

# UNIVERSAL LIMITS OF SUBSTITUTION-CLOSED PERMUTATION CLASSES

FRÉDÉRIQUE BASSINO, MATHILDE BOUVEL, VALENTIN FÉRAY, LUCAS GERIN,  
MICKAËL MAAZOUN, AND ADELIN PIERROT

**ABSTRACT.** We consider uniform random permutations in proper substitution-closed classes and study their limiting behavior in the sense of permutons.

The limit depends on the generating series of the simple permutations in the class. Under a mild sufficient condition, the limit is an elementary one-parameter deformation of the limit of uniform separable permutations, previously identified as the Brownian separable permuton. This limiting object is therefore in some sense universal. We identify two other regimes with different limiting objects. The first one is degenerate; the second one is nontrivial and related to stable trees.

These results are obtained thanks to a characterization of the convergence of random permutons through the convergence of their expected pattern densities. The limit of expected pattern densities is then computed by using the substitution tree encoding of permutations and performing singularity analysis on the tree series.

## CONTENTS

1. Introduction . . . . .	2
1.1. Permutation classes and their limit . . . . .	2
1.2. The permuton viewpoint . . . . .	4
1.3. Substitution-closed classes . . . . .	5
1.4. Our results: Universality . . . . .	8
1.5. Our results: Beyond universality . . . . .	10
1.6. Limits of proportions of pattern occurrences . . . . .	11
1.7. Outline of the proof . . . . .	12
1.8. Organization of the paper . . . . .	14
2. Convergence of random permutons . . . . .	15
2.1. Deterministic permutons and extracted permutations . . . . .	15
2.2. Random permutons and convergence in distribution . . . . .	16
3. Coding permutations by trees . . . . .	20
3.1. Substitution trees . . . . .	20
3.2. Induced trees . . . . .	23
4. Exact enumeration of various families of trees . . . . .	26
4.1. Generating functions of $\mathcal{S}$ -canonical trees (possibly with marked leaves) . . . . .	26
4.2. Generating function counting trees with marked leaves inducing a given tree . . . . .	28

---

2010 *Mathematics Subject Classification.* 60C05, 05A05.

*Key words and phrases.* Combinatorial probability, scaling limits of combinatorial structures, universality phenomena, Brownian and stable objects, permutation patterns, permutons, analytic combinatorics.

5. Asymptotic analysis: The standard case $S'(R_S) > 2/(1 + R_S)^2 - 1$ . . . . .	34
5.1. Definition of the biased Brownian separable permuton and statement of the theorem . . . . .	34
5.2. Asymptotics of the generating function of trees with no or one marked leaf . . . . .	37
5.3. Asymptotics of the generating function of marked trees with a given induced tree . . . . .	40
5.4. Probability of tree patterns . . . . .	42
5.5. Back to permutations . . . . .	42
5.6. Occurrences of nonseparable patterns . . . . .	44
6. Asymptotic analysis: The degenerate case $S'(R_S) < 2/(1 + R_S)^2 - 1$ . . . . .	46
6.1. Asymptotic behavior of the main series . . . . .	46
6.2. Probability of given patterns . . . . .	48
6.3. Hypothesis (CS) and convergence of uniform random simple permutations . . . . .	51
7. Asymptotic analysis: The critical case $S'(R_S) = 2/(1 + R_S)^2 - 1$ . . . . .	53
7.1. The case $\delta \in (1, 2)$ . . . . .	54
7.2. The case $\delta > 2$ . . . . .	56
Appendix A. Complex analysis toolbox . . . . .	57
A.1. Aperiodicity and Daffodil Lemma . . . . .	57
A.2. Transfer theorem . . . . .	57
A.3. Singular differentiation . . . . .	58
A.4. Exponents of dominant singularity . . . . .	58
A.5. An analytic implicit function theorem . . . . .	60
A.6. Proof of Lemma 6.3 . . . . .	60
A.7. Proof of Lemma 7.3 . . . . .	62
Appendix B. On the simulations given in the introduction . . . . .	65
B.1. Biased Brownian permuton . . . . .	65
B.2. Stable permutons . . . . .	65
References . . . . .	66

## 1. INTRODUCTION

The aim of this paper is to study the asymptotic behavior of a permutation of large size, picked uniformly at random in a *substitution-closed permutation class* generated by a given (finite or infinite) family of *simple permutations* satisfying additional conditions. We first give a few definitions necessary to present the recent literature on related problems, and to state our results.

**1.1. Permutation classes and their limit.** For any positive integer  $n$ , the set of permutations of  $[n] := \{1, 2, \dots, n\}$  is denoted by  $\mathfrak{S}_n$ . We write permutations of  $\mathfrak{S}_n$  in one-line notation as  $\sigma = \sigma(1)\sigma(2) \dots \sigma(n)$ . For a permutation  $\sigma$  in  $\mathfrak{S}_n$ , the *size*  $n$  of  $\sigma$  is denoted by  $|\sigma|$ .

For  $\sigma \in \mathfrak{S}_n$ , and  $I \subset [n]$  of cardinality  $k$ , let  $\text{pat}_I(\sigma)$  be the permutation of  $\mathfrak{S}_k$  induced by  $\{\sigma(i) : i \in I\}$ . For example for  $\sigma = 65831247$  and  $I = \{2, 5, 7\}$  we have

$$\text{pat}_{\{2,5,7\}}(65831247) = 312$$

since the values in the subsequence  $\sigma(2)\sigma(5)\sigma(7) = 514$  are in the same relative order as in the permutation 312. A permutation  $\pi = \text{pat}_I(\sigma)$  is a *pattern* involved (or contained) in  $\sigma$ , and the subsequence  $(\sigma(i))_{i \in I}$  is an *occurrence* of  $\pi$  in  $\sigma$ . When a pattern  $\pi$  has no occurrence in  $\sigma$ , we say that  $\sigma$  *avoids*  $\pi$ . The pattern containment relation defines a partial order on  $\mathfrak{S} = \cup_n \mathfrak{S}_n$ : we write  $\pi \preceq \sigma$  if  $\pi$  is a pattern of  $\sigma$ .

A *permutation class* is a family  $\mathcal{C}$  of permutations that is downward closed for  $\preceq$ , i.e. for any  $\sigma \in \mathcal{C}$  and any pattern  $\pi \preceq \sigma$ , it holds that  $\pi \in \mathcal{C}$ . For every set  $B$  of patterns, we denote by  $\text{Av}(B)$  the set of all permutations that avoid every pattern in  $B$ . Clearly, for all  $B$ ,  $\text{Av}(B)$  is a permutation class. Conversely (see for instance [18, Paragraph 5.1.2]), every class  $\mathcal{C}$  of permutations can be defined by a set  $B$  of excluded patterns. Moreover, for any given  $\mathcal{C}$ , we can define uniquely a set  $B$  such that  $\mathcal{C} = \text{Av}(B)$ : it is enough to impose that  $B$  is chosen minimal (for set inclusion) among all  $B'$  such that  $\mathcal{C} = \text{Av}(B')$ . This  $B$  (which happens to be an antichain) is called the *basis* of the class  $\mathcal{C}$ . The basis of a permutation class may be finite or infinite (see [18, Paragraph 7.2.3]).

One in many ways permutation classes can be studied is by looking at the features of a typical large permutation  $\sigma$  in the class. A particularly interesting characteristic is the frequency of occurrence of a pattern  $\pi$ , especially when it is considered for all  $\pi$  simultaneously. Denote by  $\text{occ}(\pi, \sigma)$  the number of occurrences of a pattern  $\pi \in \mathfrak{S}_k$  in  $\sigma \in \mathfrak{S}_n$  and by  $\widetilde{\text{occ}}(\pi, \sigma)$  the *pattern density* of  $\pi$  in  $\sigma$ . More formally

$$\begin{aligned} \text{occ}(\pi, \sigma) &= \text{card}\{I \subset [n] \text{ of cardinality } k \text{ such that } \text{pat}_I(\sigma) = \pi\} \\ (1) \quad \widetilde{\text{occ}}(\pi, \sigma) &= \frac{\text{occ}(\pi, \sigma)}{\binom{n}{k}} = \mathbb{P}(\text{pat}_I(\sigma) = \pi), \end{aligned}$$

where  $I$  is randomly and uniformly chosen among the  $\binom{n}{k}$  subsets of  $[n]$  with  $k$  elements. The study of the asymptotics of  $\widetilde{\text{occ}}(\pi, \sigma_n)$ , where  $\sigma_n$  is a uniform random permutation of size  $n$  in a permutation class  $\mathcal{C}$  and  $\pi \in \mathfrak{S}$  is a fixed pattern, has been carried out in several cases.

- The behavior of  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]$  for various classes  $\mathcal{C}$  and fixed  $\pi$  was investigated by Bóna [16, 17], Homberger [29], Chang, Eu and Fu [22] and Rudolf [44].
- Janson, Nakamura and Zeilberger [33] considered higher moments and joint moments, rigourously when  $\mathcal{C} = \mathfrak{S}$ , and also empirically for various classes  $\mathcal{C}$ . A bit later, Janson [32] has given for  $\mathcal{C} = \text{Av}(132)$  the joint limit in distribution of the random variables  $\widetilde{\text{occ}}(\pi, \sigma_n)$ , properly rescaled as to yield a nontrivial limit.

A parallel line of work consists in studying the asymptotic shape of the diagram of  $\sigma_n$ . The *diagram* of a permutation  $\sigma \in \mathfrak{S}_n$  is the set of points  $\{(i, \sigma(i)), 1 \leq i \leq n\}$  in the Cartesian plane. To gain understanding on typical large permutations in  $\mathcal{C}$ , one can investigate the geometric properties of the diagram of  $\sigma_n$  as  $n \rightarrow \infty$ , possibly after rescaling this diagram so that it fits into a unit square.

- Madras and Liu [36], Atapour and Madras [10] and Madras and Pehlivan [37] considered the asymptotic shape of  $\sigma_n$  when  $\mathcal{C} = \text{Av}(\tau)$  for small patterns  $\tau$ .
- In parallel, Miner and Pak [40] described very precisely the asymptotic shape of  $\sigma_n$ , when  $\mathcal{C} = \text{Av}(\tau)$  for the 6 patterns  $\tau$  in  $\mathfrak{S}_3$ . These shapes are related to Brownian excursion, as explained by Hoffman, Rizzolo and Slivken [27, 28].

- Bevan describes the limit shape of permutations in so-called *connected monotone grid classes* [14, Chapter 6].

These two points of view may seem different, but they are in fact tightly bound together. Indeed, as we shall see in Section 2, it follows from results of [30] that the convergence of pattern densities characterizes the convergence of the diagrams, seen as permutons. This important property was actually the main motivation for the introduction of permutons in [30].

**1.2. The permuton viewpoint.** A permuton is a probability measure on the unit square  $[0, 1]^2$  with both its marginals uniform. Permutons generalize permutation diagrams in the following sense: to every permutation  $\sigma \in \mathfrak{S}_n$ , we associate the permuton  $\mu_\sigma$  with density

$$\mu_\sigma(dxdy) = n \mathbf{1}_{\sigma(\lfloor xn \rfloor) = \lfloor yn \rfloor} dxdy.$$

Note that it amounts to replacing every point  $(i, \sigma(i))$  in the diagram of  $\sigma$  (normalized to the unit square) by a square of the form  $[(i-1)/n, i/n] \times [(\sigma(i)-1)/n, \sigma(i)/n]$ , which has mass  $1/n$  uniformly distributed.

The space  $\mathcal{M}$  of permutons is equipped with the topology of weak convergence of measures, which makes it a compact metric space. This allows to define convergent sequences of permutations: we say that  $(\sigma_n)_n$  converges to  $\mu$  when  $(\mu_{\sigma_n}) \rightarrow \mu$  weakly. Similarly, one can define convergence in distribution of random permutations to a random permuton.

By definition, convergence to a permuton encodes the first-order asymptotics of the shape of a sequence of permutations. As we shall see in Section 2, it also encodes the first-order asymptotics of pattern densities: a sequence  $(\sigma_n)_n$  of random permutations converges in distribution to a random permuton if and only if the sequences of random variables  $(\widetilde{\text{occ}}(\pi, \sigma_n))_n$  converge in distribution, jointly for all  $\pi \in \mathfrak{S}$ . Moreover, for any pattern  $\pi$ , the limit distribution of the density of  $\pi$  can be expressed as a function of the limit permuton.

Our previous article [11] studies the limit of the class  $\mathcal{C} = \text{Av}(2413, 3142)$  of separable permutations, in terms of pattern densities and permutons.

**Theorem 1.1.** *Let  $\sigma_n$  be a uniform random separable permutation of size  $n$ . There exists a non-deterministic permuton  $\mu$ , called the *Brownian separable permuton*, such that  $(\mu_{\sigma_n})_n$  converges in distribution to  $\mu$ .*

The result of [11] is more precise, and describes the asymptotic joint distribution of the random variables  $\widetilde{\text{occ}}(\pi, \sigma_n)$  as a measurable functional of a *signed Brownian excursion*. This object is a normalized Brownian excursion whose strict local minima are decorated with an i.i.d. sequence of balanced signs in  $\{+, -\}$ . In another paper [35], the fifth author gives a direct construction of  $\mu$  from this signed Brownian excursion.

The class of separable permutations is the smallest nontrivial substitution-closed class, as defined in the next subsection. The present paper aims at showing a convergence result similar to Theorem 1.1 for other substitution-closed classes. We will see that in many cases the limit belongs to a one-parameter family of deformations of the Brownian separable permuton: the *biased Brownian separable permuton*  $\mu^{(p)}$  of parameter  $p \in (0, 1)$  is obtained from a biased signed Brownian excursion (defined similarly to the signed Brownian excursion but with each

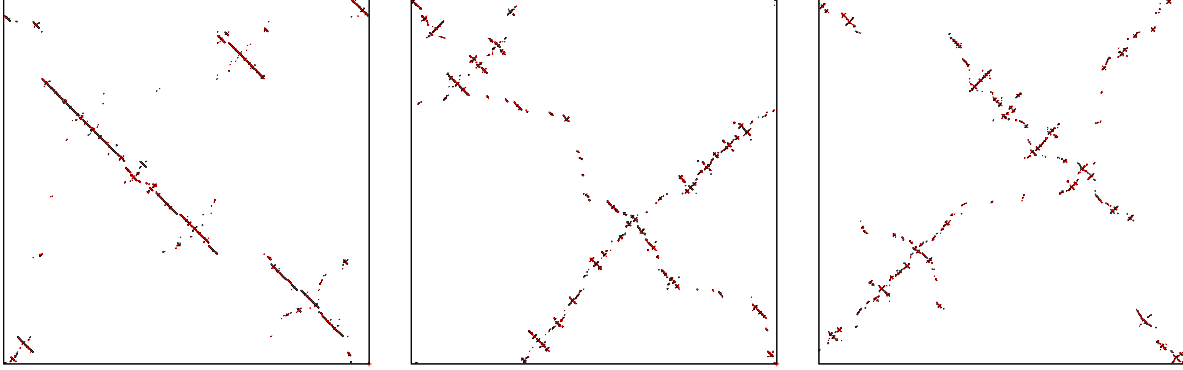


FIGURE 1. Three simulations of the biased Brownian separable permuton  $\mu^{(p)}$  (left:  $p = 0.2$ ; middle:  $p = 0.45$ ; right:  $p = 0.5$ ). How these simulations were obtained is discussed in Appendix B.

sign having probability  $p$  of being a  $+$ ). Simulations of the biased Brownian separable permuton are given in Fig. 1. A precise definition will be given in Section 5 (Definition 5.1).

Finally, we mention that although permutons are a very nice, natural, and powerful way of studying “limits of permutation classes” (in particular because they unify many earlier results, as explained above), this approach has its weaknesses. Most importantly, it gives no information beyond the first order. For instance, the results of [40, 27] describe the *canoe* shape of large permutations in classes avoiding one pattern of length three. As the width of the canoe is a  $o(n)$ , one only sees the diagonal (or antidiagonal) in the permuton limit, but has no information about the fluctuations around this limit.

### 1.3. Substitution-closed classes.

**Definition 1.2.** Let  $\theta = \theta(1) \cdots \theta(d)$  be a permutation of size  $d$ , and let  $\pi^{(1)}, \dots, \pi^{(d)}$  be  $d$  other permutations. The substitution of  $\pi^{(1)}, \dots, \pi^{(d)}$  in  $\theta$  is the permutation of size  $|\pi^{(1)}| + \cdots + |\pi^{(d)}|$  obtained by replacing each  $\theta(i)$  by a sequence of integers isomorphic to  $\pi^{(i)}$  while keeping the relative order induced by  $\theta$  between these subsequences.

This permutation is denoted by  $\theta[\pi^{(1)}, \dots, \pi^{(d)}]$ . We sometimes refer to  $\theta$  as the *skeleton* of the substitution.

When  $\theta$  is  $12 \dots k$  (resp.  $k \dots 21$ ), for any value of  $k \geq 2$ , we rather write  $\oplus$  (resp.  $\ominus$ ) instead of  $\theta$ . Note that the specific value of  $k$  does not appear in this notation, but can be recovered counting the number of permutations  $\pi^{(i)}$  which are substituted in  $\oplus$  (resp.  $\ominus$ ).

Examples of substitution (see Fig. 2 below) are conveniently presented representing permutations by their diagrams: the diagram of  $\theta[\pi^{(1)}, \dots, \pi^{(d)}]$  is obtained by inflating each point  $\theta_i$  of  $\theta$  by a square containing the diagram of  $\pi^{(i)}$ .

By definition of permutation classes, if  $\theta[\pi^{(1)}, \dots, \pi^{(d)}] \in \mathcal{C}$  for some permutation class  $\mathcal{C}$ , then  $\theta, \pi^{(1)}, \dots, \pi^{(d)} \in \mathcal{C}$ . The converse is not always true.

**Definition 1.3.** A permutation class  $\mathcal{C}$  is substitution-closed if, for every  $\theta, \pi^{(1)}, \dots, \pi^{(d)}$  in  $\mathcal{C}$ ,  $\theta[\pi^{(1)}, \dots, \pi^{(d)}] \in \mathcal{C}$ .

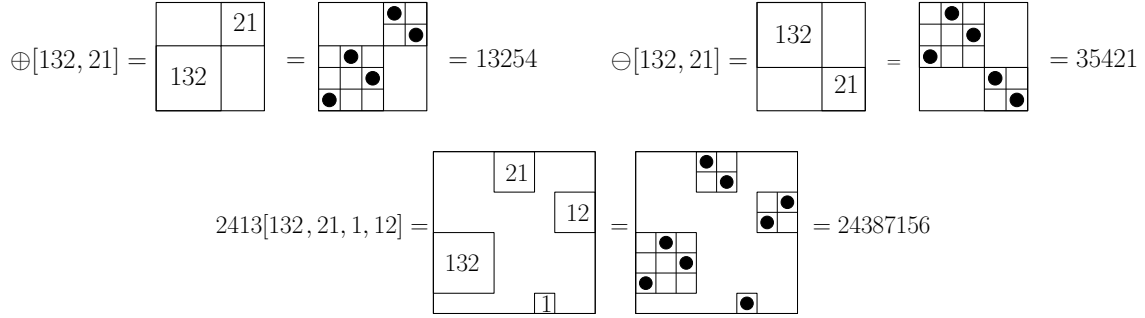


FIGURE 2. Substitution of permutations.

The focus of this paper will be substitution-closed classes. To study such classes it is essential to observe that any permutation has a canonical decomposition using substitutions, which can be encoded in a tree. This decomposition is canonical in the same sense as the decomposition of integers into products of primes. In this analogy, *simple permutations* play the role of prime numbers and the substitution plays the role of the product. We first give a simple definition: a permutation  $\sigma$  is  $\oplus$ -indecomposable (resp.  $\ominus$ -indecomposable) if it cannot be written as  $\oplus[\pi^{(1)}, \pi^{(2)}]$  (resp.  $\ominus[\pi^{(1)}, \pi^{(2)}]$ ), (or equivalently, if there is no  $d$  such that  $\sigma$  can be written as  $\oplus[\pi^{(1)}, \dots, \pi^{(d)}]$  (resp.  $\ominus[\pi^{(1)}, \dots, \pi^{(d)}]$ )).

**Definition 1.4.** A simple permutation is a permutation of size  $n > 2$  that does not map any nontrivial interval (i.e. a range in  $[n]$  containing at least two and at most  $n - 1$  elements) onto an interval.

For example, 451326 is not simple as it maps  $[3; 5]$  onto  $[1; 3]$ . The smallest simple permutations are 2413 and 3142 (there is no simple permutation of size 3).

**Remark:** Usually in the literature, the definition of a simple permutation requires only  $n > 2$  and not  $n \geq 2$ , so that 12 and 21 are considered to be simple. However in our work, 12 and 21 do not play the same role as the other simple permutations, that is why we do not consider them to be simple.

**Theorem 1.5** (Decomposition of permutations, Proposition 2 in [1]). *Every permutation  $\sigma$  of size  $n \geq 2$  can be uniquely decomposed as either:*

- $\alpha[\pi^{(1)}, \dots, \pi^{(d)}]$ , where  $\alpha$  is simple (of size  $d \geq 4$ ),
- $\oplus[\pi^{(1)}, \dots, \pi^{(d)}]$ , where  $d \geq 2$  and  $\pi^{(1)}, \dots, \pi^{(d)}$  are  $\oplus$ -indecomposable,
- $\ominus[\pi^{(1)}, \dots, \pi^{(d)}]$ , where  $d \geq 2$  and  $\pi^{(1)}, \dots, \pi^{(d)}$  are  $\ominus$ -indecomposable.

This decomposition theorem can be applied recursively inside the permutations  $\pi^{(i)}$  appearing in the items above, until we reach permutations of size 1. Doing so, a permutation  $\sigma$  can be naturally encoded by a rooted planar tree, whose internal nodes are labeled by the skeletons of the substitutions that are considered along the recursive decomposition process, and whose leaves correspond to the elements of  $\sigma$ . This construction provides a one-to-one correspondence between permutations and *canonical trees* (defined below) that maps the size to the number of leaves.

**Definition 1.6.** A canonical tree is a rooted planar tree whose internal nodes carry labels satisfying the following constraints.

- Internal nodes are labeled by  $\oplus$ ,  $\ominus$ , or by a simple permutation.
- A node labeled by  $\alpha$  has degree<sup>1</sup>  $|\alpha|$ , nodes labeled by  $\oplus$  and  $\ominus$  have degree at least 2.
- A child of a node labeled by  $\oplus$  (resp.  $\ominus$ ) cannot be labeled by  $\oplus$  (resp.  $\ominus$ ).

Canonical trees are known in the literature under several names: decomposition trees, substitution trees, . . . We choose the term canonical because we consider many variants of substitution trees in this paper, but only these canonical ones provide a one-to-one correspondence with permutations.

The representation of permutations by their canonical trees is essential in the study of substitution-closed classes. The reason is that, for any such class  $\mathcal{C}$ , the set of canonical trees of permutations in  $\mathcal{C}$  can be easily described.

**Proposition 1.7.** Let  $\mathcal{C}$  be a substitution-closed permutation class, and assume<sup>2</sup> that  $12, 21 \in \mathcal{C}$ . Denote by  $\mathcal{S}$  the set of simple permutations in  $\mathcal{C}$ . The set of canonical trees encoding permutations of  $\mathcal{C}$  is the set of all canonical trees built on the set of nodes  $\{\oplus, \ominus\} \cup \{\alpha : \alpha \in \mathcal{S}\}$ .

*Proof.* First, if a canonical tree contains a node labeled by a simple permutation  $\alpha \notin \mathcal{S}$ , then the corresponding permutation  $\sigma$  contains the pattern  $\alpha \notin \mathcal{C}$ , and hence  $\sigma \notin \mathcal{C}$ . Second, by induction, all canonical trees built on  $\{\oplus, \ominus\} \cup \{\alpha : \alpha \in \mathcal{S}\}$  encode permutations of  $\mathcal{C}$ , because  $\mathcal{C}$  is substitution-closed. If necessary, details can be found in [1, Lemma 11].  $\square$

For instance, the class  $\text{Av}(2413, 3142)$  of separable permutations studied in [11] correspond to the set of all canonical trees built on  $\{\oplus, \ominus\}$ , i.e., to  $\mathcal{S} = \emptyset$ . It is therefore the smallest nontrivial substitution-closed class.

*Observation 1.8.* Let  $\mathcal{C}$  be any substitution-closed permutation class, and let  $\mathcal{S}$  be the set of simple permutations in  $\mathcal{C}$ . Because  $\mathcal{C}$  is a class, it holds that for all  $\alpha \in \mathcal{S}$ , if  $\alpha'$  is a simple permutation such that  $\alpha' \preceq \alpha$ , then  $\alpha' \in \mathcal{S}$ . Whenever a set  $\mathcal{S}$  of simple permutations satisfies this property, we say that  $\mathcal{S}$  is *downward-closed* (implicitly: for  $\preceq$  and among the set of simple permutations).

Thanks to their encoding by families of trees, it can be proved that substitution-closed permutation classes (possibly, satisfying additional constraints) share a common behavior. For example, the canonical tree representation of their elements imply that all substitution-closed classes with finitely many simple permutations have an algebraic generating functions [1, Corollary 14]. (This is actually easy, the main contribution of [1] being to generalize this algebraicity result to all classes containing a finite number of simple permutations, again using canonical trees as a key tool.) Our work illustrates this universality paradigm in probability theory: we prove that the biased Brownian separable permutation is the limiting permutation of many substitution-closed classes (see Theorem 1.10 and to a lesser extent Theorem 7.8).

<sup>1</sup>Throughout the paper, by *degree* of a node in a tree, we mean the number of its children (which is sometimes called *arity* in other works). Note that it is different from the graph-degree: for us, the edge to the parent (if it exists) is not counted in the degree.

<sup>2</sup>Otherwise,  $\mathcal{C} = \{12 \dots k : k \geq 1\}$  or  $\mathcal{C} = \{k \dots 21 : k \geq 1\}$  or  $\mathcal{C} = \{1\}$  and these cases are trivial.

**1.4. Our results: Universality.** Let  $\mathcal{S}$  be a (finite or infinite) set of simple permutations. We denote by  $\langle \mathcal{S} \rangle_n$  the set of permutations of size  $n$  whose canonical trees use only nodes  $\oplus$ ,  $\ominus$  and  $\alpha \in \mathcal{S}$ , and we define  $\langle \mathcal{S} \rangle = \bigcup_n \langle \mathcal{S} \rangle_n$ . From Proposition 1.7 and Observation 1.8, every substitution-closed permutation class  $\mathcal{C}$  containing 12 and 21 can be written as  $\mathcal{C} = \langle \mathcal{S} \rangle$  for a downward-closed set  $\mathcal{S}$  of simple permutations (which is just the set of simple permutations in  $\mathcal{C}$ ).

*Remark 1.9.* For a generic (not necessarily downward-closed) set  $\mathcal{S}$  of simple permutations,  $\langle \mathcal{S} \rangle$  is a family of permutations more general than a substitution-closed permutation class. The results that we obtain apply not only to permutation classes but also to such sets of permutations.

Note however that our work does not consider similar sets of permutations where nodes labeled  $\oplus$  and/or  $\ominus$  would not be allowed in the underlying canonical trees. While we have seen that such a restriction would only give rise to trivial permutation classes, it is not the case in this more general setting. We point out that such sets of permutations could also be studied by the approach developed in this paper, but we prefer to leave such cases outside of our study. Indeed, to cover them, it would require to re-do all computations, modifying the combinatorial equations that we start from (see Proposition 4.1 p. 26) and all equations that follow, so as not to allow the nodes labeled  $\oplus$  and/or  $\ominus$ .

We are interested in the asymptotic behavior of a uniform permutation  $\sigma_n$  in  $\langle \mathcal{S} \rangle_n$  which we describe in terms of permutons. Let

$$S(z) = \sum_{\alpha \in \mathcal{S}} z^{|\alpha|}$$

be the generating function of  $\mathcal{S}$  and let  $R_S \in [0, +\infty]$  be the radius of convergence of  $S$ .

**Theorem 1.10** (Main Theorem: the standard case). *Let  $\mathcal{S}$  be a set of simple permutations such that*

$$(H1) \quad R_S > 0 \quad \text{and} \quad \lim_{\substack{r \rightarrow R_S \\ r < R_S}} S'(r) > \frac{2}{(1 + R_S)^2} - 1.$$

*For every  $n \geq 1$ , let  $\sigma_n$  be a uniform permutation in  $\langle \mathcal{S} \rangle_n$ , and let  $\mu_{\sigma_n}$  be the random permuton associated with  $\sigma_n$ . The sequence  $(\mu_{\sigma_n})_n$  tends in distribution in the weak convergence topology to the biased Brownian separable permuton  $\mu^{(p)}$  whose parameter  $p$  is given in (21) p. 36.*

An important point in Theorem 1.10 is that the limiting object depends on  $\mathcal{S}$  only through the parameter  $p$ . It turns out that  $p$  only depends on the quantity of occurrences of the patterns 12 and 21 in the elements of  $\mathcal{S}$ .

In the following, to lighten the notation, we write  $S'(R_S) := \lim_{\substack{r \rightarrow R_S \\ r < R_S}} S'(r)$ . Note that  $S'(R_S)$  may be  $\infty$ .

The case when Condition (H1) of Theorem 1.10 is not satisfied is discussed in the next section. When Condition (H1) is satisfied the case is called *standard* because there are natural and easy sufficient conditions to ensure this case (that are given below). Moreover, this case includes all sets  $\mathcal{S}$  studied so far in the literature on permutation classes, to our knowledge. This gives a fairly precise (and positive) answer to an important question raised in our previous article [11]: is the



Brownian separable permuton universal (in the sense that it describes the limit of a large family of substitution-closed classes)?

We now give several cases in which Condition (H1) of Theorem 1.10 is satisfied.

- If  $S$  is a generating function with radius of convergence  $R_S > \sqrt{2} - 1$ , (H1) is satisfied. Indeed, the condition  $R_S > \sqrt{2} - 1$  implies  $\frac{2}{(1+R_S)^2} - 1 < 0$ , and  $S'(R_S)$  is nonnegative since  $S'$  (like  $S$ ) is a series with nonnegative coefficients. In particular, the situation where  $R_S > \sqrt{2} - 1$  covers the cases where there are finitely many simple permutations in the class (then  $S$  is a polynomial and  $R_S = \infty$ ), and more generally where the number of simple permutations of size  $n$  grows subexponentially (in this case, if there are infinitely many simple permutations, necessarily  $R_S = 1$ ).
- If  $S'$  is divergent at  $R_S$ , (H1) is trivially verified. In particular, this happens when  $S$  is a rational generating function, or when  $S$  has a square root singularity at  $R_S$ .

In the literature, there are quite a few examples of permutations classes whose set  $\mathcal{S}$  of simple permutations has been enumerated. We can therefore ask whether Condition (H1) applies to them. In all examples we could find, it is indeed satisfied, and this follows from the discussion above. We record these examples here.

- Classes with finitely many simple permutations have attracted a fair amount of attention, see [1] and subsequently [12, 19, 21].
- Several families of simple permutations with a bounded number of elements of each size have appeared in the literature: the family of exceptional simple permutations (also called simple parallel alternations in [20]), the family of wedge simple permutations (see also [20]), the families of oscillations and quasi-oscillations (see [13]), and the families of simple permutations contained in the following three classes:  $\text{Av}(4213, 3142)$ ,  $\text{Av}(4213, 1342)$  and  $\text{Av}(4213, 3124)$  – see [5].
- The family of simple pin-permutations has a rational generating function – see [13].
- The generating function  $S$  is also rational when  $\mathcal{S}$  is the set of simple permutations contained in several permutation classes defined by the avoidance of two patterns of size 4, namely  $\text{Av}(3124, 4312)$  – see [41],  $\text{Av}(2143, 4312)$  and  $\text{Av}(1324, 4312)$  – see [2],  $\text{Av}(2143, 4231)$  – see [3],  $\text{Av}(1324, 4231)$  – see [6],  $\text{Av}(4312, 3142)$  and  $\text{Av}(4231, 3124)$  – see [5].
- The set  $\mathcal{S}$  of simple permutations of the class  $\text{Av}(4231, 35142, 42513, 351624)$  enumerated in [7] is also rational.
- When  $\mathcal{S}$  is the set of simple permutations avoiding 321 or of skew-merge simple permutations, the generating function  $S$  is computed in [8]: it has a squareroot singularity at  $R_S = \frac{1}{3}$ .

In addition to verifying Condition (H1), we have computed the numerical value of the parameter  $p$  for some of the above-mentioned sets  $\mathcal{S}$  of simple permutations; see Examples 5.3 to 5.5 (p. 37). All the values that we have found are between 0.45 and 0.55, which makes simulations indistinguishable in practice from the unbiased Brownian permuton (see Fig. 1).

Since all the sets of simple permutations studied in the literature (to our knowledge) satisfy Condition (H1), we may wonder if there exist substitution-closed classes that are not covered

by the standard case. Or are those observations just an artifact, maybe due to the set of simple permutations of a substitution-closed class being easier to compute when Condition (H1) is satisfied? We leave those questions open.

However, when leaving the case of permutation classes, one can find sets  $\mathcal{S}$  that do not satisfy Condition (H1), and that do not even fit into the universality class of the Brownian permuton. This is discussed in the next section.

**1.5. Our results: Beyond universality.** When  $R_S > 0$ , for the two remaining cases  $S'(R_S) < 2/(1 + R_S)^2 - 1$  and  $S'(R_S) = 2/(1 + R_S)^2 - 1$ , the asymptotic behavior of  $\mu_{\sigma_n}$  is qualitatively different, and the results require slight additional hypotheses and notation. As a consequence, for the moment we only briefly describe these behaviors, the results being stated with full rigor later.

- **Case  $S'(R_S) < 2/(1 + R_S)^2 - 1$ .** This is a *degenerate* case.

We first show in Theorem 6.9 that, with a small additional assumption which will be called  $(CS)$ , the sequence  $(\mu_{\sigma_n})$  of random permutons converges. If uniform simple permutations in  $\mathcal{S} \cap \mathfrak{S}_n$  have a limit (in the sense of permutons), we show that the limit of permutations in  $\langle \mathcal{S} \rangle$  is the same (Proposition 6.10). This explains the terminology “degenerate”: all permutations in the class (or set)  $\langle \mathcal{S} \rangle$  are close to the simple ones, and the “composite” structure of permutations does not appear in the limit.

- **Case  $S'(R_S) = 2/(1 + R_S)^2 - 1$ .** This *critical* case is more subtle.

We again need to assume the above mentioned hypothesis  $(CS)$ . According to the behavior of  $\mathcal{S}$  near  $R_S$ , the limiting permuton of  $(\mu_{\sigma_n})$  can either be the (biased) Brownian separable permuton (Theorem 7.8) or belong to a new family of *stable permutons* (Theorem 7.6). These stable permutons are defined using the marginals of the random *stable tree* (see [25]), which explains the terminology. Two simulations are presented in Fig. 3.

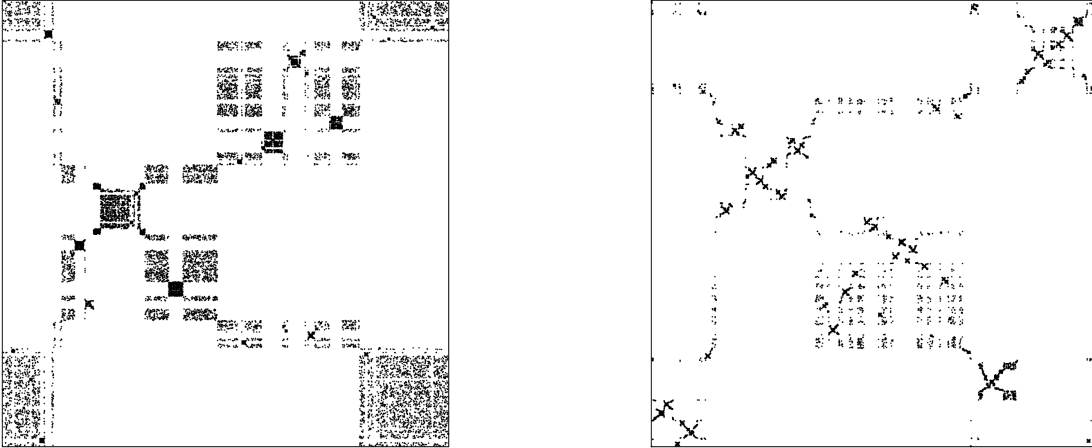


FIGURE 3. Simulations of a 1.1-stable and 1.5-stable permuton, driven by the uniform measure.

*Remark 1.11.* The variety of behaviors that we observe can be informally understood in terms of trees. We have seen in Section 1.3 that permutations in  $\langle \mathcal{S} \rangle$  can be encoded by trees. Taking a uniform element in  $\langle \mathcal{S} \rangle_n$ , we can prove that the corresponding tree is a multi-type Galton-Watson tree conditioned on having  $n$  leaves. This link with conditioned Galton-Watson trees is not used in this paper, but may give intuition on our results.

In the standard and critical case, these Galton-Watson trees are critical. It is therefore not surprising to see two different limiting behaviors. When the law of reproduction has finite variance, we get one behavior related to the Brownian excursion and the Brownian continuum random tree (that we describe as the universal one). When the law of reproduction has infinite variance, we get the behavior related to stable trees. In the standard case, the law of reproduction always has finite variance.

On the contrary, in the degenerate case, the underlying Galton-Watson tree model is subcritical. At the limit, such trees conditioned to being large have one internal node of very high degree ([31, Theorem 7.1, case (ii)]). This node corresponds to a large simple permutation in the tree encoding a uniform random permutation  $\sigma_n$  in  $\langle \mathcal{S} \rangle$ . It is therefore not surprising that  $\sigma_n$  is asymptotically close to a uniform *simple* permutation in  $\langle \mathcal{S} \rangle$ .

*Remark 1.12.* The reader may have noticed that all cases where we describe the asymptotic behavior of  $\mu_{\sigma_n}$  are such that  $R_{\mathcal{S}} > 0$ .

Observe that it is always the case for proper permutation classes (*i.e.*, permutation classes different from  $\mathfrak{S}$ ). Indeed, from the Marcus-Tardos Theorem [38], the number of permutations of size  $n$  in a proper class is at most  $c^n$ , for some constant  $c$ . For the class  $\mathfrak{S}$ , we however do have  $R_{\mathcal{S}} = 0$ , since there are asymptotically  $e^{-2}n!(1 + \mathcal{O}(1/n))$  simple permutations of size  $n$  [4, Theorem 5]. In this case, the sequence  $(\mu_{\sigma_n})_n$  of permutons associated with a uniform permutation  $\sigma_n$  in  $\mathfrak{S}$  converges in distribution to the uniform measure on  $[0, 1]^2$ . The situation where  $R_{\mathcal{S}} = 0$  may happen as well for sets  $\langle \mathcal{S} \rangle$  where  $\mathcal{S}$  is not downward-closed, but we leave these cases open.

**1.6. Limits of proportions of pattern occurrences.** Let us change our approach and discuss in this section the asymptotic behaviour (as  $n \rightarrow \infty$ ) of the proportion  $\widetilde{\text{occ}}(\pi, \sigma_n)$  of occurrences of a fixed pattern  $\pi$  in  $\sigma_n$  (as done in [16, 17, 22, 29, 32, 33, 44] for uniform random permutations in various classes). Since all known examples fit in that regime, we focus here on the standard case (when (H1) is satisfied).

As mentioned in Section 1.2 and explained in more details in Section 2, the convergence of  $\mu_{\sigma_n}$  towards  $\mu^{(p)}$  implies the (joint) convergence in distribution

$$(2) \quad (\widetilde{\text{occ}}(\pi, \sigma_n))_{\pi \in \mathfrak{S}} \rightarrow \widetilde{\text{occ}}(\pi, \mu^{(p)}).$$

The limiting random variables  $\widetilde{\text{occ}}(\pi, \mu^{(p)})$  have been studied in [11, Section 9] (for  $p = .5$ ): in particular,  $\widetilde{\text{occ}}(\pi, \mu^{(.5)})$  is non-deterministic if and only if  $\pi$  is separable of size at least 2, and it is possible to compute their moments algorithmically. These results are easily extended to the general case  $p \in (0, 1)$ . Therefore, for separable patterns  $\pi$ , (2) establishes the convergence of  $\widetilde{\text{occ}}(\pi, \sigma_n)$  to a non-deterministic limit. Since these are bounded variables, their (joint) moments also converge to the (joint) moments of the limiting vector, which can be computed algorithmically (even if in practice only low order moments can be effectively computed; see the discussion

in [11, Section 9]). Note that all these limiting moments are trivially nonzero, since these are moments of nondeterministic nonnegative random variables.

For nonseparable patterns however, the situation is different: (2) only entails the convergence of  $\widetilde{\text{occ}}(\pi, \sigma_n)$  to 0. Indeed, if  $\pi$  is nonseparable, the limiting quantity  $\widetilde{\text{occ}}(\pi, \mu^{(p)})$  is identically 0 (this is a consequence of Definition 5.1 p.36).

We can go further and ask whether  $(\widetilde{\text{occ}}(\pi, \sigma_n))_n$  has a nonzero limit in distribution with some appropriate normalization. We therefore investigate the moments of  $\widetilde{\text{occ}}(\pi, \sigma_n)$ . In Section 5.6, we establish (under hypothesis (H1)) the following asymptotic behavior<sup>3</sup>:

**Proposition 5.13.** *For each  $\pi \in \mathcal{C}$  and  $m \geq 1$ , we have  $\mathbb{E}[(\widetilde{\text{occ}}(\pi, \sigma_n))^m] = \Theta(n^{-\text{db}(\pi)/2})$ , where the permutation statistics  $\text{db}(\pi)$  is defined in Eq. (32) p.42.*

Proposition 5.13 also holds for separable patterns  $\pi$ : in that case  $\text{db}(\pi)=0$  and we have  $\mathbb{E}[(\widetilde{\text{occ}}(\pi, \sigma_n))^m] = \Theta(1)$ . No news here, since the moments of  $\widetilde{\text{occ}}(\pi, \sigma_n)$  have nonzero limits, as previously explained. For nonseparable patterns,  $\text{db}(\pi)$  is positive and measures in some sense how nonseparable  $\pi$  is. Note that the order of magnitude of  $\mathbb{E}[(\widetilde{\text{occ}}(\pi, \sigma_n))^m]$  is independent of  $m$ , which implies that there is a set of probability  $\Theta(n^{-\text{db}(\pi)/2})$  on which the variables  $\widetilde{\text{occ}}(\pi, \sigma_n)$  stays bounded away from 0 (see Corollary 5.15). This event of small probability contributes to the asymptotic behavior of moments, and thus the method of moments is inappropriate to find a nonzero limiting distribution for some appropriate normalization of  $\widetilde{\text{occ}}(\pi, \sigma_n)$ . Finding such a limiting distribution is therefore left as an open question.

**1.7. Outline of the proof.** In our previous paper [11] (*i.e.* when the family of simple permutations is  $\mathcal{S} = \emptyset$ ), the proof of the convergence to the Brownian separable permuton strongly relies on a connection to Galton-Watson trees conditioned on having a given number of leaves. This allowed us to use fine results by Kortchemski [34] or Pitman and Rizzolo [42] on such conditioned random tree models.

For a general family  $\mathcal{S}$ , generalizing this approach would require delicate results on the asymptotic behavior of conditioned *multitype* Galton-Watson trees. Moreover, there are several other steps in the main proofs of [11], in particular the subtree exchangeability argument, that are not easily adapted.

The strategy developed in the present paper is different. We strongly use the framework of permutons. Indeed, we first show that to establish the convergence in distribution of  $(\mu_{\sigma_n})_n$  to some random permuton  $\mu$ , it is enough to prove the convergence of  $\left(\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]\right)_n$  for every pattern  $\pi$  (see Theorem 2.5). By definition, if  $\pi \in \mathfrak{S}_k$  and  $n \geq k$ ,

$$(3) \quad \mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = \frac{\#\{\sigma \in \langle \mathcal{S} \rangle_n, I \subset [n] : \text{pat}_I(\sigma) = \pi\}}{\binom{n}{k} \#\langle \mathcal{S} \rangle_n}$$

The asymptotic behavior of the numerator and denominator is then obtained with analytic combinatorics, which allows us to transfer from the behavior of a generating series near its singularity to the asymptotic behavior of its coefficients. This goes in three steps.

<sup>3</sup>We say that the sequence  $(a_n)$  behaves as  $\Theta(b_n)$  if there are  $c, C > 0$  such that  $c|b_n| \leq |a_n| \leq C|b_n|$  for every  $n \geq 1$ .

**Step 1: Enumeration.** We compute (or characterize by an implicit equation) some generating series. For instance to estimate the denominator of (3) we consider  $\sum_{n \geq 1} \# \langle S \rangle_n z^n$ . We readily use the size-preserving bijection between  $\langle S \rangle$  and the class  $\mathcal{T}$  of  $S$ -canonical trees, counted by the number of leaves. Hence the generating function we want to compute is the same as that of  $\mathcal{T}$ , denoted  $T$ .

Using again the encoding of permutations by trees, the numerator can be described as a number of trees with marked leaves and some conditions on the tree induced by these marked leaves. Obtaining generating functions for such combinatorial classes is possible, and needs the introduction of several intermediary functions which count trees with various constraints, and possibly one marked leaf. This is detailed in Section 4.

**Step 2: Singularity analysis.** Then we want to know the singular behavior of the generating functions we computed so far. As it turns out, the singular behavior of some intermediate function  $T_{\text{not}\oplus}$  drives the singular behavior of all the other series. The function  $T_{\text{not}\oplus}$  is characterized by the implicit equation

$$(4) \quad T_{\text{not}\oplus}(z) = z + \Lambda(T_{\text{not}\oplus}(z)),$$

where  $\Lambda$  is a known analytic function with radius of convergence  $R_\Lambda$  that involves  $S$  and some rational functions (see Eq. (23) p. 37). Hence the behavior of  $T_{\text{not}\oplus}$  depends on whether there is a point inside the disk of convergence  $D(0, R_\Lambda)$  where  $\Lambda' = 1$ , because around such a *critical* point, the equation (4) is not invertible. Since  $\Lambda$  is a series with positive integer coefficients, it suffices to check the sign of  $\Lambda'(R_\Lambda) - 1$ , which can easily be translated in terms of the function  $S$ . This is where the sign of  $S'(R_S) - 2/(1 + R_S)^2 + 1$  appears, leading to the three different cases. More precisely<sup>4</sup>

- The standard case  $S'(R_S) > 2/(1 + R_S)^2 - 1$  is equivalent to  $\Lambda'(R_\Lambda) > 1$ . In this case there is a unique critical point  $\tau \in (0, R_\Lambda)$ . As a result, the radius of convergence of  $T_{\text{not}\oplus}$  is  $\rho = \tau - \Lambda(\tau)$ , and the analyticity of  $\Lambda$  around  $\tau = T_{\text{not}\oplus}(\rho)$  implies that  $T_{\text{not}\oplus}$  has a singularity of exponent  $1/2$  (Proposition 5.8). Such a behavior is sometimes called *branch point* in the literature:  $\Lambda$  is analytic at  $\tau$  but the equation (4) has two solutions (called branches) near  $\rho$  and one cannot find an analytic solution in a neighbourhood of  $\rho$ . The solution  $T_{\text{not}\oplus}$  is therefore singular at  $\rho$ .
- The degenerate case  $S'(R_S) < 2/(1 + R_S)^2 - 1$  is equivalent to  $\Lambda'(R_\Lambda) < 1$ . In this case there is no critical point in the disk  $D(0, R_\Lambda)$  nor at its boundary. As a result, the unique dominant singularity of  $T_{\text{not}\oplus}$  is the point  $\rho = R_\Lambda - \Lambda(R_\Lambda)$  where  $T_{\text{not}\oplus}(\rho)$  reaches the singularity  $R_\Lambda$  of  $\Lambda$ . Moreover,  $T_{\text{not}\oplus}$  has a bounded derivative at its singularity, that so has exponent  $\delta > 1$ , which is the same as the exponent of  $S$  (Lemma 6.3).
- The critical case  $S'(R_S) = 2/(1 + R_S)^2 - 1$  is equivalent to  $\Lambda'(R_\Lambda) = 1$ . In this case there is no critical point inside the disk  $D(0, R_\Lambda)$ , but the singularity  $R_\Lambda$  of  $\Lambda$  is a critical point. Once again the radius of convergence of  $T_{\text{not}\oplus}$  is  $\rho = R_\Lambda - \Lambda(R_\Lambda)$ , but  $T_{\text{not}\oplus}$  has no first derivative at its singularity. Here the exponent of the singularity of  $T_{\text{not}\oplus}$  depends on that of the singularity of  $S$ , and belongs to  $[1/2, 1)$  (see Lemma 7.3).

<sup>4</sup>In this informal description, we left out some conditions on the singularity of  $S$  that appear in the critical and degenerate cases.

Once we have found the asymptotic behavior of  $T_{\text{not}\oplus}$ , we should analyze the one of the series of trees found in Step 1. It is purely routine from an analytic point of view, but involves some combinatorial arguments, regarding the encoding of permutations by substitution trees.

**Step 3: Transfer.** Finally we use a transfer theorem of analytic combinatorics (Theorem A.3) to translate the singularity exponents we found in Step 2 into a limiting behavior for (3). Informally, a square-root singularity, which is the same as in "usual" families of trees, will lead to the Brownian separable permuton. A singularity of exponent in  $(1/2, 1)$  will lead to the  $\delta$ -stable tree, where  $\delta \in (1, 2)$  is the inverse of the exponent. A singularity of exponent  $\delta > 1$  will invariably lead to the degenerate case.

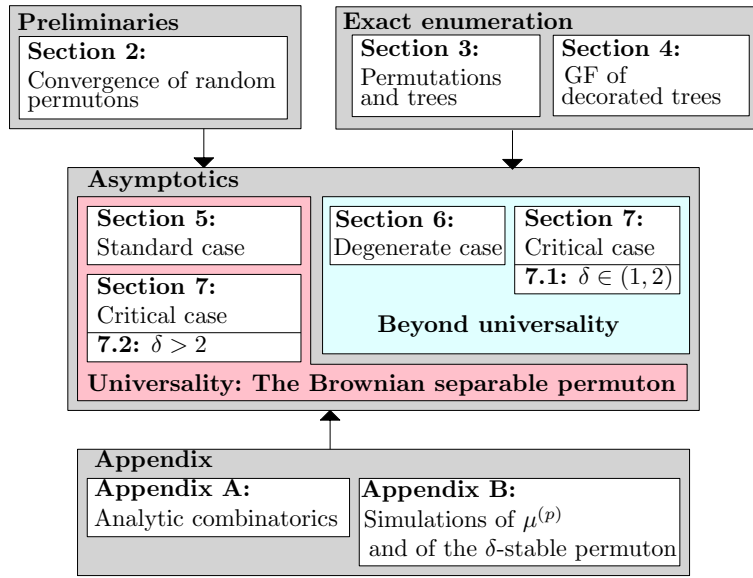


FIGURE 4. Reader's guide of the paper.

1.8. **Organization of the paper.** The paper is organized as follows (see also Fig. 4).

- Section 2 is devoted to proving useful results on the convergence of random permutons. The proofs heavily rely on previous estimates for deterministic permutons [30]. We believe that these general results regarding random permutons are interesting on their own, therefore these are presented in a self-contained way.
- In Sections 3 and 4, we prove nonasymptotic enumeration results for the number of permutations encoded by some given families of (decorated) trees. The main result is Proposition 4.5, which is the first step towards the estimation of  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]$ .
- In Sections 5, 6, 7 we prove our main results: the convergence of the sequence  $(\mu_{\sigma_n})_n$  of permutons. As already mentioned, the quantitative behavior depends on the family  $\mathcal{S}$ , more precisely on the sign of  $S'(R_S) - 2/(1 + R_S)^2 + 1$ :
  - Section 5 is devoted the standard case  $S'(R_S) > 2/(1 + R_S)^2 - 1$ . We show in Theorem 5.2 the convergence to the biased Brownian separable permuton.

- In Section 6, we consider the degenerate case  $S'(R_S) < 2/(1 + R_S)^2 - 1$ .
- In Section 7, we consider the critical case  $S'(R_S) = 2/(1 + R_S)^2 - 1$ . This case itself is divided into two subcases, according to whether the exponent  $\delta$  (defined in Definition 6.6) is smaller (Section 7.1) or greater (Section 7.2) than 2.
- We postpone to Appendix A many useful results of complex analysis.
- Finally, Appendix B exhibits simple models of random permutons that converge respectively towards the biased Brownian and stable permutons. These models were used for the simulations of Figs. 1 and 3.

## 2. CONVERGENCE OF RANDOM PERMUTONS

In this section, we first recall the terminology of (deterministic) permutons, introduced in [30]. We also adapt their results to obtain criteria for the convergence in distribution of random permutons.

**Notation.** *Since this section involves many different probability spaces, we use a superscript on  $\mathbb{P}$  (and similarly on expectation symbols  $\mathbb{E}$ ) to record the source of randomness. In the case where the event  $A(\mathbf{u}, \mathbf{v})$  (or the function  $H(\mathbf{u}, \mathbf{v})$ ) depends on two random variables  $\mathbf{u}$  and  $\mathbf{v}$ , we interpret  $\mathbb{P}^{\mathbf{u}}(A(\mathbf{u}, \mathbf{v}))$  (or  $\mathbb{E}^{\mathbf{u}}[H(\mathbf{u}, \mathbf{v})]$ ) as the conditional probability (expectation) with respect to  $\mathbf{v}$ .*

**2.1. Deterministic permutons and extracted permutations.** Recall from Section 1.2 that a permuton is a probability measure on the unit square with uniform marginals. To a permutation  $\sigma$  of size  $n$ , we can associate the permuton  $\mu_\sigma$  which is essentially the (normalized) diagram of  $\sigma$ , where each dot has been replaced with a small square of dimension  $1/n \times 1/n$  carrying a mass  $1/n$ .

Let  $\mathcal{M}$  be the set of permutons. We need to equip  $\mathcal{M}$  with a topology. We say that a sequence of (deterministic) permutons  $(\mu_n)_n$  converges *weakly* to  $\mu$  (simply denoted  $\mu_n \rightarrow \mu$ ) if

$$\int_{[0,1]^2} f d\mu_n \xrightarrow{n \rightarrow +\infty} \int_{[0,1]^2} f d\mu,$$

for every bounded and continuous function  $f : [0, 1]^2 \rightarrow \mathbb{R}$ . With this topology,  $\mathcal{M}$  is compact and metrizable by some metric  $d_\square$  (see [30, Lemmas 2.5 and 5.3]):

$$\mu_n \xrightarrow{n \rightarrow +\infty} \mu \quad \Leftrightarrow \quad d_\square(\mu_n, \mu) \xrightarrow{n \rightarrow +\infty} 0.$$

Since  $\mathcal{M}$  is compact, Prokhorov's theorem ensures that the space of probability distributions on  $\mathcal{M}$  is compact (for convergences of measure, we refer to [15]).

Recall from Section 1.1 that for  $\sigma \in \mathfrak{S}_n$  and  $\pi \in \mathfrak{S}_k$ , we have

$$\widetilde{\text{occ}}(\pi, \sigma) = \mathbb{P}^{\mathbf{I}_{n,k}} \left( \text{pat}_{\mathbf{I}_{n,k}}(\sigma) = \pi \right),$$

where  $\mathbf{I}_{n,k}$  is randomly and uniformly chosen among the  $\binom{n}{k}$  subsets of  $[n]$  with  $k$  elements. We will define the *pattern density*  $\widetilde{\text{occ}}(\pi, \mu)$  of a pattern  $\pi \in \mathfrak{S}_k$  in a *permuton*  $\mu$  by analogy with this formula.

Take a sequence of  $k$  random points  $(\vec{x}, \vec{y}) = ((x_1, y_1), \dots, (x_k, y_k))$  in  $[0, 1]^2$ , independently with common distribution  $\mu$ . Because  $\mu$  has uniform marginals and the  $x_i$ 's (resp.  $y_i$ 's) are independent, it holds that the  $x_i$ 's (resp.  $y_i$ 's) are almost surely distinct. We denote by  $(x_{(1)}, y_{(1)}), \dots, (x_{(k)}, y_{(k)})$  the  $x$ -ordered sample of  $(\vec{x}, \vec{y})$ , *i.e.* the unique reordering of the sequence  $((x_1, y_1), \dots, (x_k, y_k))$  such that  $x_{(1)} < \dots < x_{(k)}$ . Then the values  $(y_{(1)}, \dots, y_{(k)})$  are in the same relative order as the values of a unique permutation, that we denote  $\text{Perm}(\vec{x}, \vec{y})$ . Since the points are taken at random,  $\text{Perm}(\vec{x}, \vec{y})$  is a random permutation of size  $k$ . Then we set

$$\widetilde{\text{occ}}(\pi, \mu) = \mathbb{P}^{\vec{x}, \vec{y}} (\text{Perm}(\vec{x}, \vec{y}) = \pi).$$

Rewriting this probability in an integral form, we get immediately:

$$(5) \quad \widetilde{\text{occ}}(\pi, \mu) = \int_{([0, 1]^2)^k} \mathbf{1}_{\text{Perm}(\vec{x}, \vec{y}) = \pi} \mu(dx_1 dy_1) \cdots \mu(dx_k dy_k)$$

which identifies  $\widetilde{\text{occ}}(\pi, \cdot)$  as a measurable function on the space of permutons.

In the following, as we consider a random permuton  $\mu$ , we need to construct a finite sequence of points  $(x_1, y_1), \dots, (x_k, y_k)$ , which are independent with common distribution  $\mu$  *conditionally on  $\mu$* . This is possible up to considering a new probability space where the joint distribution of  $(\mu, (x_1, y_1), \dots, (x_k, y_k))$  is characterized as follows: for every positive measurable functional  $H : \mathcal{M} \times ([0, 1]^2)^k \rightarrow \mathbb{R}$ ,

$$(6) \quad \mathbb{E}^{\mu, \vec{x}, \vec{y}} [H(\mu, (x_1, y_1), \dots, (x_k, y_k))] \\ = \mathbb{E}^{\mu} \left[ \int_{([0, 1]^2)^k} \mu(dx_1 dy_1) \cdots \mu(dx_k dy_k) H(\mu, (x_1, y_1), \dots, (x_k, y_k)) \right].$$

In this new probability space, we call  $\vec{m}_k$  the vector  $(\vec{x}, \vec{y}) = (x_i, y_i)_{1 \leq i \leq k}$ , and we use the notation  $\text{Perm}(\vec{m}_k, \mu) = \text{Perm}(\vec{x}, \vec{y})$ , to highlight the two levels of randomness.

We end this section by the following two estimates, proved in [30].

**Lemma 2.1** (Occurrences in a permutation and its associated permuton [30, Lemma 3.5]). *If  $\pi \in \mathfrak{S}_k$  and  $\sigma \in \mathfrak{S}_n$ , then*

$$|\widetilde{\text{occ}}(\pi, \sigma) - \widetilde{\text{occ}}(\pi, \mu_\sigma)| \leq \frac{1}{n} \binom{k}{2}.$$

**Lemma 2.2** (Approximation of a permuton by a permutation [30, Lemma 4.2]). *There is a  $k_0$  such that if  $k > k_0$ , for any permuton  $\nu$ ,*

$$\mathbb{P}^{\vec{m}_k} [d_{\square}(\mu_{\text{Perm}(\vec{m}_k, \nu)}, \nu) \geq 16k^{-1/4}] \leq \frac{1}{2} e^{-\sqrt{k}}.$$

**2.2. Random permutons and convergence in distribution.** We now consider a sequence of random permutations  $(\sigma_n)$  (with  $\sigma_n$  of size  $n$ ) and a random permuton  $\mu$ . Taking  $I_{n,k}$  independently from  $(\sigma_n)$ , we have for every  $\pi$  of size  $k$ :

$$(7) \quad \mathbb{E}^{\sigma_n} [\widetilde{\text{occ}}(\pi, \sigma_n)] = \mathbb{E}^{\sigma_n} \left[ \mathbb{P}^{I_{n,k}} (\text{pat}_{I_{n,k}}(\sigma_n) = \pi) \right] = \mathbb{P}^{\sigma_n, I_{n,k}} (\text{pat}_{I_{n,k}}(\sigma_n) = \pi).$$



If we consider the distribution of  $(\boldsymbol{\mu}, \vec{\mathbf{m}}_k)$  described above and characterized by (6), we can apply property (6) to the measurable functional  $H(\mu, (x_1, y_1), \dots, (x_k, y_k)) = \mathbf{1}_{\text{Perm}(\vec{x}, \vec{y}) = \pi}$  and use (5) to get the following property:

$$(8) \quad \mathbb{E}^\mu[\widetilde{\text{occ}}(\pi, \boldsymbol{\mu})] = \mathbb{P}^{\boldsymbol{\mu}, \vec{\mathbf{m}}_k}(\text{Perm}(\vec{\mathbf{m}}_k, \boldsymbol{\mu}) = \pi).$$

The same argument may be applied to  $H(\nu, (x_1, y_1), \dots, (x_k, y_k)) = \mathbf{1}_{d_\square(\mu_{\text{Perm}(\vec{x}, \vec{y})}, \nu) \geq 16k^{-1/4}}$ , yielding a randomized version of Lemma 2.2.

**Lemma 2.3** (Approximation of a random permuton by a random permutation). *There is a  $k_0$  such that if  $k > k_0$ , for any random permuton  $\nu$ ,*

$$\mathbb{P}^{\vec{\mathbf{m}}_k, \nu} [d_\square(\mu_{\text{Perm}(\vec{\mathbf{m}}_k, \nu)}, \nu) \geq 16k^{-1/4}] \leq \frac{1}{2}e^{-\sqrt{k}}.$$

This result has an important consequence for the distribution of random permutons.

**Proposition 2.4** (Subpermutations characterize the distribution of  $\boldsymbol{\mu}$ ). *Let  $\boldsymbol{\mu}, \boldsymbol{\mu}'$  be two random permutons. If there exists  $k_1$  such that for  $k \geq k_1$  and every  $\pi$  of size  $k$  we have*

$$\mathbb{P}^{\boldsymbol{\mu}, \vec{\mathbf{m}}_k}(\text{Perm}(\vec{\mathbf{m}}_k, \boldsymbol{\mu}) = \pi) = \mathbb{P}^{\boldsymbol{\mu}', \vec{\mathbf{m}}_k}(\text{Perm}(\vec{\mathbf{m}}_k, \boldsymbol{\mu}') = \pi),$$

*then  $\boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{\mu}'$ .*

*Proof.* We need to prove that  $\mathbb{E}^\mu[\phi(\boldsymbol{\mu})] = \mathbb{E}^{\boldsymbol{\mu}'}[\phi(\boldsymbol{\mu}')] for every bounded and continuous function  $\phi : \mathcal{M} \rightarrow \mathbb{R}$ . Fix  $k \geq k_1$ . It holds that$

$$\begin{aligned} \mathbb{E}^\mu[\phi(\boldsymbol{\mu})] - \mathbb{E}^{\boldsymbol{\mu}'}[\phi(\boldsymbol{\mu}')] &= \mathbb{E}^{\boldsymbol{\mu}, \vec{\mathbf{m}}_k}[\phi(\boldsymbol{\mu}) - \phi(\mu_{\text{Perm}(\vec{\mathbf{m}}_k, \boldsymbol{\mu})})] \\ &\quad + \left( \mathbb{E}^{\boldsymbol{\mu}, \vec{\mathbf{m}}_k}[\phi(\mu_{\text{Perm}(\vec{\mathbf{m}}_k, \boldsymbol{\mu})})] - \mathbb{E}^{\boldsymbol{\mu}', \vec{\mathbf{m}}'_k}[\phi(\mu_{\text{Perm}(\vec{\mathbf{m}}'_k, \boldsymbol{\mu}')})] \right) \\ &\quad + \mathbb{E}^{\boldsymbol{\mu}', \vec{\mathbf{m}}'_k}[\phi(\mu_{\text{Perm}(\vec{\mathbf{m}}'_k, \boldsymbol{\mu}')}) - \phi(\boldsymbol{\mu}')], \end{aligned}$$

where  $\vec{\mathbf{m}}'_k$  denotes a sequence of  $k$  independent points with common distribution  $\boldsymbol{\mu}'$ , conditionally on  $\boldsymbol{\mu}'$ . The second term in the above display is zero by assumption. Moreover, from Lemma 2.3 the first and third terms go to zero when  $k \rightarrow +\infty$ .  $\square$

Our main theorem in this section deals with the convergence of sequences of random permutations to a random permuton. It generalizes the result of [30] which states that *deterministic* permuton convergence is characterized by convergence of pattern densities. We extend their proof to the case of *random* sequences, where permuton convergence *in distribution* is characterized by convergence of *average* pattern densities.

**Theorem 2.5.** *For any  $n$ , let  $\sigma_n$  be a random permutation of size  $n$ . Moreover, for any fixed  $k$ , let  $\mathbf{I}_{n,k}$  be a uniform random subset of  $[n]$  with  $k$  elements, independent of  $\sigma_n$ . The following assertions are equivalent.*

- (a)  $(\mu_{\sigma_n})_n$  converges in distribution for the weak topology to some random permuton  $\boldsymbol{\mu}$ .
- (b) The random infinite vector  $(\widetilde{\text{occ}}(\pi, \sigma_n))_{\pi \in \mathfrak{S}}$  converges in distribution in the product topology to some random infinite vector  $(\Lambda_\pi)_{\pi \in \mathfrak{S}}$ .

(c) For every  $\pi$  in  $\mathfrak{S}$ , there is a  $\Delta_\pi \geq 0$  such that

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] \xrightarrow{n \rightarrow \infty} \Delta_\pi.$$

(d) For every  $k$ , the sequence  $(\text{pat}_{I_{n,k}}(\sigma_n))_n$  of random permutations converges in distribution to some random permutation  $\rho_k$ .

Whenever these assertions are verified, we have  $(\Lambda_\pi)_\pi \stackrel{d}{=} (\widetilde{\text{occ}}(\pi, \mu))_\pi$  and for every  $\pi \in \mathfrak{S}_k$ ,

$$\mathbb{P}(\rho_k = \pi) = \Delta_\pi = \mathbb{E}[\Lambda_\pi] = \mathbb{E}[\widetilde{\text{occ}}(\pi, \mu)] = \mathbb{P}(\text{Perm}(\vec{\mathbf{m}}_k, \mu) = \pi).$$

*Observation 2.6.* In item (c) above, it is enough to consider all  $\pi$  of size at least 2. Indeed, for  $\pi = 1$ , the statement is trivial, since  $\widetilde{\text{occ}}(\pi, \cdot)$  is identically 1.

**Proof of (a)  $\Rightarrow$  (b).** Let  $\pi_1, \dots, \pi_r$  be a finite sequence of patterns. By [30, Lemma 5.3], the map  $\mu \mapsto (\widetilde{\text{occ}}(\pi_i, \mu))_{1 \leq i \leq r}$  is continuous. Therefore,  $\mu_{\sigma_n} \xrightarrow{d} \mu$  implies

$$(\widetilde{\text{occ}}(\pi_i, \mu_{\sigma_n}))_{1 \leq i \leq r} \xrightarrow{d} (\widetilde{\text{occ}}(\pi_i, \mu))_{1 \leq i \leq r}.$$

Using Lemma 2.1, one can replace each  $\widetilde{\text{occ}}(\pi_i, \mu_{\sigma_n})$  by  $\widetilde{\text{occ}}(\pi_i, \sigma_n)$  in the above convergence. This proves the convergence in distribution of all the finite-dimensional marginals  $(\widetilde{\text{occ}}(\pi_i, \sigma_n))_{1 \leq i \leq k}$ , and hence of  $(\widetilde{\text{occ}}(\pi, \sigma_n))_{\pi \in \mathfrak{S}}$  in the product topology (see for instance [15, ex. 2.4 p. 19]).

**Proof of (b)  $\Rightarrow$  (c).** If  $\widetilde{\text{occ}}(\pi, \sigma_n) \xrightarrow{d} \Lambda_\pi$ , as  $\widetilde{\text{occ}}$  is bounded by one we have

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[\Lambda_\pi].$$

**Proof of (c)  $\Rightarrow$  (d).** Fix  $\pi \in \mathfrak{S}_k$  and consider the sequence

$$\mathbb{P}^{\sigma_n, I_{n,k}}(\text{pat}_{I_{n,k}}(\sigma_n) = \pi) = \mathbb{E}^{\sigma_n}[\widetilde{\text{occ}}(\pi, \sigma_n)],$$

which converges if (c) holds (the equality comes from Eq. (7)). Since  $\text{pat}_{I_{n,k}}(\sigma_n)$  is a random variable taking its values in the finite set  $\mathfrak{S}_k$ , this says exactly that the sequence  $(\text{pat}_{I_{n,k}}(\sigma_n))_n$  converges in distribution.

**Proof of (d)  $\Rightarrow$  (a).** Consider a sequence of random permutations  $(\sigma_n)$  satisfying (d), i.e. for every  $k$ , there is a random permutation  $\rho_k$  such that  $\text{pat}_{I_{n,k}}(\sigma_n) \xrightarrow{d} \rho_k$ . In other terms, for every pattern  $\pi$  of size  $k$ , we have

$$\mathbb{P}^{\sigma_n, I_{n,k}}(\text{pat}_{I_{n,k}}(\sigma_n) = \pi) \rightarrow \mathbb{P}(\rho_k = \pi).$$

From Lemma 2.1 and Eq. (7), we get

$$\mathbb{E}^{\sigma_n}[\widetilde{\text{occ}}(\pi, \mu_{\sigma_n})] = \mathbb{E}^{\sigma_n}[\widetilde{\text{occ}}(\pi, \sigma_n)] + \mathcal{O}(1/n) = \mathbb{P}^{I_{n,k}, \sigma_n}(\text{pat}_{I_{n,k}}(\sigma_n) = \pi) + \mathcal{O}(1/n).$$

Set  $\theta_{k,n} = \text{Perm}(\vec{\mathbf{m}}_k, \mu_{\sigma_n})$ . Then, using Eq. (8), for every  $\pi \in \mathfrak{S}_k$ , we have

$$\begin{aligned} \mathbb{P}^{\theta_{k,n}}(\theta_{k,n} = \pi) &= \mathbb{P}^{\vec{\mathbf{m}}_k, \sigma_n}(\text{Perm}(\vec{\mathbf{m}}_k, \mu_{\sigma_n}) = \pi) = \mathbb{E}^{\sigma_n}[\widetilde{\text{occ}}(\pi, \mu_{\sigma_n})] \\ &= \mathbb{P}^{I_{n,k}, \sigma_n}(\text{pat}_{I_{n,k}}(\sigma_n) = \pi) + \mathcal{O}(1/n) \rightarrow \mathbb{P}(\rho_k = \pi). \end{aligned}$$

In other words,  $\theta_{k,n} \xrightarrow{d} \rho_k$ . Since  $\mu_{\rho_k}$  takes its values in a finite set of permutons, this implies

$$(9) \quad \mu_{\theta_{k,n}} \xrightarrow{d} \mu_{\rho_k}.$$

Let  $H : (\mathcal{M}, d_{\square}) \rightarrow \mathbb{R}$  be a bounded continuous functional. It holds that

$$\begin{aligned} |\mathbb{E}[H(\mu_{\sigma_n})] - \mathbb{E}[H(\mu_{\theta_{k,n}})]| &\leq \mathbb{E}[|H(\mu_{\sigma_n}) - H(\mu_{\theta_{k,n}})|] \\ &\leq \mathbb{E}\left[|H(\mu_{\sigma_n}) - H(\mu_{\theta_{k,n}})| \mathbf{1}_{d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}}) \leq 16k^{-1/4}}\right] \\ &\quad + \mathbb{E}\left[|H(\mu_{\sigma_n}) - H(\mu_{\theta_{k,n}})| \mathbf{1}_{d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}}) > 16k^{-1/4}}\right]. \end{aligned}$$

The first term can be bounded by introducing the modulus of continuity of  $H$ , which is defined as  $\omega(\varepsilon) = \sup_{d_{\square}(\xi, \zeta) \leq \varepsilon} |H(\xi) - H(\zeta)|$ . Since  $\mathcal{M}$  is compact, it goes to 0 when  $\varepsilon$  goes to 0. Hence,

$$\begin{aligned} &\mathbb{E}\left[|H(\mu_{\sigma_n}) - H(\mu_{\theta_{k,n}})| \mathbf{1}_{d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}}) \leq 16k^{-1/4}}\right] \\ &\leq \mathbb{E}\left[\omega(d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}})) \mathbf{1}_{d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}}) \leq 16k^{-1/4}}\right] \leq \omega(16k^{-1/4}). \end{aligned}$$

As for the second term, for  $k$  large enough, Lemma 2.3 yields

$$\begin{aligned} &\mathbb{E}\left[|H(\mu_{\sigma_n}) - H(\mu_{\theta_{k,n}})| \mathbf{1}_{d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}}) > 16k^{-1/4}}\right] \\ &\leq \mathbb{E}\left[2 \sup |H| \mathbf{1}_{d_{\square}(\mu_{\sigma_n}, \mu_{\theta_{k,n}}) > 16k^{-1/4}}\right] \leq \frac{1}{2} e^{-\sqrt{k}} 2 \sup |H|. \end{aligned}$$

Putting things together, we obtain

$$(10) \quad |\mathbb{E}[H(\mu_{\sigma_n})] - \mathbb{E}[H(\mu_{\theta_{k,n}})]| \leq \omega(16k^{-1/4}) + \frac{1}{2} e^{-\sqrt{k}} 2 \sup |H|.$$

Assume that  $(\mu_{\sigma_n})_n$  has a subsequence converging in distribution to a random permuton  $\mu'$ . Taking the limit when  $n \rightarrow \infty$  of (10) along this subsequence, we get

$$|\mathbb{E}[H(\mu')] - \mathbb{E}[H(\mu_{\rho_k})]| \leq \omega(16k^{-1/4}) + e^{-\sqrt{k}} \sup |H|.$$

(Recall indeed that  $(\theta_{k,n})_n$  converges to  $\rho_k$  in distribution.) The right-hand side tends to 0 when  $k$  tends to infinity, which proves that  $(\mu_{\rho_k})_k$  converges to  $\mu'$  in distribution as well.

Therefore, all converging subsequences of  $(\mu_{\sigma_n})_n$  converge to the same limit  $\mu'$ , which is the limit of  $(\mu_{\rho_k})_{k \geq 1}$ . Thanks to the compactness of the space of probability distributions on  $\mathcal{M}$ , this is enough to conclude that  $(\mu_{\sigma_n})$  has indeed a limit. Item (a) is proved.

**Proof of additional statements.** Assume that (a)–(d) hold. That  $(\Lambda_{\pi})_{\pi} \stackrel{d}{=} (\widetilde{\text{occ}}(\pi, \mu))_{\pi}$  follows from the proof of (a) $\Rightarrow$ (b). Fix any integer  $k$ , and any permutation  $\pi$  of size  $k$ . The above equality in distribution implies  $\mathbb{E}[\Lambda_{\pi}] = \mathbb{E}[\widetilde{\text{occ}}(\pi, \mu)]$ . That  $\Delta_{\pi} = \mathbb{E}[\Lambda_{\pi}]$  is clear from the proof of (b) $\Rightarrow$ (c). The equality  $\mathbb{P}(\rho_k = \pi) = \Delta_{\pi}$  follows from the proof of (c) $\Rightarrow$ (d). Finally,  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \mu)] = \mathbb{P}(\text{Perm}(\vec{\mathbf{m}}_k, \mu) = \pi)$  comes from Eq. (8).  $\square$

*Remark 2.7.* In some sense, Theorem 2.5 can be seen as an analogue of a theorem of Aldous for random trees [9, Theorem 18]. Both in permutations and trees, there is a natural way to construct a smaller structure from  $k$  elements of a big structure ( $k$  entries of a permutation, or  $k$  nodes of a tree). The goal is then to reduce the convergence of the big structure to the convergence, for each  $k$ , of the smaller structure constructed from  $k$  random elements. For trees, we need an extra tightness assumption (that the family of trees is “leaf-tight” in Aldous’ terminology). In our case, since the space of permutons is compact, we do not need such an assumption.

We finish this section by a comment on the existence of random permutons with prescribed marginals.

**Definition 2.8.** A family of random permutations  $(\rho_n)_n$  is consistent if

- i) for every  $n \geq 1$ ,  $\rho_n \in \mathfrak{S}_n$ ,
- ii) for every  $n \geq k \geq 1$ , if  $I_{n,k}$  is a uniform subset of  $[n]$  of size  $k$ , independent of  $\rho_n$ , then  $\text{pat}_{I_{n,k}}(\rho_n) \stackrel{d}{=} \rho_k$ .

It follows that consistent family of random permutations and random permutons are essentially equivalent:

**Proposition 2.9.** If  $\mu$  is a random permuton, then the family defined by  $\rho_k \stackrel{d}{=} \text{Perm}(\vec{\mathbf{m}}_k, \mu)$  is consistent. Conversely, for every consistent family of random permutations  $(\rho_k)_{k \geq 1}$ , there exists a random permuton  $\mu$  whose distribution is uniquely determined, such that  $\text{Perm}(\vec{\mathbf{m}}_k, \mu) \stackrel{d}{=} \rho_k$ . In that case,  $\mu_{\rho_n} \xrightarrow[n \rightarrow \infty]{d} \mu$ .

*Proof.* Set  $n \geq k \geq 1$ . The first assertion follows from the following coupled construction of  $\vec{\mathbf{m}}_n$  and  $\vec{\mathbf{m}}_k$ :  $\vec{\mathbf{m}}_k$  is a uniform random subset of  $\vec{\mathbf{m}}_n$ , chosen independently of it. It follows that  $\text{Perm}(\vec{\mathbf{m}}_k, \mu) = \text{pat}_{I_{n,k}}(\text{Perm}(\vec{\mathbf{m}}_n, \mu))$ , for some random subset  $I_{n,k}$  of  $[n]$ . By construction, the distribution of  $I_{n,k}$  is uniform and independent of  $\text{Perm}(\vec{\mathbf{m}}_n, \mu)$ . Hence the consistency follows.

The converse is immediate, by applying Theorem 2.5 to the sequence  $(\rho_k)_{k \geq 1}$ . Consistency ensures that we get the prescribed marginals, and uniqueness in distribution follows by Proposition 2.4.  $\square$

### 3. CODING PERMUTATIONS BY TREES

**3.1. Substitution trees.** As seen in Section 1.3 (Theorem 1.5), any permutation  $\sigma$  can be recursively decomposed using substitutions in a canonical way and this decomposition can be encoded in a canonical tree. However, if we do not impose conditions on  $\theta$  and the  $\pi^{(i)}$ ’s (as done in Theorem 1.5), a permutation  $\sigma$  may be represented in many ways as a substitution  $\sigma = \theta[\pi^{(1)}, \dots, \pi^{(d)}]$ , where the  $\pi^{(i)}$ ’s themselves may be further decomposed using substitutions. Such decompositions can be recorded in *substitution trees*.

**Definition 3.1.** A rooted planar tree is either a leaf, or consists of a root node  $\emptyset$  with an ordered  $k$ -tuple of subtrees attached to the root, which are themselves rooted planar trees. In our context, the size of a tree  $t$  is its number of leaves. It is denoted  $|t|$ , whereas  $\#t$  denotes the number of nodes of  $t$  (including both leaves and internal nodes).

Internal vertices of all trees considered in this paper have degree at least 2. It is natural (and also convenient for counting purposes in Section 4) to consider that the single leaf of the tree of size 1 is also its root (and is therefore also denoted  $\emptyset$ ).

Since we work with *planar* trees, we can label their leaves canonically with the integers from 1 to  $|t|$ : the leaf labeled by  $i$  is the  $i$ th leaf met in the depth-first traversal of  $t$  which choses left before right. A subset of the set of leaves of a tree  $t$  is therefore canonically represented by a subset  $I$  of  $[|t|]$ .

**Definition 3.2.** A substitution tree of size  $n$  is a labeled rooted planar tree with  $n$  leaves, where any internal node with  $k \geq 2$  children is labeled by a permutation of size  $k$ . Internal nodes with only one child are forbidden.

Internal nodes labeled by the ascending permutation  $12 \cdots r$  or the descending permutation  $r \cdots 21$  (for some  $r \geq 2$ ) will play a particular role. Therefore we replace *every* such label with a  $\oplus$  (for ascending permutations) or a  $\ominus$  (for descending permutations). Since the size of a label corresponds to the number of children and since there is exactly one ascending (resp. descending) permutation of each size, there is no loss of information in this replacement. Internal nodes labeled  $\oplus$  or  $\ominus$  are called *linear nodes*, the other nodes being called *nonlinear*. Among nonlinear nodes, the ones labeled by simple permutations are called *simple nodes*.

An example of substitution tree is shown in Fig. 5, left.

**Definition 3.3.** Let  $t$  be a substitution tree. We define inductively the permutation  $\text{perm}(t)$  associated with  $t$ :

- if  $t$  is just a leaf, then  $\text{perm}(t) = 1$ ;
- if the root of  $t$  has  $r \geq 2$  children with corresponding subtrees  $t_1, \dots, t_r$  (from left to right), and is labeled with the permutation  $\theta$ , then  $\text{perm}(t)$  is the permutation obtained as the substitution of  $\text{perm}(t_1), \dots, \text{perm}(t_r)$  in  $\theta$ :

$$\text{perm}(t) = \theta[\text{perm}(t_1), \dots, \text{perm}(t_r)].$$

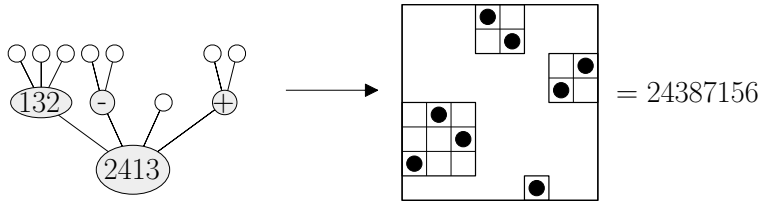


FIGURE 5. A substitution tree encoding a permutation.

Fig. 5 illustrates this construction. When  $\text{perm}(t) = \sigma$ , we say that  $t$  is a tree that *encodes*  $\sigma$ , or a tree *associated with*  $\sigma$ . Nonsimple permutations  $\sigma$  are encoded by several trees  $t$ . However, if we restrict ourselves to canonical trees (which are particular cases of substitution trees; see Definition 1.6), we have uniqueness. Indeed, from Theorem 1.5, to any permutation  $\sigma$  we can associate uniquely a canonical tree  $t$  such that  $\text{perm}(t) = \sigma$ .

The remaining of Section 3.1 is devoted to the proof of simple combinatorial lemmas on the structure of the set of substitution trees associated with a given permutation  $\sigma$ . These lemmas are useful in Section 5.

We first make the following observation. Take a substitution tree  $\tau$  of some permutation  $\pi$  with a marked node  $v$  labeled by  $\theta$ . Consider also a substitution tree  $\tau'$  of  $\theta$ . Then replacing  $v$  by the tree  $\tau'$  yields a new substitution tree  $\tau''$  of *the same permutation*  $\pi$ . (When doing this replacement the  $|\theta|$  subtrees attached to  $v$  are glued on the leaves of  $\tau'$ , respecting their order, see Fig. 6.) This operation will be referred to as *the inflation of  $v$  with  $\tau'$* .

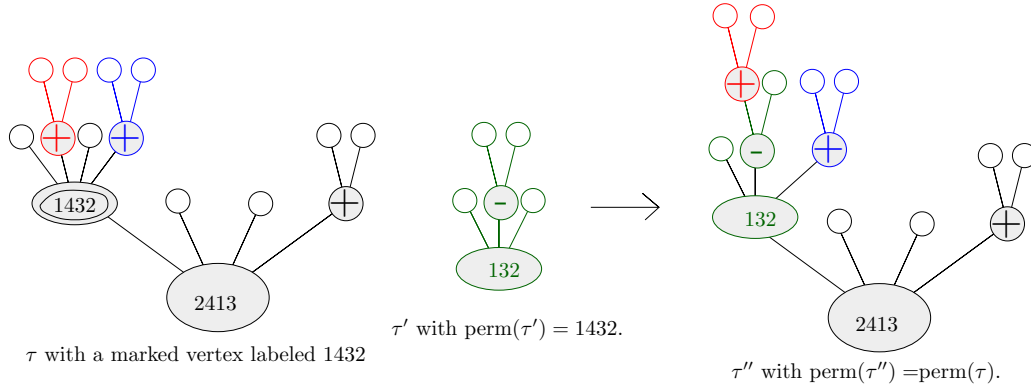


FIGURE 6. Illustration of the inflation procedure (best seen with colors).

Conversely, consider a connected set  $A$  of internal nodes in a substitution tree  $\tau''$  of  $\pi$  (i.e. if a node  $v$  of  $\tau''$  has an ancestor in  $A$  and a descendant in  $A$ , then  $v \in A$ ). From this set we build a substitution tree  $\tau'$  whose set of internal nodes is  $A$ , the ancestor-descendant relation in  $\tau'$  is inherited from the one in  $\tau''$ , and we add leaves so that the degree of each node of  $A$  is the same in  $\tau'$  than in  $\tau''$ . We denote  $\theta = \text{perm}(\tau')$ . Then merging all nodes in  $A$  into a single node labeled by  $\theta$  turns  $\tau''$  into a new substitution tree  $\tau$  of *the same permutation*  $\pi$ . We call this a *merge operation*. For example, the tree  $\tau$  of Fig. 6 can be obtained from the tree  $\tau''$  of the same figure by merging the nodes labeled 132 and  $\ominus$ .

We now consider a last family of substitution trees. An *expanded tree* is a substitution tree where nonlinear nodes are labeled by simple permutations, while linear nodes are required to be binary.

**Lemma 3.4.** *Any expanded tree of  $\pi$  is obtained from its canonical tree by inflating all nodes labeled by  $\oplus$  (resp.  $\ominus$ ) with binary trees whose internal nodes are all labeled by  $\oplus$  (resp.  $\ominus$ )*

*Proof.* Let  $\tau$  be an expanded tree of  $\pi$ . Consider, if any, two adjacent linear nodes of  $\tau$  with the same label (either both  $\oplus$  or both  $\ominus$ ) and merge them. Note that the resulting node will still have label  $\oplus$  or  $\ominus$ . We repeat this operation until there is no adjacent linear nodes with the same label. Nonlinear nodes in the resulting tree  $\tau'$  are all labeled by simple permutations: it is the case in  $\tau$  (by definition of expanded trees) and we did not create any new nonlinear nodes. Therefore  $\tau'$  satisfy all conditions of canonical trees (see Definition 1.6). By uniqueness,  $\tau'$  is *the* canonical

tree of  $\pi$ . Reversing the merge operations,  $\tau$  can be obtained from  $\tau'$  by inflating its linear nodes, which proves the proposition.  $\square$

This lemma has an interesting immediate corollary, regarding the number and properties of expanded trees.

**Corollary 3.5.** *Let  $\pi$  be a permutation and  $d_1, \dots, d_r$  (resp.  $e_1, \dots, e_s$ ) be the degrees of the nodes labeled  $\oplus$  (resp.  $\ominus$ ) in the canonical tree of  $\pi$ . Then*

- *the number  $\widetilde{N}_\pi$  of expanded trees of  $\pi$  is  $\prod_{i=1}^r \text{Cat}_{d_i-1} \prod_{j=1}^s \text{Cat}_{e_j-1}$ , where we denote by  $\text{Cat}_k := \frac{1}{k+1} \binom{2k}{k}$  the  $k$ -th Catalan number.*
- *each expanded tree of  $\pi$  has  $\sum_{i=1}^r (d_i - 1)$  nodes labeled  $\oplus$  and  $\sum_{j=1}^s (e_j - 1)$  nodes labeled  $\ominus$ .*
- *the labels of the nonlinear nodes in any expanded tree of  $\tau$  are the same as in its canonical tree.*

**Lemma 3.6.** *Any substitution tree of  $\pi$  can be obtained from some expanded tree of  $\pi$  by merge operations.*

*Proof.* The proof is similar to that of Lemma 3.4. Starting from any substitution tree of  $\pi$  and inflating every node that is neither simple nor binary by an expanded tree encoding its label, we get an expanded tree. Reversing these inflation operations, we can obtain any substitution tree from some expanded tree of  $\pi$ , using only merge operations.  $\square$

**3.2. Induced trees.** Since permutations are encoded by trees and since we are interested in patterns in permutations, we consider an analogue of patterns in trees: this leads to the notion of *induced trees*.

**Definition 3.7** (First common ancestor). *Let  $t$  be a tree, and  $u$  and  $v$  be two nodes (internal nodes or leaves) of  $t$ . The first common ancestor of  $u$  and  $v$  is the node furthest away from the root  $\emptyset$  that appears on both paths from  $\emptyset$  to  $u$  and from  $\emptyset$  to  $v$  in  $t$ .*

The following simple observation allows to read the relative order of  $\sigma_i$  and  $\sigma_j$  in any substitution tree encoding  $\sigma$ .

*Observation 3.8.* Let  $i \neq j$  be two leaves of a substitution tree  $t$  and  $\sigma = \text{perm}(t)$ . Let  $v$  be the first common ancestor of  $i, j$  in  $t$  and  $\theta$  be the permutation labeling  $v$ . We define  $k$  (resp.  $\ell$ ) such that the  $k$ -th (resp.  $\ell$ -th) child of  $v$  is an ancestor of  $i$  (resp.  $j$ ).

Then  $\sigma_i > \sigma_j$  if and only if  $\theta_k > \theta_\ell$ .

**Definition 3.9** (Induced tree). *Let  $t$  be a substitution tree, and let  $I$  be a subset of the leaves of  $t$ . The tree  $t_I$  induced by  $I$  is the substitution tree of size  $|I|$  defined as follows. The tree structure of  $t_I$  is given by:*

- *the leaves of  $t_I$  are the leaves of  $t$  labeled by elements of  $I$ ;*
- *the internal nodes of  $t_I$  are the nodes of  $t$  that are first common ancestors of two (or more) leaves in  $I$ ;*
- *the ancestor-descendant relation in  $t_I$  is inherited from the one in  $t$ ;*
- *the order between the children of an internal node of  $t_I$  is inherited from  $t$ .*

- if  $v$  is labeled by a permutation  $\theta$  in  $t$ , the label of  $v$  in  $t_I$  is given by the pattern of  $\theta$  induced by the children of  $v$  having a descendant that belongs to  $t_I$  (or equivalently, to  $I$ ).

FIGURE 7. On the left: A substitution tree  $t$  of size  $n = 26$  (which happens to be a canonical tree), where leaves are indicated both by  $\circ$  and  $\bullet$ . Among these 26 leaves,  $|I| = 9$  leaves are marked and indicated by  $\bullet$ . In green are shown the internal nodes of  $t$  which are first common ancestors of these 9 marked leaves. On the right: The substitution tree induced by the 9 marked leaves. Observe that the node  $v$  labeled by 362514 in  $t$  is labeled by 2413 in  $t_I$ . This is because only the first, second, fifth and sixth children of  $v$  have descendants that belong to  $I$ , and  $\text{pat}_{\{1,2,5,6\}}(362514) = 2413$ . The induced tree is not canonical since 132 is not simple.

$\oplus$  in  $t$ , it has also label  $\oplus$  in  $t_I$ . Indeed,  $\oplus$  nodes correspond to increasing permutations and all patterns of increasing permutations are increasing permutations. The same holds with  $\ominus$ . The converse is however not true: a node can be linear in  $t_I$  but nonlinear in  $t$  (e.g. the bottommost green node in Fig. 7).

*Observation 3.10.* By definition, for any substitution tree  $t$  with  $k$  leaves and subset  $I$  of  $[k]$ ,  $t_I$  is a substitution tree. However, if  $t$  is a canonical tree,  $t_I$  is a substitution tree which is not necessarily canonical (see for example Fig. 7).

An important feature of induced trees is the following, which follows from Observation 3.8 and is illustrated in Fig. 8.



**Lemma 3.11.** *Let  $t$  be a substitution tree with  $k$  leaves, and  $I$  be a subset of  $[k]$ . We have*

$$\text{pat}_I(\text{perm}(t)) = \text{perm}(t_I).$$

As a consequence of this formula, counting the total number of occurrences of a given pattern in some family of permutations can be reduced to counting the total number of induced trees equal to a given  $t_0$  in the corresponding family of canonical trees. This is precisely the goal of the next section.

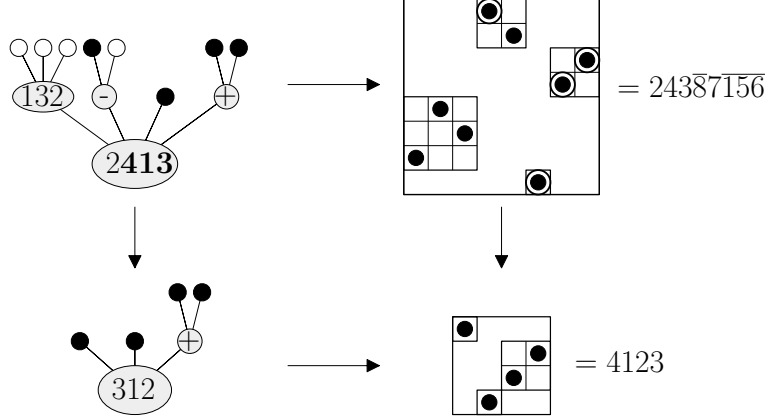


FIGURE 8. Illustration of Lemma 3.11. Top: a substitution tree  $t$  with  $|I|$  marked leaves (in this example  $k = 8$  and  $I = \{4, 6, 7, 8\}$ ), and the permutation  $\text{perm}(t)$  it encodes, with the corresponding  $|I|$  marked elements (at positions in  $I$ ). Bottom: the induced tree  $t_I$  and the induced pattern  $\text{pat}_I(\text{perm}(t)) = \text{perm}(t_I)$ .

#### 4. EXACT ENUMERATION OF VARIOUS FAMILIES OF TREES

Let  $\mathcal{S}$  be a fixed family of simple permutations. Recall that its generating function is

$$S(z) = \sum_{\alpha \in \mathcal{S}} z^{|\alpha|} = \sum_{n \geq 4} s_n z^n,$$

where  $s_n$  is the number of permutations of size  $n$  in  $\mathcal{S}$ . An  $\mathcal{S}$ -canonical tree is any canonical tree whose simple nodes carry labels in  $\mathcal{S}$ . We denote by  $\mathcal{T}$  the combinatorial class of  $\mathcal{S}$ -canonical trees, the size of  $|t|$  a tree  $t$  being its number of leaves. Recall that  $\langle \mathcal{S} \rangle$  is by definition the set of permutations whose canonical tree is in  $\mathcal{T}$ . Since canonical trees encode permutations in a unique way, perm defines a size-preserving bijection between  $\mathcal{T}$  and  $\langle \mathcal{S} \rangle$ . Both have therefore the same generating function which we denote by

$$T(z) = \sum_{t \in \mathcal{T}} z^{|t|} = \sum_{\sigma \in \langle \mathcal{S} \rangle} z^{|\sigma|}.$$

In Section 4.1 below, we explain how to compute  $T(z)$  starting from the datum  $S(z)$ . We then study families of  $\mathcal{S}$ -canonical trees with one marked leaf, with constraints on the root and/or on the marked leaf. These are building blocks for Section 4.2, where we consider the family of  $\mathcal{S}$ -canonical trees with  $k$  marked leaves, inducing a given tree  $t_0$ .

**4.1. Generating functions of  $\mathcal{S}$ -canonical trees (possibly with marked leaves).** In order to compute  $T(z)$  in terms of  $S(z)$ , we need to introduce the auxiliary family  $\mathcal{T}_{\text{not}\oplus}$  (resp.  $\mathcal{T}_{\text{not}\ominus}$ ) of  $\mathcal{S}$ -canonical trees with a root (always denoted  $\emptyset$ ) that is **not** labeled  $\oplus$  (resp.  $\ominus$ ), and its generating function  $T_{\text{not}\oplus}$  (resp.  $T_{\text{not}\ominus}$ ):

$$T_{\text{not}\oplus}(z) = \sum_{\substack{t \in \mathcal{T}; \\ \emptyset \text{ is not labeled } \oplus}} z^{|t|}.$$

Note that replacing all labels  $\ominus$  by  $\oplus$  and  $\oplus$  by  $\ominus$  defines an involution on  $\mathcal{S}$ -canonical trees. This implies in particular  $T_{\text{not}\oplus} = T_{\text{not}\ominus}$  and will be used to get other similar identities below.

**Proposition 4.1.** *The generating function  $T_{\text{not}\oplus}$  is determined by the following implicit equation*

$$(11) \quad T_{\text{not}\oplus} = z + \frac{T_{\text{not}\oplus}^2}{1 - T_{\text{not}\oplus}} + S\left(\frac{T_{\text{not}\oplus}}{1 - T_{\text{not}\oplus}}\right).$$

*The main series  $T$  is then simply given in terms of  $T_{\text{not}\oplus}$  by*

$$(12) \quad T = \frac{T_{\text{not}\oplus}}{1 - T_{\text{not}\oplus}}.$$

*Proof.* A tree of  $\mathcal{T}$  is either a leaf, or a root labeled  $\oplus$  and a sequence of at least two trees in  $\mathcal{T}_{\text{not}\oplus}$ , or a root labeled  $\ominus$  and a sequence of at least two trees in  $\mathcal{T}_{\text{not}\ominus}$ , or a root labeled by  $\alpha \in \mathcal{S}$  and a sequence of  $|\alpha|$  unconstrained trees. Therefore

$$T = z + \frac{T_{\text{not}\oplus}^2}{1 - T_{\text{not}\oplus}} + \frac{T_{\text{not}\ominus}^2}{1 - T_{\text{not}\ominus}} + S(T) = z + 2 \frac{T_{\text{not}\oplus}^2}{1 - T_{\text{not}\oplus}} + S(T)$$

Similarly,

$$(13) \quad T_{\text{not}\oplus} = T_{\text{not}\ominus} = z + \frac{T_{\text{not}\oplus}^2}{1 - T_{\text{not}\oplus}} + S(T).$$

By combining these two equations we get  $T = T_{\text{not}\oplus} + \frac{T_{\text{not}\oplus}^2}{1 - T_{\text{not}\oplus}} = \frac{T_{\text{not}\oplus}}{1 - T_{\text{not}\oplus}}$ , that is Eq. (12). Substituting it back in (13) gives Eq. (11).  $\square$

We now consider trees with a marked leaf. As before, subscripts indicate a constraint on the root. The generating function of trees with a marked leaf counted by their number of **unmarked** leaves is obtained by differentiating the generating function of trees without marked leaf:  $T'$ ,  $T'_{\text{not}\oplus}$ ,  $T'_{\text{not}\ominus}$ . Indeed,

$$T'_{\text{not}\oplus}(z) = \sum_{\substack{t \in \mathcal{T}; \\ \emptyset \text{ is not labeled } \oplus}} |t| z^{|t|-1} = \sum_{\substack{\hat{t} \text{ obtained by marking a leaf} \\ \text{from a tree of } \mathcal{T} \text{ of root not labeled } \oplus}} z^{\#\text{unmarked leaves of } \hat{t}}.$$

Accordingly, we denote by  $\mathcal{T}'$ ,  $\mathcal{T}'_{\text{not}\oplus}$  and  $\mathcal{T}'_{\text{not}\ominus}$  the families of trees counted by  $T'$ ,  $T'_{\text{not}\oplus}$  and  $T'_{\text{not}\ominus}$ .

Consistently, we use superscripts when we consider families of trees with a marked leaf that satisfies an additional constraint, and similarly for their generating function. We say that a leaf is  $\oplus$ -replaceable (resp.  $\ominus$ -replaceable) if it may be replaced by a tree whose root is labeled  $\oplus$  (resp.  $\ominus$ ) without violating the definition of canonical trees (see the third item in Definition 1.6). In other words, its parent (if it exists) should be labeled by  $\ominus$  or by a simple permutation (resp. by  $\oplus$  or by a simple permutation). We then denote  $\mathcal{T}_{\text{not}\oplus}^+$  (resp.  $\mathcal{T}_{\text{not}\oplus}^-$ ) the families of trees in  $\mathcal{T}_{\text{not}\oplus}$  with a  $\oplus$ -replaceable marked leaf (resp. a  $\ominus$ -replaceable marked leaf). Similar definitions hold for  $\mathcal{T}^+$ ,  $\mathcal{T}^-$ ,  $\mathcal{T}_{\text{not}\ominus}^+$  and  $\mathcal{T}_{\text{not}\ominus}^-$ .

As for  $T'$ , we take the convention that  $T_{\text{not}\oplus}^+$  and all generating functions with superscript count trees according to the number of **unmarked** leaves. By definition,  $T_{\text{not}\oplus}^-$  has constant coefficient 1 (corresponding to the tree consisting of a single leaf). We however take the convention that  $T_{\text{not}\oplus}^+$  has constant coefficient 0: in other terms, the single leaf is excluded from the family  $\mathcal{T}_{\text{not}\oplus}^+$  (intuitively, a single leaf cannot be replaced by a tree with root labeled  $\oplus$ , since the trees in  $\mathcal{T}_{\text{not}\oplus}$  should not have a root labeled  $\oplus$ ).

**Proposition 4.2.** *The generating functions  $T^+$ ,  $T_{\text{not}\oplus}^+$  and  $T_{\text{not}\oplus}^-$  are given by the following formulas:*

$$(14) \quad T^+ = \frac{1}{1 - WS'(T) - W - S'(T)};$$

$$(15) \quad T_{\text{not}\oplus}^+ = \frac{1}{1 + W} T^+;$$

$$(16) \quad T_{\text{not}\oplus}^- = (WS'(T) + W + S'(T))T_{\text{not}\oplus}^+$$

where  $T$  and  $T_{\text{not}\oplus}$  are given by Proposition 4.1 and  $W = (\frac{1}{1 - T_{\text{not}\oplus}})^2 - 1$ .

Quantities with a minus superscript are obtained by symmetry:  $T^- = T^+$ ,  $T_{\text{not}\ominus}^- = T_{\text{not}\oplus}^+$  and  $T_{\text{not}\oplus}^- = T_{\text{not}\ominus}^+$ .

*Proof.* Consider a tree  $t$  in  $\mathcal{T}_{\text{not}\oplus}^+$ . As explained above,  $|t| \neq 1$  and we distinguish cases according to the label of the root of  $t$ , which may be either  $\ominus$  or a simple permutation.

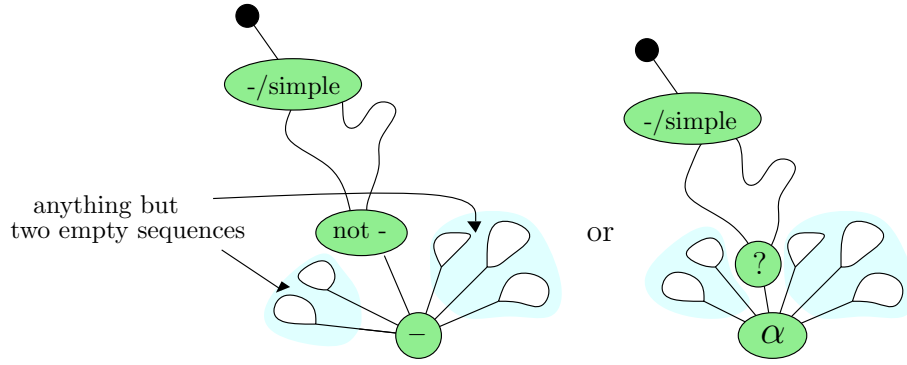


FIGURE 9. Proof of Proposition 4.2.

- i) The root of  $t$  is labeled  $\ominus$  (see left of Fig. 9). Then  $t$  can be decomposed as a tree in  $\mathcal{T}_{\text{not}\ominus}^+$  (which may be a single leaf) and a nonempty pair of sequences of unmarked trees in  $\mathcal{T}_{\text{not}\ominus}$ .
- ii) The root of  $t$  is labeled by a simple permutation  $\alpha \in \mathcal{S}$  of size  $d$  (see right of Fig. 9). Then  $t$  can be decomposed as a  $d$ -uple of unconstrained trees, with one of them having a  $\oplus$ -replaceable marked leaf.

Therefore we have

$$T_{\text{not}\oplus}^+ = WT_{\text{not}\ominus}^+ + S'(T)T^+,$$

where  $W = (\frac{1}{1-T_{\text{not}\oplus}})^2 - 1$  counts nonempty pairs of sequences of unmarked trees in  $\mathcal{T}_{\text{not}\ominus}$  (since  $T_{\text{not}\ominus} = T_{\text{not}\oplus}$ ).

Similarly, we have

$$\begin{aligned} T_{\text{not}\ominus}^+ &= 1 + WT_{\text{not}\oplus}^+ + S'(T)T^+; \\ T^+ &= 1 + WT_{\text{not}\oplus}^+ + WT_{\text{not}\ominus}^+ + S'(T)T^+. \end{aligned}$$

The above three equations form a system with three indeterminates:  $T_{\text{not}\oplus}^+$ ,  $T_{\text{not}\ominus}^+$  and  $T^+$  ( $W$  and  $T$  are known thanks to Proposition 4.1). Solving this system with a computer algebra software gives Eqs. (14) to (16).

The symmetry argument giving  $T_{\text{not}\oplus}^-$ ,  $T_{\text{not}\ominus}^-$  and  $T^-$  consist as before in exchanging  $\ominus$  and  $\oplus$  labels in  $\mathcal{S}$ -canonical trees.  $\square$

**4.2. Generating function counting trees with marked leaves inducing a given tree.** To enumerate trees with marked leaves inducing a given tree, we introduce an other kind of generating functions. Recall that for any permutations  $\alpha$  and  $\theta$ ,  $\text{occ}(\theta, \alpha)$  is the number of occurrences of  $\theta$  in  $\alpha$ . For a permutation  $\theta$ , we set

$$(17) \quad \text{Occ}_\theta(z) = \sum_{\alpha \in \mathcal{S}} \text{occ}(\theta, \alpha) z^{|\alpha| - |\theta|}.$$

*Observation 4.3.* For  $d \geq 1$  and any fixed  $\alpha$ ,  $\sum_{\theta \in \mathfrak{S}_d} \text{occ}(\theta, \alpha) = \binom{|\alpha|}{d}$ . Therefore  $\sum_{\theta \in \mathfrak{S}_d} \text{Occ}_\theta$  is related to the  $d$ -th derivative of  $S$  by  $\sum_{\theta \in \mathfrak{S}_d} \text{Occ}_\theta = \frac{S^{(d)}}{d!}$ . This implies that the radius of convergence of each  $\text{Occ}_\theta$  is at least  $R_S$ , the radius of convergence of  $S$ .

Fix a substitution tree  $t_0$  with  $k$  leaves. Let us call  $\mathcal{T}_{t_0}$  the family of  $\mathcal{S}$ -canonical trees  $t$  with  $k$  marked leaves  $\vec{\ell} = (\ell_1, \dots, \ell_k)$  such that these leaves induce  $t_0$ :

$$\mathcal{T}_{t_0} = \left\{ (t, \vec{\ell}) \text{ such that } t \in \mathcal{T} \text{ and } t_{\vec{\ell}} = t_0 \right\}.$$

We define the size of an object  $(t, \vec{\ell})$  as the number of leaves in  $t$  (**both marked and unmarked**). The corresponding generating series is denoted  $T_{t_0}(z)$

Let  $(t, \vec{\ell}) \in \mathcal{T}_{t_0}$ . As noted after the definition of induced trees, a nonlinear node of  $t_{\vec{\ell}}$  has to come from a nonlinear node of  $t$ , whereas a linear node of  $t_{\vec{\ell}}$  may come from a linear or a nonlinear node of  $t$ . In order to ease the enumeration, we partition  $\mathcal{T}_{t_0}$  according to the set of nodes of  $t_{\vec{\ell}} = t_0$  coming from nonlinear nodes of  $t$  (that is, simple nodes of  $t$  since  $t$  is canonical).

More formally, let  $\text{Int}(t)$  be the set of internal nodes of a tree  $t$ . With each  $(t, \vec{\ell}) \in \mathcal{T}_{t_0}$ , we associate the set  $\text{FCA}(\vec{\ell}) \subseteq \text{Int}(t)$  of the first common ancestors of  $\vec{\ell}$  in  $t$ . From the definition of induced tree, a node  $v$  in  $t_0$  corresponds to a unique node in  $\text{FCA}(\vec{\ell})$ , that we denote  $\varphi(v)$ . For  $V_s \subseteq \text{Int}(t_0)$ , let

$$\mathcal{T}_{t_0, V_s} = \left\{ (t, \vec{\ell}) \in \mathcal{T}_{t_0} : \{v \in \text{Int}(t_0) : \varphi(v) \text{ is simple}\} = V_s \right\}.$$

Clearly  $\mathcal{T}_{t_0, V_s}$  is nonempty if and only if  $V_s$  contains every nonlinear node of  $t_0$ . An example of a marked tree  $(t, \vec{\ell})$  with the corresponding pair  $(t_0, V_s)$  is shown on Fig. 10. In pictures, we will always circle nodes  $v$  in  $V_s$  and the corresponding nodes  $\varphi(v)$  in  $t$ .

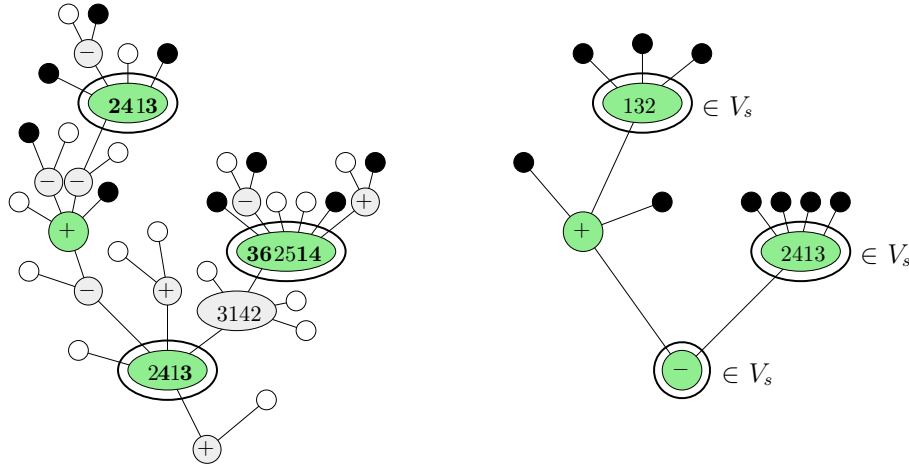


FIGURE 10. On the left: A canonical tree  $t$  with  $k = 9$  marked leaves (in black). The set  $\text{FCA}(\vec{\ell})$  of first common ancestors of the marked leaves is the set of green nodes. On the right: the corresponding decorated tree  $(t_0, V_s)$ . The tree  $t_0$  is the tree induced by the marked leaves in  $t$ .

**Definition 4.4.** A decorated tree is a pair  $(t_0, V_s)$  where

- $t_0$  is a substitution tree;

- $V_s$  is a subset of  $\text{Int}(t_0)$  that contains all nonlinear nodes.

Therefore we have the following decomposition:

$$\mathcal{T}_{t_0} = \bigcup_{V_s \text{ s.t. } (t_0, V_s) \text{ is a decorated tree}} \mathcal{T}_{t_0, V_s}.$$

Let  $(t_0, V_s)$  be a decorated tree. We consider the generating function  $T_{t_0, V_s}$  of  $\mathcal{T}_{t_0, V_s}$ , the size of  $(t, \vec{\ell})$  being its number of leaves (both marked and unmarked):

$$T_{t_0, V_s}(z) = \sum_{(t, \vec{\ell}) \in \mathcal{T}_{t_0, V_s}} z^{|t|}.$$

To compute  $T_{t_0, V_s}$ , we introduce some notation. For every internal node  $v$  of  $t_0$ , let

- $\theta_v$  be the permutation labeling  $v$ ,
- $d'_v$  be its number of children which are leaves or in  $V_s$ ,
- $d_v^+$  be its number of children which are not in  $V_s$  and are labeled by  $\oplus$ ,
- $d_v^-$  be its number of children which are not in  $V_s$  and are labeled by  $\ominus$ ,
- $d_v = d'_v + d_v^+ + d_v^-$  be its total number of children.

We also set the type of root to be  $\prime$  if the root of  $t_0$  is in  $V_s$ , and  $+$  (resp.  $-$ ) if the root is not in  $V_s$  and labeled  $\oplus$  (resp.  $\ominus$ ).

**Proposition 4.5** (Enumeration of trees with marked leaves inducing a given decorated tree). *Let  $(t_0, V_s)$  be a decorated tree and  $k$  be its number of leaves. Then*

$$(18) \quad T_{t_0, V_s} = z^k T^{\text{type of root}} \prod_{v \in \text{Int}(t_0)} A_v,$$

where

$$(19) \quad A_v = \begin{cases} \text{Occ}_{\theta_v}(T) (T')^{d'_v} (T^+)^{d_v^+} (T^-)^{d_v^-} & \text{if } v \in V_s, \\ \left( \frac{1}{1-T_{\text{not}\oplus}} \right)^{d_v+1} (T'_{\text{not}\oplus})^{d'_v} (T^+_{\text{not}\oplus})^{d_v^+} (T^-_{\text{not}\oplus})^{d_v^-} & \text{if } v \notin V_s \text{ and } \theta_v = \oplus, \\ \left( \frac{1}{1-T_{\text{not}\ominus}} \right)^{d_v+1} (T'_{\text{not}\ominus})^{d'_v} (T^+_{\text{not}\ominus})^{d_v^+} (T^-_{\text{not}\ominus})^{d_v^-} & \text{if } v \notin V_s \text{ and } \theta_v = \ominus. \end{cases}$$

*Proof.* The proof is based on a decomposition of marked  $\mathcal{S}$ -canonical trees  $(t, \vec{\ell})$  of  $\mathcal{T}_{t_0, V_s}$  followed by a study of the series  $A_v$  depending on the type of the node  $\varphi(v)$  in  $t$ .

**First step: Decomposing a tree in  $\mathcal{T}_{t_0, V_s}$ .**

We fix a decorated tree  $(t_0, V_s)$  with  $k$  leaves and a marked  $\mathcal{S}$ -canonical tree  $(t, \vec{\ell}) \in \mathcal{T}_{t_0, V_s}$ . We want to decompose  $t$  into subtrees, one for each internal node of  $t_0$  plus one attached to the root of  $t$ . Recall that  $\varphi : \text{Int}(t_0) \rightarrow \text{FCA}(\vec{\ell}) \subseteq \text{Int}(t)$  is the correspondence between the internal nodes of  $t_0$  and the set of first common ancestors of leaves  $\vec{\ell}$  in  $t$ .

For every internal node  $v$  of  $t_0$ , let  $t_v$  be the subtree of  $t$  defined as follows.

- The root of  $t_v$  is  $\varphi(v)$ .
- The nodes of  $t_v$  are descendants of  $\varphi(v)$ .

- A descendant of  $\varphi(v)$  in  $t$  belongs to  $t_v$  if and only if its first proper ancestor in  $\text{FCA}(\vec{\ell})$  is  $\varphi(v)$  (proper meaning different from the node itself).

Moreover we define  $t_\emptyset$  as the subtree of  $t$  rooted at the root of  $t$  and containing the nodes of  $t$  having no proper ancestor in  $\text{FCA}(\vec{\ell})$ . (If  $\varphi$  maps the root of  $t_0$  to the root of  $t$ , then  $t_\emptyset$  is reduced to a leaf.)

A schematic representation of the trees  $t_v$  and  $t_\emptyset$  is given in Fig. 11.

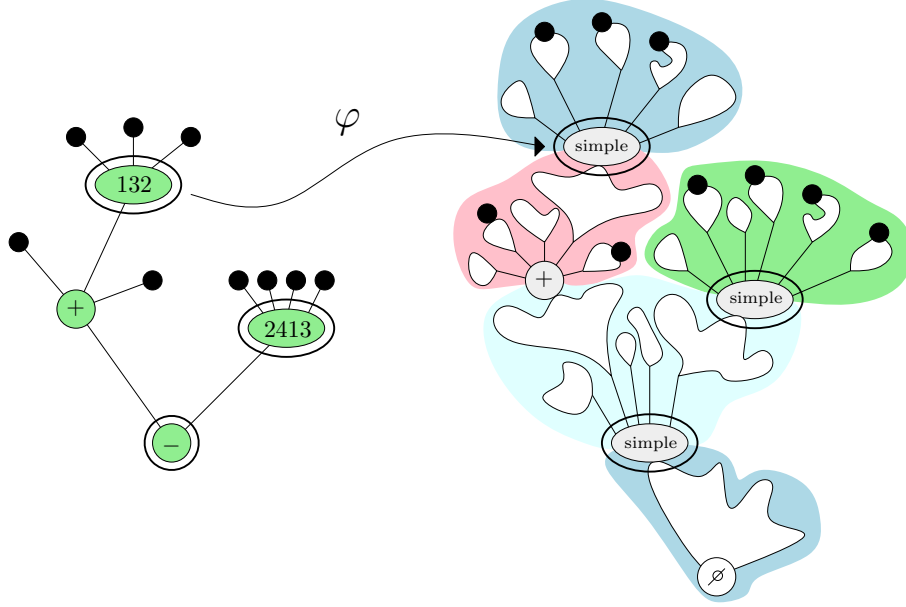


FIGURE 11. On the left: A decorated tree  $(t_0, V_s)$  with 9 leaves and 4 internal nodes. On the right: A schematic representation of a tree  $t \in \mathcal{T}_{t_0, V_s}$ . In this case,  $t$  is decomposed into  $4 + 1$  subtrees  $t_v$ . Note that a linear node in  $t_0$  correspond to either a simple or a linear node in  $t$ , depending on whether or not it belongs to  $V_s$ .

By definition, a node  $u$  of  $t$  that is not in  $\text{FCA}(\vec{\ell})$  belongs to exactly one  $t_v$ . On the contrary, if  $u$  is in  $\text{FCA}(\vec{\ell})$ , then  $u$  is the root of  $t_{\varphi^{-1}(u)}$  and is a leaf of another  $t_v$ , where  $v$  is the parent of  $\varphi^{-1}(u)$ . (If  $\varphi^{-1}(u)$  is the root of  $t_0$ , then there is no such  $v$ , and  $u$  is a leaf of  $t_\emptyset$ .)

By construction of  $t_v$  and  $t_\emptyset$ , their leaves are either leaves of  $t$  or belong to  $\text{FCA}(\vec{\ell})$ . We mark the leaves that belong to  $\text{FCA}(\vec{\ell})$  or that are marked leaves of  $t$ . In this way, the trees  $t_v$  and  $t_\emptyset$  that we have constructed are *marked trees*.

The following properties are straightforward to check.

- The tree  $t_v$  is an  $\mathcal{S}$ -canonical tree with  $d_v$  marked leaves.
- The root of  $t_v$  is nonlinear if and only if  $v \in V_s$ .
- The root of  $t_v$  is  $\oplus$  if and only if  $v \notin V_s$  and is labeled  $\oplus$ .
- The root of  $t_v$  is  $\ominus$  if and only if  $v \notin V_s$  and is labeled  $\ominus$ .

- v) The  $d_v$  marked leaves of  $t_v$  belong to  $d_v$  subtrees coming from  $d_v$  *distinct* children of the root of  $t_v$ . The pattern induced by the position of those  $d_v$  children on the permutation labeling the root of  $t_v$  is  $\theta_v$ . (For example, in Fig. 10, four marked leaves are branched on the node labeled 362514 at positions 1, 2, 5, 6. This implies that the corresponding node in  $t_0$  is labeled with  $\theta_v = 2413$ .)
- vi) Let  $w$  be the  $i$ -th child of  $v$  in  $t_0$ . If  $w \in \text{Int}(t_0) \setminus V_s$ , and its label is a  $\oplus$  (resp. a  $\ominus$ ), then the  $i$ -th marked leaf of  $t_v$  must be  $\oplus$ -replaceable (resp.  $\ominus$ -replaceable).

The combinatorial class of trees satisfying properties i) to vi) will be denoted  $\mathcal{A}_v$ .

In addition, we observe that  $t_\emptyset$  is an  $\mathcal{S}$ -canonical tree with one marked leaf; moreover, if the root of  $t_0$  is not in  $V_s$  and labeled by  $\oplus$  (resp.  $\ominus$ ), then the marked leaf of  $t_\emptyset$  must be  $\oplus$ -replaceable (resp.  $\ominus$ -replaceable).

This yields a map

$$\begin{aligned} \mathbf{T} : \mathcal{T}_{t_0, V_s} &\rightarrow \mathcal{T}^{\text{type of root}} \times \prod_{v \in \text{Int}(t_0)} \mathcal{A}_v \\ (t, \vec{\ell}) &\mapsto (t_\emptyset, (t_v)_{v \in \text{Int}(t_0)}). \end{aligned}$$

We claim that this map is one-to-one, and that the inverse map is obtained as follows. Let be given  $t_\emptyset$  and a collection of trees  $t_v$ , one for each internal node of  $t_0$ . We first take  $t_\emptyset$  and glue  $t_{\text{root of } t_0}$  on it, the root of  $t_{\text{root of } t_0}$  replacing the marked leaf of  $t_\emptyset$ . We then proceed inductively: if  $t_v$  has already been glued and  $w$  is the  $i$ -th child of  $v$ , we glue  $t_w$  on  $t_v$ , by replacing the  $i$ -th marked leaf of  $t_v$  with the root of  $t_w$ . This yields a tree  $t$  with  $k$  marked leaves, denoted  $\vec{\ell}$ . This tree is  $\mathcal{S}$ -canonical because of items i), iii), iv) and vi) of the definition of  $\mathcal{A}_v$ : we only glue trees with root  $\oplus$  (resp.  $\ominus$ ) on  $\oplus$ -replaceable leaves (resp.  $\ominus$ -replaceable leaves). By construction, the  $k$  marked leaves  $\vec{\ell}$  induce a tree having the same structure as  $t_0$  and item v) of the definition of  $\mathcal{A}_v$  ensures that the labels in the induced tree and in  $t_0$  do match. Because of item ii), the tree  $(t, \vec{\ell})$  is indeed in  $\mathcal{T}_{t_0, V_s}$ . We have therefore construct a map from  $\mathcal{T}^{\text{type of root}} \times \prod_{v \in \text{Int}(t_0)} \mathcal{A}_v$  to  $\mathcal{T}_{t_0, V_s}$ . By construction, this map indeed inverts  $\mathbf{T}$ , and  $\mathbf{T}$  is a bijection.

Let  $A_v$  be the generating function of the combinatorial class  $\mathcal{A}_v$ , counted by the number of **unmarked leaves**. If  $A_v$  verifies (19), then (18) follows from the fact that  $\mathbf{T}$  is a bijection. Note indeed that the factor  $z^k$  in (18) comes from the fact that we count marked leaves in the series in the left-hand side and but not in the series in the right hand side (the bijection  $\mathbf{T}$  leaves the number of unmarked leaves invariant).

We are left to show that the generating function  $A_v$  verifies (19).

### Second step (i): Computing $A_v$ when $v \in V_s$ .

Recall that the  $d_v$  marked leaves of any tree  $t \in \mathcal{A}_v$  belong to  $d_v$  subtrees coming from  $d_v$  *distinct* children of the root of  $t$ . Since  $v \in V_s$ , the elements  $t$  of  $\mathcal{A}_v$  can be uniquely decomposed as follows (this decomposition is illustrated on Fig. 12).

- i) The root of  $t$  should be labeled by a simple permutation  $\alpha$  in  $\mathcal{S}$ ; among the  $|\alpha|$  children of the root,  $d_v$  are marked (corresponding to the subtrees containing a marked leaf) and the pattern of  $\alpha$  corresponding to the positions of these marked children should be  $\theta_v$ . (In Fig. 12,  $\alpha = 3142$ , the marked leaves are the first, third and fourth subtrees, and the pattern of 3142 corresponding to positions  $\{1, 3, 4\}$  is indeed  $\theta_v = 231$ )



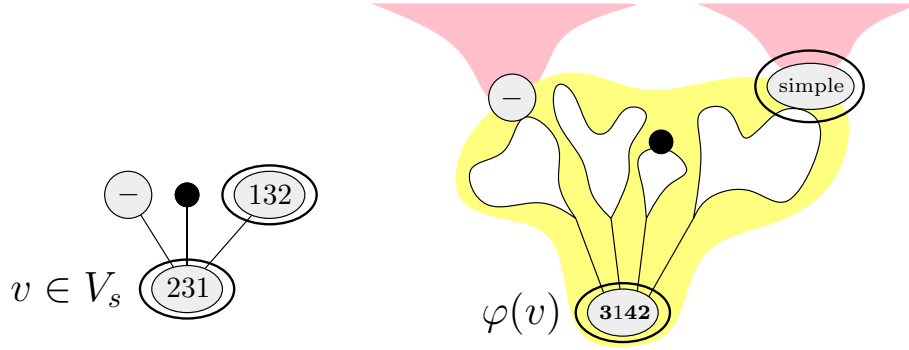


FIGURE 12. Computing  $A_v$  when  $v \in V_s$ . On the left, we only represent the node  $v$  in  $t_0$  and its children; the other parts of  $t_0$  being irrelevant to compute  $A_v$ . On the right we zoom in on the part corresponding to  $t_v$  in the schematic representation of an element in  $\mathcal{T}_{(t_0, V_s)}$ . The series  $A_v$  is here given by  $A_v = \text{Occ}_{231}(T) (T')^2 (T^-)$ .

- ii) We glue  $|\alpha| - d_v$  unmarked  $\mathcal{S}$ -canonical trees with arbitrary roots on the unmarked children of  $\alpha$ . (In Fig. 12, we have only one such tree, which is glued on the second child of the root);
- iii) We glue  $d_v$   $\mathcal{S}$ -canonical trees with one leaf marked and an arbitrary root on the marked children of  $\alpha$ . In addition,
  - for  $d_v'$  of these trees, there is no constraint on the marked leaf. (In Fig. 12, the trees glued on the third and fourth children of the root are unconstrained trees with a marked leaf.)
  - For  $d_v^+$  (resp.  $d_v^-$ ) of these trees, the marked leaf must be  $\oplus$ -replaceable (resp.  $\ominus$ -replaceable). (In Fig. 12, we must glue a tree with a  $\ominus$ -replaceable marked leaf on the first child of the root.)

The generating functions of the first two steps can be computed as

$$\sum_{\alpha \in \mathcal{S}} \text{occ}(\theta_v, \alpha) T(z)^{|\alpha| - d_v} = \text{Occ}_{\theta_v}(T(z)),$$

where  $\text{Occ}_{\theta_v}$  is defined in Eq. (17) p. 28. Indeed, once the label  $\alpha$  of the root is chosen,  $\text{occ}(\theta_v, \alpha)$  counts the number of ways to mark children of the root in step i), and  $T(z)^{|\alpha| - d_v}$  comes from step ii). Step iii) yields an additional factor  $(T')^{d_v'} (T^+)^{d_v^+} (T^-)^{d_v^-}$ . This proves the formula (19) in the case where  $v$  is in  $V_s$ .

### Second step (ii): Computing $A_v$ when $v \notin V_s$ .

When  $v$  is not in  $V_s$  and labeled by  $\oplus$ , the elements of the class  $\mathcal{A}_v$  can be uniquely decomposed as follows (this decomposition is illustrated on Fig. 13).

- i) The root is labeled by  $\oplus$ .
- ii) We attach to the root  $d_v$   $\mathcal{S}$ -canonical trees whose root is not labeled by  $\oplus$ , each with one marked leaf. In addition,

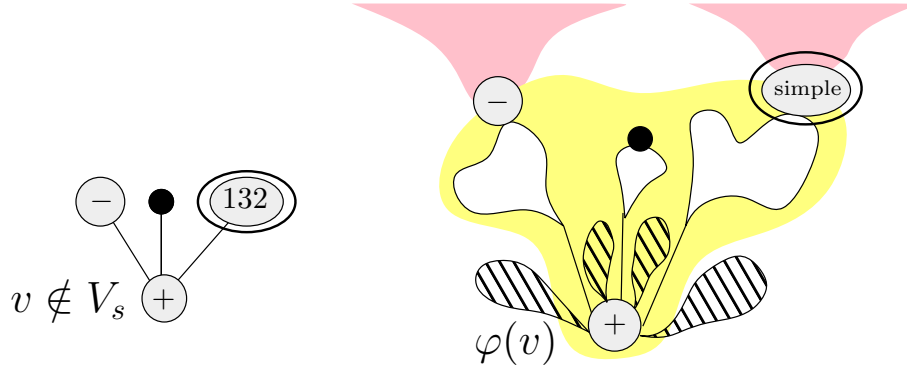


FIGURE 13. Computing  $A_v$  when  $v$  is not in  $V_s$ . On the left, we again represent only the node  $v$  in  $t_0$  and its children; the other parts of  $t_0$  being irrelevant to compute  $A_v$ . On the right we zoom in on the part corresponding to  $t_v$  in the schematic representation of an element in  $\mathcal{T}_{(t_0, V_s)}$ . The series  $A_v$  is here given by

$$A_v = \left( \frac{1}{1-T_{\text{not}\oplus}} \right)^4 (T'_{\text{not}\oplus})^2 (T^-_{\text{not}\oplus}).$$

- for  $d'_v$  of these trees, there is no constraint on the marked leaf. (In Fig. 13, the two right-most nonhatched trees attached to the root are trees with an unconstrained marked leaf.)
  - for  $d_v^+$  (resp.  $d_v^-$ ) of these trees, the marked leaf must be  $\oplus$ -replaceable (resp.  $\ominus$ -replaceable). (In Fig. 13, the left-most nonhatched tree attached to the root should have a  $\ominus$  replaceable marked leaf.)
- iii) Inbetween and around these  $d_v$  trees, we attach  $d_v + 1$  possibly empty sequences of unmarked  $\mathcal{S}$ -canonical trees whose root is not labeled by  $\oplus$ . (In Fig. 13, each of these sequences is represented by a hatched blob.)

Item i) does not involve any choice. Choices in item ii) are counted by  $(T'_{\text{not}\oplus})^{d'_v} (T^+_{\text{not}\oplus})^{d_v^+} (T^-_{\text{not}\oplus})^{d_v^-}$ , while item iii) yields a factor  $\left( \frac{1}{1-T_{\text{not}\oplus}} \right)^{d_v+1}$ . This proves the formula (19) in the case where  $v$  is not in  $V_s$  and labeled by  $\oplus$ .

The case when  $v$  is not in  $V_s$  and is labeled by  $\ominus$  follows by symmetry. This ends the proof of the combinatorial identity (19) and therefore of Proposition 4.5.  $\square$

## 5. ASYMPTOTIC ANALYSIS: THE STANDARD CASE $S'(R_S) > 2/(1 + R_S)^2 - 1$

Let  $\mathcal{S}$  be a set of simple permutations. The goal of this section is to precisely state, and then prove, Theorem 1.10 (p.8): the convergence to the biased Brownian separable permuton of uniform random permutations in  $\langle \mathcal{S} \rangle$  when  $\mathcal{S}$  satisfies Condition (H1).

**5.1. Definition of the biased Brownian separable permuton and statement of the theorem.** The (unbiased) Brownian separable permuton was defined in [11]. Because the biased Brownian separable permuton is a one-parameter deformation of it, it is useful to first recall some

facts about the substitution trees encoding separable permutations and the (unbiased) Brownian separable permuton.

As noted in Section 1.3, the canonical trees (called decomposition trees in [11]) of separable permutations are those whose internal nodes are all labeled  $\oplus$  or  $\ominus$ . If we consider more generally substitution trees, the following implication still holds: if  $\tau$  is a substitution tree whose all nodes are labeled  $\oplus$  or  $\ominus$ , then  $\text{perm}(\tau)$  is a separable permutation.

Recall from Section 3.1 that an *expanded tree* is a substitution tree where nonlinear nodes are labeled by simple permutations, while linear nodes are required to be binary. In the case of separable permutations, we do not have simple nodes, so that *expanded tree* are binary trees labeled with  $\oplus$  and  $\ominus$ . These are also referred to as *separation trees* in the literature. Fig. 14 shows a separable permutations together with two separation trees associated with it.

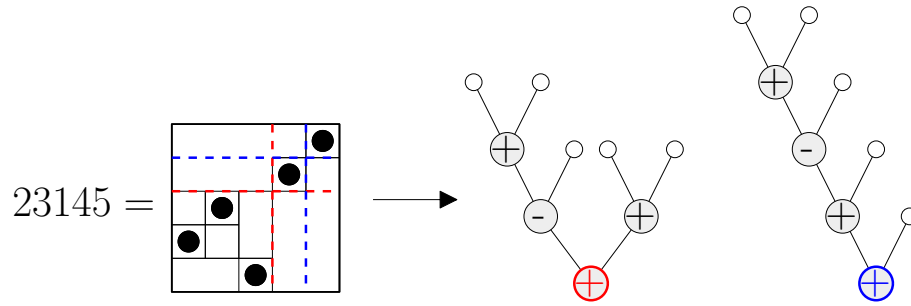


FIGURE 14. Two examples of separation trees associated with the separable permutation  $\pi = 23145$ . In this case  $r_+(\pi) = 3$ ,  $r_-(\pi) = 1$ .

For any separable permutation  $\pi$ , we denote by  $N_\pi$  its number of separation trees. If  $\pi$  is not separable, we set  $N_\pi = 0$ . It is shown in [11, Prop. 9.1] that the Brownian separable permuton  $\mu$  satisfies the following property: for any  $k \geq 2$  and any  $\pi \in \mathfrak{S}_k$ ,

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \mu)] = \frac{N_\pi}{2^{k-1} \text{Cat}_{k-1}},$$

where, as before, we denote by  $\text{Cat}_k := \frac{1}{k+1} \binom{2k}{k}$  the  $k$ -th Catalan number. In other words, the random permutation of size  $k$  extracted from  $\mu$  is distributed like the permutation encoded by a uniform binary tree with  $k$  leaves, and hence  $k - 1$  internal nodes, whose signs are chosen uniformly and independently in  $\{\oplus, \ominus\}$ . In light of Proposition 2.4 and Eq. (8) (p. 17), this characterizes the Brownian separable permuton  $\mu$  among random permutons.

The biased Brownian separable permuton of parameter  $p \in (0, 1)$  has a similar characterization, except that the signs are now chosen with a bias. For a separable  $\pi$ , let  $r_+(\pi)$  (resp.  $r_-(\pi)$ ) be the number of internal nodes labeled  $\oplus$  (resp.  $\ominus$ ) in a separation tree of  $\pi$ . Even if this is not relevant for the present paper, let us observe that  $r_+(\pi)$  (resp.  $r_-(\pi)$ ) is simply the number of ascents (resp. descents) of  $\pi^5$ . In particular,  $r_+(\pi)$  and  $r_-(\pi)$  do not depend on the choice of a separation tree (this is also a particular case of Corollary 3.5).

<sup>5</sup> To see this, observe that each internal node  $v$  of a separation tree is the first common ancestor of exactly one pair of consecutive leaves (the right-most leaf of its left subtree and the left-most leaf of its right subtree). This two

**Definition 5.1.** *The biased Brownian separable permuton of parameter  $p \in (0, 1)$  is the random permuton  $\mu^{(p)}$  characterized by the following relations: for all  $k \geq 2$  and all  $\pi \in \mathfrak{S}_k$ ,*

$$(20) \quad \mathbb{E}[\widetilde{\text{occ}}(\pi, \mu^{(p)})] = \frac{N_\pi}{\text{Cat}_{k-1}} p^{r_+(\pi)} (1-p)^{r_-(\pi)}.$$

(Note that the right-hand side is zero if and only if  $\pi$  is not separable.)

Several remarks are in order.

- For  $p = 1/2$ , we get the unbiased Brownian separable permuton.
- This characterization of  $\mu^{(p)}$  is equivalent to the following: for every  $k \geq 1$ ,

$$\text{Perm}(\vec{\mathbf{m}}_k, \mu^{(p)}) \stackrel{d}{=} \text{perm}(\mathbf{b}_k^{(p)}),$$

where  $\mathbf{b}_k^{(p)}$  is a uniform binary planar tree with  $k$  leaves, where each internal node is labeled  $\oplus$  (resp.  $\ominus$ ) with probability  $p$  (resp.  $1-p$ ), independently from each other.

- The existence of  $\mu^{(p)}$  is not immediate from this definition, but according to Proposition 2.9, it suffices to show that  $\text{perm}(\mathbf{b}_k^{(p)})$  forms a consistent family of random permutations. This is indeed the case, and follows from the fact that a uniform induced subtree of  $\mathbf{b}_n^{(p)}$  of size  $k$  is distributed like  $\mathbf{b}_k^{(p)}$  (this is, *e.g.*, a consequence of Rémy's algorithm to generate uniform random binary trees [43]). An alternative proof of the existence of  $\mu^{(p)}$  is an explicit construction starting from a Brownian excursion; see [35].
- Knowing a priori that such a permuton exists is not necessary for the proof of our main theorem. Indeed, we will prove that the quantity  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]$  converges to the right-hand side of Eq. (20) (for all patterns  $\pi$ ), where  $\sigma_n$  is a uniform permutation in  $\langle \mathcal{S} \rangle_n$ , and the parameter  $p$  depends on  $S$ . From Theorem 2.5, this implies the existence of a random permuton  $\mu^{(p)}$  satisfying (20) (only for the relevant value of  $p$ ; not for all  $p$ ) and the convergence in distribution of  $(\mu_{\sigma_n})_n$  to  $\mu^{(p)}$ .

Now, we have all the necessary definitions to make explicit the parameter  $p$  of the statement of Theorem 1.10, that we restate in a full version. (Recall also the definition of  $\text{Occ}_\theta(z)$  from Eq. (17), p. 28.)

**Theorem 5.2.** *Let  $\mathcal{S}$  be a set of simple permutations such that*

$$(H1) \quad R_S > 0 \quad \text{and} \quad \lim_{\substack{r \rightarrow R_S \\ r < R_S}} S'(r) > \frac{2}{(1+R_S)^2} - 1.$$

*For every  $n \geq 1$ , let  $\sigma_n$  be a uniform permutation in  $\langle \mathcal{S} \rangle_n$ , and let  $\mu_{\sigma_n}$  be the random permuton associated with  $\sigma_n$ . The sequence  $(\mu_{\sigma_n})_n$  tends in distribution in the weak convergence topology to the biased Brownian separable permuton  $\mu^{(p)}$  of parameter  $p$ , where*

$$(21) \quad p = \frac{(1+\kappa)^3 \text{Occ}_{12}(\kappa) + 1}{(1+\kappa)^3 (\text{Occ}_{12}(\kappa) + \text{Occ}_{21}(\kappa)) + 2}$$

*and  $\kappa$  is the unique point in  $(0, R_S)$  such that  $S'(\kappa) = \frac{2}{(1+\kappa)^2} - 1$ .*

---

consecutive leaves, corresponding to consecutive elements of the permutation, form an ascent (resp. a descent) if and only if  $v$  is labeled by  $\oplus$  (resp.  $\ominus$ ).

The existence and uniqueness of  $\kappa$  is immediate from the fact that  $t \mapsto S'(t) - \frac{2}{(1+t)^2} + 1$  is increasing on  $[0, R_S)$ , going from  $-1$  to a positive number (possibly  $+\infty$ ).

*Example 5.3.* In many cases  $\text{Occ}_{12} = \text{Occ}_{21}$ , and then  $p = 1/2$  and  $\mu^{(p)}$  is the unbiased Brownian separable permuton. This is the case with separable permutations ( $\mathcal{S} = \emptyset$ ), with  $\mathcal{S} = \{2413\}$  or  $\mathcal{S} = \{3142\}$ , and with any set of simple permutations stable by taking reverse or complement.

*Example 5.4.* When  $\mathcal{S}$  is the family of increasing oscillations (see for instance [13]), we can compute

$$S(z) = \frac{2z^4}{1-z}; \quad \text{Occ}_{12}(z) = \frac{2z^2(3-3z+z^2)}{(1-z)^3}; \quad \text{Occ}_{21}(z) = \frac{2z^2(3-2z)}{(1-z)^2}.$$

We get through numerical approximation  $\kappa \approx 0.2708$  and deduce  $p \approx 0.5353$ .

*Example 5.5.* Taking  $\mathcal{S}$  to be the family of simple permutations in  $\text{Av}(321)$ , we are interested in the class  $\mathcal{C} = \langle \mathcal{S} \rangle$  which is the substitution-closure of  $\text{Av}(321)$ . In this case, [8] gives

$$S(z) = \frac{1-z-2z^2-2z^3-\sqrt{1-2z-3z^2}}{2+2z}.$$

We get through numerical approximation  $\kappa \approx 0.2486$ . It seems hard to compute the generating series  $\text{Occ}_{12}$  and  $\text{Occ}_{21}$ , but they can be estimated at  $\kappa$  using a Boltzmann sampler for the class  $\mathcal{S}$ . We get  $\text{Occ}_{12}(\kappa) \approx 1.440$  and  $\text{Occ}_{21}(\kappa) \approx 1.414$  and deduce  $p \approx 0.5034$ .

The remainder of Section 5 is devoted to the proof of Theorem 5.2, using generating functions from Section 4 and methods of analytic combinatorics. More precisely, using Theorem 2.5 we are interested in the limit of  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]$ , that we will express in terms of probability that a tree with marked leaves induces a given subtree. This probability itself will be expressed as the ratio of the coefficients of the generating function of trees with marked leaves inducing a given subtree and of the generating function of trees without marked leaves. We begin with the study of the asymptotics of these generating functions.

**5.2. Asymptotics of the generating function of trees with no or one marked leaf.** From Eq. (11) p.26, we have

$$(22) \quad T_{\text{not}\oplus} = z + \Lambda(T_{\text{not}\oplus})$$

where

$$(23) \quad \Lambda(u) = \frac{u^2}{1-u} + S\left(\frac{u}{1-u}\right).$$

We denote by  $R_\Lambda$  the radius of convergence of  $\Lambda$ . Note that  $R_\Lambda = \frac{R_S}{1+R_S} \leq 1$ . In the following, to lighten the notation, we write  $\Lambda'(R_\Lambda) := \lim_{r \rightarrow R_\Lambda} \Lambda'(r)$ . Note that  $\Lambda'(R_\Lambda)$  may be  $\infty$ .

*Observation 5.6.* Differentiating Eq. (23), it follows from elementary computations that the condition (H1) is equivalent to  $R_\Lambda > 0$  and  $\Lambda'(R_\Lambda) > 1$ .

*Observation 5.7.* Since  $S$  is analytic at 0 with nonnegative coefficients, the same holds for  $\Lambda$ . Moreover, the series expansion of  $\Lambda$  is

$$\Lambda(u) = u^2 + \sum_{i \geq 3} \lambda_i u^i, \text{ with } \lambda_i \geq 1 \text{ for all } i \geq 3.$$

In particular it is aperiodic, in the sense given in Appendix A.1.

**Proposition 5.8** (Asymptotics of the generating function of  $\mathcal{S}$ -canonical trees with no marked leaf). *Assume that (H1) holds, and recall that  $\kappa$  is defined by  $S'(\kappa) = \frac{2}{(1+\kappa)^2} - 1$ . There is a unique  $\tau \in (0, R_\Lambda)$  such that  $\Lambda'(\tau) = 1$ , and we have  $\tau = \frac{\kappa}{1+\kappa}$ . The generating functions  $T$  and  $T_{\text{not}\oplus}$  have the same radius of convergence  $\rho = \tau - \Lambda(\tau) \in (0, \tau)$  and have a unique dominant singularity<sup>6</sup> in  $\rho$ . Their asymptotic expansions near  $\rho$  are:*

$$(24) \quad T(z) = \frac{\tau}{1-\tau} - \beta \lambda \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho}),$$

$$(25) \quad T_{\text{not}\oplus}(z) = \tau - \beta \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho}).$$

where  $\beta = \sqrt{\frac{2\rho}{\Lambda''(\tau)}}$  and  $\lambda = \frac{1}{(1-\tau)^2}$ . In particular,  $T$  and  $T_{\text{not}\oplus}$  are convergent at  $z = \rho$  and

$$T(\rho) = \frac{\tau}{1-\tau}, \quad T_{\text{not}\oplus}(\rho) = \tau.$$

This type of behavior with a square-root dominant singularity is classical for series defined by an implicit equation (such as  $T_{\text{not}\oplus}$ , which is characterized by  $T_{\text{not}\oplus}(z) = z + \Lambda(T_{\text{not}\oplus}(z))$ ), that belong to the *smooth implicit function schema* [26, Def. VII.4]. This schema is defined by the existence of a solution to some *characteristic equation*, which in our case reduces to hypothesis (H1), as explained in Section 1.7. Our result is a special case of [19, Th. 1], a general result for equations of the form  $U(z) = z + \Lambda(U(z))$ . Implicit equations of this form characterize generating functions of weighted trees counted by their number of leaves, and are also considered in [42, Prop. 8].

*Proof.* From Observation 5.7,  $\Lambda'$  is strictly increasing in the real interval  $(0, R_\Lambda)$ . Together with the fact that  $\Lambda'(0) = 0$  and the assumption  $\Lambda'(R_\Lambda) > 1$  (see Observation 5.6), this proves the existence and uniqueness of  $\tau > 0$  such that  $\Lambda'(\tau) = 1$ .

Note that differentiating Eq. (23) gives  $\Lambda'(\frac{u}{1+u}) = (1+u)^2(1+S'(u)) - 1$ , from which it follows that  $\Lambda'(\frac{\kappa}{1+\kappa}) = 1$ . Therefore  $\tau = \frac{\kappa}{1+\kappa}$ .

We now consider the expansion of  $T_{\text{not}\oplus}$  and deduce afterwards the one of  $T$ . From Eq. (22), we have  $T_{\text{not}\oplus}(z) = z + \Lambda(T_{\text{not}\oplus}(z))$ . Then Theorem 1 in [19] gives that  $T_{\text{not}\oplus}$  is analytic at 0 and has a unique dominant singularity of exponent  $\frac{1}{2}$  in  $\rho = \tau - \Lambda(\tau)$ , with the expansion given in Eq. (25):  $T_{\text{not}\oplus}(z) = \tau - \beta \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho})$ .

<sup>6</sup>For the reader who is not familiar with complex analysis, recall that all useful definitions and results are given in Appendix A. In particular, "near  $\rho$ " means "in a  $\Delta$ -neighborhood of  $\rho$ ", where " $\Delta$ -neighborhood" is defined in Definition A.2. The formal definition of (*unique*) *dominant singularity* is given in Eq. (54) p.58.

The next step is to justify that  $\rho \in (0, \tau)$ . That  $\rho < \tau$  follows from  $\Lambda(\tau) > 0$  (since  $\tau > 0$ ). Moreover, since  $\Lambda$  has nonnegative coefficients and no constant term, we have  $\Lambda(\tau) < \tau\Lambda'(\tau) = \tau$ , so that  $\rho > 0$ .

Finally, we look at the series  $T = \frac{T_{\text{not}\oplus}}{1 - T_{\text{not}\oplus}}$  (see Eq. (12)). Observe that  $T_{\text{not}\oplus}(\rho) = \tau < R_\Lambda \leq 1$ . Consequently, the dominant singularity of  $T$  is the same as  $T_{\text{not}\oplus}$ , i.e.  $\rho$ , and is still unique – indeed this singularity is reached before that the denominator vanishes; more formally, this is a particular case of subcritical composition (Lemma A.6). The asymptotic expansion of  $T$  near  $\rho$  is obtained through the following computation:

$$T(z) = \frac{\tau - \beta\sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho})}{1 - \tau + \beta\sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho})} = \frac{\tau}{1 - \tau} - \frac{\beta}{(1 - \tau)^2} \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho}). \quad \square$$

**Proposition 5.9** (Asymptotics of the generating function of  $\mathcal{S}$ -canonical trees with marked leaves). *All generating functions  $T'$ ,  $T'_{\text{not}\oplus}$ ,  $T'_{\text{not}\ominus}$ ,  $T^+$ ,  $T^+_{\text{not}\oplus}$ ,  $T^+_{\text{not}\ominus}$ ,  $T^-$ ,  $T^-_{\text{not}\oplus}$  and  $T^-_{\text{not}\ominus}$  have a unique dominant singularity in  $\rho$ . They diverge at the singularity  $z = \rho$  and behave as  $\text{cst}(1 - \frac{z}{\rho})^{-1/2}$ , where the constants are given in the table below:*

<i>superscript</i> \ <i>subscript</i>	$\emptyset$	$\text{not}\oplus$	$\text{not}\ominus$
'	$\gamma\lambda^2$	$\gamma\lambda$	$\gamma\lambda$
+	$\gamma\lambda$	$\gamma$	$\gamma$
−	$\gamma\lambda$	$\gamma$	$\gamma$

with  $\gamma = \frac{\beta(1-\tau)^2}{2\rho}$  (recall that  $\lambda = \frac{1}{(1-\tau)^2}$  was defined in Proposition 5.8).

Note that the table is in fact a rank 1 matrix. Namely, passing from a root different from  $\oplus$  (resp.  $\ominus$ ) to a nonconditioned root always adds a factor  $\lambda$ , independently of the condition on the marked leaf. Similarly, removing the leaf condition always yields the same factor  $\lambda$  independently of the conditions on roots.

*Proof.* By singular differentiation (see Theorem A.4) of Eqs. (24) and (25), we have (near  $\rho$ )

$$T'(z) = \frac{\beta}{2\rho(1-\tau)^2} (1 - \frac{z}{\rho})^{-1/2} + \mathcal{O}(1) = \gamma\lambda^2 (1 - \frac{z}{\rho})^{-1/2} + \mathcal{O}(1),$$

$$T'_{\text{not}\oplus}(z) = \frac{\beta}{2\rho} (1 - \frac{z}{\rho})^{-1/2} + \mathcal{O}(1) = \gamma\lambda (1 - \frac{z}{\rho})^{-1/2} + \mathcal{O}(1).$$

Since  $T_{\text{not}\ominus} = T_{\text{not}\oplus}$ , we have obtained the constants in the first line of the table. We now turn to the two last lines. In order to use Eqs. (14) to (16) p.27, we first compute the expansions of all intermediate quantities appearing in these formulas. From Eq. (25), we obtain the following expansion near  $\rho$ :

$$(26) \quad 1 + W(z) = \left( \frac{1}{1 - T_{\text{not}\oplus}(z)} \right)^2 = \frac{1}{(1 - \tau)^2} - \frac{2\beta}{(1 - \tau)^3} \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho}).$$

We turn to the expansion of  $S'(T)$ . Differentiating Eq. (23) yields

$$\Lambda'(u) = \frac{1}{(1-u)^2} \left( 1 + S' \left( \frac{u}{1-u} \right) \right) - 1.$$

Putting  $u = T_{\text{not}\oplus}$  and using Eq. (12) gives

$$(27) \quad 1 + S'(T) = 1 + S'\left(\frac{T_{\text{not}\oplus}}{1 - T_{\text{not}\oplus}}\right) = (1 - T_{\text{not}\oplus})^2 (\Lambda'(T_{\text{not}\oplus}) + 1).$$

Recall that  $T_{\text{not}\oplus}(\rho) = \tau < R_\Lambda$ . Therefore, the composition  $\Lambda' \circ T_{\text{not}\oplus}$  is subcritical (see Lemma A.6). This implies that  $\Lambda' \circ T_{\text{not}\oplus}$  has a unique dominant singularity at  $\rho$ , and plugging the asymptotic expansion (25) of  $T_{\text{not}\oplus}$  at  $\rho$ , we obtain

$$(28) \quad \begin{aligned} \Lambda'(T_{\text{not}\oplus}(z)) &= \Lambda'(\tau) + \Lambda''(\tau)(T_{\text{not}\oplus}(z) - \tau) + O((T_{\text{not}\oplus}(z) - \tau)^2) \\ &= 1 - \frac{2\rho}{\beta} \sqrt{1 - \frac{z}{\rho}} + O(1 - \frac{z}{\rho}), \end{aligned}$$

where we used the equalities  $\Lambda'(\tau) = 1$  (by definition of  $\tau$ ) and  $\Lambda''(\tau) = \frac{2\rho}{\beta^2}$  (by definition of  $\beta$ ). Combining Eq. (25) and Eq. (28) into Eq. (27), we obtain, after simplification:

$$(29) \quad 1 + S'(T) = 2(1 - \tau)^2 - \left[ \frac{2\rho}{\beta}(1 - \tau)^2 - 4\beta(1 - \tau) \right] \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho})$$

The expansion of  $WS'(T) + W + S'(T)$  then follows from Eq. (26) and Eq. (29):

$$(30) \quad \begin{aligned} WS'(T) + W + S'(T) &= (1 + W)(1 + S'(T)) - 1 \\ &= 2\frac{(1-\tau)^2}{(1-\tau)^2} - 1 - \left[ \frac{4\beta(1-\tau)^2}{(1-\tau)^3} + \frac{2\rho(1-\tau)^2}{\beta(1-\tau)^2} - \frac{4\beta(1-\tau)}{(1-\tau)^2} \right] \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho}) \\ &= 1 - \frac{1}{\gamma\lambda} \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}(1 - \frac{z}{\rho}) \text{ by definition of } \gamma \text{ and } \lambda. \end{aligned}$$

We can now derive the expansions of our generating functions, using Eqs. (14) to (16). First,

$$T^+ = \frac{1}{1 - WS'(T) - W - S'(T)} = \gamma\lambda \left(1 - \frac{z}{\rho}\right)^{-1/2} + \mathcal{O}(1).$$

Then,

$$T_{\text{not}\ominus}^+ = \frac{1}{1 + W} T^+ = \frac{\gamma\lambda}{\lambda} \left(1 - \frac{z}{\rho}\right)^{-1/2} + \mathcal{O}(1) = \gamma \left(1 - \frac{z}{\rho}\right)^{-1/2} + \mathcal{O}(1).$$

Since  $WS'(T) + W + S'(T)$  takes value 1 in  $\rho$ , the series  $T_{\text{not}\oplus}^+$  has the same first-order expansion:

$$T_{\text{not}\oplus}^+ = (WS'(T) + W + S'(T))T_{\text{not}\oplus}^+ = \gamma \left(1 - \frac{z}{\rho}\right)^{-1/2} + \mathcal{O}(1).$$

By symmetry, we have  $T^- = T^+$ ,  $T_{\text{not}\ominus}^- = T_{\text{not}\oplus}^+$  and  $T_{\text{not}\oplus}^- = T_{\text{not}\ominus}^+$ , and this completes the proof of the proposition.  $\square$

**5.3. Asymptotics of the generating function of marked trees with a given induced tree.** We recall some notation introduced in Section 4.2. Let  $t_0$  be a substitution tree with  $k \geq 2$  leaves and  $e$  edges.

Let  $V_*$  (resp.  $V_+$ ,  $V_-$ ) be the set of nonlinear nodes (resp. nodes labeled  $\oplus$ ,  $\ominus$ ) in  $\text{Int}(t_0)$ . Recall that, for  $v \in \text{Int}(t_0)$ ,  $d_v$  is the degree of  $v$  and  $\theta_v$  the permutation labeling  $v$ , and that  $\mathcal{T}_{t_0}$  is the set of  $\mathcal{S}$ -canonical trees  $t$  with  $k$  marked leaves such that these leaves induce  $t_0$ . Denote by  $T_{t_0}$  the generating function of  $\mathcal{T}_{t_0}$  (where the size is the number of leaves, both marked and unmarked).



**Proposition 5.10.** *The series  $T_{t_0}$  has a unique dominant singularity in  $\rho$ , with the asymptotic expansion  $T_{t_0} = B_{t_0}(1 - \frac{z}{\rho})^{-(e+1)/2}(1 + o(1))$ , where the constant  $B_{t_0}$  is*

$$(31) \quad B_{t_0} = \rho^k (\gamma \lambda^2)^{e+1} \prod_{v \in V_*} \text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) \times \prod_{v \in V_+ \cup V_-} (\text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) + (1-\tau)^{d_v+1}).$$

*Proof.* By definition,  $T_{t_0} = \sum_{V_s} T_{t_0, V_s}$ , where the sum runs over sets  $V_s$  such that  $(t_0, V_s)$  is a decorated tree. We start from the formula for  $T_{t_0, V_s}$ , which is given by Proposition 4.5. From Proposition 5.9, the nine series  $T', \dots, T_{\text{not}\ominus}^-$  of trees with one marked leaf all have unique dominant singularities in  $\rho$ . This is also the case for the functions  $\text{Occ}_{\theta_v}(T)$  and  $\left(\frac{1}{1-T_{\text{not}\oplus}}\right)^{d_v}$  by subcritical composition (see Lemma A.6). Indeed, from Observation 4.3 the radius of convergence of  $\text{Occ}_{\theta}$  is at least  $R_S$  and from Proposition 5.8,  $T$  and  $T_{\text{not}\oplus}$  are convergent at  $\rho$  with  $T(\rho) = \frac{\tau}{1-\tau} < \frac{R_\Lambda}{1-R_\Lambda} = R_S$  and  $T_{\text{not}\oplus}(\rho) = \tau < 1$ . As a consequence,  $T_{t_0}$  has a unique dominant singularity in  $\rho$  (see Lemma A.5).

For exact asymptotics near  $\rho$ , note that  $\text{Occ}_{\theta_v}(T)$  and  $\frac{1}{1-T_{\text{not}\oplus}}$  converge respectively to  $\text{Occ}_{\theta_v}(\frac{\tau}{1-\tau})$  and  $\frac{1}{1-\tau}$  in  $\rho$  (see Proposition 5.8), while the nine series  $T', \dots, T_{\text{not}\ominus}^-$  behave as  $\text{cst}(1 - \frac{z}{\rho})^{-1/2}$ , where the constant are given in Proposition 5.9. We thus get using the notation of Proposition 4.5:

$$\begin{aligned} z^k T^{\text{type of root}} &= \rho^k \gamma \lambda^{1+\mathbf{1}_{\text{root} \in V_s}} (1 - \frac{z}{\rho})^{-1/2} (1 + o(1)); \\ A_v &= (1 - \frac{z}{\rho})^{-d_v/2} (1 + o(1)) \cdot \begin{cases} \text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) (\gamma \lambda^2)^{d'_v} (\gamma \lambda)^{d_v^+ + d_v^-} & \text{if } v \in V_s; \\ (\frac{1}{1-\tau})^{d_v+1} (\gamma \lambda)^{d'_v} \gamma^{d_v^+ + d_v^-} & \text{if } v \notin V_s. \end{cases} \end{aligned}$$

The asymptotic of  $T_{t_0, V_s}$  near  $\rho$  is then obtained by multiplying the above expressions. The formula can be simplified by observing that  $\sum_{v \in \text{Int}(t_0)} (d'_v + d_v^+ + d_v^-) = \sum_{v \in \text{Int}(t_0)} d_v = e$  and  $\sum_{v \in \text{Int}(t_0)} d'_v + \mathbf{1}_{\text{root} \in V_s} = |V_s| + k$ , and we obtain:

$$\begin{aligned} T_{t_0, V_s} &= (1 + o(1)) (1 - \frac{z}{\rho})^{-(e+1)/2} \rho^k \gamma^{e+1} \lambda^{1+k+|V_s|} \prod_{v \in V_s} \lambda^{d_v} \text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) \prod_{v \notin V_s} \left(\frac{1}{1-\tau}\right)^{d_v+1} \\ &= (1 + o(1)) (1 - \frac{z}{\rho})^{-(e+1)/2} \rho^k (\gamma \lambda^2)^{e+1} \prod_{v \in V_s} \text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) \prod_{v \notin V_s} \left[\left(\frac{1}{1-\tau}\right)^{d_v+1} \lambda^{-d_v-1}\right]. \end{aligned}$$

To write the second line, we have used that

$$\sum_{v \in V_s} d_v + \sum_{v \notin V_s} (d_v + 1) + 1 + k + |V_s| = \sum_{v \in \text{Int}(t_0)} d_v + |\text{Int}(t_0)| + k + 1 = 2e + 2.$$

Now we have that  $T_{t_0}$  is the sum of  $T_{t_0, V_s}$  over sets  $V_s$  such that  $(t_0, V_s)$  is a decorated tree. By definition, such  $V_s$  can be written as  $V_* \cup \tilde{V}_s$  for some  $\tilde{V}_s \subset V_+ \cup V_-$  (recall that  $V_*$  is the set of

nonlinear nodes of  $t_0$ ). This change of variables leads to

$$T_{t_0} = (1 + o(1))(1 - \frac{z}{\rho})^{-(e+1)/2} \rho^k (\gamma \lambda^2)^{e+1} \prod_{v \in V_*} \text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) \\ \times \sum_{\widetilde{V}_s \subset V_+ \cup V_-} \left( \prod_{v \in \widetilde{V}_s} \text{Occ}_{\theta_v}(\frac{\tau}{1-\tau}) \right) \left( \prod_{\substack{v \in V_+ \cup V_- \\ v \notin \widetilde{V}_s}} \left[ \left( \frac{1}{1-\tau} \right)^{d_v+1} \lambda^{-d_v-1} \right] \right).$$

We first observe that since  $\lambda = (1-\tau)^{-2}$ , the last factor simplifies as  $(1-\tau)^{d_v+1}$ . The proposition then follows by writing the sum of products on the second line as a product of sums.  $\square$

**5.4. Probability of tree patterns.** Recall that  $\mathcal{T}$  is the set of  $\mathcal{S}$ -canonical trees (*i.e.* canonical trees of permutations in  $\langle \mathcal{S} \rangle$ ). We take a uniform random tree with  $n$  leaves in  $\mathcal{T}$  and mark  $k$  of its leaves, also chosen uniformly at random. We denote by  $\mathbf{t}_k^{(n)}$  the tree induced by the  $k$  marked leaves.

**Proposition 5.11.** *Let  $k \geq 2$ , and let  $t_0$  be any substitution tree with  $k$  leaves. Then*

$$\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) = k! \frac{2\sqrt{\pi}}{\Gamma(\frac{e(t_0)+1}{2})} \frac{(1-\tau)^2}{\beta} B_{t_0} n^{e(t_0)/2+1-k} (1 + o(1)),$$

where  $B_{t_0}$  is given by Eq. (31), and  $e(t_0)$  is the number of edges of  $t_0$ .

*Proof.* Directly from the definition, we have:

$$\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) = \frac{[z^n]T_{t_0}(z)}{\binom{n}{k}[z^n]T(z)}.$$

The Transfer Theorem (Theorem A.3) gives us the asymptotic behavior of  $[z^n]T(z)$  and  $[z^n]T_{t_0}(z)$  from the asymptotic expansions in Proposition 5.8 (Eq. (24)) and Proposition 5.10. Deriving the result from there is a routine exercise.  $\square$

### 5.5. Back to permutations.

Let  $\pi$  be a permutation of size  $k$ . Recall from Section 3.1 that an *expanded tree* is a substitution tree where nonlinear nodes are labeled by simple permutations, while linear nodes are required to be binary. As in Corollary 3.5, we denote  $\widehat{N}_\pi$  the number of expanded trees of  $\pi$ . We know (see Corollary 3.5) that they have all the same number of linear nodes labeled  $\oplus$  (resp.  $\ominus$ ), this number being denoted by  $r_+$  (resp.  $r_-$ ) and they all contain the same  $r_*$  simple nodes, whose labels will be denoted as  $\theta_1, \dots, \theta_{r_*}$ .

We introduce the *default of binarity* of the permutation  $\pi$ :

$$(32) \quad \text{db}(\pi) = \sum_{i=1}^{r_*} (|\theta_i| - 2).$$

Observe that  $\text{db}(\pi) = 0$  if and only if  $\pi$  is separable.

Finally, to state the next proposition, we also need to introduce the quantities

$$\nu_+ = \text{Occ}_{12} \left( \frac{\tau}{1-\tau} \right), \quad \nu_- = \text{Occ}_{21} \left( \frac{\tau}{1-\tau} \right) \quad \text{and} \quad p = \frac{\nu_+ + (1-\tau)^3}{\nu_+ + \nu_- + 2(1-\tau)^3},$$

Note that this is the same  $p$  as in Theorem 5.2.

**Proposition 5.12.** *Let  $\pi \in \mathfrak{S}_k$  with  $k \geq 2$  and let  $\sigma_n$  be a uniform random permutation in  $\langle \mathcal{S} \rangle_n$ . With notation as above, we have*

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = B_\pi n^{-\text{db}(\pi)/2} (1 + o(1)),$$

$$\text{where } B_\pi = \widetilde{N}_\pi \frac{k! \sqrt{\pi}}{2^{2k-2-\text{db}(\pi)} \Gamma(k - \frac{\text{db}(\pi)+1}{2})} p^{r_+} (1-p)^{r_-} \prod_{i=1}^{r_*} \left[ \rho^{-1} \left( \frac{\beta}{(1-\tau)^2} \right)^{|\theta_i|} \text{Occ}_{\theta_i} \left( \frac{\tau}{1-\tau} \right) \right].$$

*Proof.* We denote by  $\mathbf{I}$  a uniform random  $k$ -element subset of  $[n]$  and by  $\mathbf{t}^{(n)}$  a uniform random  $\mathcal{S}$ -canonical tree with  $n$  leaves. It holds that  $\sigma_n \stackrel{d}{=} \text{perm}(\mathbf{t}^{(n)})$ . As a consequence of Eq. (7) and Lemma 3.11, we have

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = \mathbb{P}(\text{pat}_{\mathbf{I}}(\sigma_n) = \pi) = \mathbb{P}(\text{perm}(\mathbf{t}^{(n)}) = \pi) = \sum_{t_0: \text{perm}(t_0) = \pi} \mathbb{P}(\mathbf{t}^{(n)} = t_0).$$

After plugging the estimate of Proposition 5.11, we get

$$(33) \quad \mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = (1 + o(1)) \sum_{t_0: \text{perm}(t_0) = \pi} k! \frac{2\sqrt{\pi}}{\Gamma(\frac{e(t_0)+1}{2})} \frac{(1-\tau)^2}{\beta} B_{t_0} n^{-\text{db}(t_0)/2}$$

where  $\text{db}(t_0) = 2k - 2 - e(t_0)$  is the default of binarity of the tree  $t_0$ .

We claim that if  $t_0$  is a substitution tree of  $\pi$ , then  $\text{db}(t_0) \geq \text{db}(\pi)$  with equality if and only if  $t_0$  is an expanded tree. Indeed,

$$\text{db}(t_0) = e(t_0) + 2(k - 1 - e(t_0)) = \sum_{v \in \text{Int}(t_0)} (d_v - 2).$$

Moreover from Lemma 3.6 any substitution tree can be obtained from an expanded tree of  $\pi$  by merging some internal nodes along edges connecting them and such merges always increase (strictly) the considered sum, which proves the claim.

It follows that, in the sum of Eq. (33), only expanded trees appear asymptotically. Moreover,  $e(t_0)$  and the constant  $B_{t_0}$  does not depend on the choice of an expanded tree  $t_0$  of  $\pi$ . As a result, we get  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = (1 + o(1)) B_\pi n^{-\text{db}(\pi)/2}$  where

(34)

$$\begin{aligned} B_\pi &= \widetilde{N}_\pi \frac{k! 2\sqrt{\pi}}{\Gamma(\frac{e+1}{2})} \frac{(1-\tau)^2}{\beta} \rho^k (\gamma \lambda^2)^{e+1} \left( \prod_{i=1}^{r_*} \text{Occ}_{\theta_i} \left( \frac{\tau}{1-\tau} \right) \right) \left( \nu_+ + (1-\tau)^3 \right)^{r_+} \left( \nu_- + (1-\tau)^3 \right)^{r_-} \\ &= \widetilde{N}_\pi \frac{k! 2\sqrt{\pi}}{\Gamma(\frac{e+1}{2})} p^{r_+} (1-p)^{r_-} \left( \prod_{i=1}^{r_*} \text{Occ}_{\theta_i} \left( \frac{\tau}{1-\tau} \right) \right) \times \rho^k (\gamma \lambda^2)^{e+1} \frac{(1-\tau)^2}{\beta} \left( \nu_+ + \nu_- + 2(1-\tau)^3 \right)^{r_+ + r_-} \end{aligned}$$

Differentiating Eq. (23) p.37 twice yields the identity  $\Lambda''(\tau) = \frac{4}{1-\tau} + \frac{1}{(1-\tau)^4} S'''(\frac{\tau}{1-\tau})$  since  $\Lambda'(\tau) = 1$ . Moreover, since  $\text{Occ}_{12} + \text{Occ}_{21} = \frac{S''}{2}$  (see Observation 4.3), this gives us

$$\nu_+ + \nu_- + 2(1-\tau)^3 = \frac{(1-\tau)^4}{2} \Lambda''(\tau) = \frac{(1-\tau)^4 \rho}{\beta^2}$$

Finally, after collecting everything together, we get

$$\begin{aligned} & \rho^k (\gamma \lambda^2)^{e+1} \frac{(1-\tau)^2}{\beta} \left( \nu_+ + \nu_- + 2(1-\tau)^3 \right)^{r_+ + r_-} \\ &= \rho^k \left( \frac{\beta}{2\rho(1-\tau)^2} \right)^{e+1} \frac{(1-\tau)^2}{\beta} \left( \frac{\rho(1-\tau)^4}{\beta^2} \right)^{r_+ + r_-} = \frac{1}{\rho^{r_*} 2^{e+1}} \left( \frac{\beta}{(1-\tau)^2} \right)^{2r_* + \text{db}(\pi)}, \end{aligned}$$

where the last equality above has been obtained using that, for any expanded tree  $t_0$  of  $\pi$ , we have  $r_+ + r_- + r_* = |\text{Int}(t_0)| = e - k + 1$  and  $\text{db}(\pi) = 2k - 2 - e$ . This allows us to simplify Eq. (34) and yields the desired value of  $B_\pi$ .  $\square$

We can now conclude the proof of Theorem 5.2. Let  $\sigma_n$  be a uniform random permutation in  $\langle S \rangle_n$ . Our goal is to show that  $\mu_{\sigma_n}$  converges to the biased Brownian separable permutation of parameter  $p$ . Let  $\pi$  be any permutation of size  $k \geq 2$ . As a consequence of Theorem 2.5 (with Observation 2.6) and Definition 5.1, we just have to show that

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] \xrightarrow{n \rightarrow \infty} \frac{N_\pi}{\text{Cat}_{k-1}} p^{r_+(\pi)} (1-p)^{r_-(\pi)}.$$

Assume first that  $\pi$  is not separable. In this case, we have  $N_\pi = 0$ . It also holds that  $\text{db}(\pi) > 0$ , and Proposition 5.12 implies that  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] \rightarrow 0$ .

Assume on the contrary that  $\pi$  is separable. In this case,  $\text{db}(\pi) = 0$ ,  $\widetilde{N}_\pi = N_\pi$  and  $r_* = 0$ . Therefore, from Proposition 5.12 we get that

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] \xrightarrow{n \rightarrow \infty} N_\pi p^{r_+(\pi)} (1-p)^{r_-(\pi)} \frac{k! \sqrt{\pi}}{2^{2k-2} \Gamma(k - \frac{1}{2})} = \frac{N_\pi}{\text{Cat}_{k-1}} p^{r_+(\pi)} (1-p)^{r_-(\pi)},$$

where we have used the identity  $\Gamma(k - \frac{1}{2}) = \frac{2^{3-2k} \sqrt{\pi} \Gamma(2k-2)}{\Gamma(k-1)} = \frac{2^{3-2k} \sqrt{\pi} (2k-3)!}{(k-2)!}$  coming from the duplication formula of the Gamma function. This concludes the proof.  $\square$

**5.6. Occurrences of nonseparable patterns.** Since  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]$  tends to 0 whenever  $\pi$  is a nonseparable pattern and the random variable takes only nonnegative values,  $\widetilde{\text{occ}}(\pi, \sigma_n)$  tends to 0 in probability.

Here, we discuss more precisely the asymptotic behavior of  $\widetilde{\text{occ}}(\pi, \sigma_n)$  in this case. The first result gives the order of magnitude of its moments; we then present a consequence for the random variable itself.

**Proposition 5.13.** *For  $\pi \in \langle S \rangle$  and  $m \geq 1$ ,  $\mathbb{E}[(\widetilde{\text{occ}}(\pi, \sigma_n))^m] = \Theta(n^{-\text{db}(\pi)/2})$ .*

*Remark 5.14.* This result does also consider separable patterns  $\pi$ , but in this case it is a direct consequence of our main theorem (as explained in Section 1.6).

*Proof.* By definition,  $\widetilde{\text{occ}}(\pi, \sigma_n) = \binom{n}{k}^{-1} \sum_{I \subset [n], |I|=k} \mathbf{1}_{\text{pat}_I(\sigma_n)=\pi}$ . Consequently,

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)^m] = \binom{n}{k}^{-m} \mathbb{E} \left[ \sum_{\substack{I_1, \dots, I_m \subset [n] \\ \forall i, |I_i|=k}} \mathbf{1}_{\forall i, \text{pat}_{I_i}(\sigma_n)=\pi} \right].$$

We split the sum according to the different possible values of  $K = \bigcup_i I_i$  and  $j = |K|$ . Denoting  $B_{k,m}^K$  the set of possible covers of  $K$  by  $m$  sets of size  $k$ , this gives

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)^m] = \binom{n}{k}^{-m} \mathbb{E} \left[ \sum_{j=k}^{mk} \sum_{\substack{K \subset [n] \\ |K|=j}} \sum_{(I_1, \dots, I_m) \in B_{k,m}^K} \mathbf{1}_{\forall i, \text{pat}_{I_i}(\sigma_n)=\pi} \right].$$

Let us now remark that the unique increasing bijection between  $K$  and  $[j]$  induces a bijection between  $B_{k,m}^K$  and  $B_{k,m}^{[j]}$ . Let  $(J_i)_{1 \leq i \leq m}$  denote the image of  $(I_i)_{1 \leq i \leq m}$  by this bijection. Clearly,

$$(35) \quad \text{pat}_{I_i}(\sigma_n) = \pi \iff \text{pat}_{J_i}(\text{pat}_K(\sigma_n)) = \pi.$$

The sum can now be decomposed according to the different values of  $\rho = \text{pat}_K(\sigma_n)$  yielding

$$\begin{aligned} \mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)^m] &= \binom{n}{k}^{-m} \mathbb{E} \left[ \sum_{j=k}^{mk} \sum_{\substack{K \subset [n] \\ |K|=j}} \sum_{(J_1, \dots, J_m) \in B_{k,m}^{[j]}} \sum_{\rho \in \mathfrak{S}_j} \mathbf{1}_{\text{pat}_K(\sigma_n)=\rho} \mathbf{1}_{\forall i, \text{pat}_{J_i}(\rho)=\pi} \right] \\ &= \sum_{j=k}^{mk} \sum_{(J_1, \dots, J_m) \in B_{k,m}^{[j]}} \sum_{\substack{\rho \in \mathfrak{S}_j \\ \forall i, \text{pat}_{J_i}(\rho)=\pi}} \binom{n}{k}^{-m} \binom{n}{j} \mathbb{E}[\widetilde{\text{occ}}(\rho, \sigma_n)]. \end{aligned}$$

Since the summation index sets do not depend on  $n$ , it is enough to consider each summand separately to get the asymptotics. Proposition 5.12 implies that the summand  $\binom{n}{k}^{-m} \binom{n}{j} \mathbb{E}[\widetilde{\text{occ}}(\rho, \sigma_n)]$  is of order  $n^{j-km-\text{db}(\rho)/2}$ .

Whenever Eq. (35) holds,  $\pi$  is a pattern of  $\rho = \text{pat}_K(\sigma_n)$ . As a consequence, an expanded tree of  $\rho$  must have a substitution tree of  $\pi$  as an induced tree. Since the default of binarity may only decrease when taking induced trees, this implies that  $\text{db}(\rho) \geq \text{db}(\pi)$ . Since additionally  $j \leq km$ , we deduce that  $j - km - \text{db}(\rho)/2 \leq -\text{db}(\pi)/2$  which gives  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)^m] = \mathcal{O}(n^{-\text{db}(\pi)/2})$ .

To prove that  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)^m] = \Theta(n^{-\text{db}(\pi)/2})$ , it is then enough to find one summand, which grows as  $n^{-\text{db}(\pi)/2}$  for large  $n$ . This is achieved considering the summand indexed by

$$j = km; J_i = \{mq + i : 0 \leq q \leq k-1\}; \rho = \pi[1 \cdots m, \dots, 1 \cdots m].$$

Indeed in this case,  $\text{db}(\rho) = \text{db}(\pi)$ , so that  $j - km - \text{db}(\rho)/2 = -\text{db}(\pi)/2$ , which concludes the proof of the proposition.  $\square$

**Corollary 5.15.** For  $\pi \in \langle \mathcal{S} \rangle$  and  $\varepsilon > 0$  small enough,  $\mathbb{P}(\widetilde{\text{occ}}(\pi, \sigma_n) > \varepsilon) = \Theta(n^{-\text{db}(\pi)/2})$ .

*Proof.* The upper bound is an immediate consequence of Markov's inequality. For the lower bound, let  $X$  be a random variable in  $[0, 1]$ , we have

$$\begin{aligned} \mathbb{P}(X \geq \varepsilon) &\geq \int_{\varepsilon}^1 \mathbb{P}(X \geq u) du \geq \int_{\varepsilon}^1 (u - \varepsilon) \mathbb{P}(X \geq u) du \\ &\geq \int_0^1 (u - \varepsilon) \mathbb{P}(X \geq u) du = \frac{1}{2} \mathbb{E}[X^2] - \varepsilon \mathbb{E}[X]. \end{aligned}$$

The corollary follows by taking  $X = \widetilde{\text{occ}}(\pi, \sigma_n)$  and  $\varepsilon$  small enough.  $\square$

## 6. ASYMPTOTIC ANALYSIS: THE DEGENERATE CASE $S'(R_S) < 2/(1 + R_S)^2 - 1$

In this section, we are interested in the case where the generating function  $S$  of simple permutations in  $\mathcal{S}$  satisfies the following condition.

**Definition 6.1** (Hypothesis (H2)). *The generating function  $S$  of a family  $\mathcal{S}$  of simple permutations is said to satisfy hypothesis (H2) if  $S$  meets the following conditions at its radius of convergence  $R_S > 0$ :*

i)  $S'$  is convergent at  $R_S$  and

$$(36) \quad S'(R_S) < \frac{2}{(1 + R_S)^2} - 1;$$

ii)  $S$  has a dominant singularity of exponent  $\delta > 1$  in  $R_S$ .

Item ii means that, around the singularity  $R_S$ , one has

$$S(z) = g_S(z) + (C_S + o(1))(R_S - z)^\delta,$$

for some analytic function  $g_S$  and constant  $C_S \neq 0$ . We refer to Appendix A.4 for a precise definition. Clearly, under (H2), it holds that  $R_S < \infty$ . Note also that the assumption  $\delta > 1$  is redundant with the convergence of  $S'$  at  $R_S$ .

**6.1. Asymptotic behavior of the main series.** As in Section 5.2, the first step is to derive the asymptotic behavior of all generating functions for marked trees around their common dominant singularity. In this section, we will not compute constants explicitly, but only focus on the singularity exponent. Indeed, keeping track only of singularity exponents is here sufficient to determine the limiting permuton.

The function  $\Lambda$  is defined in (23) by:

$$\Lambda(u) = \frac{u^2}{1 - u} + S\left(\frac{u}{1 - u}\right).$$

**Lemma 6.2.** *Assume that  $S$  satisfies hypothesis (H2). Then  $\Lambda$  has a unique dominant singularity of exponent  $\delta$  in  $R_\Lambda := \frac{R_S}{1 + R_S} < 1$ . Moreover,  $\Lambda'$  is convergent at  $R_\Lambda$  and  $\Lambda'(R_\Lambda) < 1$ .*

*Proof.* The first assertion follows from Lemma A.6 (Supercritical case), using also that  $R_S < \infty$ . The convergence of  $\Lambda'$  at  $R_\Lambda$  follows from that of  $S'$  at  $R_S$ . And the inequality  $\Lambda'(R_\Lambda) < 1$  is a straightforward computation from (36).  $\square$

Recall that from Proposition 4.1 (p. 26)  $T_{\text{not}\oplus}$  is implicitly defined by the equation

$$(37) \quad T_{\text{not}\oplus}(z) = z + \Lambda(T_{\text{not}\oplus}(z)).$$

As explained in Section 1.7, the condition  $\Lambda'(R_\Lambda) < 1$  implies that the singularity of  $T_{\text{not}\oplus}(z)$  is not a *branch point*, but is inferred from the singularity of  $\Lambda$ .

**Lemma 6.3.** *Assume that  $S$  satisfies hypothesis (H2). Then there is a unique  $\rho > 0$  such that  $T_{\text{not}\oplus}(\rho) = R_\Lambda$ . Moreover,  $\rho$  is the radius of convergence of  $T_{\text{not}\oplus}$  and  $T_{\text{not}\oplus}$  has a unique dominant singularity of exponent  $\delta$  in  $\rho$ .*

The proof is rather technical and postponed to Appendix A.6.

Now since from Eq. (12):  $T = \frac{T_{\text{not}\oplus}}{1 - T_{\text{not}\oplus}}$ , at the singularity  $\rho$  of  $T_{\text{not}\oplus}$ , the denominator of  $T(\rho)$  is  $1 - T_{\text{not}\oplus}(\rho) = 1 - R_\Lambda > 0$ , hence the singularity of  $T$  is inherited from the one of  $T_{\text{not}\oplus}$ . More precisely, from Lemma A.6 (Subcritical case), we have the following corollary.

**Corollary 6.4.** *Assume that  $S$  satisfies hypothesis (H2). Then  $T$  has a unique dominant singularity of exponent  $\delta$  in  $\rho$ , with  $T(\rho) = \frac{R_\Lambda}{1 - R_\Lambda} = R_S$ .*

We now turn to the behavior of generating function of trees with one marked leaf.

**Lemma 6.5.** *Assume that  $S$  satisfies hypothesis (H2). Then each of the nine generating functions  $T', T^+, \dots, T_{\text{not}\ominus}^-$  has a unique dominant singularity of exponent  $\delta - 1$  in  $\rho$ .*

*Proof.* For  $T'$  (resp.  $T'_{\text{not}\oplus} = T'_{\text{not}\ominus}$ ) this follows immediately from Corollary 6.4 (resp. Lemma 6.3) by singular differentiation (Lemma A.7).

Recall that  $T^+$  is explicitly given in Proposition 4.2 as

$$T^+ = \frac{1}{1 - WS'(T) - W - S'(T)}$$

where  $W = (\frac{1}{1 - T_{\text{not}\oplus}})^2 - 1$ . We examine  $W$  and  $S'(T)$  to determine the exponent of the singularity of  $T^+$  at its radius of convergence (which we will prove to be  $\rho$ ).

Since  $T_{\text{not}\oplus}(\rho) = R_\Lambda < 1$ , again by subcritical composition, Lemma 6.3 gives that  $W$  has a unique dominant singularity of exponent  $\delta$  in  $\rho$ . Moreover,  $W(\rho) = (\frac{1}{1 - R_\Lambda})^2 - 1$ .

As for  $S'(T)$  we need to analyse  $S'$  and  $T$  and determine whether the composition is critical, supercritical or subcritical (see Lemma A.6).

- By singular differentiation (Lemma A.7), it follows from (H2) that  $S'$  has a dominant singularity of exponent  $\delta - 1$  in  $R_S$ . In addition,  $S'(R_S) < \frac{2}{(1 + R_S)^2} - 1$  by (H2).
- Recall that from Corollary 6.4  $T$  has a unique dominant singularity of exponent  $\delta$  in  $\rho$ , with  $T(\rho) = \frac{R_\Lambda}{1 - R_\Lambda} = R_S$ . Moreover,  $T$  is aperiodic since it counts a superset of the separable permutations.
- In  $\rho$ , we have

$$T(\rho) = \frac{T_{\text{not}\oplus}(\rho)}{1 - T_{\text{not}\oplus}(\rho)} = \frac{R_\Lambda}{1 - R_\Lambda} = R_S,$$

so that the composition  $S' \circ T$  is critical.

From Lemma A.6 (Critical case-A), we obtain that  $S'(T)$  has a unique dominant singularity of exponent  $\delta - 1$  in  $\rho$ , and therefore so does  $WS'(T) + W + S'(T)$ .

In  $\rho$ , the value of the series  $WS'(T) + W + S'(T) = (W + 1)(S'(T) + 1) - 1$  is less than  $(\frac{1}{1-R_A})^2 \frac{2}{(1+R_S)^2} - 1 = 1$ . Therefore, by subcritical composition,  $T^+ = \frac{1}{1-WS'(T)-W-S'(T)}$  has a unique dominant singularity of exponent  $\delta - 1$  in  $\rho$ . Since  $\frac{1}{1+W}$  and  $(WS'(T) + W + S'(T))$  have unique dominant singularities in  $\rho$  of respective exponents  $\delta$  and  $\delta - 1$ , the other cases follow by Lemma A.5, using the formulas given in Proposition 4.2.  $\square$

**6.2. Probability of given patterns.** Recall that the function  $\text{Occ}_\theta$  was defined in Eq. (17) by  $\text{Occ}_\theta(z) = \sum_{\alpha \in \mathcal{S}} \text{occ}(\theta, \alpha) z^{|\alpha| - |\theta|}$ .

Unlike in the previous section, the functions  $\text{Occ}_\theta(z)$  will appear in the asymptotic behaviors, thus we need some additional assumptions on them. First, as noticed in Observation 4.3, we have

$$(38) \quad \sum_{\theta \in \mathfrak{S}_k} \text{Occ}_\theta(z) = \frac{1}{k!} S^{(k)}(z),$$

which, under (H2), has a dominant singularity of exponent  $\delta - k$  in  $R_S$  (see Lemma A.7, about singular differentiation). The following hypothesis is thus reasonable.

**Definition 6.6** (Hypothesis (CS)). *Let  $S$  have a dominant singularity of exponent  $\delta > 1$  in  $R_S$ . The family of simple permutations  $\mathcal{S}$  satisfies the hypothesis (CS) if, for each pattern  $\theta$  of size  $k$ , the corresponding series  $\text{Occ}_\theta(z)$  has a dominant singularity of exponent at least  $\delta - k$  in  $R_S$ .*

Let  $t_0$  be a substitution tree with  $k$  leaves. Recall that  $\text{Int}(t_0)$  is the set of internal nodes of  $t_0$ , and that for any node  $v \in \text{Int}(t_0)$ ,  $d_v$  denotes its number of children. Recall also from Section 4.2 (p. 28) that  $\mathcal{T}_{t_0}$  is the family of canonical trees with  $k$  marked leaves which induce a tree equal to  $t_0$ , and that  $T_{t_0}$  is its generating function.

Combining Proposition 4.5 and the above results, we obtain the following.

**Proposition 6.7.** *For any substitution tree  $t_0$  of size  $k \geq 2$ , assuming (H2) and (CS), the series  $T_{t_0}$  has a unique dominant singularity of exponent at least  $\tilde{e}_{t_0}$  in  $\rho$ , where*

- $\tilde{e}_{t_0} = \sum_{v: d_v > \delta} (\delta - d_v)$ , if there is at least one node  $v$  such that  $d_v > \delta$ ;
- $\tilde{e}_{t_0} = \delta - \max_{v \in \text{Int}(t_0)} d_v$  otherwise<sup>7</sup>.

*Proof.* First recall that  $\mathcal{T}_{t_0}$  can be decomposed as a union  $\bigcup_{V_s} \mathcal{T}_{t_0, V_s}$ , where  $V_s$  are subsets of  $\text{Int}(t_0)$  which contain all nonlinear nodes (see Section 4.1). It is therefore enough to prove that each series  $T_{t_0, V_s}$  has a unique dominant singularity in  $\rho$  with at least the desired exponent.

Recall from Proposition 4.5 the following formula for  $T_{t_0, V_s}(z)$ :

$$T_{t_0, V_s} = z^k T^{\text{type of root}} \prod_{v \in \text{Int}(t_0)} A_v,$$

where each  $A_v$  is given by (19) and depends on the type of  $v$ . Using hypothesis (CS), Corollary 6.4 and Lemma A.6 (Critical case-A), we know that  $\text{Occ}_{\theta_v}(T)$  has a dominant singularity of exponent at least  $\delta - d_v$  in  $\rho$ , and, from Lemma 6.5, that it is the term with the lowest exponent

<sup>7</sup>Note for future reference that the two expressions for  $\tilde{e}_{t_0}$  are equal when there is a unique  $v$  such that  $d_v > \delta$ .



and the only possibly divergent term arising in  $A_v$ . It then follows from Lemma A.5 that  $A_v$  has a dominant singularity of exponent at least  $\delta - d_v$  in  $\rho$ .

The series  $T_{t_0, V_s}(z)$  is then the product of the  $A_v$ 's and of some convergent series. The result of the proposition is obtained from Lemma A.5:  $T_{t_0, V_s}(z)$  has a dominant singularity of exponent at least  $\sum_{v: d_v > \delta} (\delta - d_v)$  in  $\rho$ , as soon as this sum is nonempty (i.e., when one of the series is divergent) and of the smallest singularity exponent among them, that is  $\delta - \max_{v \in \text{Int}(t_0)} d_v$  otherwise, i.e., if all factors are convergent.  $\square$

As in Section 5.4 (p. 42), we take a uniform random tree with  $n$  leaves in  $\mathcal{T}$  with  $k$  marked leaves (chosen also uniformly at random). Denote as before by  $\mathbf{t}_n^k$  the tree induced by the  $k$  marked leaves.

**Corollary 6.8.** *Assume (H2) and (CS). For any substitution tree  $t_0$  with  $k \geq 2$  leaves, the probability  $\mathbb{P}(\mathbf{t}_k^{(n)} = t_0)$  tends to 0, unless  $t_0$  has only one internal node.*

*Proof.* As in Section 5.4, we use the formula

$$\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) = \frac{[z^n]T_{t_0}(z)}{\binom{n}{k}[z^n]T(z)}$$

From Theorem A.3 and Proposition 6.7 (and using the notation  $\tilde{e}_{t_0}$  herein defined), we get that

$$[z^n]T_{t_0}(z) = (\tilde{C}_{t_0} + o(1))\rho^{-n}n^{-1-\tilde{e}_{t_0}}$$

for some constant  $\tilde{C}_{t_0}$ , possibly equal to 0. On the other hand, Theorem A.3 and Corollary 6.4 imply that

$$[z^n]T(z) = C\rho^{-n}n^{-\delta-1}(1 + o(1)),$$

for some constant  $C \neq 0$ . Putting everything together, we obtain

$$\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) = \frac{k! (\tilde{C}_{t_0} + o(1))}{C} n^{e_{t_0}},$$

where  $e_{t_0} = \delta - k - \tilde{e}_{t_0}$ .

For any subset  $A$  of the internal nodes of a tree  $t$  with  $k$  leaves, we claim that the following inequality holds:  $\sum_{v \in A} d_v \leq |A| + k - 1$ . It is clear when  $k = 1$ . For  $k > 1$ , we decompose  $t$  as a root  $\emptyset$  with  $d \geq 2$  subtrees  $t_1, \dots, t_d$ . The chosen set  $A$  of nodes of  $t$  determines a set  $A_i$  of nodes in each tree  $t_i$  that has  $k_i$  leaves. Assume that  $\sum_{v \in A_i} d_v \leq |A_i| + k_i - 1$  for all  $i$ . Then, we have

$$\begin{aligned} \sum_{v \in A} d_v &= d \cdot \mathbf{1}_{\emptyset \in A} + \sum_{i=1}^d \sum_{v \in A_i} d_v \leq d \cdot \mathbf{1}_{\emptyset \in A} + \sum_{i=1}^d |A_i| + k_i - 1 \\ &= d \cdot \mathbf{1}_{\emptyset \in A} + |A| - \mathbf{1}_{\emptyset \in A} + k - d = |A| + k - 1 + (d - 1)(\mathbf{1}_{\emptyset \in A} - 1), \end{aligned}$$

and with the observation that  $(d - 1)(\mathbf{1}_{\emptyset \in A} - 1) \leq 0$ , this proves our claim. Set  $A = \{v \in \text{Int}(t_0) : d_v > \delta\}$ ; if  $|A| \geq 1$ , one has

$$e_{t_0} = \delta - k - \delta|A| + \sum_{v \in A} d_v \leq \delta - k - \delta|A| + |A| + k - 1 = (1 - |A|)(\delta - 1),$$

which is negative for  $|A| \geq 2$  (indeed,  $\delta > 1$  by (H2)). When  $|A| = 0$  or  $|A| = 1$ , we have (see also Footnote 7)  $\tilde{e}_{t_0} = \delta - \max_{v \in \text{Int}(t_0)} d_v$  and, therefore,  $e_{t_0} = \max_{v \in \text{Int}(t_0)} d_v - k$  is negative unless  $t_0$  has exactly one internal node (which is then of degree  $k$ ).  $\square$

It is now straightforward to translate this result in terms of the probability to find a given pattern in a random permutation in the set  $\langle S \rangle$ . As recalled in Appendix A.4, the hypothesis (CS) is equivalent to the following: for every  $k \geq 1$  and every permutation  $\theta$  of size  $k$ , there exists an analytic function  $g_\theta$  and a constant  $C_\theta$  (possibly equal to 0) such that, on an  $\Delta$ -neighborhood of  $R_S$ , it holds that

$$(39) \quad \text{Occ}_\theta(z) = g_\theta(z) + (C_\theta + o(1))(R_S - z)^{\delta-k}.$$

The quantities  $C_\theta$  are involved in the statement of the following theorem.

**Theorem 6.9.** *Let  $\sigma_n$  be a uniform random permutation in  $\langle S \rangle_n$ . We assume hypotheses (H2) and (CS). Then, for any  $k \geq 2$  and for any  $\pi \in \mathfrak{S}_k$ ,*

$$(40) \quad \lim_{n \rightarrow \infty} \mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = \frac{C_\pi}{\sum_{\theta \in \mathfrak{S}_k} C_\theta}.$$

And there exists a random permutation  $\mu_S$  with

$$(41) \quad \mathbb{E}[\widetilde{\text{occ}}(\pi, \mu_S)] = \frac{C_\pi}{\sum_{\theta \in \mathfrak{S}_k} C_\theta}$$

such that  $(\mu_{\sigma_n})_n$  tends to  $\mu_S$  in distribution.

*Proof.* The starting point is the same as in the standard case (Proposition 5.12). As before we denote by  $I$  a random uniform  $k$ -element subset of  $[n]$ , and  $\mathbf{t}_k^{(n)}$  is the tree of size  $k$  induced by  $k$  uniform leaves in a uniform canonical tree of size  $n$  in  $\mathcal{T}$ . From Lemma 3.11, we have

$$\mathbb{E}(\widetilde{\text{occ}}(\pi, \sigma_n)) = \mathbb{P}(\text{pat}_I(\sigma_n) = \pi) = \mathbb{P}(\text{perm}(\mathbf{t}_k^{(n)}) = \pi) = \sum_{t_0: \text{perm}(t_0) = \pi} \mathbb{P}(\mathbf{t}_k^{(n)} = t_0),$$

where the sum runs over all substitution trees encoding  $\pi$ . Denote by  $t_0^\pi$  the substitution tree with only one internal node labeled by  $\pi$ . When  $n$  tends to infinity, using Corollary 6.8, we know that every term in the above sum vanishes, but the term corresponding to  $t_0^\pi$ :

$$(42) \quad \lim_{n \rightarrow \infty} \mathbb{E}(\widetilde{\text{occ}}(\pi, \sigma_n)) = \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{t}_k^{(n)} = t_0^\pi).$$

Now we can compute directly from Proposition 4.5 that

$$T_{t_0^\pi} = \mathbf{1}_{\pi \text{ linear}} z^k T^{+} \left( \frac{1}{1 - T_{\text{not}\oplus}} \right)^{k+1} (T')^k + z^k \text{Occ}_\pi(T) (T')^{k+1}.$$

The first term has a dominant singularity of exponent at least  $\delta - 1$ , while the dominant singularity exponent of the second term is at least  $\delta - k$ . We therefore focus on the second term and get by an easy computation that, around  $z = \rho$ , we have

$$\text{Occ}_\pi(T(z)) = g(z) + (C_\pi T'(\rho)^{k-\delta} + o(1)) (\rho - z)^{\delta-k},$$

and thus

$$T_{t_0^\pi}(z) = h(z) + (\rho^k T'(\rho)^{\delta+1} C_\pi + o(1)) (\rho - z)^{\delta-k},$$

where  $g$  and  $h$  are analytic functions. Then we can apply the Transfer Theorem (Theorem A.3) to  $T$  and  $T_{t_0^\pi}$  (as in the proof of Corollary 6.8) and obtain that  $\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{t}_k^{(n)} = t_0^\pi)$  is proportional to  $C_\pi$  for  $\pi \in \mathfrak{S}_k$ . Since the left-hand side of Eq. (42) sums to one (when summed over  $\pi \in \mathfrak{S}_k$ , for a fixed  $k$ ), this proves Eq. (40).

The rest of the statement follows immediately, using Theorem 2.5 (with Observation 2.6).  $\square$

**6.3. Hypothesis (CS) and convergence of uniform random simple permutations.** We may wish to replace the hypothesis (CS) with a less technical hypothesis, such as the convergence of a random *simple* permutation in our set  $\mathcal{S}$  to the random permuton  $\mu_{\mathcal{S}}$  uniquely determined by Eq. (41). We show here that, though not equivalent, these hypotheses are strongly related. Remark that we do not assume (H2) here, so that the existence of  $\mu_{\mathcal{S}}$  cannot be inferred from Theorem 6.9 and needs a separate proof.

**Proposition 6.10.** *Suppose that  $\mathcal{S}$  has a dominant singularity of exponent  $\delta > 1$  and assume condition (CS). Then there exists a permuton  $\mu_{\mathcal{S}}$  such that*

$$(43) \quad \mathbb{E}[\widetilde{\text{occ}}(\pi, \mu_{\mathcal{S}})] = \frac{C_\pi}{\sum_{\theta \in \mathcal{S}_k} C_\theta}.$$

*Let  $\alpha_n$  be a uniform random permutation of size  $n$  in  $\mathcal{S}$ . If  $(\mu_{\alpha_n})$  converges in distribution, then its limit is  $\mu_{\mathcal{S}}$ . Conversely, if we assume that  $\mathcal{S}$  and all series  $\text{Occ}_\theta$  have a unique dominant singularity, then  $(\mu_{\alpha_n})$  converges in distribution (and the limit must be  $\mu_{\mathcal{S}}$ , using the first part of the proposition).*

*Proof.* We start with the existence of  $\mu_{\mathcal{S}}$ . For every  $k \geq 1$ , let  $\rho_k$  be a random permutation in  $\mathfrak{S}_k$  such that  $\mathbb{P}(\rho_k = \pi) = C_\pi / (\sum_{\theta \in \mathcal{S}_k} C_\theta)$ . By Proposition 2.9, we only need to show that  $(\rho_k)_k$  forms a consistent family. Let  $1 \leq k \leq n$ , then for  $\pi \in \mathfrak{S}_k$ ,

$$(44) \quad \mathbb{P}(\text{pat}_{\mathbf{I}_{n,k}}(\rho_n) = \pi) = \frac{1}{\binom{n}{k} \sum_{\theta \in \mathcal{S}_k} C_\theta} \sum_{\sigma \in \mathfrak{S}_n} \text{occ}(\pi, \sigma) C_\sigma.$$

On the other hand, the following combinatorial identity can be derived from the definition of the  $(\text{Occ}_\theta)_{\theta \in \mathfrak{S}}$ :

$$(45) \quad \frac{1}{(n-k)!} \text{Occ}_\pi^{(n-k)}(z) = \sum_{\sigma \in \mathfrak{S}_n} \text{occ}(\pi, \sigma) \text{Occ}_\sigma(z).$$

Indeed, the left-hand side is the series of simple permutations in  $\mathcal{S}$  whose entries are partitioned into a set of  $k$  marked entries forming a pattern  $\pi$  and a set of  $n-k$  marked entries. The right-hand side counts the same object, according to the pattern  $\sigma$  formed by all the  $n$  marked entries. To distinguish the marked entries of the first set from the ones of the second set, we need to specify a subpattern  $\pi$  inside the pattern  $\sigma$ , which explains the factor  $\text{occ}(\pi, \sigma)$ .

We now differentiate both sides of Eq. (45)  $m$  times so that  $\delta - n - m < 0$ , and replace all series with their asymptotic estimates obtained thanks to hypothesis (CS), Eq. (39) and singular

differentiation (Theorem A.4)<sup>8</sup>.

$$\begin{aligned} g_\pi^{(m)}(z) + (-1)^{m+n-k}(\delta - k)_{m+n-k}(C_\pi + o(1))(R_S - z)^{\delta-m-n} \\ = (n-k)! \sum_{\sigma \in \mathfrak{S}_n} \text{occ}(\pi, \sigma) [g_\sigma^{(m)}(z) + (-1)^m(\delta - n)_m(C_\sigma + o(1))(R_S - z)^{\delta-m-n}] \end{aligned}$$

As only the singular parts diverge, taking the limit in  $z \rightarrow R_S$  allows to identify the constants, yielding  $\sum_{\sigma \in \mathfrak{S}_n} \text{occ}(\pi, \sigma) C_\sigma = (-1)^{n-k} \binom{\delta-k}{n-k} C_\pi$ . Plugging this back in Eq. (44) yields

$$\mathbb{P}(\text{pat}_{\mathbf{I}_{n,k}}(\boldsymbol{\rho}_n) = \pi) \propto C_\pi, \pi \in \mathfrak{S}_k.$$

As probabilities sum to 1, we get  $\mathbb{P}(\text{pat}_{\mathbf{I}_{n,k}}(\boldsymbol{\rho}_n) = \pi) = \mathbb{P}(\boldsymbol{\rho}_k = \pi)$ , proving the consistency of  $(\boldsymbol{\rho}_k)_k$ .

As noticed in the proof of Theorem 6.9, the convergence of  $(\mu_{\alpha_n})$  to  $\mu_S$  is equivalent to the following: for any fixed  $k \geq 2$  and any  $\pi \in \mathfrak{S}_k$ , the limit  $\lim_{n \rightarrow \infty} \mathbb{E}[\widetilde{\text{occ}}(\pi, \alpha_n)]$  exists and is proportional to  $C_\pi$  for  $\pi \in \mathfrak{S}_k$ . Furthermore, by consistency, we only need to show it for large  $k$ . Directly from the definitions, we have

$$(46) \quad \mathbb{E}[\widetilde{\text{occ}}(\pi, \alpha_n)] = \frac{k! [z^{n-k}] \text{Occ}_\pi(z)}{[z^{n-k}] S^{(k)}(z)}.$$

For the proof of the direct implication, we assume that  $\mu_{\alpha_n}$  converges in distribution. From Theorem 2.5, this means that  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \alpha_n)]$  has a limit  $\Delta_\pi$  for every  $\pi$  of size  $k \geq 2$ . Then when  $n$  goes to infinity,

$$[z^n] \text{Occ}_\pi(z) = \frac{\Delta_\pi + o(1)}{k!} [z^n] S^{(k)}(z).$$

As a consequence, for any fixed  $\pi$  and  $\varepsilon > 0$ , there exists polynomials  $g_-, g_+$  such that for any real  $z > 0$ ,

$$(47) \quad \left( \frac{\Delta_\pi}{k!} - \varepsilon \right) S^{(k)}(z) + g_-(z) \leq \text{Occ}_\pi(z) \leq \left( \frac{\Delta_\pi}{k!} + \varepsilon \right) S^{(k)}(z) + g_+(z).$$

Hypothesis (CS) implies that in  $R_S$  we have  $\text{Occ}_\pi(z) = g_\pi(z) + (C_\pi + o(1))(R_S - z)^{\delta-k}$  for some analytic function  $g_\pi$ . Also  $S^{(k)}$  has a dominant singularity of exponent  $\delta - k$  in  $R_S$  so  $S^{(k)}(z) = g_{S^{(k)}}(z) + (C_{S^{(k)}} + o(1))(R_S - z)^{\delta-k}$  for some analytic function  $g_{S^{(k)}}$  and constant  $C_{S^{(k)}} > 0$ . Plugging these asymptotic estimates into (47) yields

$$\begin{aligned} \left( \frac{\Delta_\pi}{k!} - \varepsilon \right) [(C_{S^{(k)}} + o(1))(R_S - z)^{\delta-k} + g_{S^{(k)}}] + g_- \\ \leq (C_\pi + o(1)) [(R_S - z)^{\delta-k} + g_\pi] \\ \leq \left( \frac{\Delta_\pi}{k!} + \varepsilon \right) [(C_{S^{(k)}} + o(1))(R_S - z)^{\delta-k} + g_{S^{(k)}}] + g_+^{(m)}. \end{aligned}$$

<sup>8</sup>For  $x \in \mathbb{C}$  and  $r \in \mathbb{N}$ , we denote by  $(x)_r$  the falling factorial  $x(x-1) \cdots (x-r+1)$

Let  $k$  be such that  $\delta - k < 0$ , so that the singular parts are the only diverging quantities when  $z \rightarrow R_S$ . After taking the limit we get  $\left| C_\pi - \frac{C_{S(k)}}{k!} \Delta_\pi \right| \leq \varepsilon$  for every  $\varepsilon$  and hence equality. We have proven that  $(\Delta_\pi)_{\pi \in \mathfrak{S}_k}$  is proportional to  $(C_\pi)_{\pi \in \mathfrak{S}_k}$  for large  $k$ , as desired.

For the converse, Theorem A.3 (which we can apply because of the hypotheses on  $S$  and  $\text{Occ}_\theta$ ) gives the following asymptotic behavior when  $n \rightarrow \infty$ :

$$\mathbb{E}[\widetilde{\text{occ}}(\pi, \alpha_n)] = \frac{n^{-k}}{k!} \frac{(C_\pi + o(1)) R_S^{-n+k} n^{-\delta+k-1}}{C_S R_S^{-n} n^{-\delta+1}}.$$

For fixed  $k$ , the limit of the right-hand side is proportional to  $C_\pi$ .  $\square$

## 7. ASYMPTOTIC ANALYSIS: THE CRITICAL CASE $S'(R_S) = 2/(1 + R_S)^2 - 1$

The goal of this section is to describe the limiting permuton of a uniform permutation in a substitution-closed class  $\mathcal{C}$ , whose set of simple permutations satisfy the following hypothesis.

**Definition 7.1** (Hypothesis (H3)). *A family  $\mathcal{S}$  of simple permutations is said to satisfy hypothesis (H3) if the generating function  $S$  meets the following conditions at its radius of convergence  $R_S > 0$ :*

- $S$  has a dominant singularity of exponent  $\delta > 1$  in  $R_S$ ;
- $S'$  is convergent at  $R_S$  and

$$S'(R_S) = \frac{2}{(1 + R_S)^2} - 1.$$

This hypothesis implies the following behavior of  $\Lambda$  near its singularity.

**Lemma 7.2.** *Assume that  $S$  satisfies hypothesis (H3) and set, as before,*

$$\Lambda(u) = \frac{u^2}{1 - u} + S\left(\frac{u}{1 - u}\right).$$

*Then  $\Lambda$  has a unique dominant singularity of exponent  $\delta$  in  $R_\Lambda := \frac{R_S}{1 + R_S} < 1$ . Moreover,  $\Lambda'$  is convergent at  $R_\Lambda$  and  $\Lambda'(R_\Lambda) = 1$ .*

*Proof.* Follows from Lemma A.6 (Supercritical case).  $\square$

In the following, we denote  $\delta_* = \min(\delta, 2)$ . The behavior of  $T_{\text{not}\oplus}$  is given in the following lemma, whose proof is postponed to Appendix A.7. Note that the exponent is different from the one observed in Section 6 (the singularity comes here from a mixture of a branch point and of the singularity of  $\Lambda$ , as explained in Section 1.7).

**Lemma 7.3.** *Assume that  $S$  satisfies hypothesis (H3). Then there is a unique  $\rho > 0$  such that  $T_{\text{not}\oplus}(\rho) = R_\Lambda$ . Moreover,  $\rho$  is the radius of convergence of  $T_{\text{not}\oplus}$  and  $T_{\text{not}\oplus}$  has a unique dominant singularity of exponent  $1/\delta_*$  in  $\rho$ .*

From here, the strategy is similar as the previous sections, we therefore skip unnecessary details. As in the previous section, we deduce immediately the asymptotic behavior of all generating functions of marked trees:

**Corollary 7.4.** *Assume that  $S$  satisfies hypothesis (H3) and define  $\rho$  as above. Then  $T$  has a unique dominant singularity of exponent  $1/\delta_*$  in  $\rho$ , with  $T(\rho) = R_S$ . Moreover each of the generating functions  $T', T'_{\text{not}\oplus}, \dots, T'_{\text{not}\ominus}$  has a unique dominant singularity of exponent  $1/\delta_* - 1$  in  $\rho$ .*

To go further, we have to assume hypothesis (CS), i.e. that, for any  $\theta$  of size  $k$ , the series  $\text{Occ}_\theta(z)$  has a unique dominant singularity of exponent at least  $\delta - k$  in  $R_S$ . The composition  $\text{Occ}_\theta \circ T$  is then critical, and from Lemma A.6 (Critical case-B),  $\text{Occ}_\theta(T(z))$  has a unique singularity of exponent at least  $\frac{1}{\delta_*}(\delta - k)$  in  $\rho$ .

By a straightforward computation from Proposition 4.5, we get the following singularity for the series  $T_{t_0, V_S}$ .

**Lemma 7.5.** *Fix a tree  $t_0$  and a subset  $V_S$  of its internal nodes. Then the series  $T_{t_0, V_S}$  has a singularity of exponent at least*

$$\tilde{e}_{(t_0, V_S)} := (|E| + 1)(\frac{1}{\delta_*} - 1) + \sum_{v \in V_S} \frac{1}{\delta_*} \min(\delta - d_v, 0).$$

The behavior of the uniform permutation in  $\langle S \rangle_n$  now depends on whether  $\delta$  is smaller or larger than 2.

### 7.1. The case $\delta \in (1, 2)$ .

**Theorem 7.6.** *Let  $\mathcal{S}$  be a family of simple permutations verifying hypothesis (H3) and (CS), with  $\delta \in (1, 2)$ . We consider the permuton  $\mu_{\mathcal{S}}$  as in Proposition 6.10 and denote for  $\pi \in \mathfrak{S}_k$*

$$\Delta_\pi = \mathbb{E}[\widetilde{\text{occ}}(\pi, \mu_{\mathcal{S}})] = \frac{C_\pi}{\sum_{\theta \in \mathfrak{S}_k} C_\theta}.$$

Let also

$$\nu_{\delta, k}(t_0) = \frac{k!}{(\delta - 1) \cdots ((k - 1)\delta - 1)} \prod_{v \in \text{Int}(t_0)} \frac{(d_v - 1 - \delta) \cdots (2 - \delta)(\delta - 1)}{d_v!}$$

be the probability distribution of the  $k$ -th marginal of the  $\delta$ -stable tree (see [25, Thm 3.3.3]).

If  $\sigma_n$  is a uniform permutation in  $\langle S \rangle_n$ , then

$$(48) \quad \lim_{n \rightarrow \infty} \mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)] = \sum_{t_0: \text{perm}(t_0) = \pi} \nu_{\delta, k}(t_0) \prod_{v \in \text{Int}(t_0)} \Delta_{\theta_v}.$$

As a consequence,  $\mu_{\sigma_n}$  converges in distribution to a random permuton, whose average pattern densities are determined by Eq. (48) and depend only on  $\delta$  and  $\mu_{\mathcal{S}}$ . We call it the  $\delta$ -stable permuton driven by  $\mu_{\mathcal{S}}$ .

A construction of the  $\delta$ -stable permuton driven by  $\nu$  for every  $\delta \in (1, 2)$  and random permuton  $\nu$  is given in Lemma B.2.

**Remark 7.7.** In this case, all possible patterns, in particular nonseparable ones, appear with positive probability in the limit (as long as they appear with positive probability in a uniform simple permutation in the class). More precisely, the proof will show the following:  $k$  random

leaves in a uniform canonical tree induce substitution trees with arbitrary large node degrees, and the first common ancestors of those leaves are all simple permutations with probability tending to 1.

*Proof.* We start from the estimate of Lemma 7.5. In the present case  $\delta_* = \delta$  and  $\delta - d_v$  is always negative. If we fix  $(t_0, V_s)$  and apply Theorem A.3 to  $T$  and  $T_{t_0, V_s}$ , we get

$$\frac{[z^n]T_{t_0, V_s}}{\binom{n}{k}[z^n]T} = (\text{cst} + o(1))n^{e(t_0, V_s)},$$

with

$$e(t_0, V_s) = \frac{1}{\delta} - k - \tilde{e}(t_0, V_s) = \frac{1}{\delta} - k - (|E| + 1) \left( \frac{1}{\delta} - 1 \right) + \sum_{v \in V_s} \left( \frac{d_v}{\delta} - 1 \right) = - \sum_{v \in \text{Int}(t_0) \setminus V_s} \left( \frac{d_v}{\delta} - 1 \right),$$

where the last identity uses  $|E| - |\text{Int}(t_0)| - k + 1 = 0$ . Therefore  $e(t_0, V_s) < 0$  unless  $V_s = \text{Int}(t_0)$ , *i.e.* all internal nodes of  $t_0$  are in  $V_s$ . As a consequence, if  $\mathbf{t}^{(n)}$  is a uniform canonical tree of size  $n$  and  $\mathbf{I}$  is a uniform subset of the leaves of length  $k$ ,

$$\mathbb{P}(\mathbf{t}_{\mathbf{I}}^{(n)} = t_0) = \sum_{V_s \subset \text{Int}(t_0)} \frac{[z^n]T_{t_0, V_s}}{\binom{n}{k}[z^n]T} = \frac{[z^n]T_{t_0, \text{Int}(t_0)}}{\binom{n}{k}[z^n]T} + o(1).$$

Now we want to compute the exact asymptotics, so we look at the constant in front of  $T_{t_0, \text{Int}(t_0)}$ . In what follows, we will always denote by  $C_A$  the constant in front of the singular part of the expansion of the analytic function  $A$ , *i.e.*  $A(z) = g_A(z) + C_A(R_A - z)^{\delta_A}$ , with  $g_A$  analytic at  $R_A$ . An easy computation in the formula given by Proposition 4.5 yields

$$\begin{aligned} T_{t_0, \text{Int}(t_0)}(z) &= z^k (T')^{|E|+1} \prod_v \text{Occ}_{\theta_v}(T) \\ &= \left[ \rho^k \left( \frac{-C_T}{\delta} \right)^{|E|+1} \prod_{v \in \text{Int}(t_0)} C_{\theta_v}(-C_T)^{\delta - d_v} + o(1) \right] (\rho - z)^{1/\delta - k}. \end{aligned}$$

From the Transfer Theorem applied to  $T$  and  $T_{t_0}$  we deduce

$$\begin{aligned} \mathbb{P}(\mathbf{t}_{\mathbf{I}}^{(n)} = t_0) &= \frac{\Gamma(-1/\delta) \rho^{-n-(1/\delta-k)} n^{-(1/\delta-k+1)} \left[ \rho^k \left( \frac{-C_T}{\delta} \right)^{|E|+1} \prod_{v \in \text{Int}(t_0)} C_{\theta_v}(-C_T)^{\delta - d_v} + o(1) \right]}{\binom{n}{k} \Gamma(k-1/\delta) \rho^{-n-1/\delta} n^{-(1/\delta+1)} [C_T + o(1)]} \\ &= \frac{-k! \Gamma(-1/\delta)}{\Gamma(k-1/\delta)} \left[ \frac{(-C_T)^{|E|}}{\delta^{|E|+1}} \prod_{v \in \text{Int}(t_0)} C_{\theta_v}(-C_T)^{\delta - d_v} + o(1) \right] \\ (49) \quad &= \frac{-k! \Gamma(-1/\delta)}{\Gamma(k-1/\delta) \delta^k} \prod_{v \in \text{Int}(t_0)} \frac{1}{\delta} C_{\theta_v}(-C_T)^{\delta} + o(1). \end{aligned}$$

The recursive property of the Gamma function gives  $\frac{-k!\Gamma(-1/\delta)}{\Gamma(k-1/\delta)\delta^k} = \frac{k!}{(\delta-1)\cdots((k-1)\delta-1)}$ . Furthermore, by definition of  $\Delta_{\theta_v}$ , relation (38) and singular differentiation of  $S$ , we get

$$C_{\theta_v} = \Delta_{\theta_v} \frac{C_{S(d_v)}}{d_v!} = \Delta_{\theta_v} \frac{C_S}{d_v!} (-1)^{d_v} \delta(\delta-1) \cdots (\delta-d_v+1) = \Delta_{\theta_v} \frac{C_S}{d_v!} \delta(\delta-1)(2-\delta) \cdots (d_v-1-\delta).$$

This allows us to rewrite (49) as

$$(50) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{t}_I^{(n)} = t_0) = \frac{k!}{(\delta-1)\cdots((k-1)\delta-1)} \prod_{v \in \text{Int}(t_0)} \frac{(\delta-1)(2-\delta) \cdots (d_v-1-\delta)}{d_v!} \Delta_{\theta_v} C_S (-C_T)^\delta.$$

Now the proof of Lemma 7.3 (see Eq. (64)) yields  $-C_{\text{not}\oplus} = C_\Lambda^{-1/\delta}$ . By definition of  $\Lambda$  and  $T$ , we can easily show that

$$C_\Lambda = \frac{C_S}{(1-R_\Lambda)^{2\delta}} \text{ and } C_T = \frac{C_{\text{not}\oplus}}{(1-R_\Lambda)^2}.$$

From there we deduce  $-C_T = C_S^{-1/\delta}$ . Therefore Eq. (50) rewrites

$$(51) \quad \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{t}_I^{(n)} = t_0) = \frac{k!}{(\delta-1)\cdots((k-1)\delta-1)} \prod_{v \in \text{Int}(t_0)} \frac{(\delta-1)(2-\delta) \cdots (d_v-1-\delta)}{d_v!} \Delta_{\theta_v}.$$

Summing over trees  $t_0$  with  $\text{perm}(t_0) = \pi$  gives the theorem.  $\square$

## 7.2. The case $\delta > 2$ .

**Theorem 7.8.** *Let  $\mathcal{S}$  be a family of simple permutations verifying hypotheses (H3) and (CS), with  $\delta > 2$ . If  $\sigma_n$  is a uniform permutation in  $\langle \mathcal{S} \rangle_n$ , then  $\sigma_n$  converges in distribution to the biased Brownian separable permuton of parameter  $p$ , where*

$$p = \frac{(1+R_S)^3 \text{Occ}_{12}(R_S) + 1}{(1+R_S)^3 (\text{Occ}_{12}(R_S) + \text{Occ}_{21}(R_S)) + 2}.$$

*Remark 7.9.* While the limiting permuton in this case is independent of  $\delta > 2$  and is the same than in the standard case, the fine details of this convergence might be different. In particular, if  $\pi$  is a nonseparable pattern, the order of magnitude of  $\mathbb{E}[\widetilde{\text{occ}}(\pi, \sigma_n)]$  depends on  $\delta$  and is in general bigger than in the standard case – compare Eq. (52) to Proposition 5.13.

*Proof.* Let  $(t_0, V_s)$  be a decorated tree with  $k$  leaves. Once again, applying Theorem A.3 to  $T$  and  $T_{t_0, V_s}$  leads to

$$\frac{[z^n] T_{t_0, V_s}}{\binom{n}{k} [z^n] T} = (\text{cst} + o(1)) n^{e(t_0, V_s)},$$



But in this case,

$$(52) \quad e_{(t_0, V_s)} = \frac{1}{2} - k - \tilde{e}_{(t_0, V_s)} = \frac{1}{2} + -k + \frac{1}{2}(|E| + 1) + \frac{1}{2} \sum_{v \in V_s; d_v > \delta} (d_v - \delta) \\ \leq \frac{1}{2}|E| + 1 - k + \frac{1}{2} \sum_{v \in \text{Int}(t_0)} (d_v - 2) = 0.$$

The last equality again follows from  $|E| - |\text{Int}(t_0)| - k + 1 = 0$ . We have equality if and only if  $d_v = 2$  for all internal nodes of  $t_0$ , *i.e.* if the tree is binary. This implies that if  $t_0$  is not binary,

$$(53) \quad \mathbb{P}(\mathbf{t}_I^n) = o(1).$$

We can show that, up to replacing  $\tau$  with  $R_\Lambda$  and  $\kappa$  with  $R_S$ , the estimates of the singular parts of  $T, T_{\text{not}\oplus}, T', T^+, \dots, T_{\text{not}\ominus}^-$  in Propositions 5.8 and 5.9 still hold. Indeed, the proofs can be transposed verbatim up to replacing some calls to the standard case of Lemma A.6 with the critical case.

Proposition 5.10 does not however hold in its generality anymore, because  $\text{Occ}_\theta$  is not necessarily convergent at  $\frac{\tau}{1-\tau}$  for every  $\theta$ . It is nonetheless convergent when  $|\theta| = 2$ , since the singularity of  $\text{Occ}_\theta$  is in  $\delta - 2$ . This is enough to show that Proposition 5.10 still holds for binary trees. Moreover nonbinary trees still disappear in the limit according to Eq. (53). This allows us to conclude as in Section 5.  $\square$

## APPENDIX A. COMPLEX ANALYSIS TOOLBOX

**A.1. Aperiodicity and Daffodil Lemma.** To study the asymptotic behavior of combinatorial generating functions, it is important to locate dominant singularities. The following lemma is useful to this purpose.

Recall that a function  $A$  analytic at 0 is *aperiodic* if there do not exist two integers  $r \geq 0$  and  $d \geq 2$  and a function  $B$  analytic at 0 such that  $A(z) = z^r B(z^d)$ .

**Lemma A.1** (Daffodil Lemma). *Let  $A$  be a generating function (with nonnegative coefficients) analytic in  $|z| < R_A$ . If  $A$  is aperiodic, then  $|A(z)| < A(|z|) \leq A(R_A)$  for  $|z| \leq R_A$  and  $z \neq |z|$ . (The case  $|z| = R_A$  can only be considered if  $A(R_A) < \infty$ .)*

This lemma can be found in [26, Lemma IV.1, p. 266]. Note that this reference does not consider the case of  $z$  on the circle of convergence, *i.e.*  $|z| = R_A$  (although this case is used later in the book, *e.g.* in the proof of Theorem VI.6, p. 405); the proof of the lemma in this case is similar to  $|z| < R_A$ . The complete statement of Daffodil Lemma in [26] also deals with cases where the function  $A$  is periodic, but we do not need these cases in our work.

**A.2. Transfer theorem.** We start by defining the notion of  $\Delta$ -domain.

**Definition A.2** ( $\Delta$ -domain and  $\Delta$ -neighborhood). *A domain  $\Delta$  is a  $\Delta$ -domain at 1 if there exist two real numbers  $R > 1$  and  $0 < \phi < \frac{\pi}{2}$  such that*

$$\Delta = \{z \in \mathbb{C} \mid |z| < R, z \neq 1, |\arg(z - 1)| > \phi\}.$$

By extension, for a complex number  $\rho \neq 0$ , a domain is a  $\Delta$ -domain at  $\rho$  if it is the image by the mapping  $z \rightarrow \rho z$  of a  $\Delta$ -domain at 1. A  $\Delta$ -neighborhood of  $\rho$  is the intersection of a neighborhood of  $\rho$  and a  $\Delta$ -domain at  $\rho$ .

We will make use of the following family of  $\Delta$ -neighborhoods: for  $\rho \neq 0 \in \mathbb{C}$ ,  $0 < r < |\rho|$ ,  $\varphi < \pi/2$ , set  $\Delta(\varphi, r, \rho) = \{z \in \mathbb{C}, |\rho - z| < r, \arg(\rho - z) > \varphi\}$ .

When a function  $A$  is analytic on a  $\Delta$ -domain at some  $\rho$ , the asymptotic behavior of its coefficients is closely related to the behavior of the function near the singularity  $\rho$ . The following theorem is a corollary of [26, Theorem VI.3 p. 390].

**Theorem A.3** (Transfer Theorem). *Let  $A$  be a function analytic on a  $\Delta$ -domain  $\Delta$  at  $R_A$ ,  $\delta$  be an arbitrary real number in  $\mathbb{R} \setminus \mathbb{Z}_{\geq 0}$  and  $C_A$  a constant possibly equal to 0.*

*Suppose  $A(z) = (C_A + o(1))(1 - \frac{z}{R_A})^\delta$  when  $z$  tends to  $R_A$  in  $\Delta$ . Then the coefficient of  $z^n$  in  $A$  satisfies*

$$[z^n]A(z) = (C_A + o(1)) \frac{1}{R_A^n} \frac{n^{-(\delta+1)}}{\Gamma(-\delta)}.$$

**A.3. Singular differentiation.** The next result is also useful to us.

**Theorem A.4** (Singular differentiation). *Let  $A$  be an analytic function in a  $\Delta$ -neighborhood of  $R_A$  with the following singular expansion near its singularity  $R_A$*

$$A(z) = \sum_{j=0}^J C_j (R_A - z)^{\delta_j} + \mathcal{O}((R_A - z)^\delta),$$

where  $\delta_j, \delta \in \mathbb{C}$ .

*Then, for each  $k > 0$ , the  $k$ -th derivative  $A^{(k)}$  is analytic in some  $\Delta$ -domain at  $R_A$  and*

$$A^{(k)}(z) = (-1)^k \sum_{j=0}^J C_j \delta_j (\delta_j - 1) \cdots (\delta_j - k + 1) (R_A - z)^{\delta_j - k} + \mathcal{O}((R_A - z)^{\delta - k}).$$

We refer the reader to [26, Theorem VI.8 p. 419] for a proof of this theorem (this reference considers functions defined on a  $\Delta$ -domain, but the proof still works with functions defined on a  $\Delta$ -neighborhood).

**A.4. Exponents of dominant singularity.** In this section, we introduce some compact terminology and easy lemmas to keep track of the exponent  $\delta$  of the singularities and of the shape of the domain of analyticity without computing the functions explicitly.

Recall that the radius of convergence  $R_A$  of an analytic function  $A$  is the modulus of the singularities closest to the origin, called *dominant singularities*. Recall also that for series with positive real coefficients, by Pringsheim's theorem [26, Th. IV.6 p. 240],  $R_A$  is necessarily a dominant singularity. This justifies the following definition:

Let  $\delta$  be a real, which is not an integer. We say that a series  $A$  with radius of convergence  $R_A$  has a *dominant singularity of exponent  $\delta$  in  $R_A$*  (resp. *of exponent at least  $\delta$* ) if  $A$  has an analytic continuation on a  $\Delta$ -neighborhood  $\Delta_A$  of  $R_A$  and, on  $\Delta_A$ , we have

$$(54) \quad A(z) = g_A(z) + (C_A + o(1)) (R_A - z)^\delta,$$

where  $g_A(z)$  is an analytic function on a neighbourhood of  $R_A$  (called the *analytic part*), and  $C_A$  a nonzero constant (resp. any constant);  $(C_A + o(1))(R_A - z)^\delta$  is sometimes referred to as the *singular part*.

If furthermore,  $A$  has no other singularity on the disk of convergence, we say that it has a *unique* dominant singularity of exponent  $\delta$  (resp. at least  $\delta$ ) in  $R_A$ . Since we assumed that  $A$  has an analytic continuation on a  $\Delta$ -neighborhood  $\Delta_A$  of  $R_A$ , by a standard compactness argument, this is equivalent to say that  $A$  can be extended to a  $\Delta$ -domain in  $R_A$ .

We make the following observation: for  $\delta > 1$ , the above hypotheses (in any of their variant) imply that  $A$  itself and its derivative  $A'$  have finite limits  $A(R_A)$  and  $A'(R_A)$  when  $z$  tends to  $R_A$  from below and that, on a  $\Delta$ -neighborhood of  $R_A$ , we have

$$(55) \quad A(z) = A(R_A) + A'(R_A)(z - R_A) + \mathcal{O}((R_A - z)^{\delta_*}),$$

where  $\delta_* = \min(\delta, 2)$ .

We now record a few easy lemmas to manipulate these notions. First consider the stability by product.

**Lemma A.5.** *Let  $F$  and  $G$  be series with nonnegative coefficients and the same radius of convergence  $R = R_F = R_G \in (0, \infty)$ . Assume they have each a dominant singularity of exponent  $\delta_F$  and  $\delta_G$  respectively in  $R$ . Then  $F \cdot G$  has a dominant singularity in  $R$  of exponent  $\delta$  defined by*

- $\delta = \delta_F + \delta_G$  if both  $\delta_F$  and  $\delta_G$  are negative;
- $\delta = \min(\delta_F, \delta_G)$  otherwise.

Moreover, if both  $F$  and  $G$  have unique dominant singularities, so has  $F \cdot G$ .

*Proof.* The proof is easy. The analytic function  $F \cdot G$  can be extended to the intersection of the domain of  $F$  and  $G$ . The exponent of the singular expansion around  $R$  is obtained by multiplying singular expansion of  $F$  and  $G$ : note that, if  $\delta_F$  is negative, the series  $F$  is divergent and the singular part is the dominant part around  $R$ . On the opposite, when  $\delta_F$  is positive, the dominant part of the expansion is the value  $F(R)$  of the analytic part at point  $R$ , which is always positive, since the series has nonnegative coefficients. The same holds of course for  $G$ , which explains the case distinction in the lemma.  $\square$

We now consider the composition  $F \circ G$ . We should differentiate cases where  $G(R_G) > R_F$ ,  $G(R_G) < R_F$  or  $G(R_G) = R_F$  (called sometimes supercritical, subcritical and critical cases [26, Sec.VI.9]).

**Lemma A.6** (Dominant singularity of  $F \circ G$ ). *Let  $F$  and  $G$  be series with nonnegative coefficients with radii of convergence  $R_F, R_G$  in  $(0, \infty)$ .*

**Supercritical case:** *Assume that  $G(0) < R_F < G(R_G)$  ( $G(R_G)$  may be finite or infinite).*

*Call  $\rho$  the unique positive number with  $G(\rho) = R_F$ .*

*We assume that  $F$  has a dominant singularity of exponent  $\delta_F$  in  $R_F$ . Then:*

- i)  $F \circ G$  has also a dominant singularity of exponent  $\delta_F$  in  $\rho$ .
- ii) Moreover, if  $G$  is aperiodic, then the dominant singularity of  $F \circ G$  is unique.

**Subcritical case:** *Assume that  $G(R_G) < R_F$ .*

*We assume that  $G$  has a dominant singularity of exponent  $\delta_G$  in  $R_G$ . Then:*

- i)  $F \circ G$  has also a dominant singularity of exponent  $\delta_G$  in  $R_G$ .
- ii) Moreover, if the dominant singularity of  $G$  is unique, then the dominant singularity of  $F \circ G$  is unique.

**Critical case-A:** Assume that  $G(R_G) = R_F$ .

We assume that  $F$  and  $G$  both have a dominant singularity of respective exponents  $\delta_F$  and  $\delta_G$ . Suppose furthermore  $\delta_G > 1$ . Then:

- i)  $F \circ G$  has also a dominant singularity of exponent  $\min(\delta_G, \delta_F)$  in  $R_G$ .
- ii) Moreover, if  $G$  is aperiodic, then the dominant singularity of  $F \circ G$  is unique.

**Critical case-B:** Assume again that  $G(R_G) = R_F$ . We assume that  $F$  and  $G$  both have a dominant singularity of respective exponents  $\delta_F$  and  $\delta_G$ . Suppose furthermore  $\delta_G \in (0, 1)$ . Then:

- i)  $F \circ G$  has a dominant singularity of exponent at least  $\delta_F \delta_G$  in  $R_G$ .
- ii) Moreover if  $G$  is aperiodic then the singularity is unique.

**Proof. Supercritical case:** It is clear that  $F \circ G$  is analytic around any  $r \in [0, \rho)$  and has nonnegative coefficients, hence it has radius of convergence at least  $\rho$ .

To show that  $F \circ G$  is defined in a  $\Delta$ -neighborhood  $\Delta$  of  $\rho$ , we show that  $G(\Delta)$  is included in  $\Delta_F$ . This follows easily from the fact that  $G$  is analytic in  $\rho$  and has a derivative  $G'(\rho)$  which is a positive real number.

The fact that  $F \circ G$  has a dominant singularity of exponent  $\delta_F$  in  $\rho$  is immediate by composing the (singular) expansion of  $F$  around  $R_F$  and the (regular) expansion of  $G$  around  $\rho$ .

The last statement is also easy. In the case where we assume  $G$  aperiodic, we need Lemma A.1, which ensures that  $|G(\zeta)| < R_F$  for  $|\zeta| \leq |\rho|$ ,  $\zeta \neq \rho$ .

**Subcritical case.** Immediate: the dominant singularity is only induced by  $G$ .

**Critical cases.** The proof is similar to the supercritical case. □

Finally, we state the following result, which follows from Theorem A.4.

**Lemma A.7** (Singular differentiation). *If  $F$  has a (unique) dominant singularity of exponent (at least)  $\delta$  in  $\rho$ , then its  $k$ -th derivative  $F^{(k)}$  has a (unique) dominant singularity of exponent (at least)  $\delta - k$  in  $\rho$ .*

**A.5. An analytic implicit function theorem.** The following theorem allows to locate the dominant singularity of series defined by an implicit equation.

**Lemma A.8** (Analytic Implicit Functions). *Let  $F(z, w)$  be a bivariate function analytic at  $(z_0, w_0)$ . If  $F(z_0, w_0) = 0$  and, denoting  $F_w = \frac{\partial F}{\partial w}$ ,  $F_w(z_0, w_0) \neq 0$ , then there exists a unique function  $\phi(z)$  analytic in a neighbourhood of  $z_0$  such that  $\phi(z_0) = w_0$  and  $F(z, \phi(z)) = 0$ .*

We refer the reader to [26, Lemma VII.2, p. 469] for a proof of this result.

**A.6. Proof of Lemma 6.3.** Let  $R_{\text{not}\oplus}$  be the radius of convergence of  $T_{\text{not}\oplus}$ . If  $T_{\text{not}\oplus}(R_{\text{not}\oplus}) \geq R_\Lambda$ , then by intermediate value theorem, we know that there exists  $\rho$  as in the lemma.

We will prove this by contradiction. Assume  $T_{\text{not}\oplus}(R_{\text{not}\oplus}) < R_\Lambda$ . We apply Lemma A.8. The bivariate function we consider is  $(z, w) \mapsto z - w + \Lambda(w)$ . It vanishes at  $(R_{\text{not}\oplus}, T_{\text{not}\oplus}(R_{\text{not}\oplus}))$

and the derivative with respect to  $w$  at that point is nonzero since

$$\Lambda'(T_{\text{not}\oplus}(R_{\text{not}\oplus})) < \Lambda'(R_\Lambda) < 1.$$

Therefore,  $T_{\text{not}\oplus}$  has an analytic continuation on a neighborhood of  $R_{\text{not}\oplus}$ . Since it has positive coefficients, by Pringsheim's theorem [26, Th. IV.6 p. 240], this is in contradiction with the fact that  $R_{\text{not}\oplus}$  is the radius of convergence of  $T_{\text{not}\oplus}$ .

We have therefore proved that there exists  $\rho \leq R_{\text{not}\oplus}$  such that  $T_{\text{not}\oplus}(\rho) = R_\Lambda$ . Note that it implies the relation  $\rho = R_\Lambda - \Lambda(R_\Lambda)$ .

We now consider  $T_{\text{not}\oplus}$  around  $z = \rho$ . Equation (37) defining  $T_{\text{not}\oplus}(z)$  can be rewritten as  $T_{\text{not}\oplus}(z) = G(z, T_{\text{not}\oplus}(z))$ , where

$$G(z, w) = w + \frac{1}{1 - \Lambda'(R_\Lambda)}(z - w + \Lambda(w))$$

Since  $\Lambda$  has a dominant singularity of exponent  $\delta > 1$  in  $R_\Lambda$ , Equation (55), together with elementary computations, yield the following: for  $w$  in a  $\Delta$ -neighborhood  $D_\Lambda$  of  $R_\Lambda$ ,

$$(56) \quad G(z, w) = R_\Lambda + \frac{z - \rho}{1 - \Lambda'(R_\Lambda)} + \mathcal{O}((R_\Lambda - w)^{\delta_*}).$$

We now use Picard's method of successive approximants to show the existence and analyticity of  $T_{\text{not}\oplus}$  on a  $\Delta$ -neighborhood  $D_T$  of  $\rho$ . We refer to [26, Appendix B.5 p. 753] for a synthetic description of the method in the case where  $\Lambda$  is analytic in  $R_\Lambda$ ; we have to adapt it carefully to our setting.

Define  $\phi_0(z) = R_\Lambda$  and  $\phi_{j+1}(z) = G(z, \phi_j(z))$  whenever  $\phi_j(z)$  is in  $D_\Lambda$ . We have  $\phi_1(z) - \phi_0(z) = \frac{z - \rho}{1 - \Lambda'(R_\Lambda)}$ . Besides Theorem A.4 of singular differentiation applied to Eq. (56) implies that

$$\frac{\partial G(z, w)}{\partial w} = \mathcal{O}((R_\Lambda - w)^{\delta_* - 1}).$$

Therefore<sup>9</sup> when  $j \geq 1$  and when all relevant quantities are defined,

$$\phi_{j+1}(z) - \phi_j(z) = \mathcal{O}(\eta^{\delta_* - 1} |\phi_j(z) - \phi_{j-1}(z)|),$$

where  $\eta = \sup_{w \in D_\Lambda} |R_\Lambda - w|$ . Fix  $\varepsilon > 0$ . Up to reducing the radius of  $D_\Lambda$ , we can therefore assume that

$$(57) \quad |\phi_{j+1}(z) - \phi_j(z)| \leq \varepsilon |\phi_j(z) - \phi_{j-1}(z)|.$$

Thus, if  $\phi_M(z)$  is defined, we have

$$(58) \quad |(\phi_M(z) - R_\Lambda) - \frac{z - \rho}{1 - \Lambda'(R_\Lambda)}| = |\phi_M(z) - \phi_1(z)| \leq \frac{\varepsilon}{1 - \varepsilon} |\phi_1(z) - \phi_0(z)| = \frac{\varepsilon}{1 - \varepsilon} \left| \frac{z - \rho}{1 - \Lambda'(R_\Lambda)} \right|.$$

<sup>9</sup>There is a slight subtlety here: we would like to apply the classical inequality  $|f(w) - f(w')| \leq \|f'\|_\infty |w - w'|$ , but this is not possible since the domain  $D_\Lambda$  is not convex. Note however that a  $\Delta$ -neighborhood  $D$  is always a *quasi-convex set*, in the sense that we can always find a path between  $w$  and  $w'$  whose length is smaller than  $K|w - w'|$ , where  $K$  depends on the angle defining  $D$  but not on  $w$  and  $w'$ . Therefore the following weaker inequality holds:  $|f(w) - f(w')| \leq K\|f'\|_\infty |w - w'|$ , which is good enough for our purpose (the constant  $K$  disappears in the  $\mathcal{O}$  symbol).

If we take  $\varepsilon$  small enough, the argument of  $\phi_M(z) - R_\Lambda$  is close to the one of  $z - \rho$ . Furthermore if the modulus of  $z - \rho$  is small so is the one of  $\phi_M(z) - R_\Lambda$ . This ensures the existence of a  $\Delta$ -neighborhood  $D_T$  (not depending on  $M$  and  $z$ ), such that for  $z \in D_T$  and  $M \geq 1$ ,  $\phi_M(z)$  is in  $D_\Lambda$  as long as it is defined. In particular,  $\phi_{M+1}(z)$  is also defined and by immediate induction, all  $\phi_j$  are defined and analytic on  $D_T$ .

Eq. (57) also implies that  $\phi_j$  converges locally uniformly on  $D_T$ . The limit is the unique solution  $w$  in  $D_\Lambda$  of the fixed point equation  $w = G(z, w)$  (the uniqueness of the solution comes from the fact that for every  $z \in D_T$ ,  $w \mapsto G(z, w)$  is a contraction for  $w$  in  $D_\Lambda$ ). This limit is therefore an analytic continuation of  $T_{\text{not}\oplus}(z)$  to  $D_T$ . Note also that from Eq. (58), the following estimate holds on  $D_T$ :

$$T_{\text{not}\oplus}(z) - R_\Lambda = \frac{z - \rho}{1 - \Lambda'(R_\Lambda)} + o(|z - \rho|).$$

Using the expansion given in Eq. (54) of  $\Lambda$  around  $R_\Lambda$ , we have for  $z \in D_T$ ,

$$T_{\text{not}\oplus}(z) = \rho + (z - \rho) + g_\Lambda(T_{\text{not}\oplus}(z)) + (C_\Lambda + o(1)) (T_{\text{not}\oplus}(z) - R_\Lambda)^\delta.$$

As  $\text{id} - g_\Lambda$  is analytic at  $R_\Lambda$  with a nonzero derivative  $1 - \Lambda'(R_\Lambda)$ , it can be inverted analytically around  $R_\Lambda$  by an analytic function  $h_\Lambda$  and hence

$$T_{\text{not}\oplus}(z) = h_\Lambda \left( z + (C_\Lambda + o(1)) (T_{\text{not}\oplus}(z) - R_\Lambda)^\delta \right)$$

As  $T_{\text{not}\oplus}(z) - R_\Lambda = \frac{1+o(1)}{1-\Lambda'(R_\Lambda)}(z - \rho)$ , it follows from the Taylor expansion of  $h_\Lambda$  up to exponent  $\lceil \delta \rceil$  that  $T_{\text{not}\oplus}$  has a singularity of exponent exactly  $\delta$  in  $\rho$ . In particular  $T_{\text{not}\oplus}$  has a singularity of exponent  $\delta$  in  $\rho$  and hence  $\rho = R_{\text{not}\oplus}$ .

We now prove that  $T_{\text{not}\oplus}$  has no singularity  $\zeta$  with  $|\zeta| \leq \rho$ , except  $\zeta = \rho$ . By a classical compactness argument (see *e.g.* [24, end of proof of Theorem 2.19]), this implies that  $T_{\text{not}\oplus}$  is analytic on a  $\Delta$ -domain at  $\rho$ .

Take such a singularity. Since  $T_{\text{not}\oplus}$  has nonnegative coefficients, the triangular inequality gives  $|T_{\text{not}\oplus}(\zeta)| \leq T_{\text{not}\oplus}(\rho)$  and since  $T_{\text{not}\oplus}(z)$  has no periods, from Lemma A.1 we have a strict inequality unless  $\zeta = \rho$ . Therefore, if  $|\zeta| \leq \rho$  and  $\zeta \neq \rho$ , we have  $|\Lambda'(T_{\text{not}\oplus}(\zeta))| < \Lambda'(R_\Lambda) < 1$  and we can apply Lemma A.8 as above to argue that  $\zeta$  cannot be a singularity.  $\square$

**A.7. Proof of Lemma 7.3.** As in the proof of Lemma 6.3, the existence of  $\rho$  and the fact that the convergence radius of  $T_{\text{not}\oplus}$  is at least  $\rho$  is straightforward. The key point is to prove that  $T_{\text{not}\oplus}$  has an analytic continuation to a  $\Delta$ -neighborhood of  $\rho$ .

By assumption,  $\Lambda$  is analytic on a  $\Delta$ -neighborhood  $D_\Lambda = \Delta(\varphi_\Lambda, r_\Lambda, R_\Lambda)$  of  $R_\Lambda$ , and the following approximation holds:

$$\Lambda(w) = \Lambda(R_\Lambda) - (R_\Lambda - w) + C'_\Lambda (R_\Lambda - w)^{\delta_*} (1 + \varepsilon(w)),$$

where as before,  $\delta_* = \min(\delta, 2)$ ;  $C'_\Lambda$  is  $C_\Lambda$  or  $\Lambda''(R_\Lambda)$  depending on whether  $\delta$  is smaller or bigger than 2; and  $\varepsilon(w)$  is an analytic function on  $D_\Lambda$  tending to 0 in  $R_\Lambda$ .

Fix  $z$  in a  $\Delta$  neighborhood  $D_T$  of  $\rho$ , whose parameters  $r_T$  and  $\varphi_T$  will be precised later. The equation  $w = z + \Lambda(w)$  then rewrites as

$$(59) \quad \rho - z = C'_\Lambda (R_\Lambda - w)^{\delta_*} (1 + \varepsilon(w)),$$

or, as a fixed point equation  $w = G(z, w)$  for

$$G(z, w) := R_\Lambda - \left( \frac{1}{C'_\Lambda}(\rho - z) \cdot \frac{1}{1 + \varepsilon(w)} \right)^{1/\delta_*}.$$

We again use Picard's method of successive approximants to find an analytic solution  $w(z)$  for (59), which will be the analytic continuation of  $T_{\text{not}\oplus}(z)$  that we are looking for. For  $z \in D_T$ , set  $\phi_0(z) = R_\Lambda$  and, whenever  $\phi_i(z)$  lies in  $D_\Lambda \cup \{R_\Lambda\}$ , set  $\phi_{i+1}(z) = G(z, \phi_i(z))$ . In particular,

$$R_\Lambda - \phi_1(z) = \left( \frac{1}{C'_\Lambda}(\rho - z) \right)^{1/\delta_*}.$$

Since  $1/\delta_* < 1$ , we have  $\text{Arg}(R_\Lambda - \phi_1(z)) = \frac{1}{\delta_*} \text{Arg}(\rho - z)$ . Therefore if we choose the parameters defining the  $\Delta$ -neighborhood  $D_T$  to be  $\varphi_T = \varphi_\Lambda$  and  $r_T = (\frac{C'_\Lambda r_\Lambda}{2})^{\delta_*}$ , then  $\phi_1(z)$  lives in a domain  $\widetilde{D}_\Lambda = \Delta(\widetilde{\varphi}_\Lambda, \frac{r_\Lambda}{2}, R_\Lambda)$ , with  $\widetilde{\varphi}_\Lambda > \varphi_\Lambda$ .

We define an intermediate  $\Delta$ -neighborhood  $D'_\Lambda = \Delta(\frac{\varphi_\Lambda + \widetilde{\varphi}_\Lambda}{2}, \frac{3r_\Lambda}{4}, R_\Lambda)$ . This ensures that we have a constant  $0 < r_0 < 1$ , depending only on  $\varphi_\Lambda$  and  $\widetilde{\varphi}_\Lambda$ , such that the circle  $\gamma_w$  of center  $w$  and radius  $r_0 |R_\Lambda - w|$  is contained in  $D_\Lambda$  for every  $w \in D'_\Lambda$  and in  $\widetilde{D}_\Lambda$  for every  $w \in \widetilde{D}_\Lambda$  (cf. Fig. 15).

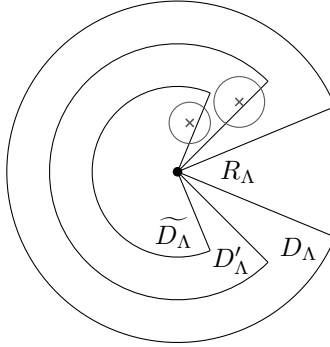


FIGURE 15. Illustration of  $D_\Lambda, D'_\Lambda, \widetilde{D}_\Lambda$ , with two exemples of circles  $\gamma_w$  represented in gray.

Consider the partial derivative

$$(60) \quad \frac{\partial G}{\partial w}(z, w) = \frac{\varepsilon'(w)}{\delta_*} \cdot \left( \frac{1}{C'_\Lambda}(\rho - z) \right)^{1/\delta_*} \cdot \left( \frac{1}{1 + \varepsilon(w)} \right)^{1/\delta_* - 1}.$$

We take  $w$  in the domain  $D'_\Lambda$ . The quantity  $\varepsilon'(w)$  can now be evaluated through a contour integral on  $\gamma_w \subset D_\Lambda$ :

$$\varepsilon'(w) = \frac{1}{2\pi i} \oint_{\gamma_w} \frac{\varepsilon(u) du}{(u - w)^2}.$$

This yields the inequality

$$|\varepsilon'(w)| = \mathcal{O} \left( \frac{\sup_{u \in D_\Lambda} |\varepsilon(u)|}{|R_\Lambda - w|} \right).$$

Plugging this back in Eq. (60), we get, for  $w$  in  $D'_\Lambda$

$$(61) \quad \left| \frac{\partial G}{\partial w}(z, w) \right| = \mathcal{O} \left( \frac{|z - \rho|^{1/\delta_*} \cdot \sup_{u \in D_\Lambda} |\varepsilon(u)|}{|R_\Lambda - w|} \right).$$

Now we shall find a domain where we have enough control on  $|\frac{\partial G}{\partial w}(z, w)|$  as to guarantee the stability of the iterates. A subtlety here is that this control is impossible near  $\phi_0(z) = R_\Lambda$ . So we need to consider a domain around  $\phi_1(z)$ , hence that depends on  $z$ . For every  $z \in D_T$ , we have  $\phi_1(z) \in \widetilde{D_\Lambda}$ , so the disk

$$\Gamma_z := \{w : |w - \phi_1(z)| \leq \frac{1}{r_0} |\phi_1(z) - R_\Lambda|\}$$

is included in  $D'_\Lambda$ . For  $w$  in  $\Gamma_z$ , we have

$$|R_\Lambda - w| = \Theta(|\phi_1(z) - R_\Lambda|) = \Theta(|\rho - z|^{1/\delta_*}),$$

which implies after plugging back into Eq. (61)

$$\left| \frac{\partial G}{\partial w}(z, w) \right| = \mathcal{O} \left( \sup_{u \in D_\Lambda} |\varepsilon(u)| \right).$$

By possibly reducing the radius  $r_\Lambda$  of  $D_\Lambda$ , we can make  $\sup_{u \in D_\Lambda} |\varepsilon(u)|$  as small as wanted: for any  $w$  in  $\Gamma$ ,

$$(62) \quad \left| \frac{\partial G}{\partial w}(z, w) \right| \leq \frac{1}{r_0+1}.$$

Similarly,

$$|\phi_2(z) - \phi_1(z)| = \left( \frac{1}{C'_\Lambda} |\rho - z| \right)^{1/\delta_*} \cdot \left| \left( \frac{1}{1 + \varepsilon(\phi_1(z))} \right)^{1/\delta_*} - 1 \right|$$

can be made smaller than  $\frac{1}{r_0+1} |\phi_1(z) - R_\Lambda|$  by reducing  $r_\Lambda$ . In particular,  $\phi_2(z)$  is in  $\Gamma_z$ .

For  $m \geq 2$ , assume that  $\phi_1(z), \dots, \phi_m(z)$  lie in  $\Gamma_z$ . Then for each  $i \leq m$ , using (62),

$$(63) \quad |\phi_{i+1}(z) - \phi_i(z)| \leq \left( \frac{1}{r_0+1} \right) |\phi_i(z) - \phi_{i-1}(z)| \leq \dots \leq \left( \frac{1}{r_0+1} \right)^{i-1} |\phi_2(z) - \phi_1(z)|.$$

Since  $\phi_m(z)$  lies in  $\Gamma_z \subset D_\Lambda$ , the next term  $\phi_{m+1}(z)$  is defined and

$$|\phi_{m+1}(z) - \phi_1(z)| \leq \sum_{i=1}^m |\phi_{i+1}(z) - \phi_i(z)| \leq \left[ \sum_{i=1}^m \left( \frac{1}{r_0+1} \right)^{i-1} \right] |\phi_2(z) - \phi_1(z)| \leq \frac{1}{r_0} |\phi_1(z) - R_\Lambda|.$$

In particular,  $\phi_{m+1}(z)$  also lies in  $\Gamma_z$  and an immediate induction shows that this is indeed the case for all  $m \geq 1$ .

By (63), the series  $\sum_{i \geq 0} \phi_{i+1}(z) - \phi_i(z)$  is uniformly bounded by a geometric series and converges towards an analytic function  $\phi$  on  $D_T$ . The limit  $\phi(z)$  is a solution of  $\phi(z) = z + \Lambda(\phi(z))$  and is the analytic continuation of  $T_{\text{not } \oplus}(z)$  that we were looking for.

A small modification of the above argument shows that, when  $r_T$ , or equivalently  $r_\Lambda$ , tends to 0, the quotient

$$\frac{|\phi_{m+1}(z) - \phi_1(z)|}{|\phi_1(z) - R_\Lambda|}$$



also tends to 0. This proves that

$$(64) \quad \phi(z) - R_\Lambda = (\phi_1(z) - R_\Lambda)(1 + o(1)) = - \left( \frac{1}{C_\Lambda}(\rho - z) \right)^{1/\delta_*} (1 + o(1)).$$

The proof that  $T_{\text{not}\oplus}(z)$  has no other singularities than  $\rho$  on the circle of convergence is similar to that of Lemma 6.3.  $\square$

## APPENDIX B. ON THE SIMULATIONS GIVEN IN THE INTRODUCTION

In this appendix, we explain how the simulations in Figs. 1 and 3 have been obtained.

**B.1. Biased Brownian permuton.** Fix  $p$  in  $(0, 1)$  and consider a uniform binary planar tree  $\mathbf{b}_n^{(p)}$  with  $n$  leaves, where each internal node is labeled  $\oplus$  (resp.  $\ominus$ ) with probability  $p$  (resp.  $1 - p$ ), independently from each other. As mentioned in Section 5.1,  $\tau_n^{(p)} := \text{perm}(\mathbf{b}_n^{(p)})$  forms a consistent family of random permutations. Therefore, from Proposition 2.9, we have the following lemma (using the notation  $\text{Perm}(\cdot, \cdot)$  defined in Section 2.1)

**Lemma B.1.** *There exist a random permuton  $\mu^{(p)}$ , whose marginals are the  $\tau_n^{(p)}$  (i.e. for all  $n$ ,  $\tau_n^{(p)} \stackrel{d}{=} \text{Perm}(\vec{\mathbf{m}}_n, \mu^{(p)})$ ) and we have the convergence in distribution*

$$\mu_{\tau_n^{(p)}} \xrightarrow{n \rightarrow +\infty} \mu^{(p)}.$$

By definition,  $\mu^{(p)}$  is the biased Brownian separable permuton with parameter  $p$  (indeed  $\tau_k^{(p)} \stackrel{d}{=} \text{Perm}(\vec{\mathbf{m}}_k, \mu^{(p)})$  implies Eq. (20) of Definition 5.1). This lemma is not needed to prove the results of this paper, but was used in the simulations. The three pictures in Fig. 1 p. 5 are obtained by drawing the diagram of a random permutation distributed as  $\tau_n^{(p)}$ , for  $p = 0.2$  and  $n = 11\,629$ ,  $p = 0.45$  and  $n = 12\,666$ ,  $p = 0.5$  and  $n = 17\,705$ .

**B.2. Stable permutons.** Fix  $\delta \in (1, 2)$ . For every  $k$  the following probability distribution on unlabeled plane trees with  $k$  leaves was introduced in [25, Thm 3.3.3] and is called the  $k$ -th marginal of the  $\delta$ -stable tree.

$$\nu_{\delta,k}(t_0) = \frac{k!}{(\delta - 1) \cdots ((k - 1)\delta - 1)} \prod_{v \in \text{Int}(t_0)} \mathbf{1}_{d_v \geq 2} \frac{(d_v - 1 - \delta) \cdots (2 - \delta)(\delta - 1)}{d_v!}.$$

Now if we fix the distribution of a random permuton  $\nu$ , for every  $n \geq 1$ , we build a random substitution tree  $\mathbf{t}_n^{(\delta, \nu)}$  as such: the tree is chosen according to  $\nu_{\delta,n}$ , and conditional on that choice, all internal nodes  $v$  are independently labeled by a permutation distributed like  $\text{Perm}(\vec{\mathbf{m}}_{d_v}, \nu)$  (the notation  $\text{Perm}(\cdot, \cdot)$  is defined in Section 2.1).

Now we can define the permutations  $\tau_n^{(\delta, \nu)} = \text{perm}(\mathbf{t}_n^{(\delta, \nu)})$ . This family of permutations is consistent: we omit the proof of this fact, which follows from the consistency of the family  $(\text{Perm}(\vec{\mathbf{m}}_k, \nu))_k$  (Proposition 2.9) and from Marchal's algorithm [39] to generate trees of distribution  $\nu_{\delta,k}$ . We deduce the following lemma.

**Lemma B.2.** *For every  $\delta \in (1, 2)$  and random permuton  $\nu$ , there exists a random permuton  $\mu^{(\delta, \nu)}$ , whose marginals are the  $\tau_n^{(\delta, \nu)}$  ( i.e. for all  $n$ ,  $\tau_n^{(\delta, \nu)} \stackrel{d}{=} \text{Perm}(\vec{m}_n, \mu^{(\delta, \nu)})$ ) and we have the convergence in distribution*

$$\mu_{\tau_n^{(\delta, \nu)}} \xrightarrow{n \rightarrow +\infty} \mu^{(\delta, \nu)}.$$

We call  $\mu^{(\delta, \nu)}$  the  $\delta$ -stable permuton driven by  $\nu$ .

Once again, this lemma is used in our simulations. The pictures in Fig. 3 are the rescaled diagrams of realizations of  $\tau_n^{(\delta, \nu)}$  for  $n = 20\,000$  and  $\delta \in \{1.1, 1.5\}$ , with  $\nu$  the (nonrandom) uniform measure on  $[0, 1]^2$ .

## REFERENCES

- [1] M. H. Albert, M. D. Atkinson. Simple permutations and pattern restricted permutations. *Discrete Mathematics*, 300(1):1–15, 2005.
- [2] M. H. Albert, M. D. Atkinson, R. Brignall. The enumeration of three pattern classes using monotone grid classes. *Electronic Journal of Combinatorics*, 19(3) #P20, 2012.
- [3] M. H. Albert, M. D. Atkinson, R. Brignall. The enumeration of permutations avoiding 2143 and 4231. *Pure Mathematics and Applications*, 22(2):87–98, 2011.
- [4] M.H.Albert, M.D. Atkinson, M. Klazar. The enumeration of simple permutations. *Journal of Integer Sequences* vol.6 (2003), article 03.4.4.
- [5] M. H. Albert, M. D. Atkinson, V. Vatter. Inflations of geometric grid classes: three case studies. *Australasian Journal of Combinatorics*, vol. 58 (2014), p. 27–47.
- [6] M. H. Albert, M. D. Atkinson, V. Vatter. Counting 1324, 4231-Avoiding Permutations. *Electronic Journal of Combinatorics*, 16(1) #R136, 2009.
- [7] M. H. Albert, R. Brignall. Enumerating indices of Schubert varieties defined by inclusions. *Journal of Combinatorial Theory, Series A*, 123(1): 154–168, 2014.
- [8] M. H. Albert, V. Vatter. Generating and Enumerating 321-Avoiding and Skew-Merged Simple Permutations. *Electronic Journal of Combinatorics*, 20(2) #P44, 2013.
- [9] D. Aldous. The Continuum Random Tree III. *Annals of Probability*, 21(1): 248—289, 1993.
- [10] M. Atapour, N. Madras. Large deviations and ratio limit theorems for pattern-avoiding permutations. *Combinatorics, Probability and Computing*, 23(2): 160–200, 2014.
- [11] F. Bassino, M. Bouvel, V. Féray, L. Gerin, A. Pierrot. The Brownian limit of separable permutations, *Annals of Probability*, arXiv:1602.04960, to appear.
- [12] F. Bassino, M. Bouvel, A. Pierrot, D. Rossin. An algorithm for deciding the finiteness of the number of simple permutations in permutation classes. *Advances in Applied Mathematics*, vol. 64 (2015), p. 124–200.
- [13] F. Bassino, M. Bouvel, D. Rossin. Enumeration of Pin-Permutations. *Electronic Journal of Combinatorics*, vol. 18 (2011), #P57.
- [14] D. Bevan. *On the growth of permutation classes*. PhD thesis (2015), Open University, arXiv:1506.06688.
- [15] P. Billingsley. *Convergence of probability measures* (2d edition). John Wiley & Sons (1999).
- [16] M. Bóna. The absence of a pattern and the occurrences of another. *Discrete Mathematics & Theoretical Computer Science*, 12(2): 89–102, 2010.
- [17] M. Bóna. Surprising Symmetries in Objects Counted by Catalan Numbers. *The Electronic Journal of Combinatorics*, 19(1) #62, 2012.
- [18] M. Bóna. *Combinatorics of permutations* (2d edition). Chapman-Hall and CRC Press (2012).
- [19] M. Bouvel, M. Mishna, C. Nicaud. Some families of trees arising in permutation analysis. Preprint, arXiv:1609.09586, (2016).
- [20] R. Brignall. A survey of simple permutations. S. Linton, N. Ruškuc and V. Vatter, Eds., vol. 376 of London Mathematical Society Lecture Note Series, Cambridge University Press, p. 41–65.

- [21] R. Brignall, N. Ruškuc, V. Vatter. Simple permutations: decidability and unavoidable substructures. *Theoretical Computer Science*, 391(1-2): 150-163, 2008.
- [22] S.-E. Cheng, S.-P. Eu, T.-S. Fu, Area of Catalan paths on a checkerboard. *European Journal of Combinatorics*, 28(4): 1331-1344, 2007.
- [23] T. Dokos, I. Pak. The expected shape of random doubly alternating Baxter permutations. *Online Journal of Analytic Combinatorics*, vol. 9 (2014), #5.
- [24] M. Drmota, *Random Trees, An Interplay between Combinatorics and Probability*, Springer (2009).
- [25] T. Duquesne, J. F. Le Gall, Random trees, Lévy processes and spatial branching processes, vol. 281 of Astérisque, Société mathématique de France (2002).
- [26] Ph. Flajolet, R. Sedgewick. *Analytic combinatorics*. Cambridge University Press (2009).
- [27] C. Hoffman, D. Rizzolo, E. Slivken. Pattern Avoiding Permutations and Brownian Excursion Part I: Shapes and Fluctuations. *Random Structures and Algorithms*, 50(3): 394-419, 2017.
- [28] C. Hoffman, D. Rizzolo, E. Slivken. Pattern Avoiding Permutations and Brownian Excursion Part II: Fixed Points. *Probability Theory and Related Fields*, p. 1-48, 2016, Springer.
- [29] C. Homberger, Expected patterns in permutation classes. *The Electronic Journal of Combinatorics*, 19:3 #43, 2012.
- [30] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Rath, R. M. Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory, Series B*, 103(1): 93-113, 2013.
- [31] S. Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probability Surveys*, vol. 9 (2012), p. 103-252.
- [32] S. Janson. Patterns in random permutations avoiding the pattern 132. *Combinatorics Probability and Computing*, 26(1): 24-51, 2017.
- [33] S. Janson, B. Nakamura, D. Zeilberger. On the Asymptotic Statistics of the Number of Occurrences of Multiple Permutation Patterns. *Journal of Combinatorics*, 6(1-2): 117-143, 2015.
- [34] I. Kortchemski. Invariance principles for Galton-Watson trees conditioned on the number of leaves. *Stochastic Processes and Applications*, 122(9): 3126-3172, 2012.
- [35] M. Maazoun, On the Brownian separable permuton. In preparation.
- [36] N. Madras, H. Liu. Random pattern-avoiding permutations. *Algorithmic, Probability and Combinatorics*, vol. 520 of Contemp. Math., p. 173-194. Amer. Math. Soc., 2010.
- [37] N. Madras, L. Pehlivan. Structure of Random 312-Avoiding Permutations. *Random Structures and Algorithms*, 49(3): 599-631, 2016.
- [38] A. Marcus and G. Tardos. Excluded permutation matrices and the Stanley-Wilf conjecture. *Journal of Combinatorial Theory, Series A*, 107(1): 153-160, 2004.
- [39] P. Marchal. A note on the fragmentation of a stable tree. In *Fifth Colloquium on Mathematics and Computer Science*, DMTCS Proceedings (2008), p. 489-500.
- [40] S. Miner, I. Pak. The shape of random pattern-avoiding permutations. *Advances in Applied Mathematics*, vol. 55 (2014), p. 86-130.
- [41] J. Pantone. The Enumeration of Permutations Avoiding 3124 and 4312. *Annals of Combinatorics*, 21(2): 293-315, 2017.
- [42] J. Pitman, D. Rizzolo. Schröder's problems and scaling limits of random trees. *Transactions of the American Mathematical Society*, 367(10): 6943-6969, 2015.
- [43] J.-L. Rémy. Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. Théor.*, 19(2): 179-195, 1985.
- [44] K. Rudolph. Pattern popularity in 132-avoiding permutations. *The Electronic Journal of Combinatorics*, 20:1 #8, 1985.

(FB) UNIVERSITÉ PARIS 13, SORBONNE PARIS CITÉ, LIPN, CNRS UMR 7030, F-93430 VILLETANEUSE, FRANCE

*E-mail address:* bassino@lipn.univ-paris13.fr

(MB,VF) INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH, WINTERTHURERSTR. 190, CH-8057 ZÜRICH, SWITZERLAND

*E-mail address:* mathilde.bouvel@math.uzh.ch

*E-mail address:* valentin.feray@math.uzh.ch

(LG) CMAP, ÉCOLE POLYTECHNIQUE, CNRS, ROUTE DE SACLAY, F-91128 PALAISEAU CEDEX, FRANCE

*E-mail address:* gerin@cmap.polytechnique.fr

(MM) ÉCOLE NORMALE SUPÉRIEURE DE LYON, UMPA UMR 5669 CNRS, 46 ALLÉE D'ITALIE, F-69364 LYON CEDEX 07, FRANCE

*E-mail address:* mickael.maazoun@ens-lyon.fr

(AP) LRI, UNIVERSITÉ PARIS-SUD, BAT. 650 ADA LOVELACE, F-91405 ORSAY CEDEX, FRANCE

*E-mail address:* adeline.pierrot@lri.fr