Lecture 16

Why Binary Search Trees?

What is the height of a BST with n nodes?

Height varies from $\lfloor \log_2 n \rfloor$ to $n - 1$

Time complexity of Insert and Delete is O(H) so the lower bound of H is the best case scenario.

best case	$\lfloor \log_2 n \rfloor$
avg case	$\log_2 n$
worst case	$n - 1$

Proof of $\log_2 n$ on avg

For n elements there exist `n!` ways of constructing said binary tree All elements are coming in random order

strong("Note: ") body

Assumption: A corner element does not occur 3 consecutive times, or every 3 consecutive elements include at least one middle element.

The middle 98% is middle elements, and the left and right 1% are the corner elements.

Let us assume the first element is a middle element, with L subtree bigger than R subtree.

Let us assume the next two elements are corner elements, i.e don't have two subtrees.

After 3 steps, the remaining nodes are at most $\frac{99}{100}n$

After 3k steps, the remaining nodes are at most $\left(\frac{99}{100^k}\right)n$

$k = \log_n\left(\frac{100}{99}\right)$

Issues with this proof

- Assumption does not cover all cases
- Abuse of the Big O notation. In most practical cases the constant is bigger than the function itself
- Grains of rice proof

$$X_{i,j} = \begin{cases} 1 \text{ if node } \ i \ \text{ is an ancestor of node } \ j \\ \\ 0 \ \text{ otherwise.} \end{cases}$$

In how many of the 10! BSTs of 1-10 is node 7 an ancestor of node 6?

Relationship of all other nodes with 6 and 7 is identical.

Thus, for all permutations where 6 appears before 7, 6 is an ancestor of 7 ($X_{6,7} = 1$) and vice versa.

Thus, $Pr\left[X_{7,6} = 1\right] = \frac{1}{2}$ because for all permutations where 6 is before 7, there exists a permutation where the 6 and 7 are swapped and 7 comes before 6

$Pr\left[X_{8,6} = 1\right]$?

Out of 3! permutations 867 and 876 are valid orders for ancestry

So, $Pr\left[X_{8,6} = 1\right] = \frac{1}{2}$

$Pr\left[X_{9,6} = 1\right]$?

If from 678 any occur before, 9 and 6 are split apart into different subtrees from that element. Thus probability is 1/4

What is the expected depth of node 6?

$\frac{2+3+4+2+4+\dots}{10!}$

D_j : Depth of node j

Expected depth = Number of ancestors

$D_6 = X_{1,6} + X_{2,6} + \dots + X_{10,6}$

$$\mathbb{E}[D_6] = \mathbb{E}\left[X_{1,6} + \cdots + X_{10,6}\right]$$

$$= \mathbb{E}\left[X_{1,6}\right] + \mathbb{E}\left[X_{2,6}\right] + \cdots + \mathbb{E}\left[X_{10,6}\right]$$

$$= \Pr\left[X_{1,6} = 1\right] + \Pr\left[X_{2,6} = 1\right] + \cdots + \Pr\left[X_{10,6} = 1\right]$$

$$= \left(\frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2}\right) + \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right)$$

Harmonic Sum (H_n) = $log_e n + \gamma$ = $log_e n + 0.5772$

$$\text{Average depth} \leq 2\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n/2}\right)$$

$$\leq 2\ln\left(\frac{n}{2}\right) + 1.15$$

$$\leq 2.88(\log_2 n - \log_2 2)$$

$$\leq 2.88 \log_2 n$$