Average Case Analysis of Quicksort

CLRS Reading: Sections 7.3, 7.4, pages 179 – 190

On Average?

- How unbalanced can the partitions become before the costs go up?
- For example, $T(n) = T(9n/10) + T(n/10) + n$
\[ T(n) = T(n/10) + T(9n/10) + n \]

The running time of Quicksort is dominated by time spent in Partition.

\[ H - 3 \]
**Focusing on Partition**

**Lemma 7.1.** The running time of Quicksort is $O(n + X)$, where $X$ is the number of comparisons performed in all calls to Partition.

```plaintext
Partition(A, p, r)
    x = A[r]
    i = p - 1
    for j = p to r - 1
        if A[j] ≤ x
            i = i + 1
            exchange A[i] with A[j]
        exchange A[i+1] with A[r]
    return i + 1
```

**Calculating $X^*$**

*Remark.* Over the entire execution of Quicksort, each pair of elements is compared at most once.

*The total number of comparisons performed by Partition over the entire execution of Quicksort.*
**Number of Comparisons**

<table>
<thead>
<tr>
<th>Q</th>
<th>U</th>
<th>I</th>
<th>Z</th>
<th>O</th>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

**When are two values not compared?**

*In general, \(z_i\) and \(z_j\) are compared if and only if the first element to be chosen as a pivot from \([z_i \ldots z_j]\) is either \(z_i\) or \(z_j\).*

... but after the partition, no two elements separated by \(T\) are ever compared.

Pivot \(T\) is compared with every element, ...

\[\begin{align*}
&Q \ U \ I \ Z \ O \ R \ T \\
&Q \ I \ O \ R \ T \ Z \ U \\
&Q \ I \ O \ R \ T \ U \ Z \\
&I \ O \ Q \ R \ T \ U \ Z \\
&I \ O \ Q \ R \ T \ U \ Z
\end{align*}\]
\[ X = \sum_{1 \leq i \leq n-1} \sum_{1 \leq j \leq n} X_{ij} \]

where \( X_{ij} \) is the indicator function

\[
X_{ij} = \begin{cases} 
1 & \text{if } z_i \text{ is compared with } z_j \\
0 & \text{otherwise}
\end{cases}
\]

*and \( z_1, z_2, \ldots, z_n \) are the elements of \( \mathcal{A} \) in sorted order.

---

**Average Value of \( X \)**

\[
E[X] = E[\sum_{1 \leq i \leq n-1} \sum_{1 \leq j \leq n} X_{ij}] \\
= \sum_{1 \leq i \leq n-1} \sum_{1 \leq j \leq n} E[X_{ij}] \\
= \sum_{1 \leq i \leq n-1} \sum_{1 \leq j \leq n} \Pr[z_i \text{ is compared to } z_j]
\]

*Here we use the fact that the average of the sum is the sum of averages.*
\[ E[X] = \sum_{1 \leq i \leq n-1} \sum_{i+1 \leq j \leq n} \frac{2}{j - i + 1} \]

Pr[\(z_i\) is compared to \(z_j\)] = Pr[\(z_i\) or \(z_j\) is 1st pivot chosen from \([z_i..z_j]\)]

= Pr[\(z_i\) is 1st pivot chosen from \([z_i..z_j]\)]

+ Pr[\(z_j\) is 1st pivot chosen from \([z_i..z_j]\)]

= \frac{2}{j - i + 1}

\[ H_n = O(\log n) \]