On Symmetries of Non-Plane Trees in a Non-Uniform Model

Jacek Cichoń, Abram Magner, Wojciech Szpankowski, and Krzysztof Turowski

JCI: Wrocław University of Science and Technology, Poland
AMA: University of Illinois at Urbana-Champaign, USA
WSZ: Purdue University, USA
KTU: Gdańsk University of Technology, Poland

SODA/ANALCO,January, 2017

Considered Objects

Class of trees

- We consider non-plane unlabeled binary trees (each node has either zero or exactly two children) [Otter trees]
- Wedderburn–Etherington numbers:

$$a_n \sim 0.3188 \frac{1}{n^{3/2}} \left(\frac{1}{0.4027} \right)^n$$

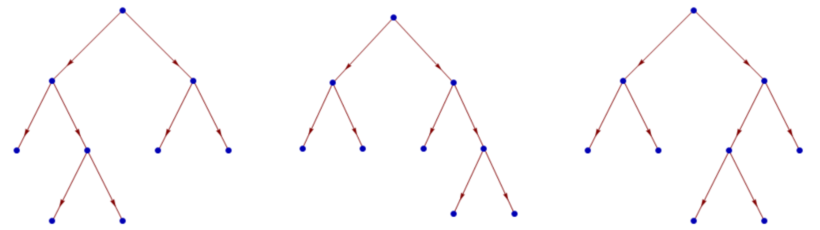
- W-E numbers count the number of Otter trees with n leaves

Remark: The numbers b_n of binary plane trees with n leaves satisfy

$$b_n \sim \frac{1}{2\sqrt{\pi}} \frac{1}{n^{3/2}} 4^n$$

Plane and Non-Plane Trees

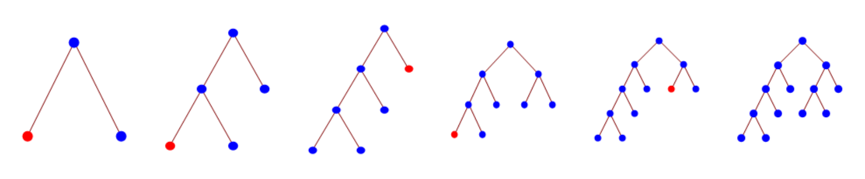
Each of this plane trees corresponds with one Non-Plane tree



- T_n - set of binary plane trees with n leaves
- S_n - set of binary non-plane trees with n leaves
- $[s]_\sim = \{t \in T_n : t \sim s\}$, for $s \in S_n$

Probability

Generation of plane trees



Basic formula

For $t \in T_n$ we have

$$\Pr[T_n = t] = \prod_{v \in t^o} \frac{1}{\Delta(v) - 1}$$

where t^o is the set of interval nodes of t, $\Delta(v)$ is the number of leaves of a tree rooted at v.

Probability - II

Remark: generate randomly binary search tree from random permutation, make "de-labelization"; we get the same probability model.

Recurrence

If $t = t_1 \star t_2 \in T_n$, then

$$\Pr[T_n = t] = \frac{1}{n-1} \Pr[T_{\Delta(t_1)} = t_1] \Pr[T_{\Delta(t_2)} = t_2]$$

where $\Delta(s)$ is the number of leaves in s

Connection between S and T

$$\Pr[S_n = s] = card([s]_\sim) \cdot \Pr[T_n = t], \quad t \in [s]_\sim$$

Symmetries

Definition

$sym(t)$ = the number of non-leaf (internal) nodes v of tree t such that the two subtrees stemming from v are isomorphic.

Basic property

$$card([s]_\sim) = 2^{n-1-sym(s)}$$

Basic recurrence

$$sym(s_1 \star s_2) = \begin{cases} sym(t_1) + sym(t_2) + 1 & : t_1 = t_2 \\ sym(t_1) + sym(t_2) & : t_1 \neq t_2 \end{cases}$$

Generating functions

Two basic generating functions

- $F(u,z) = \sum_{t \in T} \Pr[T = t] u^{sym(t)} z^{|t|}$
- $B(u,z) = \sum_{t \in T} \Pr[T = t]^2 u^{sym(u)} z^{|t|-1}$

Theorem

Let $f(u,z) = \frac{F(u,z)}{z}$. *Then*

$$\frac{\partial f(u,z)}{\partial z} = f(u,z)^2 + (u-1)B(u^2, z^2)$$

(Riccati differential equation)

Number of symmetries

Definition

$$\mathcal{E}(z) = \sum_{n\geq 1} \mathbb{E}[sym(S_n)]z^n$$

Theorem

Let $B(z) = \sum_{t\in T} \Pr[T = t]^2 z^{|t|-1} \quad (= \sum_n b_n z^n)$. *Then*

$$\mathcal{E}'(z) = \frac{2\mathcal{E}(z)}{z(1-z)} + B(z^2)$$

We should know the behavior of $B(z) = \sum_n b_n z^n$. We can calculate $b_1, b_2, b_3, \ldots$:

$$1, 1, \frac{1}{2}, \frac{2}{9}, \frac{13}{144}, \frac{7}{200}, \frac{851}{64800}, \frac{13}{2700}, \frac{1199}{691200}, \frac{2071}{3359232}$$

Extraction of coefficients of function $B(z)$

We put

$$C(z) = zB(z)$$

Differential equation

$$C(z) - zC'(z) + z^2 C''(z) = C^2(z)$$

Recurrence

$$c_n = \frac{1}{(n-1)^2} \sum_{k=1}^{n-1} c_k c_{n-k}$$

Solution of recurrence

Recurrence

$$c_n = \frac{1}{(n-1)^2} \sum_{k=1}^{n-1} c_k c_{n-k}$$

Numerical computations: $b_n = c_{n+1} \approx (\frac{1}{3.14})^n \cdot 6n$

Solution of recurrence

Recurrence

$$c_n = \frac{1}{(n-1)^2} \sum_{k=1}^{n-1} c_k c_{n-k}$$

Numerical computations: $b_n = c_{n+1} \approx (\frac{1}{3.14})^n \cdot 6n$

SOLUTION !!!

H-H Chern, M. Fernández-Camacho, H-K. Hwang, and C. Martinez, *Psi-series method for equality of random trees and quadratic convolution recurrences*, 2012:

$$b_n = \rho^n \left(6n - \frac{22}{5} + O(n^{-5})\right)$$

where $\rho = 0.3183843834378459\ldots$.

Solution of differential equation

- we defined: $\mathcal{E}(z) = \sum_{n\geq 1} \mathbb{E}[sym(S_n)]z^n$
- we know that: $\mathcal{E}'(z) = \frac{2\mathcal{E}(z)}{z(1-z)} + B(z^2)$
- we know a lot about $B(z) = \sum_n b_n z^n$

Theorem

$$\mathrm{E}[sym(S_n)] = n \sum_{k=1}^{\left\lfloor \frac{n+1}{2} \right\rfloor} \frac{b_k}{(2k-1)k(2k+1)} + (-1)^{n+1} b_{\left\lfloor \frac{n+1}{2} \right\rfloor}$$

hence

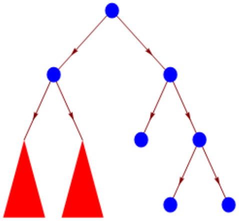
$$\mathrm{E}[sym(S_n)] = n \cdot (0.3725463659 \pm 10^{-10})$$

Application - compression

We know that $\mathrm{E}[sym(S_n)] \approx 0.3725 \cdot n$

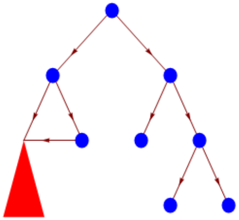
Simple compression algorithm

If you find a symmetric inner node, replace one of its sub-trees by a pointer. Let $size(S_n)$ denote the size of generated structure.



$$\mathbb{E}[size(S_n)] = n \sum_{k=1}^{\lfloor \frac{n+1}{2} \rfloor} \frac{b_k}{(2k-1)(2k+1)}$$

$$\approx 0.4190 \cdot n$$



Application - entropy

We know that

- $H[S_n] = H[T_n] - H[T_n|S_n]$
- $H[T_n] = \log_2(n-1) + 2n\sum_{k=2}^{n-1} \frac{\log_2(k-1)}{k(k+1)}$
- $2\sum_{k=2}^{n-1} \frac{\log_2(k-1)}{k(k+1)} \approx 1.736$ (for $n \geq 10^5$)
- $H[T_n|S_n] = \ldots = \sum_{s\in S_n} \Pr[S_n = s] \log_2(card([s]_\sim)) = \ldots n - 1 - E[sym(T_n)]$

Theorem

$$\lim_{n\to\infty} \frac{H[S_n]}{n} = 1.109\ldots$$

This is the end

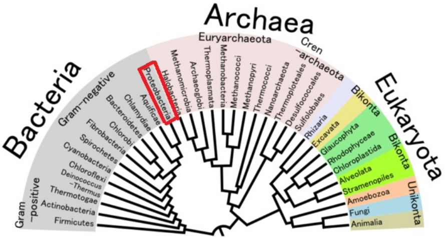


Figure 1: Phylogenetic (evolutionary) Tree

Thank You