Analysis of Quick Sort

Yu-Tai Ching
Department of Computer Science
National Chiao Tung University

Quicksort Pseudo Code

```
QUICKSORT (A,p,r)
    if p < r
            q = PARTITION (A,p,r)
            QUICKSORT (A,p,r)
            QUICKSORT (A,q + 1,r)

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
```

Randomized Quick Sort

RABDOMIZED-PARTITION (A,p,r)

1. $i = \text{RANDOM}(p,r)$
2. exchange $A[r]$ with $A[i]$
3. return PARTITION (A,p,r)

RANDOMIZED-QUICKSORT (A,p,r)

1. if $p < r$
2. $\quad q = \text{RANDOMIZED-PARTITION}(A,p,r)$
3. $\quad \text{RANDOMIZED-QUICKSORT}(A,p,r)$
4. $\quad \text{RANDOMIZED-QUICKSORT}(A,q+1,r)$

- Run time for QUICKSORT $=$ time spent in the PARTITION
  - PARTITION, select a pivot, compare pivot with all of the others.
  - pivot is never included in any future recursive call.
- Thus at most n calls to PARTITION,
- each PARTITION takes a constant time plus the time required for the for loop (lines 3-6),
- we count the number of times line 4 executed, comparing an element with pivot.

Lemma 7.1

Lemma 7.1 Let X be the number of comparisons performed in line 4 of PARTITION over the entired execution of QUICKSORT on an n-element array. Then the running time of QUICKSORT is $O(n + X)$.

Proof There are at most n calls of PARTITION, each call take constant time and time for line 4. Total time for line 4 is X.

Our goal is to compute X.

- Rename the elements of the array A as $z_1, z_2, \ldots, z_n$.
- Define $Z_{ij} = \{z_i, \ldots, z_j\}$ the set of element between z_i and z_j.
- When does the algorithm compares z_i and z_j?
- Define an indicator random variable

$$X_{ij} = I\{z_i \text{ is compared to } z_j\}.$$

.

- Since each pair is compared at most once, total number of comparisons performed by the algorithm is

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

Taking expectation of both sides,a and then using linearity of expectation

$$
\begin{aligned}
E[X] &= E\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_{ij}\right] \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[X_{ij}] \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Pr\{z_i \text{ is compared with } z_j\}
\end{aligned}
$$

- We assumed that each pivot selected randomly.
- Observe that
  - Once a pivot x is choosen, $z_i < x < z_j$, z_i and z_j cannot be compared any more.
  - if $x == z_i$, z_i is compared with all of the others.
  - if $x == z_j$, z_j is compared with all of the others.
- Thus z_i and z_j are compared iff z_i or z_j are selected as pivot.
- Probability z_i (z_j) is selected as pivot

$$\frac{1}{j - i + 1}$$

We conclude

$$
\begin{aligned}
&Pr\{z_i \text{ is compared with } z_j\}\\
&\qquad = Pr\{z_i \text{ or } z_j \text{ is selected as pivot from } Z_{ij}\}\\
&\qquad = Pr\{z_i \text{ is pivot from } Z_{ij}\} + Pr\{z_j \text{ is pivot from } Z_{ij}\}\\
&\qquad = \frac{1}{j-i+1} + \frac{1}{j-i+1}\\
&\qquad = \frac{2}{j-i+1}
\end{aligned}
$$

We get

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} \quad (\text{ let } k = j - i) \\
&< \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \\
&= \sum_{i=1}^{n-1} O(\lg n) \quad (\text{ by equation A. 7}) \\
&= O(n \log n)
\end{aligned}
$$