Algorithmique - Analyse du tri rapide randomisé

Le but est de montrer le théorème suivant, qui analyse le temps d'exécution moyen du tri rapide randomisé.

Théorème 1. *L'espérance du nombre de comparaisons du tri rapide randomisé d'un ensemble à n éléments est au plus $2nH_n$ où H_n est le terme général de la série harmonique.*

Démonstration. Soit T un tableau de n éléments. Pour $1 \leqslant i \leqslant n$, on note t_i le i-ième plus petit élément de T. Ainsi, $t_1 = \min T$ et $t_n = \max T$. On pose X_{ij} la variable aléatoire valant 1 si t_i et t_j sont comparés au cours de l'exécution, et 0 sinon.

On remarque alors que deux éléments t_i et t_j sont comparés au plus une fois. En effet, si t_i et t_j sont comparés , l'un des deux était un pivot et n'est comparé avec plus personne ensuite.

Ainsi, le nombre total de comparaisons est $N = \sum_{1 \leqslant i < j \leqslant n} X_{ij}$.

par linéarité de l'espérance, on a alors

$$\mathbb{E}(N) = \sum_{1 \leqslant i < j \leqslant n} \mathbb{E}(X_{ij})$$

En notant $p_{ij} = Pr[X_{ij} = 1]$, on a directement $\mathbb{E}(X_{ij}) = p_{ij}$.

Lemme 1. *Pour tout* $1 \leqslant i < j \leqslant n$, *on a* $p_{ij} = \frac{2}{j-i+1}$.

On peut voir l'exécution du tri rapide comme un arbre binaire, avec comme nœuds les pivots successifs. On note alors π la permutation de T obtenu en faisant un parcours en largeur de l'arbre obtenu.

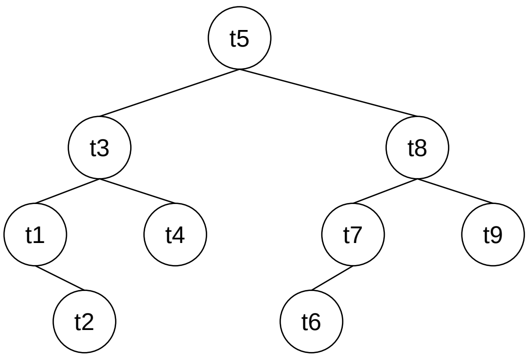

FIGURE 1 – Exemple d'exécution du tri rapide sur un tableau à 9 éléments. On a ici $\pi = (4, 8, 2, 5, 1, 9, 6, 3, 7)$.

Démonstration. On peut faire deux remarques :

1. t_i et t_j sont comparés ssi pour tout $i < l < j$, $\pi(t_i) \leqslant \pi(t_l)$ et $\pi(t_j) \leqslant \pi(t_l)$. En effet, t_i et t_j ne sont pas comparés ssi un pivot entre les deux les a séparés précédemment.
2. Chaque élément $t_i, ..., t_j$ a la même probabilité d'être le premier d'entre eux choisi pour être un pivot, et donc d'apparaître avant dans π.

Ainsi, en combinant les deux remarques, p_{ij} est exactement la probabilité que t_i ou t_j soit choisi en premier parmi $t_i, ..., t_j$ (1), et cette probabilité vaut $\frac{2}{j-i+1}$ (2). □

On peut alors revenir à notre formule,

$$
\begin{aligned}
\mathbb{E}(N) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \\
&\leqslant 2 \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{1}{k} \\
&= 2nH_n
\end{aligned}
$$

Or, on sait que $H_n \sim \ln n + o(1)$, le temps d'exécution de l'algorithme est $\mathcal{O}(n \log n)$. $\square$