David Luo
5/5
Exercise 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 know from lecture that the number of internal nodes with two external children is $N/4$. This means that $N/2$ external nodes come from these nodes with two children. There are $N+1$ external nodes, subtraction will give us the remaining number of external nodes, which must belong to internal nodes that have one children. This gives us $\frac{N+2}{2}$ internal nodes with 1 external child. With total N internal nodes and the number of nodes with 2 and 1 children known, we can also find the number of internal nodes without external children. This ends up being $N-\frac{N}{4}-\frac{N+2}{2}=\frac{N-4}{4}$ nodes. Roughly speaking, the proportions are a quarter with two, a quarter with none, and half with one.

would have preferred if you'd re-derived result about nodes w/ two external children

Below is me attempting to find the external path length of a random binary tree. I figured out that it was not relevant to the problem after I had finished the proof.

As before, our counting GF is the number of binary trees of size N: $T(z) \sim \frac{4^N}{\sqrt{\pi N^3}}$

Our cumulative cost GF is: $Q(z)=\sum_{t \in T} xpl(t) z^{\{t\}}$ where $\{t\} = \{t_L\} + \{t_R\}$ and $xpl(t) = xpl(t_L) + xpl(t_R) + \{t\}$

Using the lemmas $xpl(t) = ipl(t) + 2|t|$ and $\{t\} = |t|+1$, we can rewrite Q as: $Q(z)=\sum_{t \in T} ipl(t) z^{|t|+1} + 2\sum_{t \in T} |t| z^{|t|+1}$

We define the average xpl of a random N-node binary tree as: $\frac{[z^N]Q(z)}{T_N}$

Now let's decompose Q:

$$Q(z)=\sum_{t \in T} ipl(t) z^{|t|+1} + 2\sum_{t \in T} |t| z^{|t|+1}$$

$$= 1 + \sum_{t_L \in T}\sum_{t_R \in T} (ipl(t_L) + ipl(t_R) + |t|) z^{|t_L|+|t_R|+1} + 2\sum_{t \in T} |t| z^{|t|+1}$$

$$= 1 + 2zQ(z)T(z) + 2z^2 T'(z)T(z) + 2\frac{z}{(1-z)^2}$$

$$Q(z) = \frac{1+2z^2T'(z)T(z)+\frac{2z}{(1-z)^2}}{1-2zT(z)}$$

$$= \frac{1+2z^2\left(\frac{-1+\sqrt{1-4z}}{2z2}+\frac{1}{z\sqrt{1-4z}}\right)\left(\frac{1-\sqrt{1-4z}}{2z}\right)+\frac{2z}{(1-z)^2}}{\sqrt{1-4z}}$$

$$= \frac{2+\frac{2z-1+\sqrt{1-4z}}{z}+\frac{2z}{(1-z)^2}}{\sqrt{1-4z}} - 1$$

$$zQ(z) = \frac{4z-1+\sqrt{1-4z}+\frac{2z^2}{(1-z)^2}}{\sqrt{1-4z}} - 1$$

$$= \frac{4z}{\sqrt{1-4z}} - \frac{1}{\sqrt{1-4z}} + \frac{2z^2}{(1-z)^2\sqrt{1-4z}}$$

Now apply the radius of convergence theorem on the terms:

$$[z^N]Q(z) \sim \frac{4(\frac{1}{4})}{\sqrt{\pi}}(\tfrac{1}{4})^{-N}N^{-\frac{1}{2}} - \frac{1}{\sqrt{\pi}}(\tfrac{1}{4})^{-N}N^{-\frac{1}{2}} + \frac{2(\frac{1}{4})^2}{(1-\frac{1}{4})^2\sqrt{\pi}}(\tfrac{1}{4})^{-N}N^{-\frac{1}{2}}$$

$$= \frac{2}{9}\frac{4^N}{\sqrt{\pi N}}$$

$$Q_N/T_N \sim \frac{2}{9}N$$