ECS 122B Spring 2019 - Randomized Algorithm Analysis

Eric Li

May 16, 2019

This analysis adapts the idea in Introduction to Algorithms, third edition, chapter 7.4.2.

Recall the selection algorithm. The non-base case is first partitioning the array, then call RandomizedSelect on the partition of interest.

The only time the algorithm compares is when partitioning the array, and similar to Lemma 7.1, we can see that the running time of RandomizedSelect is $O(n+X)$, where X is the number of comparisons by the Partition algorithm.

Suppose we can sort the input array as $z_1, z_2, ..., z_n$, where z_1 is the smallest. Also, for simplicity, assume that $\forall i \neq j, z_i \neq z_j$.

From similar reason for Quicksort, the time z_i and z_j are compared is at most 1. Let $X_{ij} = I\{z_i \text{ is compared to } z_j\}$. There is

$$X = \sum_{i=0}^{n} \sum_{j=i+1}^{n} X_{ij}$$

And since they are independent from each other,

$$E[X] = E\left[\sum_{i=0}^{n} \sum_{j=i+1}^{n} X_{ij}\right] = \sum_{i=0}^{n} \sum_{j=i+1}^{n} Pr\{z_i \text{ is compared to } z_j\}$$

So we need to compute $Pr\{z_i \text{ is compared to } z_j\}$. The difference to Quicksort is that this time we have another number k, which is the number of element in the list that we want to find.

There are two cases. When $i \leq k \leq j$ (Note that there is $i < j$), z_i and z_j get compared only when z_i or z_j is the first vertex chosen among $z_i, z_{i+1}, ..., z_j$. There are $j - i + 1$ vertices in this list, so $Pr\{z_i \text{ is compared to } z_j\} = \dfrac{2}{j-i+1}$.

The second case is when k is not between i and j. Due to symmetricity, we only discuss the case where $k < i < j$ (case 2.a). Among $z_k, z_{k+1}, ..., z_i, z_{i+1}, ..., z_j$, if one of them other than z_i and z_j is chosen, z_i and z_j will not be compared. The reason is that if a number between k and i is chosen, the partition z_i and z_j belong to will not be recursed on; if a number between i and j is chosen, z_i and z_j will belong to different partitions. Now, there are $j - k + 1$ numbers to choose from, and $Pr\{z_i \text{ is compared to } z_j\} = \dfrac{2}{j-k+1}$.

Symmetrically, when $i < j < k$ (case 2.b), $Pr\{z_i \text{ is compared to } z_j\} = \dfrac{2}{k-i+1}$.

Note that when $i = k$ or $j = k$, the first case is equivalent to the second case (2.a and 2.b, respectively).

So

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{n} \sum_{j=i+1}^{n} Pr\{z_i \text{ cmp. to } z_j\} \\
&= \sum_{i=0}^{k-1} \sum_{j=i+1}^{n} Pr\{z_i \text{ cmp. to } z_j\} + \sum_{i=k}^{n} \sum_{j=i+1}^{n} Pr\{z_i \text{ cmp. to } z_j\} \\
&= \sum_{i=0}^{k-1} \left(\sum_{j=i+1}^{k} Pr\{z_i \text{ cmp. to } z_j\} + \sum_{j=k+1}^{n} Pr\{z_i \text{ cmp. to } z_j\} \right) + \sum_{i=k}^{n} \sum_{j=i+1}^{n} \frac{2}{j-k+1} \\
&= \sum_{i=0}^{k-1} \left(\sum_{j=i+1}^{k} \frac{2}{k-i+1} + \sum_{j=k+1}^{n} \frac{2}{j-i+1} \right) + \sum_{i=k}^{n} \sum_{j=i+1}^{n} \frac{2}{j-k+1} \\
&= \sum_{i=0}^{k-1} \sum_{j=i+1}^{k} \frac{2}{k-i+1} + \sum_{i=0}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} + \sum_{i=k}^{n} \sum_{j=i+1}^{n} \frac{2}{j-k+1} \\
&= \sum_{i=0}^{k-1} \frac{2\,(k-i)}{k-i+1} + \sum_{i=0}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} + \sum_{j=k+1}^{n} \sum_{i=k}^{j-1} \frac{2}{j-k+1} \\
&= \sum_{i=0}^{k-1} \frac{2\,(k-i)}{k-i+1} + \sum_{i=0}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} + \sum_{j=k+1}^{n} \frac{2\,(j-k)}{j-k+1} \\
&\leq \sum_{i=0}^{k-1} \frac{2\,(k-i+1)}{k-i+1} + \sum_{i=0}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} + \sum_{j=k+1}^{n} \frac{2\,(j-k+1)}{j-k+1} \\
&= \sum_{i=0}^{k-1} 2 + \sum_{i=0}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} + \sum_{j=k+1}^{n} 2 \\
&= 2\,n + \sum_{i=0}^{k-1} \sum_{j=k+1}^{n} \frac{2}{j-i+1} \\
&= 2\,n + \sum_{i=0}^{k-1} \sum_{j=1}^{n-k} \frac{2}{j-k-i+1} \\
&= 2\,n + \sum_{i=-k}^{-1} \sum_{j=1}^{n-k} \frac{2}{j-i+1} \\
&= 2\,n + \sum_{i=1}^{k} \sum_{j=1}^{n-k} \frac{2}{j+i+1} \\
&= 2\,n + 2\,\sum_{i=1}^{k} \sum_{j=1}^{n-k} \frac{1}{j+i+1}
\end{aligned}
$$

Now, to prove that $X = O(n)$, we only need to prove that

$$A = \sum_{i=1}^{k} \sum_{j=1}^{n-k} \frac{1}{j+i+1} = O(n)$$

Due to symmetricity, we can assume that $k \leq n/2$, which implies $k \leq n-k$. The summation becomes

$$
\begin{aligned}
A &= \sum_{i=1}^{k} \sum_{j=1}^{n-k} \frac{1}{j+i+1} \\
&= \sum_{i=1}^{k} \sum_{j=1+i}^{n-k+i} \frac{1}{j+1} \\
&= \sum_{i=1}^{k} \left(\sum_{j=1+i}^{k} \frac{1}{j+1} + \sum_{j=k}^{n-k+1} \frac{1}{j+1} + \sum_{j=n-k+1}^{n-k+i} \frac{1}{j+1} \right) \\
&= \sum_{i=1}^{k} \sum_{j=1+i}^{k} \frac{1}{j+1} + \sum_{i=1}^{k} \sum_{j=k}^{n-k+1} \frac{1}{j+1} + \sum_{i=1}^{k} \sum_{j=n-k+1}^{n-k+i} \frac{1}{j+1} \\
&= \sum_{j=2}^{k} \sum_{i=1}^{j-1} \frac{1}{j+1} + \sum_{j=k}^{n-k+1} \frac{k}{j+1} + \sum_{j=n-k+1}^{n-k+k} \sum_{i=j-n+k}^{k} \frac{1}{j+1} \\
&= \sum_{j=2}^{k} \frac{j-1}{j+1} + \sum_{j=k}^{n-k+1} \frac{k}{j+1} + \sum_{j=n-k+1}^{n} \frac{n-j+1}{j+1} \\
&\leq \sum_{j=2}^{k} 1 + \sum_{j=k}^{n-k+1} 1 + \sum_{j=n-k+1}^{n} 1 \\
&\leq \sum_{j=2}^{n} 1 \\
&\leq n
\end{aligned}
$$

So $A = O(n)$.
So $E[X] = O(n)$.
So the expected running time of RandomizedSelect is $O(n + X) = O(n)$.