

LE TRI RAPIDE

Complexité moyenne

Objectif

Il s'agit ici d'obtenir un équivalent simple de la complexité moyenne du tri rapide.

Préliminaire

On considère un espace probabilisé (Ω, \mathcal{A}, P) . Soit X et Y deux variables aléatoires définies sur cet espace et telles que :

- $X(\Omega) \subset \llbracket 0; N \rrbracket$ où N désigne un entier naturel.
- $Y(\Omega) \subset \llbracket 0; n-1 \rrbracket$ et $\forall j \in \llbracket 0; n-1 \rrbracket, P(Y = j) = \frac{1}{n}$.

Pour tout entier j dans $\llbracket 0; n-1 \rrbracket$, on désigne par $E(X | Y = j)$ l'espérance mathématique de la variable aléatoire X pour la probabilité conditionnelle $P_{(Y=j)}$. $E(X | Y = j)$ est donc l'espérance de X dans l'espace probabilisé $(\Omega, \mathcal{A}, P_{(Y=j)})$.

1. Calculer $E(Y)$.

$$\text{On a : } E(Y) = \sum_{j=0}^{n-1} j \times P(Y = j) = \sum_{j=1}^{n-1} j \times \frac{1}{n} = \frac{1}{n} \sum_{j=1}^{n-1} j = \frac{1}{n} \times \frac{n(n-1)}{2} = \frac{n-1}{2}.$$

$$E(Y) = \frac{n-1}{2}$$

2. Montrer que $E(X) = \frac{1}{n} \sum_{j=0}^{n-1} E(X | Y = j)$.

On a :

$$\begin{aligned}
 E(X) &= \sum_{i=0}^N i \times P(X=i) = \sum_{i=0}^N i \times \left(\sum_{j=0}^{n-1} P_{(Y=j)}(X=i) \times P(Y=j) \right) \\
 &= \sum_{i=0}^N i \times \left(\sum_{j=0}^{n-1} P_{(Y=j)}(X=i) \times \frac{1}{n} \right) = \frac{1}{n} \times \sum_{i=0}^N i \times \left(\sum_{j=0}^{n-1} P_{(Y=j)}(X=i) \right) \\
 &= \frac{1}{n} \times \sum_{i=0}^N \left(\sum_{j=0}^{n-1} i \times P_{(Y=j)}(X=i) \right) = \frac{1}{n} \times \sum_{j=0}^{n-1} \left(\sum_{i=0}^N i \times P_{(Y=j)}(X=i) \right) \\
 &= \frac{1}{n} \times \sum_{j=0}^{n-1} E(X | Y=j)
 \end{aligned}$$

Le résultat est ainsi établi.

$$E(X) = \frac{1}{n} \times \sum_{j=0}^{n-1} E(X | Y=j)$$

Calcul de la complexité moyenne

Soit à trier une liste $L = [l_1 l_2 l_3 \dots l_n]$ uniformément choisie dans l'ensemble des $n!$ permutations de l'ensemble $\llbracket 1; n \rrbracket$.

Notons C_n la variable aléatoire correspondant au nombre de comparaisons requises pour trier cette liste à l'aide de l'algorithme du tri rapide.

On cherche $E(C_n)$.

Pour une liste donnée, on utilise comme pivot le premier élément de cette liste.

On note G_n (respectivement D_n) la longueur de la liste de gauche L_n^G (respectivement droite L_n^D) obtenue à l'issue de la première étape du tri de la liste L et on pose $R_n = C_n - (n-1)$ correspondant aux nombres de comparaisons à effectuer après la première étape.

3. Déterminer la loi de G_n .

Nous effectuons $n-1$ comparaisons lors de la première étape. La variable aléatoire G_n peut donc prendre toutes les valeurs de l'ensemble $\llbracket 0; n-1 \rrbracket$ et toutes ces valeurs sont équiprobables. En effet, parmi les entiers de la liste L , il y en a $l_1 - 1$ qui sont inférieurs à l_1 . Comme l_1 prend uniformément ses valeurs dans $\llbracket 1; n \rrbracket$, on en déduit que G_n prend uniformément ses valeurs dans $\llbracket 0; n-1 \rrbracket$.

$$G_n \text{ suit la loi uniforme sur } \llbracket 0; n-1 \rrbracket.$$

4. Montrer que $\forall n \geq 2, \forall j \in \llbracket 0; n-1 \rrbracket, E(R_n | G_n = j) = E(C_j + C_{n-j-1})$

Si la liste de gauche L_n^G comporte j éléments ($G_n = j$) alors la liste de droite L_n^D en comportera $n-1-j$. Ainsi, le nombre de comparaison restant à effectuer, R_n est égal à $C_j + C_{n-j-1}$. On a donc, sachant que $G_n = j : R_n = C_j + C_{n-j-1}$ et, en considérant les espérances :

$$E(R_n | G_n = j) = E(C_j + C_{n-j-1})$$

5. Montrer que l'on a : $\forall n \geq 2, E(C_n) = n-1 + \frac{2}{n} \sum_{j=0}^{n-1} E(C_j)$

En tenant compte du résultat préliminaire, le résultat précédent donne :

$$\begin{aligned} E(R_n) &= \frac{1}{n} \sum_{j=0}^{n-1} E(R_n | G_n = j) \\ &= \frac{1}{n} \sum_{j=0}^{n-1} E(C_j + C_{n-j-1}) = \frac{1}{n} \sum_{j=0}^{n-1} [E(C_j) + E(C_{n-j-1})] \\ &= \frac{1}{n} \left[\sum_{j=0}^{n-1} E(C_j) + \sum_{j=0}^{n-1} E(C_{n-j-1}) \right] = \frac{1}{n} \left[\sum_{j=0}^{n-1} E(C_j) + \sum_{j=0}^{n-1} E(C_j) \right] \\ &= \frac{2}{n} \sum_{j=0}^{n-1} E(C_j) \end{aligned}$$

Or, $R_n = C_n - (n-1)$, d'où : $E(R_n) = E(C_n) - (n-1)$ puis $E(C_n) = n-1 + E(R_n)$ et, finalement :

$$E(C_n) = n-1 + \frac{2}{n} \sum_{j=0}^{n-1} E(C_j)$$

6. Montrer que l'on a : $\forall n \geq 2, E(C_{n+1}) = \frac{2n}{n+1} + \frac{n+2}{n+1} E(C_n)$

Poser alors $X_n = C_n - 2$ et en déduire : $\forall n \geq 2, E(X_{n+1}) = 2 + \frac{n+2}{n+1} E(X_n)$.

Multiplions l'égalité obtenue précédemment par n : $nE(C_n) = n(n-1) + 2 \sum_{j=0}^{n-1} E(C_j)$.

A l'ordre $n+1$: $(n+1)E(C_{n+1}) = n(n+1) + 2 \sum_{j=0}^n E(C_j) = n(n+1) + 2 \sum_{j=0}^{n-1} E(C_j) + 2E(C_n)$.

En soustrayant alors membre à membre ces deux égalités, on obtient :

$$\begin{aligned}
 (n+1)E(C_{n+1}) - nE(C_n) &= n(n+1) + 2\sum_{j=0}^{n-1} E(C_j) + 2E(C_n) - \left\{ n(n-1) + 2\sum_{j=0}^{n-1} E(C_j) \right\} \\
 \Leftrightarrow (n+1)E(C_{n+1}) &= 2n + (n+2)E(C_n) \\
 \Leftrightarrow E(C_{n+1}) &= \frac{2n}{n+1} + \frac{n+2}{n+1}E(C_n)
 \end{aligned}$$

Posons alors $X_n = C_n - 2$.

On a :

$$\begin{aligned}
 E(X_{n+1}) &= E(C_{n+1}) - 2 \\
 &= \frac{2n}{n+1} + \frac{n+2}{n+1}E(C_n) - 2 \\
 &= \frac{2n}{n+1} + \frac{n+2}{n+1}(E(X_n) + 2) - 2 \\
 &= \frac{-2}{n+1} + \frac{2(n+2)}{n+1} + \frac{n+2}{n+1}E(X_n) \\
 &= 2 + \frac{n+2}{n+1}E(X_n)
 \end{aligned}$$

7. Démontrer alors par récurrence que l'on a : $\forall n \geq 2, E(X_n) = 2(n+1)(S_{n+1} - 2)$

où, pour tout entier naturel n non nul : $S_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$.

Pour $n = 2$, on a : $E(X_n) = E(X_2) = E(C_2 - 2) = E(C_2) - 2 = 1 - 2 = -1$.

Par ailleurs :

$$2(n+1)(S_{n+1} - 2) = 2(2+1)(S_3 - 2) = 6\left(1 + \frac{1}{2} + \frac{1}{3} - 2\right) = 6\left(\frac{11}{6} - 2\right) = 6 \times \left(-\frac{1}{6}\right) = -1$$

L'égalité est donc vraie pour $n = 2$.

Supposons maintenant qu'elle soit vraie pour un rang n supérieur ou égal à 2.

On a : $E(X_n) = 2(n+1)(S_{n+1} - 2)$.

Il vient alors, en utilisant le deuxième résultat de la question précédente :

$$\begin{aligned}
 E(X_{n+1}) &= 2 + \frac{n+2}{n+1} E(X_n) \\
 &= 2 + \frac{n+2}{n+1} \times 2(n+1)(S_{n+1} - 2) \\
 &= 2 + 2(n+2)(S_{n+1} - 2) \\
 &= 2(n+2) \left[\frac{1}{n+2} + (S_{n+1} - 2) \right] \\
 &= 2(n+2)(S_{n+2} - 2)
 \end{aligned}$$

La propriété est donc vraie au rang $n+1$.

On en déduit finalement que la propriété est vraie pour tout entier naturel n supérieur ou égal à 2.

$$\boxed{\forall n \geq 2, E(X_n) = 2(n+1)(S_{n+1} - 2)}$$

8. Dédurre du résultat précédent un équivalent simple de $E(C_n)$ en $+\infty$.

On a classiquement : $S_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \sim \ln n$.

D'où : $S_{n+1} = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} + \frac{1}{n+1} \sim \ln(n+1) \underset{+\infty}{\sim} \ln n$ et $S_{n+1} - 2 \underset{+\infty}{\sim} \ln n$.

Il vient alors, en utilisant le résultat précédent et en tenant compte de $n+1 \underset{+\infty}{\sim} n$:

$$E(X_n) = 2(n+1)(S_{n+1} - 2) \underset{+\infty}{\sim} 2n \ln n.$$

Comme $X_n = C_n - 2$, on a : $E(C_n) = E(X_n) + 2$ puis $E(C_n) \underset{+\infty}{\sim} E(X_n)$ et, finalement :

$$\boxed{E(C_n) \underset{+\infty}{\sim} 2n \ln n}$$