Some simple varieties of trees arising in permutation analysis

Mathilde Bouvel
LaBRI & CNRS
Université Bordeaux
mathilde.bouvel@labri.fr

Marni Mishna
Dept. Mathematics
Simon Fraser University
mmishna@sfu.ca

Cyril Nicaud
LIGM & CNRS
Université Paris-Est
nicaud@univ-mlv.fr

Challenge

Understand the convergence of a nested family of trees towards permutations

Permutations as a limit of tree classes

In this work we define a nested family of parameterized trees $\mathcal{P}^{(k)}$ whose limit $\mathcal{P}$ is in a straightforward bijection with permutations. The trees are decorated with subclasses of simple permutations and are known as the strong interval trees of permutations. We give a specification for $\mathcal{P}^{(k)}$, and hence have easy access to asymptotic counting and parameter formulas, as well as random generation. The limits in k and n do not commute, and this leads to some interesting behaviour in the asymptotics.

$$\lim_{k\to\infty} P_n^{(k)} = P_n = n!$$

Simple Permutations

Simple permutation property: no set of consecutive numbers are together in a block.

$$\sigma = 7\,9\,4\,2\,5\,1\,3\,8\,6 \implies \text{not simple}$$

$$\pi = 8\,4\,2\,7\,1\,3\,6\,9\,5 \implies \text{simple}$$

$$\mathcal{S} = \text{set of all simple permutations}$$
$$= \{2413, 3142, 24153, 25314, 35142, 31524, \dots\}$$
$$S(z) = 2z^4 + 6z^5 + 46z^6 + 338z^7 + 2926z^8 + 28146z^9 + \dots$$

The number of simple permutations of size n: $s_n \sim n!/e^2$. [2]
oeis.org/A111111

Open problem: Describe a "nice" generation scheme for simple permutations (i.e. not rejection).

$\mathcal{P}$: The class of Strong Interval Trees

1. Size is the number of leaves
2. Three colours for internal vertices:

   Linear nodes: $\color{red}\bullet\color{blue}\bullet$ children are totally ordered.

   Prime nodes: $\bullet$ children are totally ordered, and are decorated with a simple permutation
3. There is only choice between linear nodes at the root, or as a child of a prime node.

$$\mathcal{P} = \mathcal{Z}_{\square} + \mathcal{N}_{\color{red}\bullet} \cdot \mathrm{SEQ}_{\geq 2}\, \mathcal{U}_{\color{red}\bullet} + \mathcal{N}_{\color{blue}\bullet} \cdot \mathrm{SEQ}_{\geq 2}\, \mathcal{U}_{\color{blue}\bullet} + \mathcal{N}_{\bullet} \cdot S(\mathcal{P}),$$
$$\mathcal{U}_{\color{red}\bullet} = \mathcal{Z}_{\square} + \mathcal{N}_{\color{blue}\bullet} \cdot \mathrm{SEQ}_{\geq 2}\, \mathcal{U}_{\color{blue}\bullet} + \mathcal{N}_{\bullet} \cdot S(\mathcal{P}),$$
$$\mathcal{U}_{\color{blue}\bullet} = \mathcal{Z}_{\square} + \mathcal{N}_{\color{red}\bullet} \cdot \mathrm{SEQ}_{\geq 2}\, \mathcal{U}_{\color{red}\bullet} + \mathcal{N}_{\bullet} \cdot S(\mathcal{P}).$$

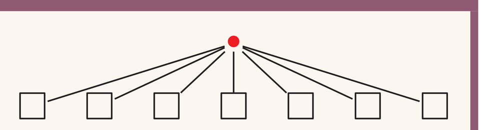
Lemma. (Simplification of system)

$$\mathcal{P} \equiv \mathrm{SEQ}_{\geq 1}\, \mathcal{U} \qquad \text{with } \Lambda(z) = \frac{z^2}{1-z} + S\Big(\frac{z}{1-z}\Big)$$
$$\mathcal{U} \equiv \mathcal{Z} + \Lambda(\mathcal{U})$$

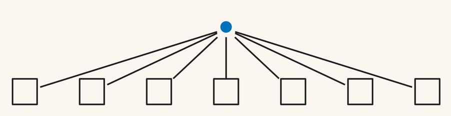
⚠ $\Lambda(z)$ is not analytic

What do the trees look like?

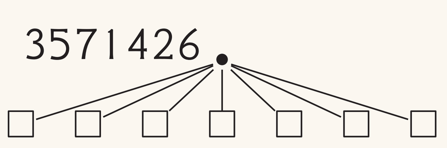
$\sigma = 1\,2\,3\,4\,5\,6\,7$ (identity)



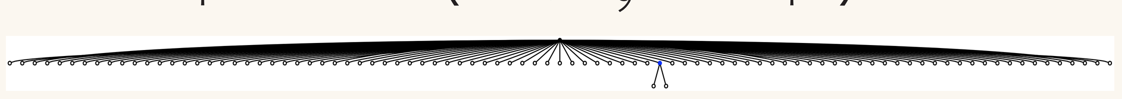
$\sigma = 7\,6\,5\,4\,3\,2\,1$ (reverse identity)



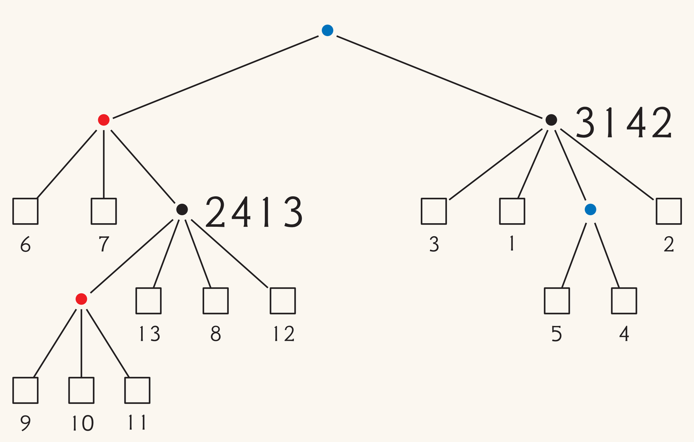
$\sigma = 3\,5\,7\,1\,4\,2\,6$ (simple)



A random permutation: (recall, $\sim \frac{1}{9}$ are simple)



$\sigma = 6\ 7\ 9\ 10\ 11\ 13\ 8\ 12\ 3\ 1\ 5\ 4\ 2$



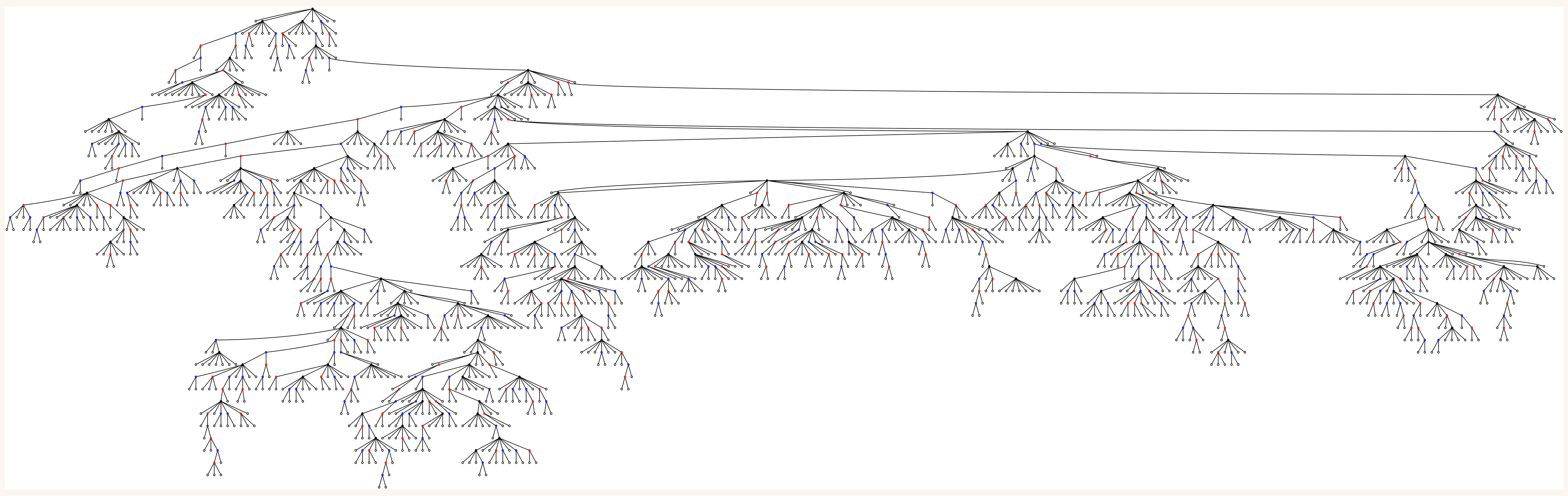
This tree models instances of perfect sorting by reversals [4]

Prime node degree restricted trees: A simple variety of trees

$\mathcal{P}^{(k)}$-trees: The sub-class of $\mathcal{P}$ where prime nodes have at most k children. This class is a simple variety of trees for each case and hence we have asymptotic enumeration and random generation for "free".

$$\mathcal{P}^{(k)} \equiv \mathrm{SEQ}_{\geq 1}\, \mathcal{U}^{(k)}$$
$$\mathcal{U}^{(k)} \equiv \mathcal{Z} + \mathrm{SEQ}_{\geq 2}(\mathcal{U}^{(k)}) + 4\,(\mathcal{P}^{(k)})^4 + \dots + s_k\,(\mathcal{P}^{(k)})^k$$

A random tree from $\mathcal{P}^{(7)}$ of size approx 1000 constructed from a Boltzmann generator



Asymptotic enumeration

Use adapted inversion formula [8] on $\mathcal{U}^{(k)} = \mathcal{Z} + \Lambda_k(\mathcal{U}^{(k)})$ with $\Lambda_k(z) = \frac{z}{1-z^2} + \sum_{j=4}^{k} s_j (\frac{z}{1-z})^j$

Theorem.

$$U_n^{(k)} \sim P_n^{(k)} \sim \sqrt{\frac{\rho_k}{2\pi\Lambda_k''(\tau_k)}} \cdot \frac{\rho_k^{-n}}{n^{3/2}}$$

Here:

$$\Lambda_k'(\tau_k) = 1 \qquad \rho_k = \Lambda_k(\tau_k) \qquad \rho_k < \tau_k < \frac{e}{k}$$

Remark $\lim_{k\to\infty} \rho_k = 0$

Tree Anatomy

Theorem. Average parameter values for trees in $\mathcal{P}_n^{(k)}$

# vertices of arity j	$\frac{\lambda_j \tau_k^j}{\rho_k} \cdot n$
# of internal vertices	$\frac{\tau_k - \rho_k}{\rho_k} \cdot n$
Sum of subtree size	$\sqrt{\frac{\pi}{2\rho_k \Lambda_k''(\tau_k)}} \cdot n^{3/2}$

Application to genome reconstruction

This permutation data structure models the analysis of perfect sorting by reversals, which is used in genome comparison. Tree parameters such as the number of internal nodes have direct interpretations on the evolutionary scenarios they represent [6]. We aspire to describe better model for the permutations arising in actual genome comparisons, and this is a first step towards that goal.

Bounds on ρ_k and a familiar limit

Theorem.

$$\rho_k = \frac{e}{k}\left(1 - \frac{5}{2}\frac{\log k}{k} + \Theta\big(\tfrac{1}{k}\big)\right)$$

We see how trees become permutations when $k = n$:

$$\sqrt{\frac{\rho_k}{2\pi\Lambda_k''(\tau_k)}} \cdot \rho_k^{-n} n^{-3/2} \leq \left(\frac{e}{4k\pi}\right)^{\frac{1}{2}} \left(\frac{k}{e}\right)^n \left(1 + \frac{5}{2}\frac{\log k}{k} + \Theta\left(\frac{1}{k}\right)\right)^n n^{-3/2}.$$

Trees		Permutations
$\gamma\rho^{-n} n^{-3/2}$	$\to$	$(n/e)^n \sqrt{2\pi n}$

Subtleties in the asymptotics

For any fixed k, as $n \to \infty$ this asymptotic formula is very accurate. However, $P_n^{(2n)} = n!$, but the tree formula produces an extra exponential factor of 2. Taking additional terms in the expansion does not help. Rather, new contours are required in the intermediate integrals, demonstrating the limits of the inversion formula.

References

[1] M.H. Albert and M.D. Atkinson. Simple permutations and pattern restricted permutations. *Discrete Math.*, 300:1–15, 2005.

[2] M.H. Albert, M.D. Atkinson, and M. Klazar. The enumeration of simple permutations. *J. Integer Seq.*, 6:03.4.4, 2003.

[3] S. Bérard, A. Bergeron, C. Chauve, and C. Paul. Perfect sorting by reversals is not always difficult. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 4:4–16, 2007.

[4] A. Bergeron, C. Chauve, F. de Montgolfier, and M. Raffinot. Computing common intervals of K permutations, with applications to modular decomposition of graphs. *Proc. 13th Annual European Symposium on Algorithms, in Lecture Notes in Comput. Sci.*, 3669:779–790, 2005.

[5] K. S. Booth and G. S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using *PQ*-tree algorithms. *J. Comput. System Sci.*, 13(3):335–379, 1976.

[6] M. Bouvel, C. Chauve, M. Mishna, and D. Rossin. Average-case analysis of perfect sorting by reversals. *Discrete Math. Algorithms Appl.*, 3(3):369–392, 2011.

[7] G. Chapuy, A. Pierrot, D. Rossin. On growth rate of wreath-closed permutation classes. Talk at the conference *Permutation Patterns*, 2011.

[8] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

Acknowledgements. MM gratefully acknowledges the support of NSERC Discovery grant, and both LIGM and LaBRI for hosting during the work. MB + CN were partially supported by ANR project Magnum (2010-BLAN-0204). We are also thankful to Cedric Chauve, Carine Pivoteau, and Rosemary McCloskey for work related to this project.