Randomized Quicksort

In the randomized version of quicksort, we pick a pivot uniformly at random from all possibilities. We will now show that the expected number of comparisons made in randomized quicksort is equal to $2n \ln n + O(n) = \Theta(n \log n)$.

**Theorem 1.** For any input array of size $n$, the expected number of comparisons made by randomized quicksort is

$$2n \ln n + O(n) = \Theta(n \log n).$$

**Proof.** Let $y_1, y_2, \ldots, y_n$ be the elements in the input array $A$ in sorted order. Let $X$ be the random variable denoting the total number of pair-wise comparisons made between elements of $A$. Let $X_{i,j}$ be the random variable denoting the total number of times elements $y_i$ and $y_j$ are compared during the algorithm.

We make the following observations:

- Comparisons between elements in the input array are done only in the function $\text{Partition}$.
- There are $n - 1$ distinct pivots chosen over the course of the algorithm, and hence $n - 1$ calls to $\text{Partition}$.
- Two elements are compared if and only if one of them is a pivot.

Let $X_{i,j}^k$ be an indicator random variable that is 1 if and only if elements $y_i$ and $y_j$ are compared in the $k$-th call to $\text{Partition}$. Then we have:

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j} \quad \text{and} \quad X_{i,j} = \sum_{k=1}^{n-1} X_{i,j}^k$$

We will now calculate $E[X_{i,j}]$. By the linearity of expectation, we have

$$E[X_{ij}] = \sum_{k=1}^{n-1} E[X_{ij}^k] = \sum_{k=1}^{n-1} \Pr[X_{ij}^k = 1]$$

Let $t$ be the iteration of the first call to $\text{Partition}$ during which one of the elements from $y_i, y_{i+1}, \ldots, y_j$ is used as the pivot. From our observations above, note that for all times before $t$, $y_i$ and $y_j$ are never compared, so $X_{i,j}^k = 0$ for all $k < t$. If one of $y_i$ or $y_j$ is chosen as the $t$-th pivot, then $X_{i,j}^t = 1$, otherwise $X_{i,j}^t = 0$ and $y_i$ and $y_j$ will be separated into different sublists and hence will never be compared again. Hence $X_{i,j}^k = 0$ for all $k > t$. 


Now, since there are \( j - i + 1 \) elements in the list \( y_i, y_{i+1}, ..., y_j \), and the pivot is chosen randomly, the probability that one of \( y_i \) or \( y_j \) is chosen as the pivot is \( \frac{2}{j - i + 1} \). Hence:

\[
E[X_{i,j}] = \Pr[X_{i,j}^t = 1] = \frac{2}{j - i + 1}
\]

Now we use this result and apply the linearity of expectation to get:

\[
E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{i,j}]
\]

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

\[
= \sum_{k=2}^{n} \frac{2}{k} (n - k + 1) \quad \text{(see the note below)}
\]

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

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

\[
= 2(n + 1) (\ln n + c) - 4n \quad \text{where } 0 \leq c < 1
\]

\[
= 2n \ln n + O(n)
\]

Here we have used the fact that the harmonic function, \( H(n) = \sum_{k=1}^{n} \frac{1}{k} \) is at most \( \ln n + c \) for some constant \( 0 \leq c < 1 \).

Also, the third equality follows by expanding the sum as

\[
\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j - i + 1} = \frac{2}{2} + \frac{2}{3} + \ldots + \frac{2}{n-2} + \frac{2}{n-1} + \frac{2}{n}
\]

\[
+ \frac{2}{2} + \frac{2}{3} + \ldots + \frac{2}{n-2} + \frac{2}{n-1}
\]

\[
+ \frac{2}{2} + \frac{2}{3} + \ldots + \frac{2}{n-2}
\]

\[
+ \frac{2}{2}
\]

and grouping the columns together.