<u>Average Case Analysis of QuickSelect</u>

Let T(n) represent the average amount of time it takes QuickSelect to run. The very first step is always the partition, which takes O(n) time. We'll simplify this and simply call this n.

Next, our two sides may split in one of n ways:

Left Side (number of items)	Right Side (number of items)
0	n-1
1	n-2
2	n-3
…	…
n-1	0

Each of these splits occurs with probability 1/n.

BUT, once we have the split, the probability of going to the left side vs. the probability of going to the right side. The value for which we are looking for can be anyone of n values. The chance that the value is on the side with k values is simply k/n. (This means the chance that the value we are looking for is chosen is 1/n. In this case, we have no more work left. This leads us to the following equation for T(n):

$$T(n) = n + \frac{1}{n}\sum_{i=0}^{n-1}(\frac{i}{n}T(i) + \frac{n-1-i}{n}T(n-1-i) + \frac{1}{n}(0))$$

We first notice that each term in the sum appears twice (and that we can drop the 0 term). This simplifies our equation to:

$$T(n) = n + \frac{1}{n}\sum_{i=1}^{n-1}(\frac{2i}{n}T(i))$$

Multiply this equation through by n^2 to yield the following

$$n^2T(n) = n^3 + 2\sum_{i=1}^{n-1}(iT(i))$$

Now, plug in n-1 in this equation to yield:

$$(n-1)^2T(n-1) = (n-1)^3 + 2\sum_{i=1}^{n-2}(iT(i))$$

Subtracting equations gives us:

$$n^2T(n) - (n-1)^2T(n-1) = 3n^2 - 3n + 1 + 2(n-1)T(n-1)$$

By combining T(n-1) terms we get:

$$n^2T(n) = (n^2 - 1)T(n-1) + 3n^2 - 3n + 1$$

Divide this equation by n(n+1):

$$\frac{n^2T(n)}{n(n+1)} = \frac{(n+1)(n-1)T(n-1)}{n(n+1)} + \frac{3n^2 - 3n + 1}{n(n+1)}$$

Simplifying, we get:

$$\frac{nT(n)}{(n+1)} = \frac{(n-1)T(n-1)}{n} + \frac{3n(n-1)}{n(n+1)} + \frac{1}{n(n+1)}$$

At this point, let $S(n) = \frac{nT(n)}{n+1}$:

$$S(n) = S(n-1) + \frac{3(n-1)}{(n+1)} + \frac{1}{n} - \frac{1}{n+1}$$

Due to the form of this recurrence relation, we can represent S(n) as a sum:

$$S(n) = \sum_{i=1}^{n} (\frac{3(i-1)}{(i+1)} + \frac{1}{i} - \frac{1}{i+1})$$

We can rewrite the fraction (i – 1)/(i + 1) as 1 – 2/(i+1), giving us the following:

$$S(n) = \sum_{i=1}^{n} (3(1 - \frac{2}{i+1}) + \frac{1}{i} - \frac{1}{i+1})$$

The latter two terms form a telescoping sum while the first two can be handled fairly easily:

$$S(n) = (\sum_{i=1}^{n} 3 - \sum_{i=1}^{n} \frac{6}{i+1}) + 1 - \frac{1}{n+1}$$

Using the simplification that the second sum is roughly equal to ln n, we get the following:

$$S(n) = 3n - 6\ln n + 1 - \frac{1}{n+1}$$

Now, since $S(n) = \frac{nT(n)}{n+1}$, it follows that $T(n) = \frac{(n+1)S(n)}{n}$. Solving for T(n), we get:

$$T(n) = 3(n+1) - \frac{6(n+1)lnn}{n} + 1$$

In a slightly simplified form, we have

$$T(n) = 3n + 4 - \frac{6(n+1)lnn}{n}$$

Note: This number represents the sum of the sizes of the arrays encountered in the set of recursive calls to solve the QuickSelect problem on an array of size n, on average.

It follows that this sum is roughly three times the size of the original input array to the problem, on average.