Proving the correctness of Context-Free Grammars.

How to show that the constructed grammar does indeed generate some required language ? If L is some language, and G is a CFG, then to prove that $L = L(G)$ you have to prove two things:

- Every string w generated by G is in L. Typically, this is a proof by induction on the number of steps in the derivation of w.
- Every string w in L can be generated by G. Typically, this is a proof by induction on the length, or some other parameter, of w.

Example: Let $L = \{0^i 1^j \mid 2i \leq j \leq 3j\}$. In the class I suggested that the following context-free grammar G generates L:

$$S \to 0S11 \mid 0S111 \mid \varepsilon$$

Now, lets prove that G is indeed the grammar that generates L, i.e. $L = L(G)$.

Claim 1 If a string w is generated by G, then $w \in L$.

Proof: induction on the number of steps in the derivation of w.

- *Base case:* derivation with 1 step. The only derivation with 1 step is $S \Rightarrow \varepsilon$, thus $w = \varepsilon$, and so $w \in L$.
- *Inductive hypothesis:* assume that for every derivation $S \stackrel{*}{\Rightarrow} w$ with $n \geq 1$ steps, w is in L.
- *Inductive step:* prove that every derivation with $n+1$ steps generates a string in L.

  Let $S \stackrel{*}{\Rightarrow} w$ be a derivation with $n+1$ steps. Since $n+1 > 1$, the first step in this derivation cannot be $S \Rightarrow \varepsilon$. Thus, the derivation starts with either $S \Rightarrow 0S11$, or $S \Rightarrow 0S111$.

  In the first case, w must be of the form $0w_1 11$, where w_1 is a string derived from the nonterminal S in the remaining n steps. By the inductive hypothesis, $w_1 \in L$, i.e. $w_1 = 0^i 1^j$, with $2i \leq j \leq 3i$. Therefore, $w = 00^i 1^j 11 = 0^{i+1} 1^{j+2} \in L$, because if $2i \leq j \leq 3i$, then $2i + 2 \leq j + 2 \leq 3i + 2$, and so $2(i+1) \leq j + 2 < 3(i+1)$.

  Similarly, in the second case, w must be of the form $0w_1 111$, where w_1 is a string derived from the nonterminal S in the remaining n steps. By the inductive hypothesis, $w_1 \in L$, i.e. $w_1 = 0^i 1^j$, with $2i \leq j \leq 3i$. Therefore, $w = 00^i 1^j 111 = 0^{i+1} 1^{j+3} \in L$, because if $2i \leq j \leq 3i$, then $2i + 3 \leq j + 3 \leq 3i + 3$, i.e. $2(i+1) < j + 2 \leq 3(i+1)$.

This completes the proof of the claim. □

Claim 2 Every string $w \in L$ can be generated by G.

Proof: we want to do induction, but we cannot do induction on the length of w. Lets list the first few strings in L: $\varepsilon, 011, 0111, 001111, 0011111, 00111111, \ldots$. You can see already that L does not have strings of of certain length (e.g. 1,2,5), so we cannot quite do induction on the length of w. Instead we can do induction on the number of 0's in w.

Note, that for any given number of 0's, $n \geq 0$, L contains exactly the strings $0^n1^{2n}, 0^n1^{2n+1}, \ldots, 0^n1^{2n+n}$, and so the claim that we will prove by induction will cover all of these strings at once.

So, we're proving the following claim: for any given $n > 0$ (let's skip ε, its clear that it can be generated by G), and any given k, $0 \leq k \leq n$, the string $w = 0^n1^{2n+k}$ can be generated by G. Proof by induction on n.

- *Base case:* $n = 1$. Then, k is either 0 or 1. In the first case, $w = 011$, and is generated as follows: $S \Rightarrow 0S11 \Rightarrow 011$. In the second case, $w = 0111$, and is generated as follows: $S \Rightarrow 0S111 \Rightarrow 0111$.
- *Inductive hypothesis:* assume that the claim is true for $n = i$, that is for any k, $0 \leq k \leq i$, the string $w = 0^i1^{2i+k}$ can be generated by G.
- *Inductive step:* prove the claim for $n = i + 1$. Let $w = 0^{i+1}1^{2(i+1)+k}$, with $0 \leq k \leq i + 1$. We want to show that w can be generated by G. We have two cases:
  - $0 \leq k \leq i$. Then $w = 0^{i+1}1^{2(i+1)+k} = 00^i1^{2i+2+k} = 00^i1^{2i+k}11$. By the inductive hypothesis, the string 0^i1^{2i+k} for $0 \leq k \leq i$ can be generated by G, that is there is a derivation $S \stackrel{*}{\Rightarrow} 0^i1^{2i+k}$, and so $S \Rightarrow 0S11 \stackrel{*}{\Rightarrow} 00^i1^{2i+k}11 = w$ is a derivation for w.
  - $k = i + 1$. Then $w = 0^{i+1}1^{2(i+1)+i+1} = 00^i1^{3i+3} = 00^i1^{3i}111$. By the inductive hypothesis, the string 0^i1^{3i} can be generated by G, that is there is a derivation $S \stackrel{*}{\Rightarrow} 0^i1^{3i}$, and so $S \Rightarrow 0S111 \stackrel{*}{\Rightarrow} 00^i1^{3i}111 = w$ is a derivation for w.

This completes the proof of Claim 2. □