QUICK SORT - AVERAGE COMPLEXITY

JACEK CICHOŃ

ABSTRACT. This note constains an analysis of avarage complexity of the Quick Sort algorithm.

1. INTRODUCTION

The Quick Sort algorithm, developed by C. A. R. Hoare, is one of the most popular sorting algorithm. Let us recall its psudo-code:

```
1: QS(Tab,left,right)
2: if left < right then
3:     select a PivotIndex
4:     PivotNewIndex := partition(Tab, left, right, PivotIndex)
5:     QS(Tab, left, pivotNewIndex - 1)
6:     QS(Tab, pivotNewIndex + 1, right)
7: end if
```

We shall discuss in this the average number of comparisons over all permutations of the input sequence. We assume that an input consists of pairwise different element and we assume that all permutations of the input sequence have equal probability. Let Q_n denotes the average number of comparison used by Quick Sort when the input is an array of size n. Clearly $Q_0 = Q_1 = 0$. Suppose hence that $n \geq 2$. After a selection of the pivot element we need to compare the pivot with all remaining elements (line 4 of the psedo-code) - this needs $n-1$ comparisons. The pivot element divides the original array into two parts: left and right. The left part can be of any size from 0 to $n-1$. If the left part has a size k then the part has size $n-1-k$. Moreover, each $k \in \{0, \ldots, n-1\}$ are equiprobable. This leads to the following equality:

$$Q_n = (n-1) + \frac{1}{n}\sum_{k=0}^{n-1}(Q_k + Q_{n-1-k})$$

But $\sum_{k=0}^{n-1} Q_{n-1-k} = \sum_{k=0}^{n-1} Q_k$, so we may rewrite te last equation into the following more compact and readible form:

$$Q_n = (n-1) + \frac{2}{n}\sum_{k=0}^{n-1} Q_k \tag{1}$$

This equation is called the **Quick Sort Equation**.

2. Solvution of the Quick Sort Equation

If we multiply both sides of the Equation 1 by n then we get

$$nQ_n = n(n-1) + 2\sum_{k=0}^{n-1} Q_k\ . \tag{2}$$

Let us assume that $n \geq 2$ and let us substitute n by $n-1$. Then we get

$$(n-1)Q_{n-1} = (n-1)(n-2) + 2\sum_{k=0}^{n-2} Q_k\ . \tag{3}$$

If we substract Equation 3 from Equation 2 side by side then we obtain

$$nQ_n - (n-1)Q_{n-1} = n(n-1) - (n-1)(n-2) + 2Q_{n-1}$$

and after simplification we get

$$nQ_n - (n-1)Q_{n-1} = 2(n-1) + 2Q_{n-1}$$

or

$$nQ_n = 2(n-1) + (n+1)Q_{n-1}$$

Let us divide both sides of this equation by $n(n+1)$. Then we obtain

$$\frac{Q_n}{n+1} = 2\frac{n-1}{n(n+1)} + \frac{Q_{n-1}}{n}\ .$$

Let us introduce a new variable $F_n = \frac{Q_n}{n+1}$. Then the last equation obtains a form

$$F_n = 2\frac{n-1}{n(n+1)} + F_{n-1}\ .$$

This immediately gives us

$$F_n = 2\sum_{k=1}^{n} \frac{k-1}{k(k+1)}$$

Notice that

$$\frac{k-1}{k(k+1)} = \frac{2}{1+k} - \frac{1}{k}\ .$$

Hence

$$F_n = 4\sum_{k=1}^{n} \frac{1}{1+k} - 2\sum_{k=1}^{n} \frac{1}{k}\,.$$

Let us recall that the nth Harmonic number H_n is defined as

$$H_n = \sum_{k=1}^{n} \frac{1}{k}\,.$$

Using Harmonic numbers we can rewrite the equation for F_n as follows:

$$F_n = 4(H_{n+1} - 1) - 2H_n = 4H_{n+1} - 2H_n - 4\,.$$

Therefore we obtained the solution of Quick Sort Equation in terms of Harmonic numbers:

$$Q_n = (n+1)(4H_{n+1} - 2H_n - 4) \tag{4}$$

Shown solution is very tricky. I do not know its author. Another solutions that I know of the Quick Sort Equation (based for example on ordinary generating function) are more tedious. Using the trick shown above we may solve a series of variants of this problem and investigate other parameters of the Quick Sort algoritm.

3. Asymptotics

To obtain an asymptotic of the terms Q_n we shall use the well known asymptotic of harmonic number:

$$H_n = \ln n + \gamma + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)\,,$$

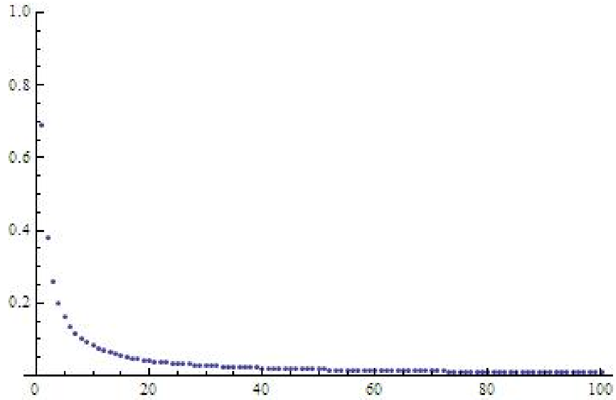
where $\gamma = 0.577216...$ is the Euler-Mascheroni constant. Putting this formula into equation 4 we get

$$\frac{Q_n}{n+1} = 4\left(\ln(n+1) + \gamma + \frac{1}{2(n+1)} + O(\frac{1}{n^2})\right) - 2\left(\ln n + \gamma + \frac{1}{2n} + O(\frac{1}{n^2})\right) - 4$$

Using equalities $\frac{1}{n+1} = \frac{1}{n} + O(\frac{1}{n^2})$ and $\ln(n+1) = \ln n + \frac{1}{n} + O(\frac{1}{n^2})$ after some easy transformations we finally get

$$Q_n = 2n(\ln n + \gamma - 2) + 2\ln n + 2\gamma + 1 + O(\tfrac{1}{n})$$

This approximation is very precise even for small numbers n In the followin picture we show the difference between numbers Q_n and the approximation $Q_n^* = 2n(\ln n + \gamma - 2) + 2\ln n + 2\gamma + 1$



4. REMARKS

If we compare the asymptotics of numbers Q_n with numbers

$$\log_2(n!) \sim 0.721348 \cdot n \cdot \ln n$$

then we see that average number of comparison in Quick Sort algorithm is quite close to the lower theoretical bound.