4

En utilisant l'inégalité de Markov :

$$\mathbb{P}[S > 4n] \leq \frac{\mathbb{E}[S]}{4n}$$

Question 4

En utilisant l'inégalité de Markov :

$$\mathbb{P}[S > 4n] \leq \frac{\mathbb{E}[S]}{4n} < \frac{2n}{4n} = \frac{1}{2}$$

Question 4

En utilisant l'inégalité de Markov :

$$\mathbb{P}[S > 4n] \leq \frac{\mathbb{E}[S]}{4n} < \frac{2n}{4n} = \frac{1}{2}$$

Question 5

On tire une fonction h jusqu'à avoir $\sum n_i^2 \leq 4n$, puis on tire les h_i de tailles n_i^2 .

Exercice 3, Question 4, 5 et 6

Question 4

En utilisant l'inégalité de Markov :

$$\mathbb{P}[S > 4n] \leq \frac{\mathbb{E}[S]}{4n} < \frac{2n}{4n} = \frac{1}{2}$$

Question 5

On tire une fonction h jusqu'à avoir $\sum n_i^2 \leq 4n$, puis on tire les h_i de tailles n_i^2 .

Question 6

Structure de données statique : si on change une clé, il faut **tout** recommencer !

Exercice 4

Question 1

On parcourt la moitié en conservant le max des valeurs déjà vues :
complexité linéaire.

Exercice 4

Question 1

On parcourt la moitié en conservant le max des valeurs déjà vues : complexité linéaire.

Question 2

Le pire cas correspond à ne prendre que des éléments parmi la moitié la plus petite :

- une chance sur deux à chaque tirage,
- k tirages indépendants.

$$\mathbb{P}[\text{echec}] \leq \left(\frac{1}{2}\right)^k$$

Exercice 4

Question 1

On parcourt la moitié en conservant le max des valeurs déjà vues : complexité linéaire.

Question 2

Le pire cas correspond à ne prendre que des éléments parmi la moitié la plus petite :

- une chance sur deux à chaque tirage,
- k tirages indépendants.

$$\mathbb{P}[\text{echec}] \leq \left(\frac{1}{2}\right)^k$$

$$\mathbb{P}[\text{succès}] \geq 1 - \left(\frac{1}{2}\right)^k$$

Exercice 5, Questions 1,2 et 3



Figure: Choix du pivot en fin de tableau.

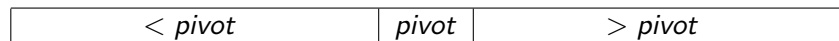


Figure: Placement des éléments par rapport au pivot.

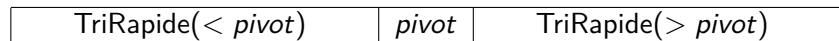


Figure: Appels récursifs.

Question 1

Le pire cas correspond au pire découpage : si le pivot est le plus petit ou le plus grand élément du tableau. Par exemple, si le tableau est déjà trié.
Complexité quadratique : $O(n^2)$.

Exercice 5, Questions 1,2 et 3

Question 1

Le pire cas correspond au pire découpage : si le pivot est le plus petit ou le plus grand élément du tableau. Par exemple, si le tableau est déjà trié. Complexité quadratique : $O(n^2)$.

Question 2

Si on compare a_i et a_j , c'est que l'un des deux est pivot. Un élément ne peut être qu'une seule fois pivot. Donc on ne peut comparer qu'une seule fois a_i et a_j .

Exercice 5, Questions 1,2 et 3

Question 1

Le pire cas correspond au pire découpage : si le pivot est le plus petit ou le plus grand élément du tableau. Par exemple, si le tableau est déjà trié. Complexité quadratique : $O(n^2)$.

Question 2

Si on compare a_i et a_j , c'est que l'un des deux est pivot. Un élément ne peut être qu'une seule fois pivot. Donc on ne peut comparer qu'une seule fois a_i et a_j .

Question 3

$$\mathbb{P}[a_i, a_j \text{ comparés}] \leq \frac{2}{j-i+1} \quad \text{avec } i < j$$

Exercice 5, Question 4

La complexité est proportionnelle au nombre de comparaisons.

Notations

- X la v.a. du nombre de comparaison,
- $X_{i,j}$ la v.a. du nombre de comparaisons entre a_i et a_j .

Exercice 5, Question 4

La complexité est proportionnelle au nombre de comparaisons.

Notations

- X la v.a. du nombre de comparaison,
- $X_{i,j}$ la v.a. du nombre de comparaisons entre a_i et a_j .

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E} \left[\sum_{i=1}^n \sum_{j=i+1}^n X_{i,j} \right] = \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{E}[X_{i,j}] \\ &= \sum_{i=1}^n \sum_{j=i+1}^n \frac{2}{j-i+1} = \sum_{i=1}^n \sum_{j=2}^n \frac{2}{j} \\ &\leq \sum_{i=1}^n 2 \log n \\ &\leq 2n \log n\end{aligned}$$