ON THE SUM OF THE SIZES OF BINARY SUBTREES OF A PERFECT BINARY TREE

CYRIL BANDERIER

Personal note. 5, June 2000
Algorithms Project, INRIA
http://algo.inria.fr/banderier

This note answers to Florent Madeleine's question (private communication) "Can the sum of the sizes of binary subtrees of a perfect binary tree of size n be polynomially bounded ?"

I will show that the answer is no.

A binary tree is a tree whose each internal node have
– either two child which are internal nodes,
– or two child which are leaves.

Let b_k be the number of binary trees with k internal nodes. The set $\mathcal{B}$ of binary trees satisfies the following recursive definition

$$\mathcal{B} = \mathcal{L} \text{ ou } \mathcal{B} \times \mathcal{N} \times \mathcal{B}$$

(a binary tree is either a leaf $\mathcal{L}$, or a node $\mathcal{N}$ with two binary subtrees, which are themselves elements of $\mathcal{B}$). Hence, giving weight 0 to leaves and 1 to internal nodes, one gets a functional equation for the generating function $B(z)$:

$$B(z) = 1 + zB(z)^2.$$

Solving it gives the closed form formula

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

The smallest binary tree, which consist of a "root which is also a leaf", has height 0. Let $B_h(x) = \sum_{k\geq 0} b_{k,h} x^k$ be the generating function (in fact it is a polynomial!) for binary trees of height $\leq h$ (and where n codes the size of the tree, i.e. its number of internal nodes).

Then, one has the following recurrence:

$$B_h(x) = 1 + xB_{h-1}(x)^2 \qquad B_0(x) = 1.$$

A perfect binary tree (some authors also call it "complete" binary tree) is a binary tree for which all leaves are at the same height, say h. Consequently, a perfect binary tree has $2^h - 1$ internal nodes.

The number of binary subtrees of this tree is the number of binary trees of height $\leq h$, that is $B_h(1)$. The sum of the size of these trees is given by

$$\sum_{b\in\mathcal{B},\mathrm{height}(b)\leq h} |b| = \sum_{k\geq 0} k b_{k,h} = B_h'(1).$$

The asymptotics of the $b_{k,h}$'s is a non trivial problem, nevertheless it has been solved by Flajolet & Odlyzko in 1984. It is related to the behavior of the Mandelbrot fractal $z_{j+1} = z_j^2 + 1$ at $z_0 = 1$. The authors give

$$\max_{k\geq 0} b_{h,k} \sim 2^{-h/2} \exp(2^h 0.407354...)\ 0.685517...$$

Taking $n := 2^h - 1$ (input size, i.e., size of the perfect binary tree) leads to the upper bound

$$B'_h(1) \leq n \max_{k\geq 0} b_{h,k} = \frac{C \exp(0.407354...n)}{\sqrt{n}}$$

In conclusion, $B'_h(1) \approx B_h(1) \approx C_1^{(2^h)}$ where $C_1 = 1.50283680104975649975293642373...$ (in fact one can be more precise and give the asymptotic equivalent $B'_h(1) \sim C_2 2^h B_h(1)$ where $C_2 = 1.58990495515926...$)

Finally, one gets the following bounds for the number of subtrees of complete binary tree of size n (and thus of height $h = \lfloor \log_2 n \rfloor$) when $n > 2$

$$1.5^n < B'_h(1) < 1.51^n.$$

This excludes any polynomial bound.

The following Maple lines easily compute the first few terms

```
B:=1;L:=[ ]:L2:=[ ]:
to 9 do
  B := 1+x*B^2; L:=[op(L),subs(x=1,B)]:L2:=[op(L2),subs(x=1,diff(B,x))]
  od:
```

Thus, the value for $h \geq 0$ of $B_h(1)$ are: 1, 2, 5, 26, 677, 458330, 210066388901,
44127887745906175987802,
1947270476915296449559703445493848930452791205,
3791862310265926082868235028027893277370233152247388584761734150717768254410341175325352026,

and for $B'_h(1)$: 1, 1, 8, 105, 6136, 8766473, 8245941529080, 3508518207951157937469961,
311594265746788494170062926869662848646207622648,
1217308491239906829392988008143949647398943617188660186130545502913055217344025410733271773705.

Reference:

Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumeration. P. Flajolet and A. M. Odlyzko. In Math. Proc. Cambridge Phil. Soc., 96 (1984). pp. 237-253.