Hands On 2: Depth of a node in a random search tree

Algorithm Design

Ferraro Domenico
559813

1 Problem

A random search tree for a set S can be defined as follows: if S is empty, then the null tree is a random search tree; otherwise, choose uniformly at random a key $k \in S$: the random search tree is obtained by picking k as root, and the random search trees on $L = \{x \in S : x < k\}$ and $R = \{x \in S : x > k\}$ become, respectively, the left and right subtrees of the root k.

Consider the Randomized Quick Sort discussed in class and analyzed with indicator variables $[CLRS\ 7.3]$, and observe that the random selection of the pivots follows the above process, thus producing a random search tree of n nodes.

1. Using a variation of the analysis with indicator variables X_{ij}, prove that the expected depth of a node (i.e., the random variable representing the distance of the node from the root) is nearly $2\ ln\ n$.

2. Prove that the expected size of its subtree is nearly $2\ ln\ n$ too, observing that it is a simple variation of the previous analysis.

3. Prove that the probability that the depth of a node exceeds $c\ 2 \ln n$ is small for any given constant $c > 2$. [Note: it can be solved with Chernoff's bounds as we know the expected value.]

2 Solution

2.1 Expected depth of a node

To compute the expected depth of a node, we just need to compute the probability that some node is a proper ancestor of some other node. Let's define the following indicator variable

$$X_{ij} = \begin{cases} 1, & if\ z_j\ is\ ancestor\ of\ z_i \\ 0, & otherwise \end{cases}$$

$$\Pr[X_{ij} = 1] = \Pr[z_j\ is\ pivot \mid z_i\ and\ z_j\ lie\ in\ the\ same\ partition] = \frac{1}{|j - i| + 1}$$

The expected number of ancestors of z_i is equal to the sum, over all other values z_j, of the probability that z_j is an ancestor of z_i:

$$E\left[\sum_{j=1}^{n} X_{ij}\right] = \sum_{j=1}^{n} E[X_{ij}] = \sum_{j=1}^{n} \Pr[X_{ij} = 1] = \sum_{j=1}^{n} \frac{1}{|j-i|+1}$$

$$= \frac{1}{i} + \frac{1}{i-1} + \cdots + 1 + \frac{1}{2} + \cdots + \frac{1}{n-i+1}$$

$$= 2\sum_{k=1}^{n} \frac{1}{k} = O(2\log n)$$

2.2 Expected size of a subtree

The expected size of a subtree of a node i is the expected number of descendants of the node. The expected number of descendants of z_i is equal to the sum, over all other values z_j, of the probability that z_j is a descendant of z_i. The analysis is like the previous problem, but we change the indicator variable into the following

$$X_{ij} = \begin{cases} 1, & if\ z_j\ is\ descendant\ of\ z_i \\ 0, & otherwise \end{cases}$$

2.3 Depth of a node

To compute the probability of $depth\ of\ i = \sum_{j=1}^{n} X_{ij} > 2c \log n$ we can use Chernoff's Bounds

$$\Pr\left[\sum_{j=1}^{n} X_{ij} > \mu + \lambda\right] \leq e^{-\frac{\lambda^2}{2\mu+\lambda}}$$

We already know that $\mu \leq 2 \log n$, then we can define λ

$$\mu + \lambda = 2c \log n$$

$$\lambda = 2c \log n - \mu \leq 2c \log n - 2 \log n = (2c - 2) \log n$$

Finally, Chernoff's Bounds becomes

$$\begin{aligned}
\Pr\left[\sum_{j=1}^{n} X_{ij} > 2c \log n\right] &\leq e^{-\frac{\lambda^2}{2\mu+\lambda}} \\
&= e^{-\frac{\left((2c-2)\log n\right)^2}{2(2\log n)+(2c-2)\log n}} \\
&= e^{-\frac{(2c-2)^2(\log n)^2}{2\log n+2c\log n}} \\
&= e^{-\frac{(2c-2)^2(\log n)^2}{(2c+2)\log n}} \\
&= e^{-\frac{(2c-2)^2\log n}{2c+2}} \\
&= n^{-\frac{(2c-2)^2}{2c+2}} \\
&= n^{-\frac{4c^2-8c+4}{2c+2}} \\
&= n^{-\frac{2\left(2c^2-4c+2\right)}{2(c+1)}} \\
&= n^{-\frac{2c^2-4c+2}{c+1}} \\
&= \frac{1}{n^{\frac{2c^2-4c+2}{c+1}}}
\end{aligned}$$

The probability that the depth of a node exceeds $2c \log n$ is small for any given constant $c > 2$.