Quicksort and Analysis:


$qsort(A, i, j)$

if $i < j$

1. Select a random element $p$ in $A[i..j]$

2. Rearrange $A[i..j]$ so that

\[
\begin{array}{c|c|c}
 i & p & j \\
\hline
< & > & > \\
\end{array}
\]

3. $qsort(A, i, k-1)$

4. $qsort(A, k+1, j)$

Call initially with $qsort(A, 1, n)$.

It is easy to see that this algorithm sorts $A$.

Run Time

Worst-Case $\Rightarrow O(n^2)$  [Homework]

Expected $\Rightarrow O(n \log n)$

We will analyze the expected run-time.
First note that if we can estimate the total number of comparisons in \textsc{qsort}, we can bound the running time - as running time is $O(\# \text{ comparisons})$.

\textbf{Question:} What is the expected number of comparisons \textsc{qsort} makes?

Let the sorted order of $A[1..n]$ be

$x_1 < x_2 < x_3 \ldots < x_n$.

(Note that $(x_1, x_2, \ldots, x_n)$ is a permutation of elements in $A$.)

Define random variables $X_{ij}$ as follows:

$$X_{ij} = \begin{cases} 1 & \text{if } x_i \text{ and } x_j \text{ are compared by } \textsc{qsort} \\ 0 & \text{if } x_i \text{ and } x_j \text{ are never compared by } \textsc{qsort} \end{cases}$$

Note that

$$E[X_{ij}] = \Pr(x_i \text{ and } x_j \text{ are compared by } \textsc{qsort}) + 0 \cdot \Pr(x_i \text{ and } x_j \text{ are not compared by } \textsc{qsort})$$

$$= \Pr(x_i \text{ and } x_j \text{ are compared by } \textsc{qsort})$$

Define $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$ = total number of comparisons made by \textsc{qsort}.
Thus \[ E(x) = E\left( \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_{ij} \right) \]
\[ = \sum_{i=1}^{n} \sum_{j=i+1}^{n} E(x_{ij}). \]

When is \( x_{ij} = 1 \)?

Consider the set of elements \( S_{ij} = \{x_i, x_{i+1}, \ldots, x_j\} \).

If in qsort execution, the first pivot chosen from the set \( S_{ij} \) is either \( x_i \) or \( x_j \), then \( x_i \) and \( x_j \) will be compared and \( x_{ij} = 1 \).

If the first pivot chosen from \( S_{ij} \) is neither \( x_i \) nor \( x_j \), then they will be in separate subarrays and hence will never be compared to each other.

What is the probability of choosing \( \overline{x}_i \) or \( \overline{x}_j \) as the first pivot in \( S_{ij} \)?

Since pivots are chosen uniformly at random, any element of \( S_{ij} \) can be a pivot is equally likely.

Thus \[ \Pr(\text{\( x_i \) or \( x_j \) is a chosen pivot in } S_{ij}) = \frac{2}{j-i+1} \]
Therefore
\[ E(X_{ij}) = \frac{2}{j-i+1} \]

and
\[ E(x) = \sum_{i=1}^{n-1} \sum_{j'=i+1}^{n} \frac{2}{j'-i+1} = 2n \ln n + \Theta(n) \]

Let us see why?

Set \( k = j-i+1 \)

then
\[ E(x) = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \]

Note that
\[ \sum_{k=2}^{n-i+1} \frac{2}{k} = \sum_{j=i+1}^{n} \frac{2}{j-i+1} \]

\[ = \sum_{k=2}^{n} \frac{2}{k} \]

\[ = \sum_{k=2}^{n} \frac{2(n+1-k)}{k} \]

\[ = (n+1) \sum_{k=2}^{n} \frac{2}{k} - 2(n-1) \]

\[ = (n+1) \sum_{k=1}^{n} \frac{2}{k} - 2(n+1) - 2(n-1) \]
\[(n+1) \sum_{k=1}^{n} \frac{1}{k} - 4n \]

\[= (2n+1) \sum_{k=1}^{n} \frac{1}{k} - 4n \]

\[\rightarrow \text{Harmonic Number} \sim \ln n + \Theta(1)\]

\[= (2n+1) \ln n + (2n+1) \Theta(n) - 4n\]

\[E(x) = 2n \ln n + \Theta(n)\]

\[\blacksquare\]