Analysis of Binary Search Trees (BST) Algorithms

Goal: to derive

$$\begin{cases} \mathbb{E}(X_n) & := \text{average \# probes per } \quad \textbf{succ} \text{ search in a random } n\text{-node BST} \\ \mathbb{E}(Y_n) & := \text{average \# probes per } \textbf{unsucc} \text{ search} \qquad\qquad \text{"} \end{cases}$$

Sample Spaces:

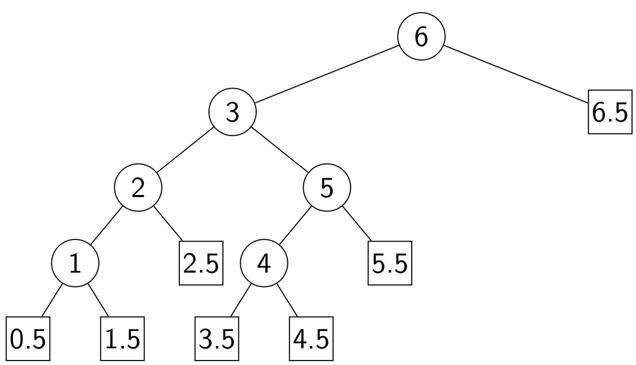
$$\begin{cases} \Omega_{X_n} & = \left\{ (x_1, x_2, \ldots, x_n; k) \,\middle|\, \begin{matrix} (x_1, x_2, \ldots, x_n) \in S_n \\ 1 \le k \le n \end{matrix} \right\}, \quad |\Omega_{X_n}| = n!\, n \\ \Omega_{Y_n} & = \left\{ (x_1, x_2, \ldots, x_n; y) \,\middle|\, \begin{matrix} (x_1, x_2, \ldots, x_n) \in S_n \\ y = 0.5,\, 1.5,\, \ldots,\, n.5 \end{matrix} \right\}, \quad |\Omega_{Y_n}| = (n+1)! \end{cases}$$

$$\begin{aligned} S_n &= \text{set of all permutations (relative order of keys) of } \{1, 2, \ldots, n\} \\ \Omega_{Y_n} &= \big\{ (x_1, x_2, \ldots, x_n; y) \mid (x_1, x_2, \ldots, x_n; y) \in S_{n+1} \big\}, \quad \text{by re-numbering} \end{aligned}$$

Random Variables:

$$\begin{cases} X_n\big((x_1, x_2, \ldots, x_n; k)\big) & := \text{\# probes to } \quad \textbf{succ} \text{ search } x_k \text{ in } T(x_1, x_2, \ldots, x_n) \\ Y_n\big((x_1, x_2, \ldots, x_n; y)\big) & := \text{\# probes to } \textbf{unsucc} \text{ search } y \text{ in } \qquad \text{"} \end{cases}$$

$$\begin{aligned} T(x_1, x_2, \ldots, x_n) &:= \text{BST formed by inserting keys } x_1, x_2, \ldots, x_n \text{ consecutively} \\ T(6, 3, 2, 1, 5, 4) &= \end{aligned}$$



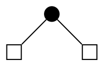
$$\begin{cases} \displaystyle\sum_{1 \le k \le 6} X_6\big((6,3,2,1,5,4;k)\big) & = (1+2+3+4+3+4) & = 17 = 6 + I\big(T(6,3,2,1,5,4)\big) \\ \displaystyle\sum_{y=0.5,\ldots,6.5} Y_6\big((6,3,2,1,5,4;y)\big) & = (4+4+3+4+4+3+1) & = 23 = \qquad E\big(T(\cdot)\big) \end{cases}$$

$$\begin{cases} I(T) & := \displaystyle\sum_{x:\text{ int node of } T} \text{path length}(x) = 11 \\ E(T) & := \displaystyle\sum_{y:\text{ ext node of } T} \text{path length}(y) = 23 = I(T) + 2 \cdot 6 \end{cases}$$

$$\mathbb{E}(X_6) = \frac{1}{6}\Big[\big(1 + \mathbb{E}(Y_0)\big) + \big(1 + \mathbb{E}(Y_1)\big) + \cdots + \big(1 + \mathbb{E}(Y_4)\big) + \big(1 + \mathbb{E}(Y_5)\big)\Big]$$

n	1	2	3	
$\mathbb{E}(X_n) = 2\frac{n+1}{n}H_n - 3$	1	$\frac{3}{2}$	$\frac{17}{9}$	$\cdots$
$\mathbb{E}(Y_n) = 2H_{n+1} - 2$	1	$\frac{5}{3}$	$\frac{13}{6}$	$\cdots$

$\boxed{n=1}$



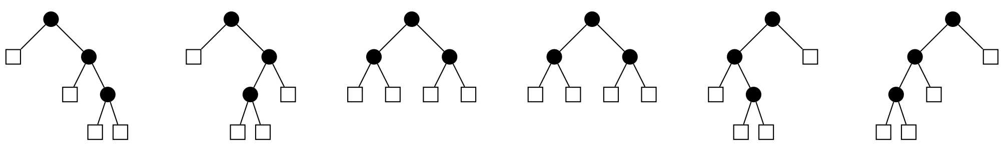
$$\begin{cases} \mathbb{E}(X_1) &= \frac{1}{1}\,[1] &= 1 \\ \mathbb{E}(Y_1) &= \frac{1}{2}\,[1+1] &= 1 \end{cases}$$

$\boxed{n=2}$

$(1,2)$ $(2,1)$

$$\begin{cases} \mathbb{E}(X_2) &= \frac{1}{2\cdot 2!}\,[(1+2)+(1+2)] &= \frac{3}{2} \\ \mathbb{E}(Y_2) &= \frac{1}{3\cdot 2!}\,[(1+2+2)+(2+2+1)] &= \frac{5}{3} \end{cases}$$

$\boxed{n=3}$



$(1,2,3)$ $(1,3,2)$ $(2,1,3)$ $(2,3,1)$ $(3,1,2)$ $(3,2,1)$

$$\begin{cases} \mathbb{E}(X_3) &= \frac{1}{3\cdot 3!}\,[(1+2+3)+(1+2+3)+(1+2+2)+(\cdot)+(\cdot)+(\cdot)] &= \frac{17}{9} \\ \mathbb{E}(Y_3) &= \frac{1}{4\cdot 3!}\,[(1+2+3+3)+(1+3+3+2)+(2+2+2+2)+(\cdot)+(\cdot)+(\cdot)] &= \frac{13}{6} \end{cases}$$

Lemma 1. An n-node binary tree T has

(1) $2n$ edges,
(2) $n+1$ external nodes (leaves),
(3) $E(T) = I(T) + 2n$.

Lemma 2. $\mathbb{E}(X_n) = 1 + \frac{1}{n} \sum_{1 \le k \le n} \mathbb{E}(Y_{k-1})$

Theorem 3.

$$\begin{cases} \mathbb{E}(X_n) & = 2\frac{n+1}{n} H_n - 3 \quad \approx 2 \ln n \\ \mathbb{E}(Y_n) & = 2H_{n+1} - 2 \qquad \approx 2 \ln n \end{cases}$$

Proof:

$$\begin{cases} \mathbb{E}(X_n) & = \sum_{w \in \Omega_{X_n}} P(w) X_n(w) = \sum_{\substack{(x_1,\ldots,x_n) \in S_n \\ 1 \le k \le n}} \frac{1}{n!\, n} X_n(x_1, \ldots, x_n; k) \\ & = \frac{1}{n!\, n} \sum_{(x_1,\ldots,x_n) \in S_n} \sum_{1 \le k \le n} X_n(x_1, \ldots, x_n; k) \\ & = \frac{1}{n!\, n} \sum_T \big(n + I(T)\big) \\ & = \frac{1}{n} \Big[n + \frac{1}{n!} \sum_T I(T) \Big] & (1) \\ \mathbb{E}(Y_n) & = \frac{1}{(n+1)!} \sum_{(x_1,\ldots,x_n) \in S_n} \sum_{y = 0.5, \ldots, n.5} Y_n(x_1, \ldots, x_n; y) \\ & = \frac{1}{(n+1)!} \sum_T E(T) \\ & = \frac{1}{(n+1) n!} \sum_T \big(2n + I(T)\big) \\ & = \frac{1}{n+1} \Big[2n + \frac{1}{n!} \sum_T I(T) \Big] & (2) \end{cases}$$

$(1) + (2) \qquad \Rightarrow (n+1)\mathbb{E}(Y_n) = n\mathbb{E}(X_n) + n$

$+$ **Lemma 2** $\Rightarrow$

$$\begin{aligned} (n+1)\,\mathbb{E}(Y_n) &= \sum_{1 \le k \le n} \mathbb{E}(Y_{k-1}) + 2n \\ -)\qquad n\,\mathbb{E}(Y_{n-1}) &= \sum_{1 \le k \le n-1} \mathbb{E}(Y_{k-1}) + 2(n-1) \end{aligned}$$

$$(n+1)\,\mathbb{E}(Y_n) = (n+1)\,\mathbb{E}(Y_{n-1}) + 2$$

$$\Rightarrow \begin{cases} \mathbb{E}(Y_n) & = \mathbb{E}(Y_{n-1}) + \frac{2}{n+1} \\ \mathbb{E}(Y_{n-1}) & = \mathbb{E}(Y_{n-2}) + \frac{2}{n} \\ & \vdots \\ \mathbb{E}(Y_2) & = \mathbb{E}(Y_1) + \frac{2}{3} \\ \mathbb{E}(Y_1) & = \frac{2}{2} \end{cases}$$

$$\Rightarrow \begin{cases} \mathbb{E}(Y_n) & = 2H_{n+1} - 2 \\ \mathbb{E}(X_n) & = \frac{n+1}{n} E(Y_n) - 1 = 2\frac{n+1}{n} H_n - 3 \end{cases} \qquad \blacksquare$$

Lemma 2. $\mathbb{E}(X_n) = 1 + \frac{1}{n} \sum_{1 \le k \le n} \mathbb{E}(Y_{k-1})$

Proof:

$$\begin{aligned}
\mathbb{E}(X_n) &= \frac{1}{n!\,n} \sum_{(x_1,\ldots,x_n)} \sum_{1\le k\le n} X_n(x_1,\ldots,x_n;k) \\
&= \frac{1}{n!\,n} \sum_{(x_1,\ldots,x_n)\in S_n} \sum_{1\le k\le n} \Big(1 + Y_n(x_1,\ldots,x_{k-1};k)\Big) \\
&= 1 + \frac{1}{n!\,n} \sum_{1\le k\le n} \sum_{(x_1,\ldots,x_k)\in S_k} n\cdots(k+1) Y_{k-1}(x_1,\ldots,x_{k-1};x_k) \\
&= 1 + \frac{1}{n} \sum_{1\le k\le n} \left[\frac{1}{k!} \sum_{(x_1,\ldots,x_k)\in S_k} Y_{k-1}(x_1,\ldots,x_{k-1};x_k)\right] \\
&= 1 + \frac{1}{n} \sum_{1\le k\le n} \mathbb{E}(Y_{k-1}) \qquad \blacksquare
\end{aligned}$$

Remark:

$$\begin{cases} \text{Dynamic model} & : \ n! \text{ trees;} \\ \text{Static} \quad\quad \text{"} & : \ \frac{1}{n+1}\binom{2n}{n} \text{ trees,} \quad (\text{BST } T(6,3,2,1,5,4) \equiv T(6,3,5,4,2,1)) \end{cases}$$