Counting Balanced Tree Shapes

Jeffrey A. Barnett*

Abstract

The number of shapes of trees tightly balanced by subtree size are counted. The branching factor is constant and the balance criterion is that sizes of sibling subtrees differ by no more than one node. The sequence of the number of tree shapes, with the branching factor fixed, for increasing tree size has a fractal-like graph. The reason for this appearance is explained.

Keywords: Balanced trees, tree shapes, combinatorial problems, data structures.

1 Preliminaries

Tree structures are used to organize data for efficient processing. Some restrictions on possible shapes must be imposed to guarantee that efficiency. One typical restriction is that sibling subtrees must have approximately the same height. Another is that sibling subtrees must have approximately the same size measured in nodes. Both restrictions, supplemented by suitable node labeling schemes, enable efficient insertion and retrieval algorithms. Such trees are said to be *balanced*. AVL trees [5], B-trees [1], and Red-Black trees [2] are some well-known examples of balanced trees. Knuth [3] provides algorithms and analyses for constructing, searching, and maintaining balanced trees of various sorts.

This article develops formulas that count the number of shapes of trees tightly balanced by subtree size. These trees will have a constant branching factor or width, w. So a (sub) tree is either "nothing" or a root node with w subtrees. A node is balanced if the difference

*The author's email is jbb@notatt.com.

in size of each pair of its subtrees is at most one node. A tree is balanced if all of its nodes are balanced. These tree are also height-balanced in the sense that variance over root-to-leaf path lengths is one node at most.

Here is the definition of '$\doteq$' the same shape predicate: 1) $\lambda \doteq \lambda$, where λ is the tree with no nodes; 2) if $u_i \doteq v_i$, where $1 \leq i \leq w$, u is a node with the ordered list of subtrees u_i, and v is a node with the ordered list of subtrees v_i, then $u \doteq v$. In other words, two shapes are the same if they look the same when drawn on paper. If $u \doteq v$, then u and v necessarily have the same number of nodes.

2 Recurrence Formula

Let $s_w(z)$, where $z \geq 0$ and $w \geq 1$, be the number of shapes of balanced trees with branching factor w and z nodes. It is easy to see that $s_1(z) = 1$ for all $z \geq 0$. Henceforth, assume $w > 1$.

Recurrence formulas used to calculate s_w are developed here. These formulas relate various values of $s_w(z)$ to $s_{w'}(z')$ where $w = w'$. In other words, there is formula for each $w > 1$. Clearly $s_w(0) = 1$ because the empty tree shape is unique and $s_w(1) = 1$ because the only shape with one node is a root with w empty (λ) children.

The recurrence relation for larger tree sizes is

$$s_w(nw + 1 + m) = \binom{w}{m} s_w(n+1)^m s_w(n)^{w-m}, \quad \text{where } 0 \leq m \leq w \text{ and } n \geq 0. \tag{1}$$

If a tree has $nw+1+m$ nodes, one node must be the root, m children of the root must have $n+1$ nodes, and the remaining $w-m$ children must have n nodes; the tree can be balanced in no other way. Now observe that 1) the m of w subtrees with $n+1$ nodes can be chosen in $\binom{w}{m}$ ways and 2) the choice of the shape of each subtree is independent of the others. So formula (1) follows. The case where $n = 0$ is of interest:

$$s_w(1+m) = \binom{w}{m} \qquad \text{where } 0 \leq m \leq w. \tag{2}$$

Sequences generated by $s_2, \ldots, s_7$ are registered on the OEIS Sequence Server [4] and initial values of $s_2, \ldots, s_6$ are shown in Table 1.

3 Divisibility Properties

The recurrence formula (1) clearly entails that $s_w(z)$ must have the form $\prod_{i=0}^{w} \binom{w}{i}^{x_i}$ for nonnegative integers x_i that depend on z. Since $\binom{2}{i}$ is 1 or 2 when $0 \leq i \leq 2$, $s_2(z)$ must be an integer power of 2. Similarly, $\binom{3}{i}$ is 1 or 3 when $0 \leq i \leq 3$ so $s_3(z)$ must be an integer power of 3. However, when $w > 3$, there are z such that $s_w(z)$ is not an integer power of w. For example, $\binom{w}{2} = w(w-1)/2$ is not a power of w when $w > 3$.

It is also straightforward to see from (1) that the sum of the $s_w(z)$ for $z = nw + 1 + m$ and $m = 0, \ldots, w$ is some integer raised to w^{th} power. To wit:

$$\begin{aligned}\sum_{m=0}^{w} s_w(nw+1+m) &= \sum_{m=0}^{w} \binom{w}{m} s_w(n+1)^m s_w(n)^{w-m} \\ &= \big(s_w(n+1) + s_w(n)\big)^w.\end{aligned}$$

4 Return to Unity

Define σ_w^n as the $n+1$ digit base-w number $1\ldots1$, i.e.,

$$\sigma_w^n = \sum_{i=0}^{n} w^i,$$

and note that $\sigma_w^{n+1} = w \cdot \sigma_w^n + 1$. Also note that $s_w(\sigma_w^0) = s_w(1) = 1$. Now assume $s_w(\sigma_w^n) = 1$ for all $0 \leq n \leq x$. So using (1)

$$\begin{aligned}s_w(\sigma_w^{x+1}) &= s_w(w\sigma_w^x + 1) \\ &= \binom{w}{0} s_w(\sigma_w^x + 1)^0 s_w(\sigma_w^x)^w \\ &= 1\end{aligned}$$

and this proves, by induction, that

Theorem 1. $s_w(\sigma_w^n) = 1$ *for all* $n \geq 0$.

The unique tree shape with σ_w^n nodes is the one where every root-to-leaf path is length n, i.e., the tree is perfectly balanced and the w children of each leaf—there are w^{n+1} in total—are all λ. As nodes are added, some of the λ are replaced with new leaves. When

w^{n+1} nodes have been added, the tree is again perfectly balanced with σ_w^{n+1} nodes and all root-to-leaf paths are now length $n+1$.

Theorem 2. *The sequence $s_w(\sigma_w^n), s_w(\sigma_w^n+1), \ldots, s_w(\sigma_w^{n+1})$ is symmetric in the sense that $s_w(\sigma_w^n+z) = s_w(\sigma_w^{n+1}-z)$ for all $n \geq 0$ and $0 \leq z \leq w^{n+1}$.*

Proof. If $z=0$ or $z=w^{n+1}$ the claim follows from Theorem 1. When $0 < z < w^{n+1}$, let $z = \sum_{i=1}^{n} z_i w^i$, where $0 \leq z_i < w$, then the claim is equivalent to the algebraic fact that

$$s_w\Big(\sigma_w^n + \sum_{i=0}^{n} z_i w^i\Big) = s_w\Big(\sigma_w^n + 1 + \sum_{i=0}^{n}(w-1-z_i)w^i\Big).$$

This proof is not, however, algebraic. The claim is established by showing a 1-to-1 correspondence between the tree shapes with $\sigma_w^n + z$ nodes and those with $\sigma_w^{n+1} - z$ nodes. Select a tree with $\sigma_w^n + z$ nodes and visit each of its w^n nodes at a distance of n from the root. Each such node has w children; in total z are leaf nodes and $w - z$ are λ. If a child is λ, replace it with a new leaf; if it is a leaf, replace it with λ. The modified tree has $\sigma_w^n + w^{n+1} - z = \sigma_w^{n+1} - z$ nodes. Any subtrees balanced before the transformation are balanced after and the transformation is clearly the sought-after 1-to-1 correspondence. □

Some implications of the fact that the sequence generated by s_w returns to unity and the symmetry demonstrated by Theorem 2 are discussed in Section 6.

5 Exact Count Formula

An exact non-recurrence formula to evaluate $s_w(z)$ is developed below. Let $n = \mathrm{L}_w(z)$, where $z > 0$ and $\mathrm{L}_w(z) = \lfloor \log_w(wz - z + 1) \rfloor - 1$. Note, if $\sigma_w^n \leq z < \sigma_w^{n+1}$ then $\mathrm{L}_w(z)$ is the minimum root-to-leaf path length in a balanced z-node tree. Represent z uniquely as

$$z = \sigma_w^n + z', \quad \text{where } z' = \sum_{i=0}^{n} z_i w^i \text{ and } 0 \leq z_i < w.$$

So $z' = z_n \ldots z_0$ is the base w representation of $z - \sigma_w^n$. Theorem 3 proves that

$$s_w(z) = \binom{w}{z_0} \prod_{i=1}^{n} \binom{w}{z_i+1}^{\mathrm{mod}(z',w^i)} \binom{w}{z_i}^{w^i - \mathrm{mod}(z',w^i)}, \tag{3}$$

where $\mathrm{mod}(z', w^i) = \sum_{j=0}^{i-1} z_j w^j$.

Theorem 3. *Formula* (3)*, where* $z > 0$*,* $n = \mathrm{L}_w(z)$*,* $z' = z - \sigma_w^n$*, and* $z_i = \mathrm{mod}(\lfloor z'/w^i \rfloor, w)$*, properly evaluates* $s_w(z)$*.*

Proof. This proof is by induction on n. For the base step let $z = \sigma_w^0 + z_0 = 1 + z_0$, where $0 \le z_0 \le w$, so $s_w(z) = \binom{w}{z_0}$ by (2) which agrees with (3) for $n = 0$. Now assume the claim is true for all $z < \sigma_w^{n+1}$ for some $n \ge 0$. Select any $\sigma_w^{n+1} \le z < \sigma_w^{n+2}$ and let $z' = z - \sigma_w^{n+1}$ so $z' = \sum_{i=0}^{n+1} z_i w^i$ where $0 \le z_i < w$, then use (1) to expand $s_w(z)$.

$$s_w(z) = \binom{w}{z_0} s_w\Big(\sigma_w^n + \sum_{i=0}^{n} z_{i+1} w^i + 1\Big)^{z_0} \times s_w\Big(\sigma_w^n + \sum_{i=0}^{n} z_{i+1} w^i\Big)^{w-z_0} \tag{4}$$

Next, the s_w terms on the right-hand side of the above are expanded then combined and simplified. The first s_w term might not straightforwardly satisfy the inductive assumption since $z_1 + 1 = w$ is possible. However, expansion can easily be accomplished via Theorem 2 followed by use of the inductive assumption. Let $z'' = \sum_{i=0}^{n} (w - 1 - z_{i+1}) w^i$, then

$$\begin{aligned}
s_w\Big(\sigma_w^n + \sum_{i=0}^{n} z_{i+1} w^i + 1\Big) &= s_w(\sigma_w^n + z'') \qquad \text{by Theorem 2} \\
&= \binom{w}{w-1-z_1} \prod_{i=1}^{n} \binom{w}{w - z_{i+1}}^{\mathrm{mod}(z'', w^i)} \prod_{i=1}^{n} \binom{w}{w-1-z_{i+1}}^{w^i - \mathrm{mod}(z'', w^i)} \\
&= \binom{w}{z_1 + 1} \prod_{i=1}^{n} \binom{w}{z_{i+1}}^{w^i - \mathrm{mod}(\lfloor z'/w \rfloor, w^i) - 1} \prod_{i=1}^{n} \binom{w}{z_{i+1}+1}^{\mathrm{mod}(\lfloor z'/w \rfloor, w^i) + 1}
\end{aligned} \tag{5}$$

The expansion of the second s_w term of (4), using the inductive assumption is

$$s_w\Big(\sigma_w^n + \sum_{i=0}^{n} z_{i+1} w^i\Big) = \binom{w}{z_1} \prod_{i=1}^{n} \binom{w}{z_{i+1}+1}^{\mathrm{mod}(\lfloor z'/w \rfloor, w^i)} \prod_{i=1}^{n} \binom{w}{z_{i+1}}^{w^i - \mathrm{mod}(\lfloor z'/w \rfloor, w^i)} \tag{6}$$

Now substitute (5) and (6) in (4) and simplify to show agreement with (3).

$$\begin{aligned}
s_w(z) &= \binom{w}{z_0} \left[\binom{w}{z_1+1} \prod_{i=1}^{n} \binom{w}{z_{i+1}+1}^{\mathrm{mod}(\lfloor z'/w \rfloor, w^i)+1} \prod_{i=1}^{n} \binom{w}{z_{i+1}}^{w^i - \mathrm{mod}(\lfloor z'/w \rfloor, w^i) - 1} \right]^{z_0} \\
&\qquad \times \left[\binom{w}{z_1} \prod_{i=1}^{n} \binom{w}{z_{i+1}+1}^{\mathrm{mod}(\lfloor z'/w \rfloor, w^i)} \prod_{i=1}^{n} \binom{w}{z_{i+1}}^{w^i - \mathrm{mod}(\lfloor z'/w \rfloor, w^i)} \right]^{w-z_0} \\
&= \binom{w}{z_0} \prod_{i=1}^{n+1} \binom{w}{z_i + 1}^{\mathrm{mod}(z', w^i)} \binom{w}{z_i}^{w^i - \mathrm{mod}(z', w^i)}
\end{aligned}$$

□

6 Discussion

Table 1 provides initial values of the sequences $s_w(z)$, where $z = 1, 2, \ldots$, for some small w and Figures 1–6 plot $\log(s_w(z))$; logarithms are used to enhance the observable detail. Each graph has a repetitive, somewhat fractal-like appearance and it is straightforward to see why. Consider the sequence $s_w(z)$, where $z = \sigma_w^{n+1}, \ldots, \sigma_w^{n+2} - 1$, i.e., the sequence $s_w(\sigma_w^{n+1} + z')$, where $z' = 0, \ldots, w^{n+2} - 1$. So z' can be written uniquely as $\sum_{i=0}^{n+1} z_i w^i$, where $0 \le z_i < w$. Now let $h = h(z) = z_{n+1}$ and $\ell = \ell(z) = \mathrm{mod}(z', w^{n+1})$ and substitute in (3).

$$\begin{aligned}
s_w(z) &= \binom{w}{z_0} \prod_{i=1}^{n+1} \binom{w}{z_i + 1}^{\mathrm{mod}(z',w^i)} \binom{w}{z_i}^{w^i - \mathrm{mod}(z',w^i)} \\
&= \binom{w}{z_0} \prod_{i=1}^{n} \binom{w}{z_i + 1}^{\mathrm{mod}(z',w^i)} \binom{w}{z_i}^{w^i - \mathrm{mod}(z',w^i)} \\
&\qquad \times \binom{w}{h+1}^{\mathrm{mod}(z',w^{n+1})} \binom{w}{h}^{w^{n+1} - \mathrm{mod}(z',w^{n+1})} \\
&= s_w(\sigma_w^n + \ell) \binom{w}{h+1}^{\ell} \binom{w}{h}^{w^{n+1} - \ell} \\
&= \left(\frac{w-h}{h+1}\right)^{\ell} \binom{w}{h}^{w^{n+1}} s_w(\sigma_w^n + \ell). \qquad (7)
\end{aligned}$$

This sequence can be divided into w sequences, each associated with a different h value. Each short sequence has the form $s_w(\sigma_w^{n+1} + hw^{n+1} + \ell)$, where $0 \le \ell < w^{n+1} - 1$; This can be written as $s_w(z) = c_1^{\ell} c_2 s_w(\sigma_w^n + \ell)$, where c_1 and c_2 are constants for a given h value according to (7). Whether the shape of $s_w(\sigma_w^n + \ell)$ is stretched upwards for increasing values of ℓ, just magnified by c_2, or stretched downwards depends on whether c_1 is greater, equal to, or less than 1, i.e., on whether $\frac{w-1}{2}$ is greater, equal to, or less than h. Each of the w pieces of this graph segment is again, a recapitulation of w magnified and possibly stretched instances of $s_w(\sigma_w^{n-1} + \mathrm{mod}(\ell, w^{n-1}))$. Decomposition can be continued until the basic building blocks of the graph are encountered: the sequence $s_w(\sigma_w^0), \ldots, s_w(\sigma_w^1 - 1) = \binom{w}{0}, \ldots, \binom{w}{w-1}$.

The graph of s_2 in Figure 1 resembles the Blancmange and Togi curves. See [6] or [7] for more information.

References

[1] R. Bayer amd E. M. McCreight, Organization and maintenance of large ordered indexes, *Acta Informatica*, 1(3), pp. 173–189, 1972.

[2] T. H. Cormen, C. E. Leiserson, R. L. Rivest, & C. Stein: *Introduction to Algorithms*, 2^{nd} Edition, MIT Press and McGraw-Hill, 2001. See Chapter 13, Red-Black Trees.

[3] D. E. Knuth, *The Art of Computer Programming, Vol. 3, Sorting and Searching*, 2^{nd} Edition, Addison Wesely, 1998. See Section 6.2, Searching by comparison of keys.

[4] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, see Sequences A110316, A131889, A131890, A131891, A131892, and A131893 at https://oeis.org, 2007.

[5] G. Adelson-Velskii and E. M. Landis, An algorithm for the organization of information, *Doklady Akademii Nauk SSSR*, 146 pp. 263-266, 1962 (Russian). English translation by Myron J. Ricci in *Soviet Math. Doklady*, 3, pp. 1259-1263, 1962.

[6] Weisstein, E. W. "Blancmange Function." From MathWorld–A Wolfram Web Resource at http://mathworld.wolfram.com/BlancmangeFunction.html, 2019.

[7] Wikipedia the Free Encyclopedia, at https://en.wikipedia.org/wiki/Blancmange_curve, 2019.

Table 1: Some values of $s_w(z)$ for small w and z.

$s_2(\sigma_2^0), \ldots, s_2(\sigma_2^4)$
1, 2, 1, 4, 4, 4, 1, 8, 16, 32, 16, 32, 16, 8, 1, 16, 64, 256, 256, 1024, 1024, 1024, 256, 1024, 1024, 1024, 256, 256, 64, 16, 1
$s_2(\sigma_2^4), \ldots, s_2(\sigma_2^5)$
1, 32, 256, 2048, 4096, 32768, 65536, 131072, 65536, 524288, 1048576, 2097152, 1048576, 2097152, 1048576, 524288, 65536, 524288, 1048576, 2097152, 1048576, 2097152, 1048576, 524288, 65536, 131072, 65536, 32768, 4096, 2048, 256, 32, 1
$s_3(\sigma_3^0), \ldots, s_3(\sigma_3^3)$
1, 3, 3, 1, 9, 27, 27, 81, 81, 27, 27, 9, 1, 27, 243, 729, 561, 19683, 19683, 59049, 59049, 19683, 177147, 531441, 531441, 1594323, 1594323, 531441, 531441, 177147, 19683, 59049, 59049, 19683, 19683, 6561, 729, 243, 27, 1
$s_4(\sigma_4^0), \ldots, s_4(\sigma_4^2)$
1, 4, 6, 4, 1, 16, 96, 256, 256, 1536, 3456, 3456, 1296, 3456, 3456, 1536, 256, 256, 96, 16, 1
$s_5(\sigma_5^0), \ldots, s_5(\sigma_5^2)$
1, 5, 10, 10, 5, 1, 25, 250, 1250, 3125, 3125, 31250, 125000, 250000, 250000, 100000, 500000, 1000000, 1000000, 500000, 100000, 250000, 250000, 125000, 31250, 3125, 3125, 1250, 250, 25, 1
$s_6(\sigma_6^0), \ldots, s_6(\sigma_6^2)$
1, 6, 15, 20, 15, 6, 1, 36, 540, 4320, 19440, 46656, 46656, 699840, 4374000, 14580000, 27337500, 27337500, 11390625, 91125000, 303750000, 540000000, 540000000, 288000000, 64000000, 288000000, 540000000, 540000000, 303750000, 91125000, 11390625, 27337500, 27337500, 14580000, 4374000, 699840, 46656, 46656, 19440, 4320, 540, 36, 1

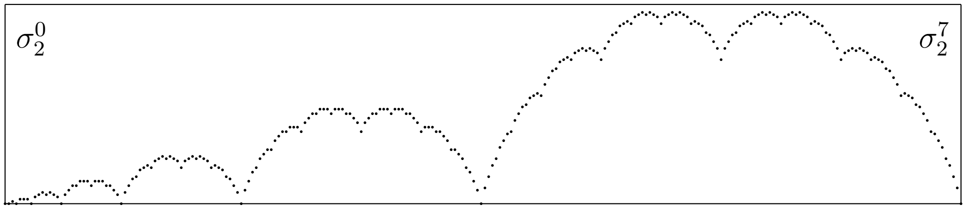

Figure 1: Initial graph of $\log(s_2(z))$

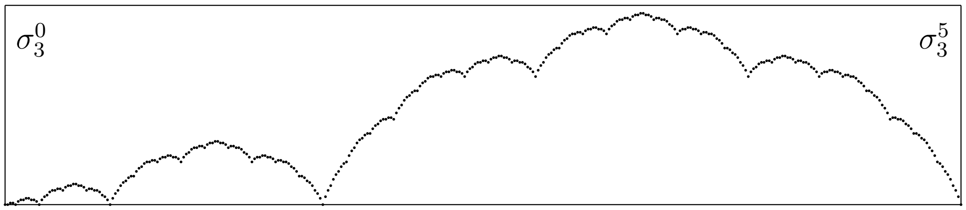

Figure 2: Initial graph of $\log(s_3(z))$

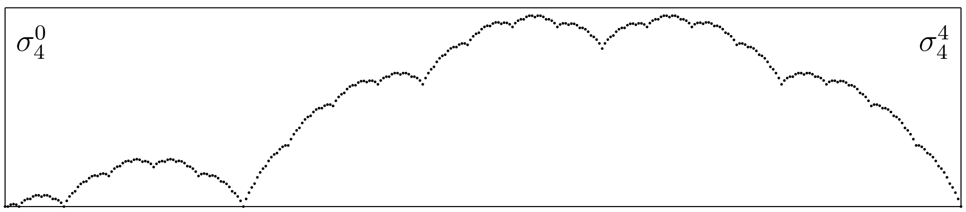

Figure 3: Initial graph of $\log(s_4(z))$

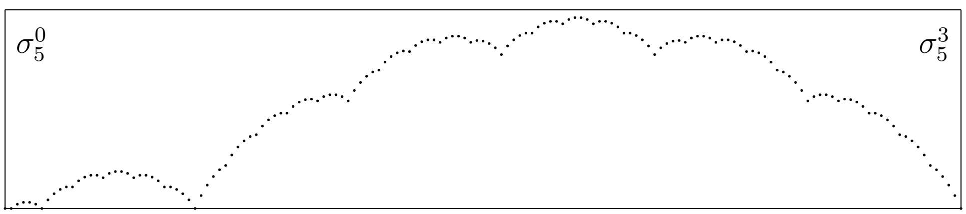

Figure 4: Initial graph of $\log(s_5(z))$

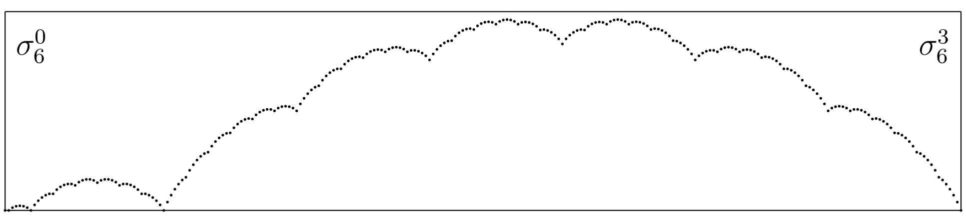

Figure 5: Initial graph of $\log(s_6(z))$

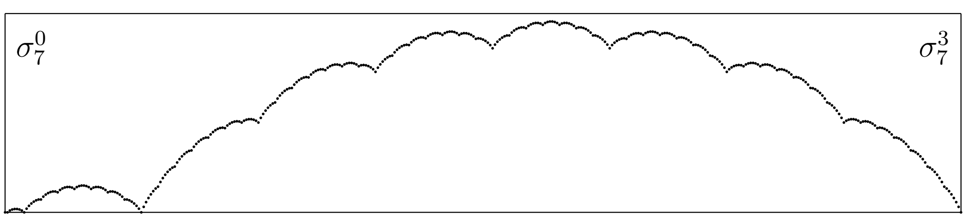

Figure 6: Initial graph of $\log(s_7(z))$