COMP3711: Design and Analysis of Algorithms

Tutorial 4

HKUST

Question 1

Recall the randomized selection algorithm. We pick a pivot randomly, divide the array into 3 segments: left, pivot, and right, and then either stops immediately or recursively solve the problem in the left or the right part of the array. As in the analysis of quicksort, we denote by $z_1, \ldots, z_n$ the elements in sorted order. Suppose the randomized selection algorithm is given the task of finding the k-th smallest element (namely, z_k). What is the probability that z_i and z_j $(i < j)$ are ever compared by the algorithm? [Hint: consider the following 3 cases separately: $k < i, i \leq k \leq j, k > j$.] Then use the indicator variable technique to show that the expected running time of the randomized selection algorithm is $O(n)$.

Solution 1

Define $X_{ij} = 1$ if z_i and z_j are compared by the algorithm, and 0 otherwise. Consider the following 3 cases:

(a) $i \leq k \leq j$: This is the same as in quicksort: if the pivot is z_i or z_j, then z_i and z_j will be compared; if the pivot is in $\{z_{i+1}, \ldots, z_{j-1}\}$, then they will not be compared. Other pivots do not decide whether z_i or z_j will be compared. So $\Pr[X_{ij} = 1] = 2/(j - i + 1)$.

(b) $i < j < k$: In this case, if the pivot is z_i or z_j, then z_i and z_j will be compared; if the pivot is in $\{z_{i+1}, \ldots, z_k\}$ but not z_j, then they will not be compared. Other pivots do not decide whether z_i or z_j will be compared. So $\Pr[X_{ij} = 1] = 2/(k - i + 1)$.

(c) $k < i < j$: In this case, if the pivot is z_i or z_j, then z_i and z_j will be compared; if the pivot is in $\{z_k, \ldots, z_{j-1}\}$ but not z_i, then they will not be compared. Other pivots do not decide whether z_i or z_j will be compared. So $\Pr[X_{ij} = 1] = 2/(j - k + 1)$.

Solution 1

By the indicator random variable technique, the expected total number of comparisons is $\sum_{i<j} E[X_{ij}] = \sum_{i<j} \Pr[X_{ij} = 1]$, so we just need to add up all these probabilities.

The probabilities in case (a) are as follows:

i:	1	2	3	4	...	k
j: k	$\frac{2}{k}$	$\frac{2}{k-1}$	$\frac{2}{k-2}$	$\frac{2}{k-3}$	...	$\frac{2}{1}$
$k+1$	$\frac{2}{k+1}$	$\frac{2}{k}$	$\frac{2}{k-1}$	$\frac{2}{k-2}$	...	$\frac{2}{2}$
$k+2$	$\frac{2}{k+2}$	$\frac{2}{k+1}$	$\frac{2}{k}$	$\frac{2}{k-1}$	...	$\frac{2}{3}$
...	...					
n	$\frac{2}{n}$	$\frac{2}{n-1}$	$\frac{2}{n-2}$	$\frac{2}{n-3}$	...	$\frac{2}{n-k+1}$

Each diagonal sums up to $O(1)$ and there are $O(n)$ diagonals, so the total is $O(n)$.

Solution 1

The probabilities in case (b) are as follows:

i:	1	2	3	4	...	$k-2$
j: 2	$\frac{2}{k}$					
3	$\frac{2}{k}$	$\frac{2}{k-1}$				
4	$\frac{2}{k}$	$\frac{2}{k-1}$	$\frac{2}{k-2}$			
...	...					
$k-1$	$\frac{2}{k}$	$\frac{2}{k-1}$	$\frac{2}{k-2}$	$\frac{2}{k-3}$	...	$\frac{2}{3}$

Each column sums up to $O(1)$ and there are $O(n)$ columns, so the total is $O(n)$.

Solution 1

The probabilities in case (c) are as follows:

	i:	$k+1$	$k+2$	$k+3$	$k+4$	$\ldots$	$n-1$
j:	$k+2$	$\frac{2}{3}$					
	$k+3$	$\frac{2}{4}$	$\frac{2}{4}$				
	$k+4$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{5}$			
	$\ldots$	$\ldots$					
	n	$\frac{2}{n-k+1}$	$\frac{2}{n-k+1}$	$\frac{2}{n-k+1}$	$\frac{2}{n-k+1}$	$\ldots$	$\frac{2}{n-k+1}$

Each row sums up to $O(1)$ and there are $O(n)$ rows, so the total is $O(n)$.

Thus, the total sum over all three cases is still $O(n)$.

Question 2

The analysis of the expected running time of randomized quicksort in lecture note assumes that all element values are distinct. In this problem, we examine what happens when they are not.

(a) Suppose that all element values are equal. What would be randomized quicksort's running time in this case?

(b) The PARTITION procedure returns an index q such that each element of $A[p...q-1]$ is less than or equal to $A[q]$ and each element of $A[q+1...r]$ is greater than $A[q]$. Modify the PARTITION procedure to produce a procedure PARTITION$'(A, p, r)$, which permutes the elements of $A[p...r]$ and returns two indices q and t, where $p \leq q \leq t \leq r$, such that

- all elements of $A[q...t]$ are equal,
- each element of $A[p...q-1]$ is less than $A[q]$, and
- each element of $A[t+1...r]$ is greater than $A[q]$

Like PARTITION, your PARTITION$'$ procedure should take $\Theta(r-p)$ time.

Question 2

(c) Modify the QUICKSORT procedure to produce QUICKSORT$'(A, p, r)$ that calls PARITITION$'$ and recurses only on partitions of elements not known to be equal to each other.

(d) Using QUICKSORT$'$, how would you adjust the analysis in lecture note to avoid the assumption that all elements are distinct?

Solution 2

(a) The partition procedure will always partition the elements into two subsets where one subset contains $n-1$ elements and the other subset contains 0 elements because all the element values are the same. So the running time is $T(n) = T(n-1) + \Theta(n)$ which is $\Theta(n^2)$.

Solution 2

(b) $\textsc{Partition}'(A, p, r)$:

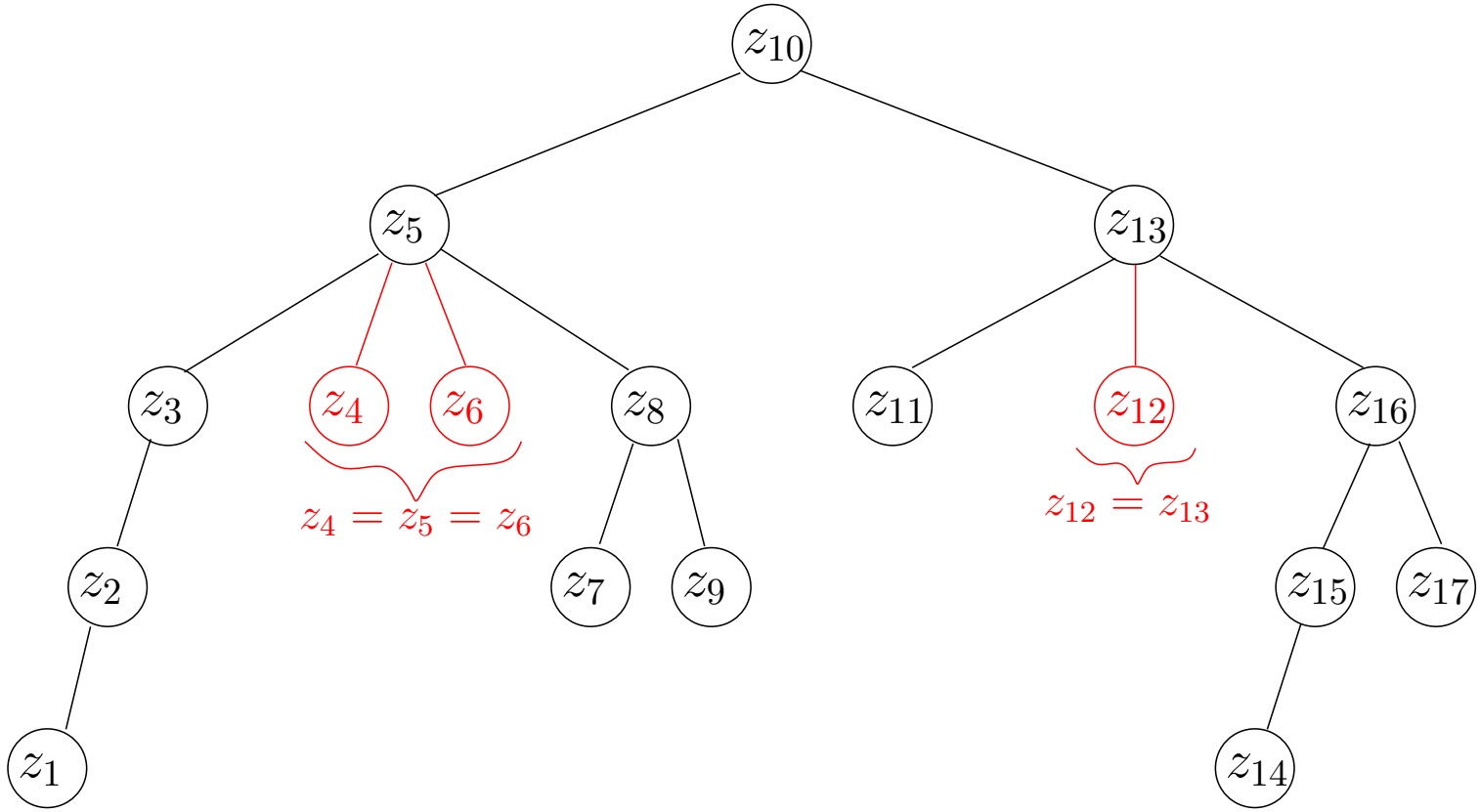
$i = \textsc{Random}(p, r)$
exchange $A[i]$ with $A[r]$
$x \leftarrow A[r]$
$i \leftarrow p - 1$
$k \leftarrow p - 1$
for $j \leftarrow p$ to $r - 1$
    if $A[j] = x$ then
        $k \leftarrow k + 1$
        exchange $A[k]$ with $A[j]$
    else if $A[j] < x$ then
        $i \leftarrow i + 1$
        $k \leftarrow k + 1$
        exchange $A[k]$ with $A[j]$
        exchange $A[i]$ with $A[k]$
exchange $A[k + 1]$ with $A[r]$
return $(i + 1, k + 1)$

Solution 2

(c) $\underline{\textsc{Quicksort}'(A, p, r):}$
if $p \geq r$ then return
$(q, t) = \textsc{Partition}'(A, p, r)$
$\textsc{Quicksort}'(A, p, q-1)$
$\textsc{Quicksort}'(A, t+1, r)$

Solution 2

(d) Relabel the elements from small to large as $z_1, z_2, \ldots, z_n$. We use a multi-ary tree representation which is similar to the binary tree representation except that the middle children of a pivot correspond to the elements equal to the pivot. Similarly, two elements are compared iff they are ancestor-descendant.



Solution 2

For any pair of z_i and z_j, every element in $\{z_i, \ldots, z_j\}$ is equally likely to be the lca of z_i and z_j. In addition, there can be some elements equal to z_i or z_j outside the range of $\{z_i, \ldots, z_j\}$, which can also be the lca of z_i and z_j. So, Pr(z_i and z_j are compared) is at most $2/(j - i + 1)$.

Let $X_{ij} = 1$ if z_i is compared with z_j.
$E(\#\text{ of comparisons}) = \sum_{i<j} E(X_{ij}) = \sum_{i<j} \Pr(z_i \text{ and } z_j \text{ are compared}) \leq \sum_{i<j} 2/(j - i + 1) = O(n \log n)$