Average Case Analysis of Quicksort

Javed Aslam, Cheng Li, Virgil Pavlu

We assume that all elements are equally likely to be chosen as the pivot element in PARTITION. When partitioning an array of size n into two subarrays, we have the following possible sizes of the subarrays.

left array	right array
$n-1$	0
$n-2$	1
$n-3$	2
.	.
.	.
.	.
1	$n-2$
0	$n-1$

So we have

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

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

$$= \frac{2}{n}\sum_{k=1}^{n-1}T(k) + n \quad \text{(no recursive call for 0-sized subproblem)} \tag{3}$$

Replacing n with $n-1$, we get

$$T(n-1) = \frac{2}{n-1}\sum_{k=1}^{n-2}T(k) + n - 1 \tag{4}$$

Multiplying (3) by n and (4) by $n-1$:

$$nT(n) = 2\sum_{k=1}^{n-1}T(k) + n^2 \tag{5}$$

$$(n-1)T(n-1) = 2\sum_{k=1}^{n-2}T(k) + (n-1)^2 \tag{6}$$

Subtracting (6) from (5), we get

$$nT(n) - (n-1)T(n-1) = 2T(n-1) + 2n - 1 \tag{7}$$

$$\Longrightarrow nT(n) = (n+1)T(n-1) + 2n - 1 \tag{8}$$

$$\Longrightarrow T(n) = \frac{n+1}{n}T(n-1) + 2 - \frac{1}{n}. \tag{9}$$

Ignoring the $1/n$ term in (9), we have

$$T(n) \leq \frac{n+1}{n}T(n-1) + 2 \tag{10}$$

Method 1: We solve (10) by iteration:

$$
\begin{aligned}
T(n) &\le 2 + \frac{n+1}{n}T(n-1) \\
&\le 2 + \frac{n+1}{n}(2 + \frac{n}{n-1}T(n-2)) \\
&= 2 + 2\frac{n+1}{n} + \frac{n+1}{n-1} + T(n-2) \\
&\le 2 + 2\frac{n+1}{n} + \frac{n+1}{n-1}(2 + \frac{n-1}{n-2}T(n-3)) \\
&= 2 + 2\frac{n+1}{n} + 2\frac{n+1}{n-1} + \frac{n+1}{n-2}T(n-3) \\
&= 2\frac{n+1}{n+1} + 2\frac{n+1}{n} + 2\frac{n+1}{n-1} + \frac{n+1}{n-2}T(n-3) \\
&. \\
&. \\
&. \\
&= 2(n+1)\sum_{i=0}^{k-1}\frac{1}{(n+1)-i} + \frac{n+1}{n-(k-1)}T(n-k) \\
&= 2(n+1)\sum_{i=0}^{n-2}\frac{1}{(n+1)-i} + \frac{n+1}{2}T(1) \\
&= 2(n+1)\sum_{j=3}^{n+1}\frac{1}{j} + \frac{n+1}{2}\Theta(1) \\
&= 2(n+1)\Theta(\ln n) + \Theta(n) \\
&= \Theta(n\log n)
\end{aligned}
$$

Method 2: Dividing (10) by $n+1$:

$$
\frac{T(n)}{n+1} \le \frac{T(n-1)}{n} + \frac{2}{n+1}
$$

Let $R(n) = \frac{T(n)}{n+1}$ (and thus, $R(n-1) = \frac{T(n-1)}{n}$). We have

$$
R(n) \le R(n-1) + \frac{2}{n+1}
$$

Note that $R(1) = \frac{T(1)}{1+1} = \Theta(1)$. We get

$$
\begin{aligned}
R(n) &\le \frac{2}{n+1} + \frac{2}{n} + \frac{2}{n-1} + \cdots + \frac{2}{3} + R(1) \\
&= 2\sum_{k=3}^{n+1}\frac{1}{k} + \Theta(1) \\
&= \Theta(\ln n)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
T(n) &= (n+1)R(n) \\
&= (n+1)\Theta(\ln n) \\
&= \Theta(n\ln n)
\end{aligned}
$$