Binary search trees of permuton samples

Benoît Corsini, Victor Dubach, Valentin Féray

AofA 2024, Bath

Full version at arXiv:2403.03151

Outline

I. BSTs and permutons

Binary Search Trees

A Binary Search Tree (BST) is a rooted labeled binary tree such that: for each vertex v, all labels of vertices in its left-subtree are smaller than that of v (resp. right-subtree, greater).

I. BSTs and permutons

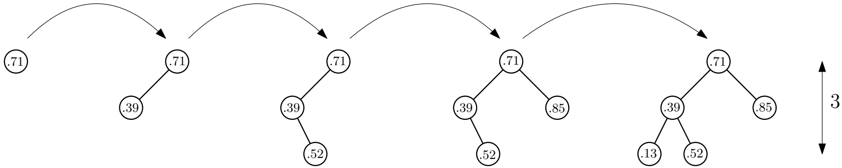
Binary Search Trees

A Binary Search Tree (BST) is a rooted labeled binary tree such that: for each vertex v, all labels of vertices in its left-subtree are smaller than that of v (resp. right-subtree, greater).

BST associated with a sequence of distinct numbers

Let $y = (y_1, \ldots, y_n)$ be a sequence of distinct numbers. We can construct a BST $\mathcal{T}\langle y \rangle$ by successively adding leaves with those labels.

Construction of $\mathcal{T}\langle y \rangle$ with $y = (.71, .39, .52, .85, .13)$:



I. BSTs and permutons

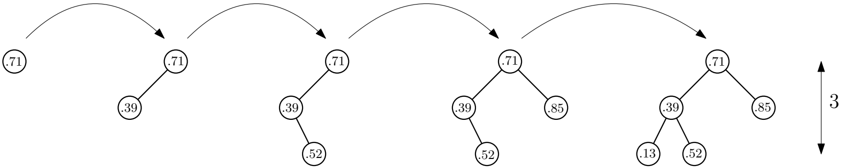
Binary Search Trees

A Binary Search Tree (BST) is a rooted labeled binary tree such that: for each vertex v, all labels of vertices in its left-subtree are smaller than that of v (resp. right-subtree, greater).

BST associated with a sequence of distinct numbers

Let $y = (y_1, \dots, y_n)$ be a sequence of distinct numbers. We can construct a BST $\mathcal{T}\langle y \rangle$ by successively adding leaves with those labels.

Construction of $\mathcal{T}\langle y \rangle$ with $y = (.71, .39, .52, .85, .13)$:



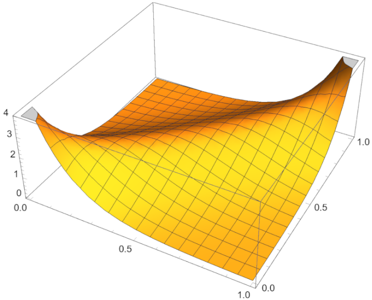
Rem: same shape as $\mathcal{T}\langle \sigma \rangle$ with $\sigma = (4, 2, 3, 5, 1)$.

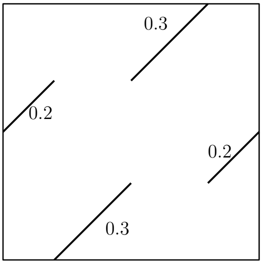
I. BSTs and permutons

Permutons

A *permuton* (or *copula*) is a probability measure μ on $[0,1]^2$ with uniform marginals:

$$\forall t \in [0,1], \quad \mu([0,t]\times[0,1]) = \mu([0,1]\times[0,t]) = t.$$





I. BSTs and permutons

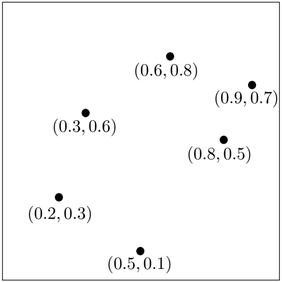
Permuton samples

Let $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.
Define a reordering $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \cdots < x_{(n)}$.
There exists a unique permutation $\sigma = \sigma\langle\mathcal{P}\rangle$ such that $(y_{(1)}, \ldots, y_{(n)})$ and $(\sigma(1), \ldots, \sigma(n))$ are in the same relative order.

I. BSTs and permutons

Permuton samples

Let $\mathcal{P} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.
Define a reordering $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \cdots < x_{(n)}$.
There exists a unique permutation $\sigma = \sigma\langle\mathcal{P}\rangle$ such that $(y_{(1)}, \dots, y_{(n)})$ and $(\sigma(1), \dots, \sigma(n))$ are in the same relative order.



$\mathcal{P}$

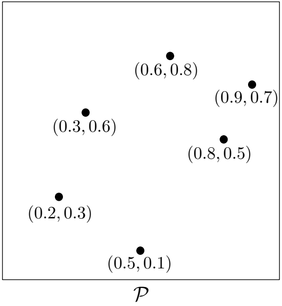
$$(y_{(1)}, \dots, y_{(6)}) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)$$

$$\sigma\langle\mathcal{P}\rangle = (2, 4, 1, 6, 3, 5)$$

I. BSTs and permutons

Permuton samples

If $\mathcal{P}$ is random i.i.d. under a permuton μ, this permutation σ_μ^n is a *permuton sample.*



$$(y_{(1)}, \ldots, y_{(6)}) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)$$

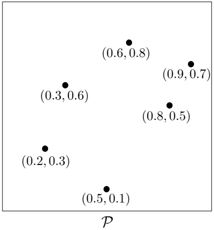
$$\sigma\langle\mathcal{P}\rangle = (2, 4, 1, 6, 3, 5)$$

Rem: if $\mu = \mathsf{Leb}_{[0,1]^2}$ then σ_μ^n is uniformly random.

I. BSTs and permutons

Permuton samples

If $\mathcal{P}$ is random i.i.d. under a permuton μ, this permutation σ_μ^n is a *permuton sample.*



$$(y_{(1)}, \dots, y_{(6)}) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)$$

$$\sigma\langle\mathcal{P}\rangle = (2, 4, 1, 6, 3, 5)$$

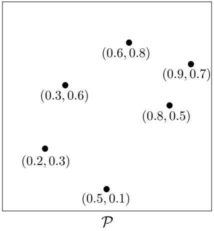
Rem: if $\mu = \text{Leb}_{[0,1]^2}$ then σ_μ^n is uniformly random.
Motivation:

- Continuous deformation of the uniform distribution

I. BSTs and permutons

Permuton samples

If $\mathcal{P}$ is random i.i.d. under a permuton μ, this permutation σ_μ^n is a *permuton sample.*



$$(y_{(1)}, \dots, y_{(6)}) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)$$

$$\sigma\langle\mathcal{P}\rangle = (2, 4, 1, 6, 3, 5)$$

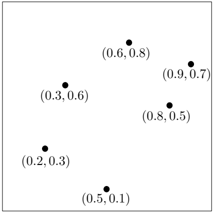
Rem: if $\mu = \text{Leb}_{[0,1]^2}$ then σ_μ^n is uniformly random.
Motivation:

- Continuous deformation of the uniform distribution
- Highly non-parametric, wide range of distributions

I. BSTs and permutons

Permuton samples

If $\mathcal{P}$ is random i.i.d. under a permuton μ, this permutation σ_μ^n is a *permuton sample.*



$\mathcal{P}$

$$(y_{(1)}, \dots, y_{(6)}) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)$$

$$\sigma\langle\mathcal{P}\rangle = (2, 4, 1, 6, 3, 5)$$

Rem: if $\mu = \mathrm{Leb}_{[0,1]^2}$ then σ_μ^n is uniformly random.
Motivation:

- Continuous deformation of the uniform distribution
- Highly non-parametric, wide range of distributions
- Insight on the links between random permutations and their limit permuton

I. BSTs and permutons

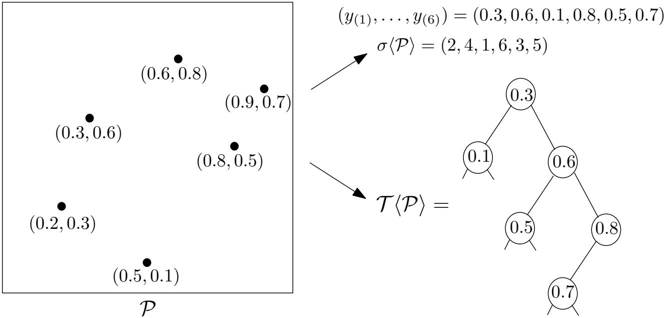
BST of a point process

Let $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.
Define a reordering $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \cdots < x_{(n)}$.
The BST of $\mathcal{P}$ is defined as $\mathcal{T}\langle\mathcal{P}\rangle = \mathcal{T}\langle y_{(1)}, \ldots, y_{(n)}\rangle$.

I. BSTs and permutons

BST of a point process

Let $\mathcal{P} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.
Define a reordering $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \cdots < x_{(n)}$.
The BST of $\mathcal{P}$ is defined as $\mathcal{T}\langle\mathcal{P}\rangle = \mathcal{T}\langle y_{(1)}, \dots, y_{(n)}\rangle$.

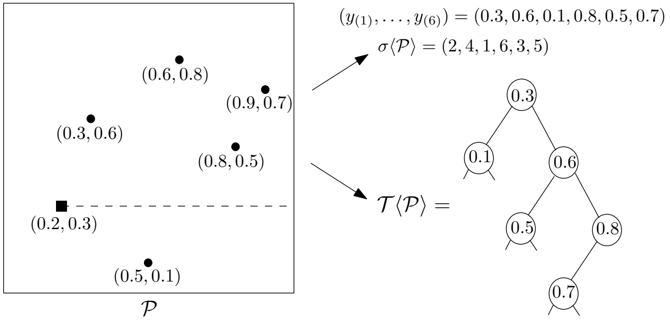


Rem: $\mathcal{T}\langle\mathcal{P}\rangle$ has the same shape as $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$.

I. BSTs and permutons

BST of a point process

Let $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.
Define a reordering $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \cdots < x_{(n)}$.
The BST of $\mathcal{P}$ is defined as $\mathcal{T}\langle\mathcal{P}\rangle = \mathcal{T}\langle y_{(1)}, \ldots, y_{(n)}\rangle$.

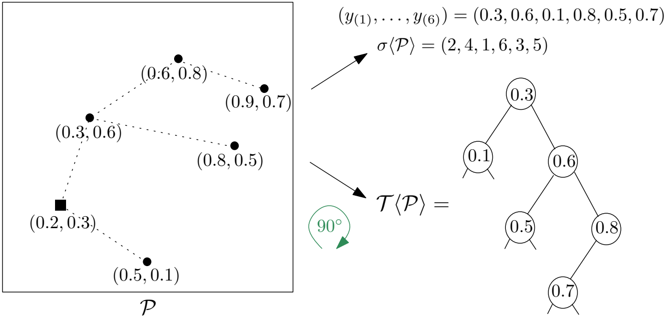


Rem: $\mathcal{T}\langle\mathcal{P}\rangle$ has the same shape as $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$.

I. BSTs and permutons

BST of a point process

Let $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.
Define a reordering $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \cdots < x_{(n)}$.
The BST of $\mathcal{P}$ is defined as $\mathcal{T}\langle\mathcal{P}\rangle = \mathcal{T}\langle y_{(1)}, \ldots, y_{(n)}\rangle$.



Rem: $\mathcal{T}\langle\mathcal{P}\rangle$ has the same shape as $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$.

II. Universality of the height

The height of a BST determines the complexity of operations such as lookup, addition or removal of data.

II. Universality of the height

The height of a BST determines the complexity of operations such as lookup, addition or removal of data.

Q: Let σ be a (random) permutation. What is the height of $\mathcal{T}\langle\sigma\rangle$?

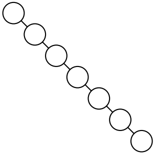
II. Universality of the height

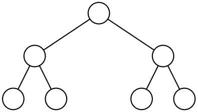
The height of a BST determines the complexity of operations such as lookup, addition or removal of data.

Q: Let σ be a (random) permutation. What is the height of $\mathcal{T}\langle\sigma\rangle$?

In general:

$$n \geq h(\mathcal{T}\langle\sigma(1),\ldots,\sigma(n)\rangle) \geq \log_2(n+1)$$





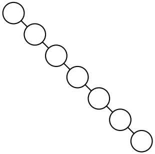
II. Universality of the height

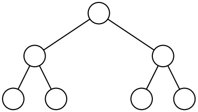
The height of a BST determines the complexity of operations such as lookup, addition or removal of data.

Q: Let σ be a (random) permutation. What is the height of $\mathcal{T}\langle\sigma\rangle$?

In general:

$$n \geq h(\mathcal{T}\langle\sigma(1),\ldots,\sigma(n)\rangle) \geq \log_2(n+1)$$





Depending on the law of σ, asymptotic results for $h(\mathcal{T}\langle\sigma\rangle)$?

II. Universality of the height

Devroye '86

If $\sigma_n \sim \mathrm{Unif}(\mathfrak{S}_n)$ then, as $n \to \infty$:

$$\frac{h(\mathcal{T}\langle\sigma_n\rangle)}{c^* \log n} \longrightarrow 1$$

in probability and in L^p for all $p \geq 1$, where $c^* \geq 2$ solves $c \log(2e/c) = 1$.

II. Universality of the height

Devroye '86

If $\sigma_n \sim \mathrm{Unif}\left(\mathfrak{S}_n\right)$ then, as $n \to \infty$:

$$\frac{h\left(\mathcal{T}\langle\sigma_n\rangle\right)}{c^* \log n} \longrightarrow 1$$

in probability and in L^p for all $p \geq 1$, where $c^* \geq 2$ solves $c\log(2e/c) = 1$.

Let $q \in [0,1]$. The Mallows distribution $M_{n,q}$ is defined by:

$$\forall \sigma \in \mathfrak{S}_n, \quad M_{n,q}(\sigma) \propto q^{\mathrm{inv}(\sigma)}.$$

Addario-Berry – Corsini '21

If $\sigma_n \sim M_{n,q_n}$ where $n(1-q_n)/\log n \to 0$, then:

$$\frac{h\left(\mathcal{T}\langle\sigma_n\rangle\right)}{c^* \log n} \longrightarrow 1$$

in probability and in L^p for all $p \geq 1$.

II. Universality of the height

Devroye '86

If $\sigma_n \sim \mathrm{Unif}\left(\mathfrak{S}_n\right)$ then, as $n \to \infty$:

$$\frac{h\left(\mathcal{T}\langle\sigma_n\rangle\right)}{c^* \log n} \longrightarrow 1$$

in probability and in L^p for all $p \geq 1$, where $c^* \geq 2$ solves $c \log(2e/c) = 1$.

Let $\theta \geq 0$. The record-biased distribution $R_{n,\theta}$ is defined by:

$$\forall \sigma \in \mathfrak{S}_n, \quad R_{n,\theta}(\sigma) \propto \theta^{\mathrm{rec}(\sigma)}.$$

Corsini '23

If $\sigma_n \sim R_{n,\theta}$ where $\theta \leq c^*$, then:

$$\frac{h\left(\mathcal{T}\langle\sigma_n\rangle\right)}{c^* \log n} \longrightarrow 1$$

in probability and in L^p for all $p \geq 1$.

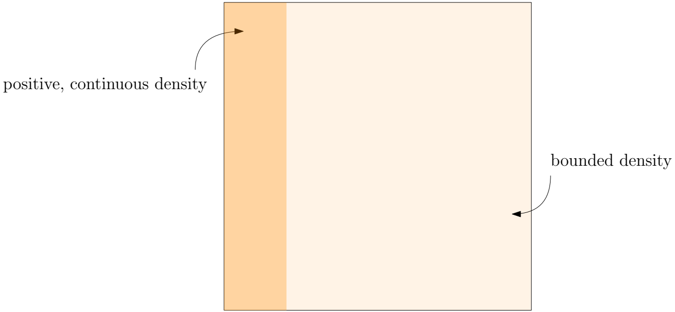
II. Universality of the height

Corsini – D. – Féray '24

Let μ be a permuton with a bounded density on $[0,1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0,1]$. Then:

$$\frac{h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right)}{c^* \log n} \longrightarrow 1$$

as $n \to \infty$, in probability and in L^p for all $p \geq 1$.



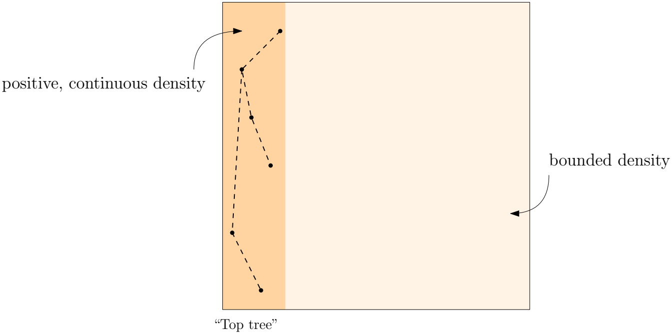
II. Universality of the height

Corsini – D. – Féray '24

Let μ be a permuton with a bounded density on $[0,1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0,1]$. Then:

$$\frac{h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right)}{c^* \log n} \longrightarrow 1$$

as $n \to \infty$, in probability and in L^p for all $p \geq 1$.



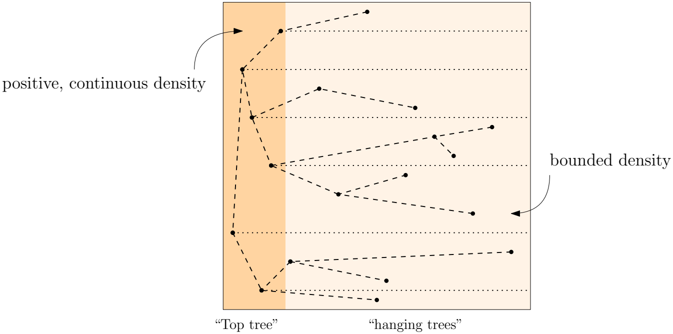
II. Universality of the height

Corsini – D. – Féray '24

Let μ be a permuton with a bounded density on $[0,1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0,1]$. Then:

$$\frac{h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right)}{c^* \log n} \longrightarrow 1$$

as $n \to \infty$, in probability and in L^p for all $p \geq 1$.



II. Universality of the height

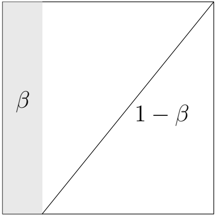
Corsini – D. – Féray '24

Let μ be a permuton with a bounded density on $[0,1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0,1]$. Then:

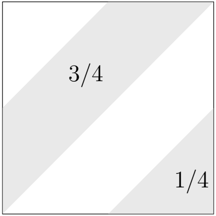
$$\frac{h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right)}{c^* \log n} \longrightarrow 1$$

as $n \to \infty$, in probability and in L^p for all $p \geq 1$.

Necessity of assumptions:

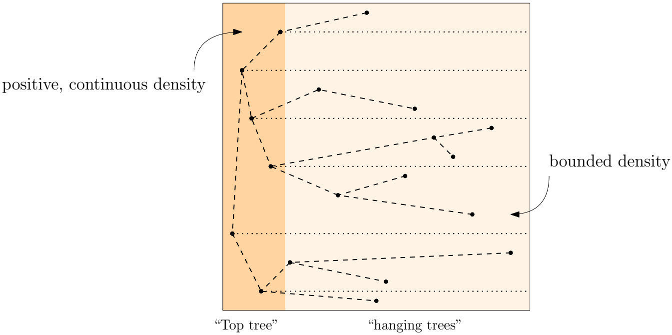


Height $\sim c_\beta \log n$ for some $c_\beta > c^*$



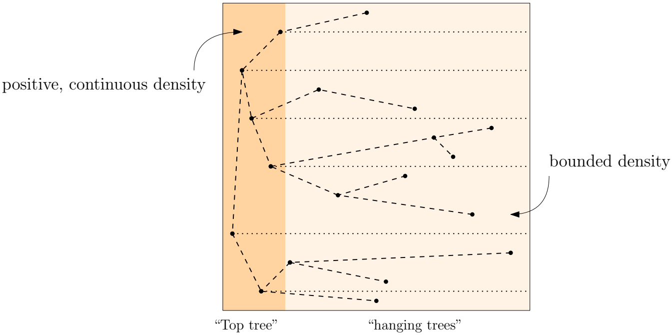
Height $\geq c\sqrt{n}$

III. Proof ideas



$$h\left(\mathcal{T}_{\mathrm{top}}\right) \;\leq\; h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right) \;\leq\; h\left(\mathcal{T}_{\mathrm{top}}\right) + 1 + \max_k h\left(\mathcal{T}^k_{\mathrm{hanging}}\right)$$

III. Proof ideas



$$h\left(\mathcal{T}_{\mathrm{top}}\right) \leq h\left(\mathcal{T}\langle\sigma_\mu^n\rangle\right) \leq h\left(\mathcal{T}_{\mathrm{top}}\right) + 1 + \max_k h\left(\mathcal{T}^k_{\mathrm{hanging}}\right)$$

Two steps:

1. $h\left(\mathcal{T}_{\mathrm{top}}\right) \sim c^* \log n$;
2. $\max_k h\left(\mathcal{T}^k_{\mathrm{hanging}}\right) = o\left(\log n\right)$.

III. Proof ideas

Two steps:

1. $h\left(\mathcal{T}_{\mathrm{top}}\right) \sim c^* \log n$;
2. $\max_k h\left(\mathcal{T}^k_{\mathrm{hanging}}\right) = o\left(\log n\right)$.

This is mainly done via coupling techniques and deviation estimates.

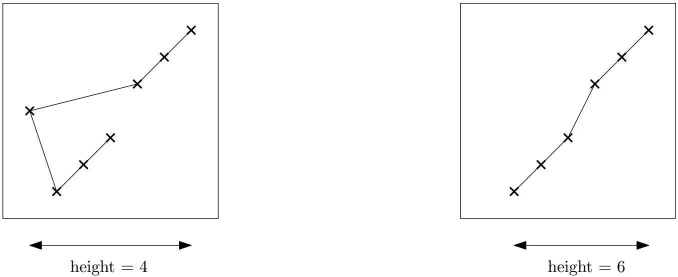
III. Proof ideas

Two steps:

1. $h\left(\mathcal{T}_{\text{top}}\right) \sim c^* \log n$;
2. $\max_k h\left(\mathcal{T}^k_{\text{hanging}}\right) = o\left(\log n\right)$.

This is mainly done via coupling techniques and deviation estimates.

Problem: taking out a point might double the height...



III. Proof ideas

Definition

A *chain* in a tree T is a subset C of vertices, which are all on a common branch of T.

III. Proof ideas

Definition

A *chain* in a tree T is a subset C of vertices, which are all on a common branch of T.

Lemma

Let $\mathcal{P}_- \subseteq \mathcal{P}_+$ be two point sets.
If $C \subseteq \mathcal{P}_+$ is a chain in $\mathcal{T}\langle\mathcal{P}_+\rangle$, then $C \cap \mathcal{P}_-$ is a chain in $\mathcal{T}\langle\mathcal{P}_-\rangle$.

III. Proof ideas

Definition

A *chain* in a tree T is a subset C of vertices, which are all on a common branch of T.

Lemma

Let $\mathcal{P}_- \subseteq \mathcal{P}_+$ be two point sets.
If $C \subseteq \mathcal{P}_+$ is a chain in $\mathcal{T}\langle\mathcal{P}_+\rangle$, then $C \cap \mathcal{P}_-$ is a chain in $\mathcal{T}\langle\mathcal{P}_-\rangle$.
If $\mathcal{C}$ is a chain of maximal size in $\mathcal{T}\langle\mathcal{P}_+\rangle$ then:

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

III. Proof ideas

Definition

A *chain* in a tree T is a subset C of vertices, which are all on a common branch of T.

Lemma

Let $\mathcal{P}_- \subseteq \mathcal{P}_+$ be two point sets.
If $C \subseteq \mathcal{P}_+$ is a chain in $\mathcal{T}\langle\mathcal{P}_+\rangle$, then $C \cap \mathcal{P}_-$ is a chain in $\mathcal{T}\langle\mathcal{P}_-\rangle$.
If $\mathcal{C}$ is a chain of maximal size in $\mathcal{T}\langle\mathcal{P}_+\rangle$ then:

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

If $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$, we can apply this lemma twice: with $\mathcal{P}_- \subseteq \mathcal{P}$, and with $\mathcal{P} \subseteq \mathcal{P}_+$. Thus:

$$h(\mathcal{T}\langle\mathcal{P}_+\rangle) - \underbrace{|\mathcal{C}_+ \cap (\mathcal{P}_+ \setminus \mathcal{P})|}_{\text{negligible?}} \leq h(\mathcal{T}\langle\mathcal{P}\rangle) \leq h(\mathcal{T}\langle\mathcal{P}_-\rangle) + \underbrace{|\mathcal{C} \cap (\mathcal{P} \setminus \mathcal{P}_-)|}_{\text{negligible?}}$$

III. Proof ideas

Hyp: μ has a density ρ, bounded above and below on $[0,\beta] \times [0,1]$.

Goal: height of the top tree.

III. Proof ideas

<u>Hyp</u>: μ has a density ρ, bounded above and below on $[0, \beta] \times [0, 1]$.

<u>Goal</u>: height of the top tree.

1. "Poissonization": easier to work with a Poisson point process $\mathcal{P}$ with intensity $n\rho$;

III. Proof ideas

Hyp: μ has a density ρ, bounded above and below on $[0, \beta] \times [0, 1]$.

Goal: height of the top tree.

1. "Poissonization": easier to work with a Poisson point process $\mathcal{P}$ with intensity $n\rho$;
2. "Thinning": there exist homogeneous Poisson point processes such that $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$
   $\longrightarrow$ estimate on $h(\mathcal{T}\langle\mathcal{P}\rangle)$;

III. Proof ideas

Hyp: μ has a density ρ, bounded above and below on $[0, \beta] \times [0, 1]$.

Goal: height of the top tree.

1. "Poissonization": easier to work with a Poisson point process $\mathcal{P}$ with intensity $n\rho$;
2. "Thinning": there exist homogeneous Poisson point processes such that $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$

   $\longrightarrow$ estimate on $h(\mathcal{T}\langle\mathcal{P}\rangle)$;
3. "dePoissonization" techniques.

III. Proof ideas

Hyp: μ has a density ρ, bounded above and below on $[0,\beta] \times [0,1]$.

Goal: height of the top tree.

1. "Poissonization": easier to work with a Poisson point process $\mathcal{P}$ with intensity $n\rho$;
2. "Thinning": there exist homogeneous Poisson point processes such that $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$

   $\longrightarrow$ (good?) estimate on $h(\mathcal{T}\langle\mathcal{P}\rangle)$;
3. "dePoissonization" techniques.

III. Proof ideas

Using Devroye's result:

If $\rho(x, y) = 1$, then σ_ρ^n is uniform and $h\left(\mathcal{T}\langle\sigma_\rho^n\rangle\right) \sim c^* \log n$.

III. Proof ideas

Using Devroye's result:

If $\rho(\mathsf{x}, \mathsf{y}) = \mathsf{f}(\mathsf{x}) \cdot \mathsf{g}(\mathsf{y})$, then σ_ρ^n is uniform and $h\left(\mathcal{T}\langle\sigma_\rho^n\rangle\right) \sim c^* \log n$.

III. Proof ideas

Using Devroye's result:

If $\rho(\mathrm{x}, \mathrm{y}) = \mathrm{f}(\mathrm{x}) \cdot \mathrm{g}(\mathrm{y})$, then σ_ρ^n is uniform and $h\left(\mathcal{T}\langle\sigma_\rho^n\rangle\right) \sim c^* \log n$.

$\longrightarrow$ if ρ on $[0, \beta] \times [0, 1]$ is close to a function that only depends on y, then the top tree is close to the BST of a uniform permutation.

III. Proof ideas

Using Devroye's result:

If $\rho(\mathrm{x},\mathrm{y}) = \mathrm{f}(\mathrm{x}) \cdot \mathrm{g}(\mathrm{y})$, then σ_ρ^n is uniform and $h\left(\mathcal{T}\langle\sigma_\rho^n\rangle\right) \sim c^* \log n$.

$\longrightarrow$ if ρ on $[0,\beta] \times [0,1]$ is close to a function that only depends on y, then the top tree is close to the BST of a uniform permutation.

- Hanging trees: the largest vertical gap of the points on the left is $\mathcal{O}(\log(n)/n)$, then deviation estimates for $h\left(\mathcal{T}^k_{\text{hanging}}\right)$ by comparison with longest monotone subsequences.

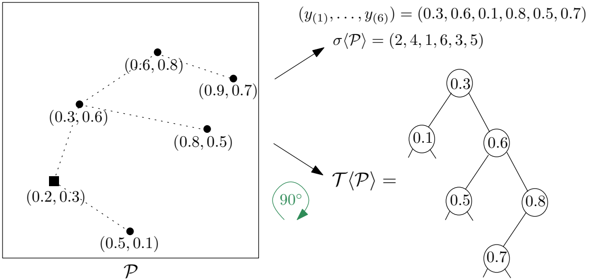
III. Proof ideas

Using Devroye's result:

If $\rho(\mathrm{x},\mathrm{y}) = \mathrm{f}(\mathrm{x})\cdot\mathrm{g}(\mathrm{y})$, then σ_ρ^n is uniform and $h\left(\mathcal{T}\langle\sigma_\rho^n\rangle\right) \sim c^* \log n$.

$\longrightarrow$ if ρ on $[0,\beta]\times[0,1]$ is close to a function that only depends on y, then the top tree is close to the BST of a uniform permutation.

- Hanging trees: the largest vertical gap of the points on the left is $\mathcal{O}(\log(n)/n)$, then deviation estimates for $h\left(\mathcal{T}^k_{\mathrm{hanging}}\right)$ by comparison with longest monotone subsequences.
- Another result: the *subtree size convergence*. The BST of a permuton sample is not "balanced" in the same way as the BST of a uniform permutation.



Thank you for your attention!