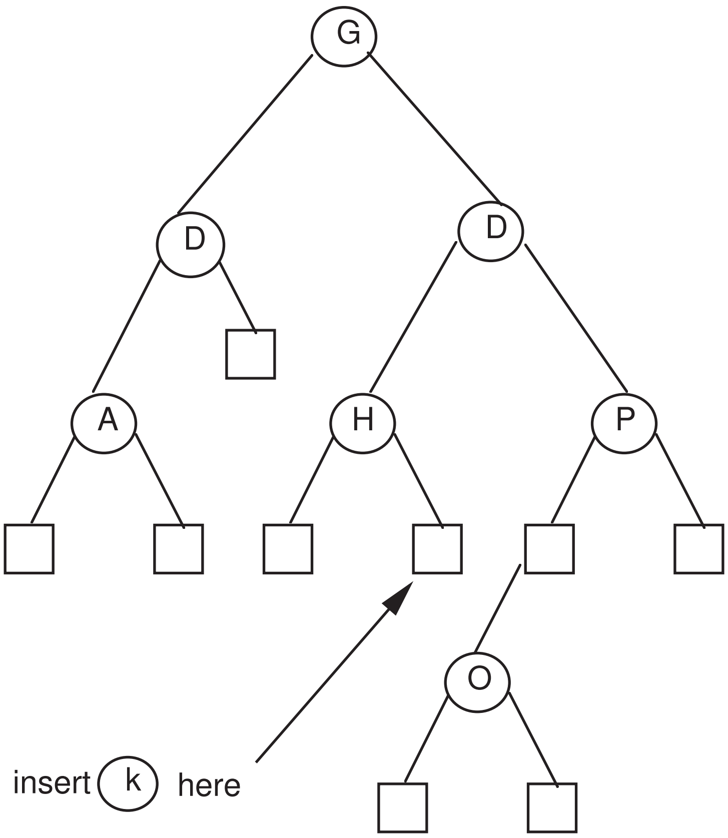
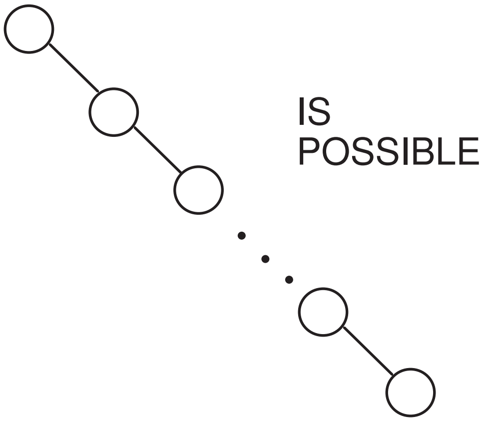
Insertions and Deletions in BST's

Insert: Put the new element node at the external node where the search ends.

Example. Insert "k" in:

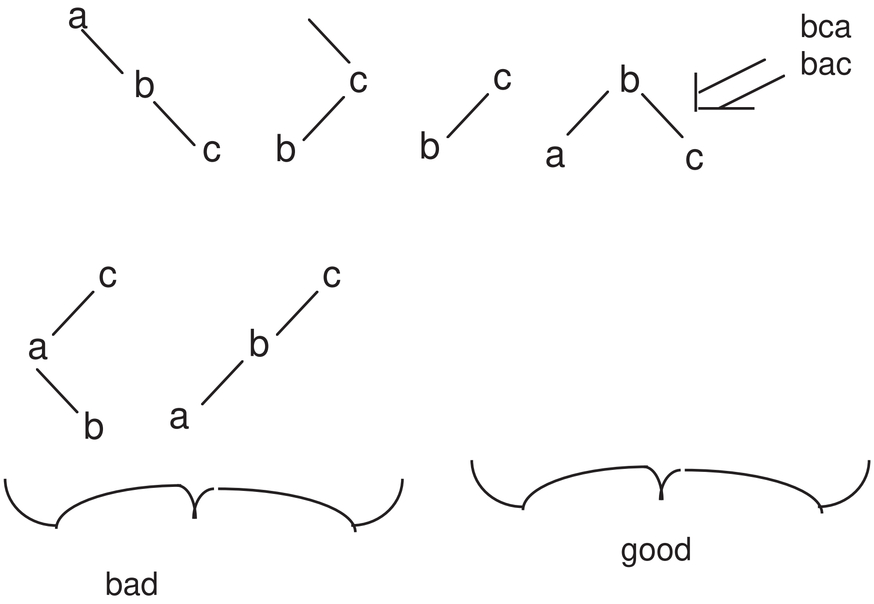


What kind of tree do we get?



- But not likely, if all input order equally likely.

• For n nodes, there are $n!$ possible orderings for the input. For $n = 3$, we have

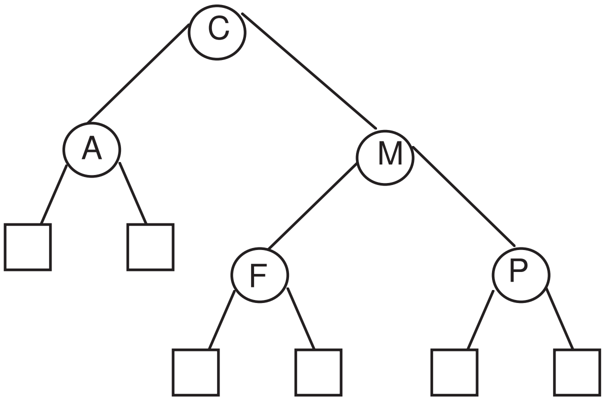


For large n: most trees are "good" (short).

- Question: What is the expected cost of a search in a tree in which n insertions were made in random order.

- insert n integers randomly to a BST. We derive the average cost for 1 search for such tree

Example. For a given tree,



$E = 2 + 2 + 3 + 3 + 3 + 3 = 16$

$I = 1 + 0 + 2 + 1 + 2 = 6$

Then the average unsuccessful search cost is 16/6, and the average successful search cost is $6/5 + 1$

Now, let E_n = expected external path length

and I_n = expected internal path length

Then for example, E_5 and I_5 is obtained by averaging all E, I's over all trees of 5 nodes.

- We show how to compute E_n

$E_0 = 0$

$$E_n = \underbrace{n+1}_{\substack{\text{root node cost} \\ \text{for each external} \\ \text{node (n+1 total)}}} + \underbrace{\frac{1}{n} \sum_{i=1}^{n} (E_{i-1}+E_{n-i})}_{\substack{\text{Avg. ext. path length of} \\ \text{left \& right subtrees over root} \\ \text{(}i^{th}\text{ element)}}}$$

After rewriting, we get

$$E_n = n + 1 + \frac{2}{n} \sum_{i=0}^{n-1} E_i$$

Then

$$\begin{aligned} nE_n - (n-1)E_{n-1} &= 2n + 2E_{n-1}, (E_0 = 0) \\ nE_n &= 2n + (n+1)E_{n-1} \end{aligned}$$

Thus,

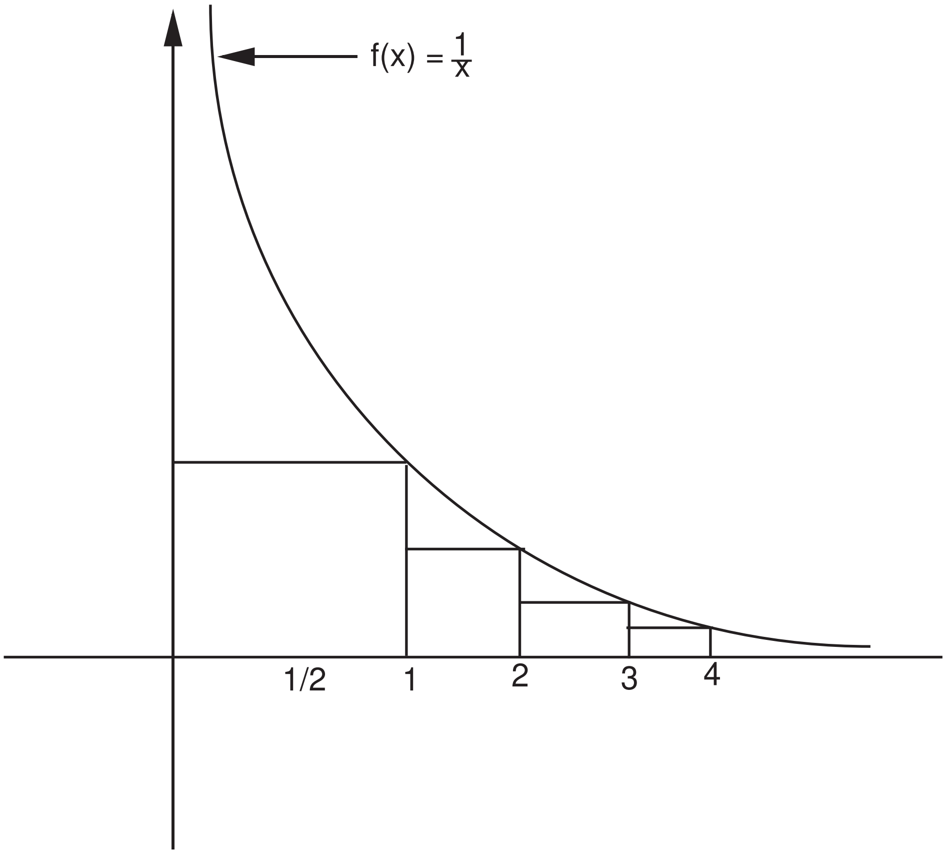
$$\begin{aligned}
E_n &= 2 + \frac{n+1}{n} E_{n-1} \\
&= 2 + 2\frac{n+1}{n} + \frac{n+1}{n} \bullet \frac{n}{n-1} E_{n-2} \\
&= 2 + 2\frac{n+1}{n} + \frac{n+1}{n-1} E_{n-2} \\
&= 2 + 2\frac{n+1}{n} + 2\frac{n+1}{n-1} + \frac{n+1}{n-2} E_{n-3}
\end{aligned}$$

$\vdots$

$$\begin{aligned}
&= 2 + 2\sum_{i=2}^{n} \frac{n+1}{i} + (n+1)E_0 \\
&= 2\sum_{i=1}^{n+1} \frac{n+1}{i} - 2(n+1) \\
&= 2(n+1)\sum_{i=1}^{n+1} \frac{1}{i} - 2(n+1)
\end{aligned}$$

+ +

- To evaluate $\sum_{i=1}^{n} \frac{1}{i}$ (Harmonic Sequence)



$$\sum_{i=1}^{n} \frac{1}{i} \approx \int_{1}^{n} \frac{1}{x} dx = \ln x|_{1}^{n} = \ln n$$

So $E_n \approx 2(n+1)\ln(n+1)$

Average Search cost of unsuccessful search is

$$\frac{E_n}{n+1} = 2\ln(n+1).$$

+

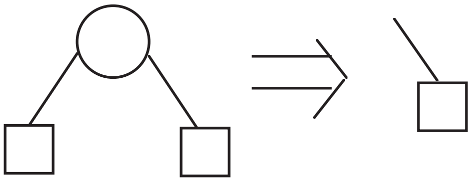
Average cost of successful search is (Cost of "unsuccessful" search when you put the value in plus looking at the value.):

$$
\begin{aligned}
& 1 + \frac{\sum_{i=0}^{n-1} \frac{E_i}{i+1}}{n} \\
\approx\ & \frac{\int_1^n 2 \ln x dx}{n} = \frac{2[x \ln x - x]|_1^n}{n} \\
\approx\ & \frac{2[n \ln n - n + 1]}{n} \approx 2 \ln n + O(1).
\end{aligned}
$$

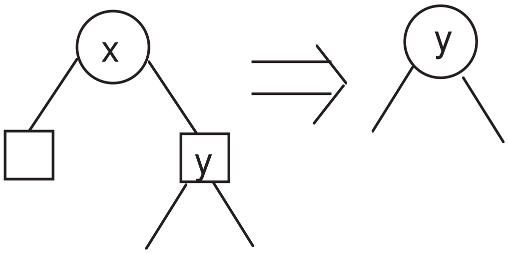
- Variance, standard deviation $\approx \sqrt{2 \ln n}$
- Expected height $\approx 4.3 \ln n$

Deletions:

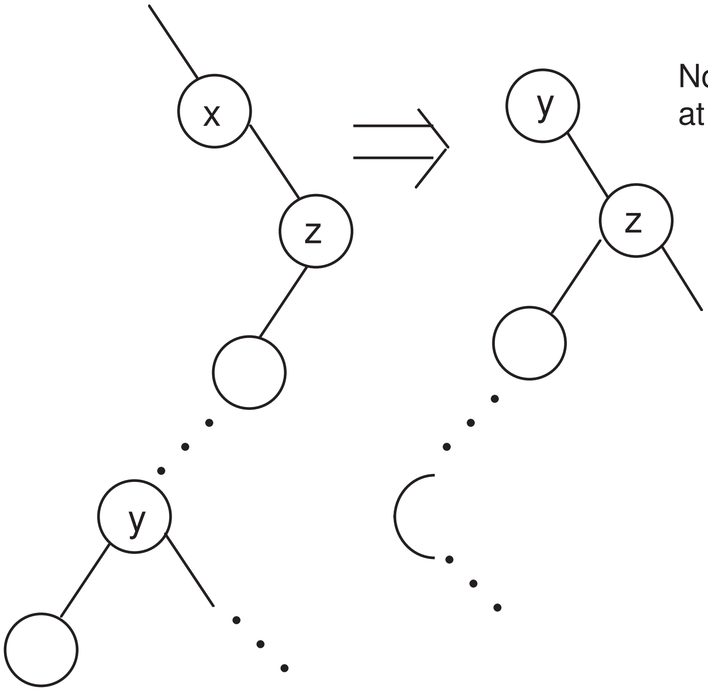
- Remove a leaf:



- Remove a node with 1 childe



- Remove a node with 2 children



Note: y can only have at most 1 child