

3. (Concentration for the running time of Randomized Quicksort)

(a) Solution: Let's consider a path that could achieve the maximum number of good nodes. Based on the information provided, we know the following: 1) Since a bad node along the path will always decrease the size of the array without adding to the number of good nodes, thus this path could only contain good nodes. 2) In addition, at each node, the path should routing in the larger size of the partition, otherwise, the larger partition path could contain a path with larger number of good nodes. 3) In order to get the largest depth of the path, at each node, the partition should give a maximum size to the larger subset, i.e. $2/3s$, where s is the size of the set at that node. 4) From point 1) to 3), we conclude that for the largest good node path possible, each partition along the path will divide the set to be of size $2/3s$ and $1/3s$, and the path will always follow the larger subset.

Now, let's compute the length of this path (i.e. the largest number of good nodes for a path). Let k be the number of good nodes along this path. Thus,

$$\left(\frac{2}{3}\right)^{k+1} n \geq 1 \quad (1)$$

$$k + 1 \leq \log_{2/3} 1/n \quad (2)$$

$$\leq 1 - \log_{2/3} n \quad (3)$$

$$\leq 1 - \frac{1}{\log_2 2/3} \log_2 n \quad (4)$$

$$k \leq -\frac{1}{\log_2 2/3} \log_2 n \approx 1.71 \log_2 n \quad (5)$$

We have shown that the maximum number of good nodes is no more than $1.71 \log_2 n$, i.e. $c = 1.71$.

(b) Solution: We get more than t nodes only if in the first t trials we get fewer than $c \log_2 n$ good nodes where c value is from (a). Then we can consider the first t nodes, and model the number of good nodes X as a binomial distribution, i.e. $X \sim \text{Bin}(t, 1/3)$. Note in order for a node to be a good node, it must choose the middle $1/3$ of the elements as pivot, and thus, $\mathbf{E}(X) = t/3$. We can set $t = c' \log_2 n$, and apply Chernoff bounds to bound the tail probability to be less than $1/n$.

First, we bound the right hand side of the Chernoff bounds,

$$e^{-\mathbf{E}(X)\delta^2/2} \leq 1/n^2 \quad (6)$$

$$\delta \geq 4/c' \quad (7)$$

Then we plug in the value of δ to the left hand side of Chernoff bounds,

$$\Pr[X \leq c \log_2 n] = \Pr[X \leq (1 - \delta)\mathbf{E}[X]] \quad (8)$$

$$= \Pr[X \leq (1 - \delta)c' \log_2 n/3] \quad (9)$$

$$\leq \Pr[X \leq (1 - 4/c')c' \log_2 n/3] \quad (10)$$

Now by comparing terms, the following inequality needs to hold

$$c \log_2 n \leq (1 - 4/c')c' \log_2 n/3 \quad (11)$$

$$c' \geq 3c + 4 \quad (12)$$

$$\geq 3 \cdot 1.71 + 4 \approx 9.13 \quad (13)$$

(c)

Proof. Let's use the results from part (b): with at most probability $1/n^2$, the number of nodes in a given root to leaf path of the tree is greater than $c' \log_2 n$, where c' is a constant we derived from part (b). Let X_i be the root-leaf distance of any path i and let N_p be the total number of path, and let $X_l = \max_{i \in [N_p]} X_i$ be the root-leaf distance of the longest path, and $N_p \leq n$. Using union bound,

$$\Pr[X_l > c' \log_2 n] \leq \sum_{i=1}^{N_p} \Pr[X_i > c' \log_2 n] \quad (14)$$

$$\leq \sum_{i=1}^n \Pr[X_i > c' \log_2 n] \quad (15)$$

$$\leq \sum_{i=1}^n 1/n^2 = 1/n \quad (16)$$

(d) Solution: Using solutions from (c): with high probability (greater than $1 - 1/n$), the number of nodes in the longest root to leaf path is not greater than $c' \log_2 n$, at each node along this path, the total comparisons of all nodes with the same depth is less than the total number of elements, n . Thus, the total running time $T(n) \leq n \cdot c' \log_2 n = c'n \log_2 n = O(n \log_2 n)$ with probability at least $1 - 1/n$.