Binary Search Trees of Permuton Samples

Benoît Corsini
Department of Mathematics and Computer Science,
Eindhoven University of Technology, The Netherlands

Victor Dubach
Université de Lorraine, CNRS, IECL, F-54000 Nancy, France

Valentin Féray
Université de Lorraine, CNRS, IECL, F-54000 Nancy, France

Abstract

Binary search trees (BST) are a popular type of structure when dealing with ordered data. They allow efficient access and modification of data, with their height corresponding to the worst retrieval time. From a probabilistic point of view, BSTs associated with data arriving in a uniform random order are well understood, but less is known when the input is a non-uniform permutation.

We consider here the case where the input comes from i.i.d. random points in the plane with law μ, a model which we refer to as a *permuton sample*. Our results show that the asymptotic proportion of nodes in each subtree only depends on the behavior of the measure μ at its left boundary, while the height of the BST has a universal asymptotic behavior for a large family of measures μ. Our approach involves a mix of combinatorial and probabilistic tools, namely combinatorial properties of binary search trees, coupling arguments, and deviation estimates.

2012 ACM Subject Classification Mathematics of computing → Trees; Mathematics of computing → Probabilistic algorithms

Keywords and phrases Binary search trees, random permutations, permutons

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.21

Related Version *Full Version*: `https://arxiv.org/abs/2403.03151` [8]

Funding *Benoît Corsini*: has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 101034253.
Valentin Féray: partially supported by the Future Leader Program of the LUE initiative.

Acknowledgements The authors are grateful to Mathilde Bouvel for pointing out the preprint [13] and for several stimulating discussions on the topic.

1 Introduction

1.1 Context and informal description of our results

A *binary search tree* (BST) is a rooted binary tree where nodes carry labels – which are real numbers – and where, for each vertex v, all labels of vertices in the left-subtree (resp. right-subtree) attached to v are smaller (resp. bigger) than that of v. Binary search trees are a popular type of data structure for storing ordered data. One key feature is that the worst-case complexity of basic operations (lookup, addition or removal of data) is proportional to the height of the tree.

Given a BST $\mathcal{T}$ and a real number x distinct from the labels of $\mathcal{T}$, there is a unique way to insert x into $\mathcal{T}$, i.e. there is a unique BST $\mathcal{T}^{+x}$ obtained from $\mathcal{T}$ by adding a new node with label x. Iterating this operation starting from the empty tree and a sequence $y = (y_1, \dots, y_n)$ of distinct values, we get a BST $\mathcal{T}\langle y\rangle$ with n nodes. An example can be found in Figure 1. The shape of $\mathcal{T}\langle y\rangle$ (i.e. the underlying binary tree without node labels)

35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 21; pp. 21:1–21:13
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

depends only on the relative order of the numbers $y_1, \dots, y_n$, and not on their actual values. We can thus assume without loss of generality that the sequence y is a permutation σ of the integers from 1 to n, and we write $\mathcal{T}\langle\sigma\rangle = \mathcal{T}\langle\sigma_1, \dots, \sigma_n\rangle$ in this case.

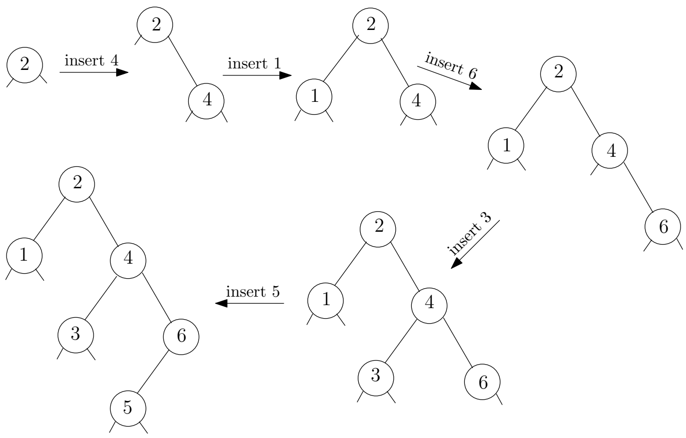

Figure 1 Iterative construction of the BST associated with the sequence $y = (2, 4, 1, 6, 3, 5)$.

In the worst case, the tree $\mathcal{T}\langle\sigma\rangle$ has height $n-1$ and further operations will have a linear complexity, which is far from optimal. However it has been proven by Devroye [10] that, if σ is a uniformly random permutation of $\{1, \dots, n\}$, then the height $h\left(\mathcal{T}\langle\sigma\rangle\right)$ is asymptotically equivalent to $c^* \log n$ for some constant c^*. Assuming that σ is uniformly distributed means that the data used to construct our BST arrived in a completely random order, which is in general unrealistic. It seems therefore natural to study BSTs associated with non-uniform random permutations, and in particular to see how Devroye's result is modified when changing the distribution of σ.

A first step in this direction has been performed in the papers [1, 7], where the BSTs associated with random Mallows and record-biased permutations are studied, showing interesting phase transition phenomena. In the current paper, we will consider some geometric models of random permutations, sampled via i.i.d. random points in the plane with some common distribution μ. These models will be referred to here as *permuton samples*, and denoted by σ_μ^n; they appear naturally in a recently developed theory of limiting objects for large permutations, called permutons [14]. The goal of studying such models is twofold. First, it is a much larger but still tractable family of models than those considered before (permuton samples are indexed by probability measures on the square, while Mallows and record-biased permutations are one-parameter families of models). Second, since permutons describe the "large-scale shape" of permutations, it enlightens the connection between this "large-scale shape" and the associated BST.

Our first result (Theorem 1) shows that, for a large family of permuton samples, the asymptotic behavior of the BST height is the same as the one found by Devroye for uniform permutations, namely that $h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right)$ is asymptotically equivalent to $c^* \log n$. Our second result (Theorem 13) studies another type of limit for the sequence of BSTs, using the formalism of *subtree size convergence* recently introduced by Grübel in [13]. In this setting and under some mild assumption, we prove convergence of the BST associated with permuton samples, where the limit object depends on the permuton only through its "derivative" at the left edge $\{0\} \times [0,1]$ of the unit square $[0,1]^2$.

In the remaining part of the introduction, we present the model of permuton samples and introduce some notation. Our main results are then stated and proved in Sections 2 and 3, and extra results are discussed in Section 4.

1.2 Our model: binary search trees of permuton samples

There is a natural way to map a (generic) finite set of points $\mathcal{P} \subset \mathbb{R}^2$ to a permutation $\sigma\langle\mathcal{P}\rangle$ and a binary search tree $\mathcal{T}\langle\mathcal{P}\rangle$, which we describe now. Let $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be a set of points in $\mathbb{R}^2$ with distinct x- and distinct y-coordinates, and let $\{(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})\}$ be its reordering such that $x_{(1)} < \ldots < x_{(n)}$. Then there exists a unique permutation $\sigma = \sigma\langle\mathcal{P}\rangle$ of $\{1, \ldots, n\}$ such that $(y_{(1)}, \ldots, y_{(n)})$ and $(\sigma_1, \ldots, \sigma_n)$ are in the same relative order. We let $\mathcal{T}\langle\mathcal{P}\rangle := \mathcal{T}\langle y_{(1)}, \ldots, y_{(n)}\rangle$ and note that the trees $\mathcal{T}\langle\mathcal{P}\rangle$ and $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$ have the same shape since their underlying data have the same relative order. These constructions are illustrated in Figure 2.

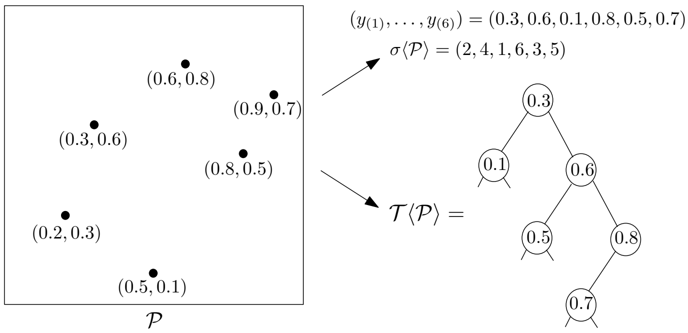

Figure 2 A set of points in $\mathbb{R}^2$ and its associated permutation and binary search tree.

Now consider a probability measure μ on $\mathbb{R}^2$ and take a set $\mathcal{P}^n_\mu$ of n i.i.d. points in $\mathbb{R}^2$ with distribution μ. In order to make sure that the associated permutation and BST are well-defined, we need the coordinates of the points to be all distinct. To this extent, we assume for the rest of this work that the projections of μ on both axes have no atom. Moreover, since the permutation and the shape of the tree only depend on the relative positions of the points, without loss of generality we can re-scale μ so that its support is in $[0, 1]^2$ and both its marginals are uniform (see [5, Remark 1.2] for details). Such measures are called *permutons*, and are natural limit objects for large permutations (see e.g. [2, 14]). The associated model of random permutations $\sigma\langle\mathcal{P}^n_\mu\rangle$ will then simply be denoted by σ^n_μ. This is a broad generalization of the uniform measure on permutations of size n, which corresponds to $\mu = \mathrm{Leb}_{[0,1]^2}$. Such models have been considered in the literature under various perspectives, see e.g. [5, 9, 11, 12, 15].

In the current paper, we are interested in the binary search tree $\mathcal{T}\langle\sigma^n_\mu\rangle$ of this random permutation model. Since we will be interested only in the shape of this tree (height in Section 2, subtree size convergence in Section 3), we may and will equivalently consider the tree $\mathcal{T}\langle\mathcal{P}^n_\mu\rangle$ instead of $\mathcal{T}\langle\sigma^n_\mu\rangle$. Furthermore, for convenience, we shall work with a Poisson point process $\mathcal{P}^N_\mu$ with intensity $n\mu$, instead of the point process $\mathcal{P}^n_\mu$. This new process has random size $N \sim \textsc{Poisson}\,(n)$, and conditionally given N it contains i.i.d. points distributed under μ. This enables useful independence properties, which make the proofs of our results easier. In the full paper [8], we explain in great detail how to "de-Poissonize" our results.

1.3 Some probabilistic notation

Throughout this paper, "with high probability" (w.h.p.) means "with probability tending to 1, as n tends to ∞". We also use the notation $X_n = o_{\mathbb{P}}(Y_n)$ to say that X_n/Y_n converges to 0 in probability, and we write $X \preceq Y$ (resp. $X \succeq Y$) to denote that X is stochastically smaller (resp. larger) than Y.

2 First main result: universal behavior of the BST height

2.1 Statement of the result and proof strategy

We denote by $h(\mathcal{T})$ the height of a tree $\mathcal{T}$, i.e. the maximal distance from a leaf to the root. As mentioned in Section 1.1, Devroye [10] proved that for uniformly random permutations σ^n of size n, the quantity $h\left(\mathcal{T}\langle\sigma^n\rangle\right) / \log n$ converges in probability and in L^p (for all $p \geq 1$) to a constant c^*, defined as the unique solution to $c \log(2e/c) = 1$ with $c \geq 2$. We provide a sufficient condition on a permuton μ, under which the same result holds for $h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right)$. In the following, a permuton μ is said to satisfy assumption (A1) if μ has a bounded density ρ on the unit square $[0,1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0,1]$.

▶ **Theorem 1** (Universality of BST height for permuton samples). *Let μ be a permuton satisfying assumption (A1), and let $\mathcal{P}_\mu^N$ be a Poisson point process with intensity $n\mu$. Then, as $n \to \infty$, the following convergence holds in probability and in L^p for all $p \geq 1$:*

$$\frac{h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right)}{c^* \log n} \longrightarrow 1\,.$$

Let us briefly overview the proof strategy of Theorem 1. We shall decompose the BST drawn from a permuton sample as a top tree, to which hanging trees are attached. To this end, consider $\beta \in (0,1)$ and set $\mathcal{P}(\beta) := \mathcal{P}_\mu^N \cap ([0,\beta] \times [0,1])$. Then set $K_\beta := |\mathcal{P}(\beta)|$ and let $y_{(1)} < \cdots < y_{(K_\beta)}$ be the ordered y-coordinates of the points in $\mathcal{P}(\beta)$. For each $0 \leq k \leq K_\beta$, define $I_k = (y_{(k)}, y_{(k+1)})$ with the convention $y_{(0)} = 0$ and $y_{(K_\beta+1)} = 1$. Finally, for each k, define $\mathcal{P}_k(\beta) := \mathcal{P}_\mu^N \cap \left((\beta,1] \times I_k\right)$. We call $\mathcal{T}\langle\mathcal{P}(\beta)\rangle$ and $\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right)_{0 \leq k \leq K_\beta}$ respectively the *top tree* and the *hanging trees* of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$. One can see that the top and hanging trees are indeed subtrees of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$. Furthermore, the entire tree can be reconstructed by grafting the hanging trees to some nodes of the top tree. In particular, this yields the following lemma:

▶ **Lemma 2.** *For any $\beta \in (0,1)$:*

$$h\left(\mathcal{T}\langle\mathcal{P}(\beta)\rangle\right) \leq h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right) \leq h\left(\mathcal{T}\langle\mathcal{P}(\beta)\rangle\right) + 1 + \max_{0 \leq k \leq K_\beta} \left\{ h\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right) \right\}.$$

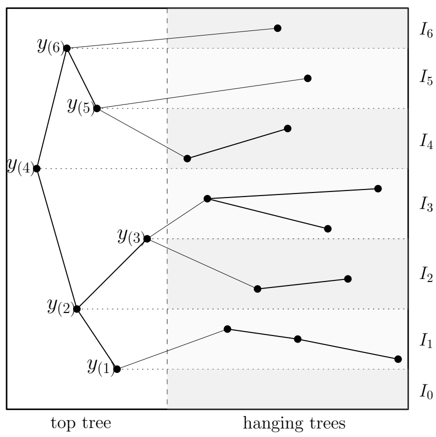

Figure 3 A sample of points and its associated BST, decomposed as top and hanging trees. The BST has been rotated of 90 degrees to the left, so that it can be drawn directly on the set of points.

See Figure 3 for an illustration. Thus, controlling the height of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$ amounts to controlling the heights of its top and hanging trees. This is done via different approaches: in Sections 2.2 and 2.3 we prove that the top tree has height $(c^* + o_{\mathbb{P}}(1))\log n$ for well chosen β, and in Sections 2.4 and 2.5 we prove that the hanging trees all have height $o_{\mathbb{P}}(\log n)$. Finally, we combine these estimates in Section 2.6 to conclude the proof of Theorem 1.

2.2 Height modification by adding/removing points

We rely on comparison arguments to prove our results: the basic idea is to (locally) compare the density of our permuton to a constant density, for which we can apply Devroye's result. However, while Poisson point processes possess nice monotonicity properties with respect to their intensities, BSTs are much trickier to handle. Indeed, one can see that adding a single point to a point set may halve the height of the associated BST. In this section, we develop adequate tools for such comparison arguments.

We start with a simple lemma about genealogies in a BST, easily derived by construction.

▶ **Lemma 3.** *Let $y = (y_1, \dots, y_n)$ be a list of distinct numbers and $\mathcal{T} = \mathcal{T}\langle y\rangle$ be the associated BST. If $i < j$ are two indices then the following are equivalent:*

- *y_i is an ancestor of y_j in $\mathcal{T}$ (the converse cannot hold);*
- *there is no $k < i$ such that y_k is between y_i and y_j, i.e. such that $(y_i - y_k)(y_j - y_k) < 0$.*

A *chain* in a tree $\mathcal{T}$ is a subset C of its nodes such that for every pair (v, w) in C, either v is an ancestor of w, or the converse. We note that the height of $\mathcal{T}$ is the maximal size of a chain, minus 1. By extension, if $\mathcal{P} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ is a generic point set, we say that $C \subseteq \mathcal{P}$ is a chain of $\mathcal{T}\langle\mathcal{P}\rangle$ if the corresponding nodes form a chain. Using Lemma 3, the following result is proved immediately.

▶ **Lemma 4.** *Let $\mathcal{P}_- \subseteq \mathcal{P}_+$ be two point sets with distinct x- and distinct y-coordinates. Then, for any chain C of $\mathcal{T}\langle\mathcal{P}_+\rangle$, the set $C \cap \mathcal{P}_-$ is a chain of $\mathcal{T}\langle\mathcal{P}_-\rangle$. Consequently, if $\mathcal{C}$ is a chain of maximal size in $\mathcal{T}\langle\mathcal{P}_+\rangle$, we have*

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \geq h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right) - \left|\mathcal{C} \cap (\mathcal{P}_+ \setminus \mathcal{P}_-)\right|.$$

Combining the above lemma with standard thinning properties of Poisson point processes, we get the following useful proposition.

▶ **Proposition 5.** *Let $\rho_- \leq \rho_+$ be two intensity functions defined on the same support $S \subseteq \mathbb{R}^2$, and $\mathcal{P}_-, \mathcal{P}_+$ be two Poisson point processes with intensities ρ_- and ρ_+. Then, we have*

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \succeq \textsc{Binomial}\left(1 + h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right), \inf_{(x,y)\in S} \frac{\rho_-(x,y)}{\rho_+(x,y)}\right) - 1.$$

Proof. Write $r := \inf_{(x,y)\in S} \frac{\rho_-(x,y)}{\rho_+(x,y)}$ where, by convention, $\frac{\rho_-(x,y)}{\rho_+(x,y)} = 1$ if $\rho_+(x,y) = 0$. We couple $\mathcal{P}_+$ and $\mathcal{P}_-$ according to the classical thinning process, meaning that $\mathcal{P}_-$ is constructed by keeping each point (x,y) of $\mathcal{P}_+$ independently with probability $\rho_-(x,y)/\rho_+(x,y) \geq r$.

Let $\mathcal{C}$ be a chain of maximal size in $\mathcal{P}_+$, and set $K := |\mathcal{C} \cap \mathcal{P}_-|$. By Lemma 4:

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \geq h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right) - \left|\mathcal{C} \cap (\mathcal{P}_+ \setminus \mathcal{P}_-)\right| = |\mathcal{C}| - 1 - \left|\mathcal{C} \cap (\mathcal{P}_+ \setminus \mathcal{P}_-)\right| = K - 1.$$

Conditionally given $\mathcal{P}_+$ we have $K \succeq \textsc{Binomial}\left(|\mathcal{C}|, r\right)$, and this concludes the proof. ◀

2.3 Controlling the height of the top tree

We can now use our tools to compare the BST of Poisson point processes with the BST of uniformly random permutations.

▶ **Proposition 6.** *Let $R = [x_1, x_2] \times [y_1, y_2]$ be a rectangle with non-empty interior and $\rho : R \to (0, \infty)$ be a continuous, positive intensity function. For each integer n, let $\mathcal{P}_\rho^N$ be a Poisson point process with intensity $n\rho$. Let $0 < m \le M < \infty$ be such that $m \le \rho \le M$, and write $\eta := \frac{M-m}{m}$. Then for any $\varepsilon > 0$ we have:*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)}{c^* \log n} - 1\right| > \eta + \varepsilon\right] = 0\,. \tag{1}$$

Moreover, for any $p > 0$, the sequence $\left(\frac{h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)}{\log n}\right)^p$ is uniformly integrable.

Proof. Write $\zeta := (x_2 - x_1)(y_2 - y_1) > 0$ for the area of R. Also note that $M/m = 1 + \eta$ and $m/M \ge 1 - \eta$. Using Proposition 5 with $\rho_- = n\rho$ and $\rho_+ = nM$ on R, we obtain

$$h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right) \succeq \textsc{Binomial}\left(1 + h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right),\, \frac{m}{M}\right) - 1\,,$$

where $\mathcal{T}\langle\mathcal{P}_+\rangle$ is the BST of a uniform permutation of random size $\textsc{Poisson}\,(n\zeta M)$. According to [10, Theorem 5.1], $h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right)$ then behaves as $c^* \log(|\mathcal{P}_+|)$ as $n \to \infty$ in probability, which is itself close to $c^* \log n$. Since $\textsc{Binomial}\,(a \log n,\, m/M)$ is concentrated around $(am/M) \log n$, we deduce:

$$h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right) \ge \frac{m}{M}\left(c^* \log n - o_{\mathbb{P}}(\log n)\right) \ge \left(1 - \eta - o_{\mathbb{P}}(1)\right)c^* \log n\,.$$

Similarly, using Proposition 5 with $\rho_- = nm$ and $\rho_+ = n\rho$ we obtain

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \succeq \textsc{Binomial}\left(1 + h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right),\, \frac{m}{M}\right) - 1\,, \tag{2}$$

where $\mathcal{T}\langle\mathcal{P}_-\rangle$ is the BST of a uniform permutation of random size $\textsc{Poisson}\,(n\zeta m)$. We proceed as before to conclude the proof of Equation (1).

For the uniform integrability claim, it suffices to establish boundedness of $\mathbb{E}\left[\frac{h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)^p}{\log(n)^p}\right]$ in n, for all $p > 0$. Conditionally given $h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)$, write $S_n{+}1$ for a random variable with distribution $\textsc{Binomial}\left(1{+}h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right),\, \frac{m}{M}\right)$. Then, using Hoeffding's inequality:

$$\mathbb{P}\left[S_n < \frac{m}{2M}\left(1{+}h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)\right){-}1 \;\middle|\; h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)\right] \le e^{-\frac{m^2}{2M^2}\left(1+h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)\right)}$$

and therefore, by discriminating according to this event for any $n \ge e$:

$$\mathbb{E}\left[\frac{h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)^p}{\log(n)^p}\right] \le \mathbb{E}\left[h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)^p e^{-\frac{m^2}{2M^2}\left(1+h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right)\right)}\right] + \mathbb{E}\left[\frac{\left((2M/m)\cdot(S_n{+}1){-}1\right)^p}{\log(n)^p}\right].$$

Since the function $x \mapsto x^p e^{-\frac{m^2}{2M^2}(1+x)}$ is bounded over $\mathbb{R}_+$, the first term is bounded in n. For the second term, we use $(a+b)^p \le 2^{p-1}(a^p + b^p)$ along with (2) to deduce:

$$\mathbb{E}\left[\frac{\left((2M/m)\cdot(S_n+1)\right)^p}{\log(n)^p}\right] \le 2^{p-1}\left(\frac{2M}{m}\right)^p\left(\mathbb{E}\left[\frac{h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right)^p}{\log(n)^p}\right] + \frac{1}{\log(n)^p}\right)$$

which is bounded in n by [10, Lemma 3.1] and Poisson estimates (indeed, recall that $\mathcal{T}\langle\mathcal{P}_-\rangle$ is the BST of a uniform permutation of random size $\textsc{Poisson}\,(n\zeta m)$). This concludes the proof. ◀

The weakness of the previous proposition is that η, which depends on the rectangle under consideration, might be large. In the next statement we show that, for continuous positive densities ρ, it is possible to choose rectangles for which the corresponding η is small.

▶ **Corollary 7.** *Let D be a compact domain in the plane and $\rho : D \to (0, \infty)$ be a continuous, positive intensity function. Then for any $\varepsilon > 0$, there exists $\beta > 0$ such that for any rectangle $R = [x_1, x_1 + \beta] \times [y_1, y_2]$ with non-empty interior contained in D:*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \cap R\rangle\right)}{c^* \log n} - 1\right| > \varepsilon\right] = 0\,.$$

In particular, taking $x_1 = y_1 = 0$ and $y_2 = 1$, the tree $\mathcal{T}\langle \mathcal{P}_\rho^N \cap R\rangle$ is the top tree $\mathcal{T}\langle \mathcal{P}(\beta)\rangle$ defined in Section 2.1. This top tree therefore has height $(c^* + \varepsilon)\log n$, for small enough β and under assumption (A1).

Proof. Let $\varepsilon > 0$ and assume that $\varepsilon < \min_D \rho$. By uniform continuity of ρ, we can find $\beta > 0$ such that for any $(x, y), (x', y') \in D$, the inequality $|x - x'| + |y - y'| \le \beta$ implies $|\rho(x, y) - \rho(x', y')| \le \varepsilon$. Then consider $R = [x_1, x_1 + \beta] \times [y_1, y_2]$ contained in D. Define

$$f : y \in [y_1, y_2] \mapsto \int_{y_1}^{y} \rho(x_1, t)dt \quad \text{ and } \quad g : y \in [y_1, y_2] \mapsto y_1 + (y_2 - y_1)f(y)/f(y_2).$$

The function g is a $\mathcal{C}^1$ increasing map from $[y_1, y_2]$ onto itself. Let $\widetilde{\mathcal{P}}$ denote the set of points obtained after applying the transformation $(x, y) \mapsto (x, g(y))$ to $\mathcal{P}_\rho^N \cap R$. This transformation does not change the relative orders of points, therefore $\mathcal{T}\langle\widetilde{\mathcal{P}}\rangle$ and $\mathcal{T}\langle \mathcal{P}_\rho^N \cap R\rangle$ have the same shape. Additionally, $\widetilde{\mathcal{P}}$ follows the law of a Poisson point process with intensity

$$n\,\frac{\rho(x, g^{-1}(y))}{g'(g^{-1}(y))} = n\,\frac{f(y_2)}{y_2 - y_1}\,\frac{\rho(x, g^{-1}(y))}{\rho(x_1, g^{-1}(y))}$$

on R. Thus we can apply Proposition 6 with $\eta = \frac{2\varepsilon}{\min_D \rho - \varepsilon}$ to obtain:

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle\tilde{\mathcal{P}}\rangle\right)}{c^* \log n} - 1\right| > \eta + \varepsilon\right] = 0\,.$$

Since this holds for any small enough $\varepsilon > 0$, and $\eta \to 0$ as $\varepsilon \to 0$, the result follows. ◀

2.4 Extreme deviation bounds, via monotone subsequences

It remains to argue that the hanging trees simultaneously all have height $o_{\mathbb{P}}(\log n)$. A "typical" horizontal band in Figure 3 contains $\mathcal{O}(1)$ points, but their maximum is actually $\mathcal{O}(\log n)$ (Proposition 11). The hanging trees are themselves BSTs of point processes, and therefore they individually have height $\mathcal{O}(\log\log n) \ll \log n$. To have this bound for all $\mathcal{O}(n)$ hanging trees simultaneously, we need adequate deviation estimates for the BST height of point processes. Such estimates are provided by Devroye for uniform BSTs [10], but the monotonicity properties of BSTs are not good enough to use direct comparison arguments. We solve this by relating the BST height of a point set to its longest monotone subsequences, for which we have good monotonicity properties and deviation bounds.

Let σ be a permutation of $\{1, \dots, n\}$. An increasing subsequence of σ is a sequence of indices $i_1 < \cdots < i_k$ such that $\sigma(i_1) < \cdots < \sigma(i_k)$. The maximum length of an increasing subsequence of σ is then denoted by $\mathrm{LIS}\,(\sigma)$. We define similarly $\mathrm{LDS}\,(\sigma)$, the maximum length of a decreasing subsequence of σ.

▶ **Lemma 8.** *For any permutation σ, we have $h\left(\mathcal{T}\langle\sigma\rangle\right) \leq \mathrm{LIS}\left(\sigma\right) + \mathrm{LDS}\left(\sigma\right)$.*

Proof. Let $i_1 < \cdots < i_k$ be a sequence of integers such that $\sigma(i_1), \ldots, \sigma(i_k)$ label nodes on a chain C of $\mathcal{T}\langle\sigma\rangle$. Define $\mathcal{I}_{\mathcal{R}}$ (resp. $\mathcal{I}_{\mathcal{L}}$) as the family of i_j's such that the node following $\sigma(i_j)$ in C lies in its right subtree (resp. left subtree). By construction, $\mathcal{I}_{\mathcal{R}} \cup \{i_k\}$ and $\mathcal{I}_{\mathcal{L}} \cup \{i_k\}$ form respectively an increasing and a decreasing subsequence of σ. The lemma follows. ◀

Combining this lemma with [6, Proposition 3.2], we get that if ρ is an integrable function then $h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right) = o_{\mathbb{P}}(n)$. We will need a more quantitative version of this, valid only for bounded functions ρ. We start with the following lemma, proved by a straightforward application of the first moment method.

▶ **Lemma 9.** *For each integer n, let σ^n be a uniform permutation of $\{1, \ldots, n\}$. Then:*

$$\mathbb{P}\left[\mathrm{LIS}\left(\sigma^n\right) \geq \frac{n}{\log n}\right] \leq \exp\left(-n + o(n)\right).$$

Using the previous two lemmas and standard techniques, we obtain a useful corollary:

▶ **Corollary 10.** *For any $M > 0$ and $\varepsilon > 0$, there exists $n_0 = n_0(M, \varepsilon)$ such that the following holds. For any $0 < \zeta \leq 1$, any function $\rho : [0,1]^2 \to [0, \infty)$ bounded by M and supported on some rectangle $[a, b] \times [c, d]$ with $(b - a)(d - c) \leq \zeta$, and for any integer $n > n_0/\zeta$:*

$$\mathbb{P}\left[h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right) > 2\varepsilon\zeta n\right] \leq 4\exp\left(-\tfrac{\varepsilon}{2}\zeta n \log(\zeta n)\right).$$

We refer to [8] for the full proof. The key argument is that LIS and LDS are, unlike the height of BSTs, monotone in their arguments: if $\mathcal{P}_- \subseteq \mathcal{P}_+$ are generic point sets, then $\mathrm{LIS}\left(\sigma\langle\mathcal{P}_-\rangle\right) \leq \mathrm{LIS}\left(\sigma\langle\mathcal{P}_+\rangle\right)$, and likewise for LDS. We can thus compare $\mathcal{P}_\rho^N$ to a *homogeneous* Poisson point process with higher intensity, and use extreme deviation bounds for the monotone subsequences of the latter.

2.5 Controlling the height of the hanging trees

Throughout the rest of this section, we use the notation of Section 2.1. For each integer $0 \leq k \leq K_\beta$, we let $\zeta_k := |I_k|$ be the vertical length of the band $(\beta, 1] \times I_k$.

▶ **Proposition 11.** *Let μ be a permuton. Assume that there exists $\beta > 0$ such that $\mu_{/[0,\beta]\times[0,1]}$ has a continuous and positive density $\rho : [0, \beta] \times [0, 1] \to (0, \infty)$. Then the following holds.*

1. *There exists $\alpha > 0$ such that $\max_k \zeta_k \leq \alpha \frac{\log n}{n}$ w.h.p. as $n \to \infty$.*
2. *All powers of $\frac{1}{\log n} \max_k \left|\mathcal{P}_k(\beta)\right|$ are uniformly integrable.*

Sketch of proof. The first item can be derived using standard results on "maximal spacings". Indeed, by a thinning procedure, $\max_k \zeta_k$ is bounded above by the largest gap among $\textsc{Poisson}\left(n\beta m\right)$ i.i.d. uniform variables in $[0, 1]$, where m is a lower bound for ρ. This is known to concentrate around $\log(n)/(n\beta m)$ [16], which proves the first item. To prove the second item, we can use that conditionally given $\mathcal{P}(\beta)$, the number $\left|\mathcal{P}_k(\beta)\right|$ has distribution $\textsc{Poisson}\left(n(1-\beta)\zeta_k\right)$. Then, conclude with item (1) and Poisson estimates. ◀

▶ **Proposition 12.** *Let μ be a permuton satisfying* (A1)*. Then for any $\beta \in (0, 1)$ we have the following convergence in probability as n goes to infinity:*

$$\frac{1}{\log n} \max_{0 \leq k \leq K_\beta} \left\{h\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right)\right\} \longrightarrow 0.$$

Proof. From Proposition 11, item (1), there exists $\alpha > 0$ such that w.h.p. $\max_k \zeta_k < \alpha \frac{\log n}{n}$. Work under this event, and conditionally given $\mathcal{P}(\beta)$. Then for each k, $\mathcal{P}_k(\beta)$ is distributed like a Poisson point process with intensity $n\rho_{/[\beta,1]\times[y_{(k)},y_{(k+1)}]}$. For each k, Corollary 10 applies with ρ restricted to $[\beta, 1] \times [y_{(k)}, y_{(k+1)}]$ and $\zeta = \alpha \frac{\log n}{n}$. With a union bound, we deduce:

$$\mathbb{P}\left[\max_{0\le k\le|\mathcal{P}(\beta)|} h\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right) > \delta \log n\right] \le 4(|\mathcal{P}(\beta)|+1)\exp\left(-\tfrac{\delta}{4\alpha}\alpha\log n\,\log(\alpha\log n)\right)$$

for $\zeta n = \alpha \log n$ large enough. But w.h.p. we have $\max_k \zeta_k < \alpha\frac{\log n}{n}$ and $|\mathcal{P}(\beta)| < n$, so the unconditioned probability tends to 0 as $n \to \infty$. This proves the proposition. ◀

2.6 Concluding the proof of the height theorem

Proof of Theorem 1. Fix $\varepsilon > 0$. Let D be a compact neighborhood of $\{0\} \times [0, 1]$ on which ρ is continuous and positive, and let $\beta = \beta(\varepsilon) > 0$ be given by Corollary 7 applied to ρ on D. Therefore, if $\mathcal{T}\langle\mathcal{P}(\beta)\rangle := \mathcal{T}\langle\mathcal{P}_\mu^N \cap ([0, \beta] \times [0, 1])\rangle$ denotes the top tree of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$:

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle\mathcal{P}(\beta)\rangle\right)}{c^* \log n} - 1\right| > \varepsilon\right] = 0\,.$$

Furthermore, by Proposition 12, the quantity

$$\frac{1}{\log n} \max_{0\le k\le K_\beta} \left\{h\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right)\right\}$$

converges in probability to 0. Combined with Lemma 2, this implies that $\frac{1}{\log n} h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right)$ converges in probability to c^*.

Together with Proposition 6 and Proposition 11, Lemma 2 also implies uniform integrability of all powers of $\frac{1}{\log n} h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right)$. Therefore the convergence holds in L^p for all $p \ge 1$, concluding the proof of Theorem 1. ◀

3 Second main result: subtree size convergence of the BSTs

3.1 Some definition, and statement of the result

Next, we state a limit theorem for $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$, in the sense of the *subtree size convergence* recently introduced by Grübel [13]. We start by recalling this notion of convergence.

Identify nodes in a binary tree with finite words in the alphabet $\{0, 1\}$ as follows: the empty word $\varnothing$ corresponds to the root, and for a node v encoded by w, the words $w0$ and $w1$ encode respectively the left and right children of v. Let $\mathbb{V} = \{0, 1\}^*$ be the set of all finite words on $\{0, 1\}$, representing all nodes of the complete infinite binary tree. A labeled tree is then identified with a function from a subset of $\mathbb{V}$ to $\mathbb{R}$, where the domain of the function is the set of nodes in the tree, and a node is mapped to its label. In particular, $\mathcal{T}(v)$ denotes the label of the node v in $\mathcal{T}$. We also write $v \in \mathcal{T}$ to indicate that the node v is in $\mathcal{T}$. Given a finite tree $\mathcal{T}$ and a node $v \in \mathbb{V}$, define

$$t(\mathcal{T}, v) := \frac{1}{|\mathcal{T}|}\left|\left\{u \in \mathcal{T} : v \preceq u\right\}\right|,$$

where $v \preceq u$ means that v is a prefix of u. In words, $t(\mathcal{T}, v)$ is the proportion of nodes in $\mathcal{T}$ which are descendants of v.

Further write Ψ for the set of functions $\psi : \mathbb{V} \to [0,1]$ such that $\psi(\varnothing) = 1$ and for any $v \in \mathbb{V}$, we have $\psi(v) = \psi(v0) + \psi(v1)$. Then a sequence of binary trees $(\mathcal{T}^n)_{n\in\mathbb{N}}$ is said to converge to a function $\psi \in \Psi$ if and only if $t(\mathcal{T}^n, v) \to \psi(v)$ for all $v \in \mathbb{V}$. If that is the case, we write $\mathcal{T}_n \xrightarrow{\text{ssc}} \psi$ and refer to this as *subtree size convergence.*

We now define two important objects before stating our result. For any complete BST $\mathcal{T} : \mathbb{V} \to (0,1)$, we define $\mathcal{T}_{\text{left}} : \mathbb{V} \to \mathbb{R}$ as follows. First, for any $v \in \{0\}^*$, let $\mathcal{T}_{\text{left}}(v) := 0$. Then if $v = v'10^k$ for some $k \geq 0$, let $\mathcal{T}_{\text{left}}(v) := \mathcal{T}(v')$. Informally, $\mathcal{T}_{\text{left}}(v)$ is the right-most ancestor of v to its left. Define similarly $\mathcal{T}_{\text{right}}$ such that $\mathcal{T}_{\text{right}}(v) := 1$ for any $v \in \{1\}^*$ and $\mathcal{T}_{\text{right}}(v) := \mathcal{T}(v')$ whenever $v = v'01^k$ for some $k \geq 0$. We note that this definition implies that $\mathcal{T}_{\text{left}}(v) < \mathcal{T}(v) < \mathcal{T}_{\text{right}}(v)$ for any $v \in \mathbb{V}$.

Given a probability measure m on $[0,1]$ without atoms, write $\psi_m \in \Psi$ for the following random object. First, let $Y = (Y_1, Y_2, \ldots)$ be an i.i.d. variables distributed according to m and write $\mathcal{T}^m := \mathcal{T}\langle Y \rangle$ for the corresponding (infinite) BST. Then, let $\psi_m := \mathcal{T}^m_{\text{right}} - \mathcal{T}^m_{\text{left}}$. This is well-defined, since $\mathcal{T}^m$ is a.s. complete [10, Theorem 6.1]. It is immediate to check that indeed $\psi_m \in \Psi$ (almost surely). See Figure 4 for an example.

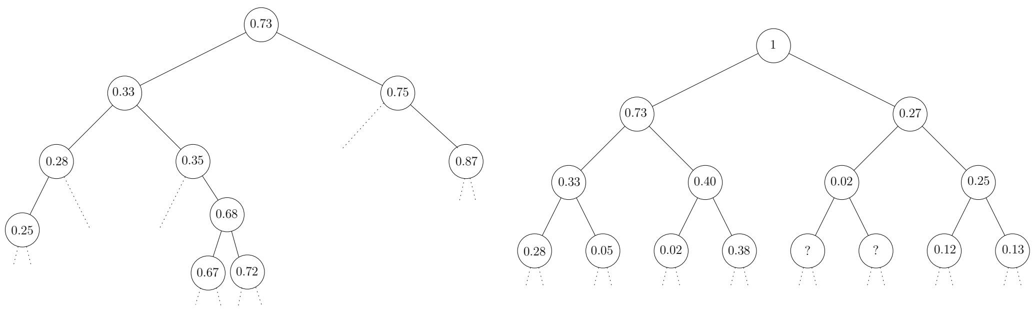

Figure 4 Example of realizations of $\mathcal{T}^m$ and ψ_m. Note that we do not have enough data to compute two of the values of ψ_m on nodes in the third level.

We can now state our second main result. A permuton is said to satisfy assumption (A2) if there exists a probability measure μ_0 on $[0,1]$, without atoms, such that

$$\tfrac{1}{x}\mu\big([0,x] \times \cdot\big) \underset{x\to 0}{\longrightarrow} \mu_0 \tag{3}$$

for the weak topology. Assumption (A2) is weaker than (A1): in particular, (A2) holds whenever μ admits a continuous density on a neighborhood of $\{0\} \times [0,1]$.

► **Theorem 13** (Subtree size convergence of BSTs for permuton samples)**.** *Let μ be a permuton satisfying* (A2)*. The following convergence in distribution holds for the subtree size topology:*

$$\mathcal{T}\langle \mathcal{P}^N_\mu \rangle \xrightarrow{\text{ssc}} \psi_{\mu_0}\,.$$

Note that the limit depends on μ only through μ_0. The assumption that μ_0 does not have atoms is important. A first difficulty when μ_0 has some atom is that the BST $\mathcal{T}\langle Y_1, Y_2, \ldots \rangle$ where Y_1, Y_2, ..., are i.i.d. variables with distribution μ_0 is ill-defined, since some of the Y_i's are equal. We can also see that, in this case, the limit of $\mathcal{T}\langle \mathcal{P}^n_\mu \rangle$ may not depend only on μ_0. Indeed, consider the permutons μ^1 and μ^2 supported by the sets $y \equiv \frac{1}{2} + x \mod 1$ and $y \equiv \frac{1}{2} - x \mod 1$. They both satisfy (3) with $\mu_0 = \delta_{1/2}$, but it is easy to see that their BSTs have different limits in the sense of subtree size convergence.

3.2 Preliminaries to the proof

We start with a variant of the Glivenko–Cantelli theorem for triangular arrays.

▶ **Proposition 14.** *Let μ be a probability measure with a finite fourth moment, and distribution function $F(x) := \mu((-\infty, x])$. For each $n \geq 1$, let $(X_{i,n})_{1\leq i\leq n}$ be i.i.d. random variables with common distribution μ and let $F_n(x) := \frac{1}{n}\left|\{i \leq n : X_{i,n} \leq x\}\right|$ be their empirical distribution function. Then F_n converges a.s. uniformly to F.*

Proof. A classical fourth moment computation, together with Borel–Cantelli lemma – see e.g. [4, Theorem 6.1] – shows that, for any fixed x, $F_n(x)$ converges a.s. to $F(x)$. The rest of the proof is similar to that of the classical Glivenko–Cantelli theorem which considers a single sequence X_i of i.i.d. random variables instead of a triangular array, but does not require a fourth moment condition; see e.g. [4, Theorem 20.6]. ◀

Under assumption (A2), we can prove convergence in distribution of the leftmost points in $\mathcal{P}_\mu^N$. The proof of the following proposition is rather technical, and can be found in [8].

▶ **Proposition 15.** *Let μ be a permuton satisfying (A2), and let $\mathcal{P}_\mu^N = \left\{\left(X_i^N, Y_i^N\right), 1\leq i\leq N\right\}$ be a Poisson point process with intensity $n\mu$. Let $\left(\left(X_{(i)}^N, Y_{(i)}^N\right)\right)_{1\leq i\leq N}$ be its reordering such that $X_{(1)}^N < \cdots < X_{(N)}^N$. Then, for any fixed $K \geq 1$, we have the following convergence in distribution:*

$$\left(Y_{(1)}^N, \ldots, Y_{(K)}^N\right) \underset{n\to\infty}{\longrightarrow} (Y_k)_{1\leq k\leq K}$$

where $(Y_k)_{1\leq k\leq K}$ is a sequence of i.i.d. random variables distributed according to μ_0.

Finally, with the notation of Section 3.1, we can use the functions $\mathcal{T}_{\text{left}}$ and $\mathcal{T}_{\text{right}}$ to describe the descendants of nodes in $\mathcal{T}$. The proof is straightforward.

▶ **Lemma 16.** *Let $y_1, \ldots, y_n$ be distinct numbers and let $\mathcal{T} := \mathcal{T}\langle y_1, \ldots, y_n\rangle$ be the corresponding BST. Let u be a node in $\mathcal{T}$ and let k be such that $\mathcal{T}(u) = y_k$. Then:*

$$t(\mathcal{T}, u) = \frac{1}{|\mathcal{T}|}\left|\{y_k, \ldots, y_n\} \cap \left(\mathcal{T}_{\text{left}}(u), \mathcal{T}_{\text{right}}(u)\right)\right|.$$

3.3 Proof of subtree size convergence

Proof of Theorem 13. Write $\mathcal{T}^N := \mathcal{T}\langle \mathcal{P}_\mu^N\rangle$. Since the subtree size topology is by definition the pointwise convergence of the function $(t(., u))_{u\in\mathbb{V}}$, we need to prove the convergence of finite-dimensional distributions. Namely we need to prove that, for any $d \geq 1$ and $u_1, \ldots, u_d \in \mathbb{V}$, we have the following convergence in distribution as $n \to \infty$:

$$\left(t(\mathcal{T}^N, u_i)\right)_{i\leq d} \longrightarrow \left(\psi_\mu(u_i)\right)_{i\leq d}. \tag{4}$$

Recall the notation of Proposition 15. Using Skorohod's representation theorem [3, Section 6], we might assume that the convergence

$$\left(Y_{(1)}^N, \ldots, Y_{(K)}^N\right) \underset{n\to\infty}{\longrightarrow} (Y_k)_{1\leq k\leq K} \tag{5}$$

holds almost surely. Since μ_0 has no atoms, the numbers $(Y_k)_{k\geq 1}$ are a.s. distinct. Moreover the tree $\mathcal{T}\langle Y_1, Y_2, \ldots\rangle$ has a.s. shape $\mathbb{V}$. Consequently, a.s. there exists a (random) threshold K such that all nodes $u_i, i \leq d$ belong to $\mathcal{T}\langle Y_1, \ldots, Y_K\rangle$. Using (5), there exists a (random) threshold n_0 such that for all $n \geq n_0$, the relative order of $(Y_{(1)}^N, \dots, Y_{(K)}^N)$ is the same as that of $(Y_1, \dots, Y_K)$. Hence the trees $\mathcal{T}_K^N := \mathcal{T}\langle Y_{(1)}^N, \dots, Y_{(K)}^N\rangle$ and $\mathcal{T}_K^\infty := \mathcal{T}\langle Y_1, \dots, Y_K\rangle$ have the same shape T_K. Moreover for any v in T_K, the labels $\mathcal{T}_K^N(v)$ and $\mathcal{T}_K^\infty(v)$ equal $Y_{(i)}^N$ and Y_i respectively, for the *same* index i. Therefore $\mathcal{T}_K^N(v) \to \mathcal{T}_K^\infty(v)$ as $n \to \infty$, a.s. in the probability space created by the application of Skorohod's representation theorem.

Now, using Lemma 16 and the fact that each u_i is filled in $\mathcal{T}^N$ before step $K = \mathcal{O}_{\mathbb{P}}(1)$:

$$t\left(\mathcal{T}^N, u_i\right) = \frac{1}{N} \left| \left\{Y_1^N, \dots, Y_N^N\right\} \cap \left(\mathcal{T}_{\text{left}}^N(u_i), \mathcal{T}_{\text{right}}^N(u_i)\right) \right| + o_{\mathbb{P}}(1)\,.$$

Consider the empirical distribution function $F_N(y) := \frac{1}{N}\left|\{Y_1^N, \dots, Y_N^N\} \cap (-\infty, y)\right|$. Then:

$$t\left(\mathcal{T}^N, u_i\right) = F_N\left(\mathcal{T}_{\text{right}}^N(u_i)\right) - F_N\left(\mathcal{T}_{\text{left}}^N(u_i)\right) + o_{\mathbb{P}}(1)\,.$$

The random variable $N \sim \textsc{Poisson}(n)$ is well-concentrated around n, and conditionally given N, the points $Y_1^N, \dots, Y_N^N$ are i.i.d. random variables in $[0,1]$. Since μ is a permuton, their common (conditional) distribution is the uniform distribution. From Proposition 14, we infer that F_n converges a.s. uniformly on $[0,1]$ to the identity function (the earlier use of Skorohod's representation theorem implies that the $(Y_i^n)_{1\leq i\leq n}$ are coupled in a nontrivial way for different values of n, but Proposition 14 applies nevertheless).

Moreover, the above discussion implies that $\mathcal{T}_{\text{right}}^N(u_i)$ and $\mathcal{T}_{\text{left}}^N(u_i)$ converge a.s. to $\mathcal{T}_{\text{right}}^\infty(u_i)$ and $\mathcal{T}_{\text{left}}^\infty(u_i)$ respectively. Therefore, a.s. in the probability space created by the application of Skorohod's representation theorem, for all $i \leq d$ we have:

$$t\left(\mathcal{T}^N, u_i\right) = \mathcal{T}_{\text{right}}^\infty(u_i) - \mathcal{T}_{\text{left}}^\infty(u_i) + o_{\mathbb{P}}(1) = \psi_{\mu_0}(u_i) + o_{\mathbb{P}}(1).$$

Since a.s. (joint) convergence implies (joint) convergence in distribution, (4) is proved. ◀

4 Extra results

In this last section, we briefly discuss some additional results and open questions. More details can be found in the full paper [8].

De-Poissonization. As mentioned in the introduction, it is possible to state Theorems 1 and 13 for $\mathcal{P}_\mu^n$ (a set of n i.i.d. points under μ) instead of $\mathcal{P}_\mu^N$ (a Poisson point process with intensity $n\mu$). In other words, it is possible to "de-Poissonize" the random size N into a deterministic size n. This is rather technical, and hinges on the comparison method of Proposition 5 along with standard estimates on the Poisson law.

Optimality of hypotheses in Theorem 1. Assumption (A1) is in some sense optimal for the upper bound on the BST height. Indeed, in [8] we exhibit two permutons which do not quite satisfy (A1), and whose BSTs are much deeper.

Universality of the lower bound for the BST height. On the other hand, we could not construct a permuton μ such that $h\left(\mathcal{T}\langle\mathcal{P}_\mu^n\rangle\right)$ is asymptotically smaller than $c^* \log(n)$. This leads us to conjecture that the BSTs of permuton samples always satisfy this lower bound. In [8], we prove a partial result in this direction.

References

1 Louigi Addario-Berry and Benoît Corsini. The height of Mallows trees. *Ann. Probab.*, 49(5):2220–2271, 2021. `doi:10.1214/20-AOP1503`.

2 Frédérique Bassino, Mathilde Bouvel, Valentin Féray, Lucas Gerin, Mickaël Maazoun, and Adeline Pierrot. Universal limits of substitution-closed permutation classes. *J. Eur. Math. Soc. (JEMS)*, 22(11):3565–3639, 2020. `doi:10.4171/JEMS/993`.

3 Patrick Billingsley. *Convergence of probability measures.* Wiley Ser. Probab. Stat. Chichester: Wiley, 2nd ed. edition, 1999.
4 Patrick Billingsley. *Probability and measure. Anniversary edition.* Hoboken, NJ: John Wiley & Sons, 2012.
5 Jacopo Borga, Sayan Das, Sumit Mukherjee, and Peter Winkler. Large deviation principle for random permutations. *International Mathematics Research Notices*, rnad096, 2023. `doi:10.1093/imrn/rnad096`.
6 Jacopo Borga, Ewain Gwynne, and Xin Sun. Permutons, meanders, and sle-decorated liouville quantum gravity, 2022. Preprint arXiv:2207.02319. `doi:10.48550/arXiv.2207.02319`.
7 Benoît Corsini. The height of record-biased trees. *Random Structures & Algorithms*, 62(3):623–644, 2023. `doi:10.1002/rsa.21110`.
8 Benoît Corsini, Victor Dubach, and Valentin Féray. Binary search trees of permuton samples, 2024. Preprint arXiv:2403.03151. `doi:10.48550/arXiv.2403.03151`.
9 Jean-Dominique Deuschel and Ofer Zeitouni. Limiting curves for i.i.d. records. *Ann. Probab.*, 23(2):852–878, 1995.
10 Luc Devroye. A note on the height of binary search trees. *Journal of the ACM (JACM)*, 33(3):489–498, 1986.
11 Victor Dubach. Increasing subsequences of linear size in random permutations and the Robinson–Schensted tableaux of permutons, 2023. Preprint arXiv:2307.05768. `doi:10.48550/arXiv.2307.05768`.
12 Victor Dubach. Locally uniform random permutations with large increasing subsequences. *Combinatorial Theory*, 3(3), 2023. `doi:10.5070/C63362784`.
13 Rudolf Grübel. A note on limits of sequences of binary trees. *Discrete Mathematics & Theoretical Computer Science*, 25(Analysis of Algorithms), 2023. `doi:10.46298/dmtcs.10968`.
14 Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio. Limits of permutation sequences. *J. Comb. Theory, Ser. B*, 103(1):93–113, 2013. `doi:10.1016/j.jctb.2012.09.003`.
15 Jonas Sjöstrand. Monotone subsequences in locally uniform random permutations. *Ann. Probab.*, 51(4):1502–1547, 2023. `doi:10.1214/23-AOP1624`.
16 Eric Slud. Entropy and maximal spacings for random partitions. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 41:341–352, 1978. `doi:10.1007/BF00533604`.