5/5

COS 488 Week 5: Q2

Dylan Mavrides

March 10, 2017

6.42) Internal nodes in binary trees fall into one of three classes: they have either two, one, or zero external children. What fraction of the nodes are of each type, in a random binary Catalan tree of N nodes?

We begin by looking for the fraction of nodes that are of the type with two external nodes (let this as a function of the tree size, be E(t)). We begin with:

$$T(z) = \sum_{t \in T} z^{|t|}$$

and

$$Q(z) = \sum_{t \in T} E(t) z^{|t|}$$

Thus if we look at a tree in terms of the right and left subtrees, we see that

$$E(t) = E(t_L) + E(t_R)$$

(note that the root node can't have two external nodes, except in the case where we have a tree with one node) Thus using our construction:

$$Q(z) = z + \sum_{t_L \in T} \sum_{t_R \in T} (E(t_L) + E(t_R)) z^{|t_L| + |t_R| + 1}$$

Note that $\sum_{t \in T} E(t_L) z^{|t_L|} \sum_{t_R \in T} z^{|t_R|} = Q(z)T(z)$. Therefore we can substitute and get:

$$Q(z) = z + 2zQ(z)T(z)$$

Solving for Q(z) we get:

$$Q(z) = \frac{z}{1 - 2zT(z)} = \frac{z}{1 - 2z(\frac{1 - \sqrt{1-4z}}{2z})} = \frac{z}{\sqrt{1-4z}}.$$

Now we use the corollary of the radius-of-convergence transfer theorem with $f(z) = z$, $\rho = 1/4$, and $\alpha = 1/2$, note that $\Gamma(1/2) = \sqrt{\pi}$ thus we have $\frac{4^n}{4\sqrt{\pi n}}$.

The total number of internal nodes of a binary catalan tree is the Nth catalan number, thus we have the fraction:

$$\frac{4^n}{4\sqrt{\pi n}} * \frac{\sqrt{\pi n^3}}{4^n} = n/4$$

Now we note that if we take a tree and remove all of the internal nodes with one external node, and compress the other subtree of the node with one external node, we're left with a tree with only nodes with zero or two external nodes. In a tree like this, we see that the nodes with two external nodes act as the leaves in a standard tree. Thus we have the number of internal nodes with two external nodes is equal to 1 + the number of internal nodes with zero external nodes. I.E. $N - n_2 = n_2 + n_0 = 2n_0 + 1$ (where the subscript is the number of external nodes, and N is the total number of nodes). Since we are dealing with asymptotics, the fraction of nodes with zero external nodes and with two external nodes is functionally the same. Thus for the case with 0 external nodes, we also have $n/4$, leaving $n/2$ of the n nodes having 1 external node on average.

Worked with Eric N, Matt T.

the equation is wrong (I'm not sure exactly how you got it, and plugging in n_2 ~ n_0 ~ n/4 it does not hold up) and there are simpler combinatorial arguments, but I assume there's a typo somewhere--this works