Wednesday 7/16: Tree and Grammar Induction

We are now going to learn how to use induction to prove claims about trees (either produced by grammars or other rules that will be spelled out in the problem)

Note how we split the larger tree of height k into smaller subtrees of height $\leq k-1$.

tree induction example: Let T be a binary tree with height h and n-many nodes. Then prove by induction that, $n \leq 2^{h+1} - 1$.

Last week Hongxuan discussed choosing the correct induction variable. When we discuss claims *about* trees, the induction variable is almost always height.

Proof:

Proof by induction on h.

Base case: When $h = 0$, there is 1 node. So $1 \leq 2^1 - 1 = 1$ is true.

Induction hypothesis: Suppose binary trees of height h with n-many nodes have $n \leq 2^{h+1} - 1$ for $h = 0...k-1$.

Goal: We want to show that for a tree of height k with n-many nodes, $n \leq 2^{k+1} - 1$.

Inductive step: Assume we have a tree of height k. There are two cases for what happens at the root.

Case 1: The root node has 2 children. Then, each of those children is the root node of a tree of height $\leq k-1$. By the inductive hypothesis, the left subtree has $\leq 2^k - 1$ nodes; the right subtree also has $\leq 2^k - 1$ nodes. Then, our tree of height k has a total of $\leq 1 + 2^k - 1 + 2^k - 1 = 2^{k+1} - 1$ nodes.

Case 2: The root node has 1 child of height $k-1$. By the IH, this subtree has $\leq 2^k - 1$ nodes. So our tree has $n \leq 1 + 2^k - 1 = 2^k$. We know $2^k \leq 2^{k+1} - 1$ since $k > 0$, so $n \leq 2^{k+1} - 1$.

Since in both cases $n \leq 2^{k+1} - 1$, we have proven our claim.

grammar tree induction example: Consider a grammar G with start symbol S and terminals a, b, with rules:

$S \quad \rightarrow \quad ab \quad | \quad SS \quad | \quad aSb$

Prove that all trees generated by G will have the same number of a nodes and b nodes.

Proof:

Proof by induction on height h.

Base case: h=0 is not possible; because a single node S is not a valid tree according to this grammar. When $h = 1$, we have a root node of S with children a and b. This tree has the same number of a's and b's.

Inductive Hypothesis: Suppose all trees generated by G of height h will have the same number of a and b nodes for $h = 1...k-1$.

Goal: We want to show that a tree of height k generated by G has an equal number of a's and b's.

Inductive step: Assume we have a tree of height k. There are two cases for what happens at the root.

Case 1: The root node S branches into two subtrees according to the second rule of the grammar. Call left subtree S_1 and the right S_2. Both have height $\leq k-1$. By the IH, S_1 has as many a's as b's, let's say m ($m \in \mathbb{Z}$), and S_2 has j, with $j \in \mathbb{Z}$. Then, our tree rooted at S has $m + j$-many a's and b's.

Case 2: The root node S has three children, aSb, according to the third rule of the grammar. The subtree rooted at S has height $= k-1$, so by the IH has m-many a's and b's ($m \in \mathbb{Z}$). Then, our tree has $m + 1$-many a's and b's from the left and right children.

In both cases, our tree has as many a's as b's, so we have proven our claim.