

Binary search trees of permuton samples

Benoît Corsini, Victor Dubach, Valentin Féray

JCB 2025, Bordeaux (LaBRI)

Outline

- I. BSTs and permutations
- II. Universality of the height
- III. Proof ideas
- IV. Subtree sizes

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

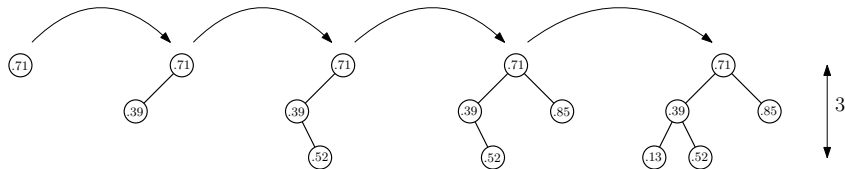
Binary Search Trees

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 $y = (y_1, \dots, y_n)$ 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 $y = (.71, .39, .52, .85, .13)$:



I. BSTs and permutons

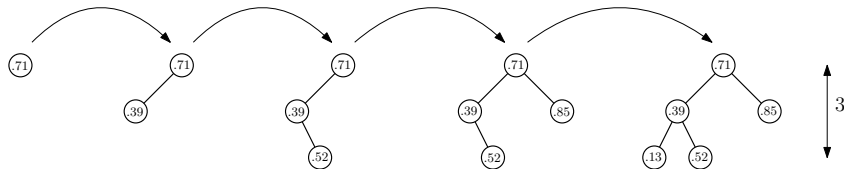
Binary Search Trees

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

Goal: Study $\mathcal{T}\langle\sigma\rangle$ where σ is a random permutation.

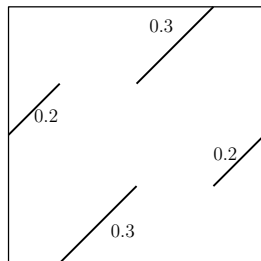
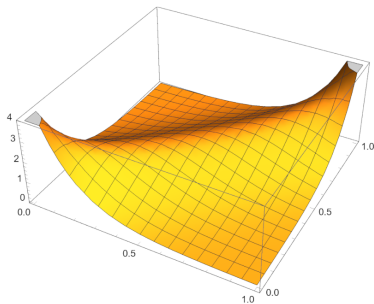
Q: What kind of randomness?

I. BSTs and permutons

Permutons

A *permuton* 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 $\mathcal{P} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.

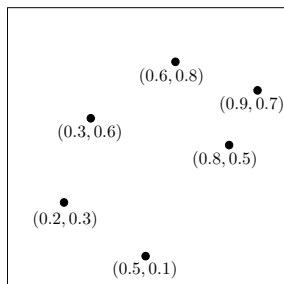
Define a reordering $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \dots < x_{(n)}$. There exists a unique permutation $\sigma = \sigma(\mathcal{P})$ such that $(y_{(1)}, \dots, y_{(n)})$ and $(\sigma(1), \dots, \sigma(n))$ are in the same relative order.

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 $x_{(1)} < \dots < x_{(n)}$. 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}

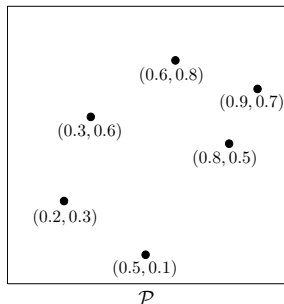
$$(y_{(1)}, \dots, y_{(6)}) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)$$

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

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*.



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

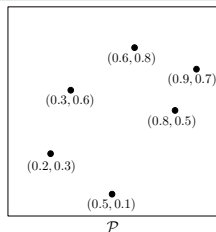
$$\sigma(\mathcal{P}) = (2, 4, 1, 6, 3, 5)$$

Rem: if $\mu = \text{Leb}_{[0,1]^2}$ 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*.



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

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

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

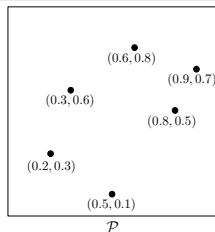
Motivation:

- 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*.



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

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

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

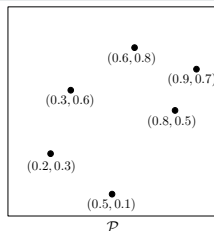
Motivation:

- 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*.



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

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

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
- Insight on the links between random permutations and their limit permuton

I. BSTs and permutons

More on permuton theory...

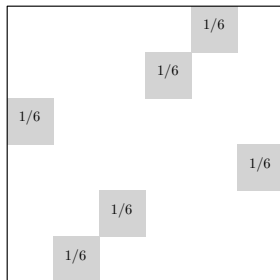
From permutations to permutons

To each permutation $\sigma \in S_n$ we can associate a permuton μ_σ with:

$$\frac{d\mu_\sigma}{d\text{Leb}_{[0,1]^2}} = n \cdot \sum_{i=1}^n \mathbf{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right]} \times \left[\frac{\sigma(i)-1}{n}, \frac{\sigma(i)}{n}\right].$$

$\sigma = 4\ 1\ 2\ 5\ 6\ 3$

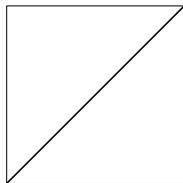
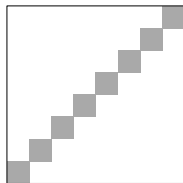
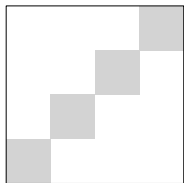
→



I. BSTs and permutons

Convergence of a sequence of permutons

If $(\sigma_n)_{n \in \mathbb{N}}$ is a sequence of permutations and μ is a permuton, we say that $(\sigma_n)_{n \in \mathbb{N}}$ converges towards μ when $\mu_{\sigma_n} \xrightarrow[n \rightarrow \infty]{} \mu$ weakly.

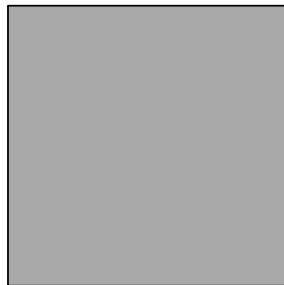
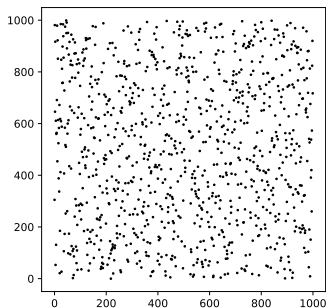


Identity permutations converge to the diagonal permuton.

I. BSTs and permutons

Convergence of a sequence of permutations

If $(\sigma_n)_{n \in \mathbb{N}}$ is a sequence of permutations and μ is a permuton, we say that $(\sigma_n)_{n \in \mathbb{N}}$ converges towards μ when $\mu_{\sigma_n} \xrightarrow[n \rightarrow \infty]{} \mu$ weakly.

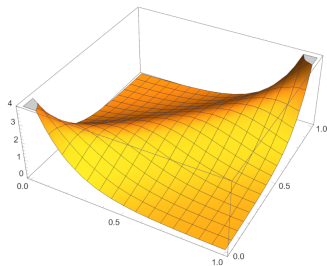
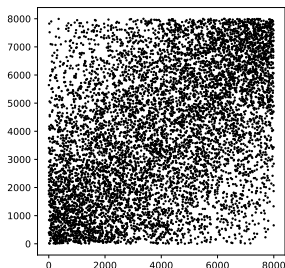


Uniform permutations converge a.s. to $\text{Leb}_{[0,1]^2}$.

I. BSTs and permutons

Convergence of a sequence of permutations

If $(\sigma_n)_{n \in \mathbb{N}}$ is a sequence of permutations and μ is a permuton, we say that $(\sigma_n)_{n \in \mathbb{N}}$ converges towards μ when $\mu_{\sigma_n} \xrightarrow[n \rightarrow \infty]{} \mu$ weakly.



Mallows permutations with parameter $q_n = 1 - \frac{\beta}{n}$ (cf later) converge in probability to the Mallows permuton μ_β .

I. BSTs and permutons

Convergence of a sequence of permutons

If $(\sigma_n)_{n \in \mathbb{N}}$ is a sequence of permutons and μ is a permuton, we say that $(\sigma_n)_{n \in \mathbb{N}}$ converges towards μ when $\mu_{\sigma_n} \xrightarrow[n \rightarrow \infty]{} \mu$ weakly.

Hoppen et al '13

For any permuton μ , the sequence of sampled permutons (σ_μ^n) converges a.s. to μ .

I. BSTs and permutons

Goal (reminder):

Study $\mathcal{T}\langle\sigma_\mu^n\rangle$ where σ_μ^n is sampled from a permuton μ .

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.

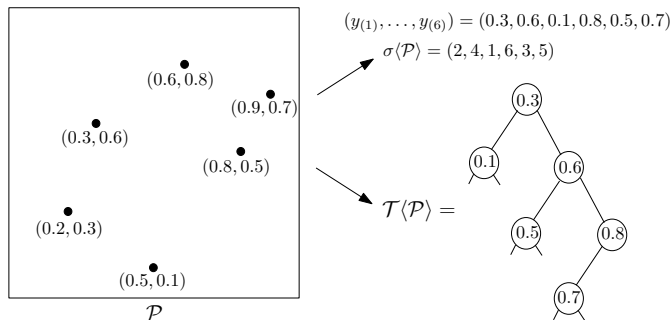
Define a reordering $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \dots < x_{(n)}$.
The BST of \mathcal{P} is defined as $\mathcal{T}(\mathcal{P}) = \mathcal{T}(y_{(1)}, \dots, y_{(n)})$.

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 $x_{(1)} < \dots < x_{(n)}$. The BST of \mathcal{P} is defined as $\mathcal{T}\langle\mathcal{P}\rangle = \mathcal{T}\langle y_{(1)}, \dots, y_{(n)}\rangle$.



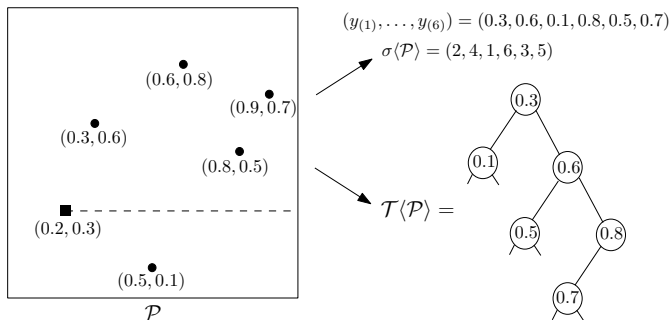
Rem: $\mathcal{T}\langle\mathcal{P}\rangle$ has the same shape as $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$.

I. BSTs and permutons

BST of a point process

Let $\mathcal{P} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.

Define a reordering $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \dots < x_{(n)}$.
The BST of \mathcal{P} is defined as $\mathcal{T}\langle\mathcal{P}\rangle = \mathcal{T}\langle y_{(1)}, \dots, y_{(n)}\rangle$.



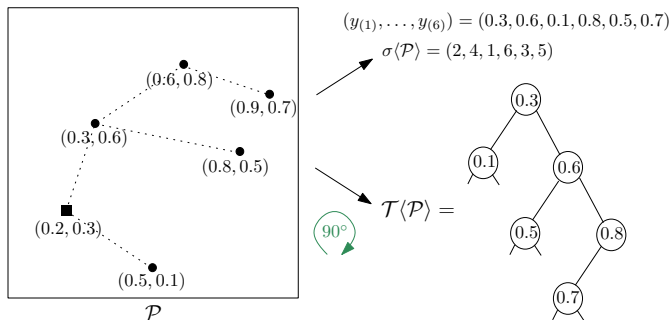
Rem: $\mathcal{T}\langle\mathcal{P}\rangle$ has the same shape as $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$.

I. BSTs and permutons

BST of a point process

Let $\mathcal{P} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset \mathbb{R}^2$ with no common x- or y-coordinate.

Define a reordering $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ by $x_{(1)} < \dots < x_{(n)}$.
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 $\mathcal{T}\langle\sigma(\mathcal{P})\rangle$.

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 $\mathcal{T}\langle\sigma\rangle$?

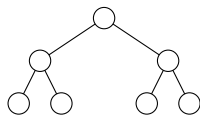
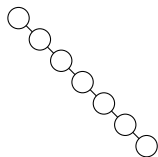
II. Universality of the height

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:

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



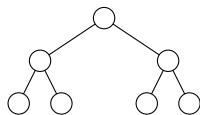
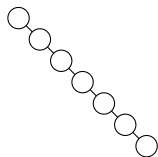
II. Universality of the height

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:

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



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

II. Universality of the height

Devroye '86

If $\sigma^n \sim \text{Unif}(\mathfrak{S}_n)$ then, as $n \rightarrow \infty$:

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

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 $\sigma^n \sim \text{Unif}(\mathfrak{S}_n)$ then, as $n \rightarrow \infty$:

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

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

Idea:

The root is labeled by $\sigma(1) \sim \text{Unif}([n])$; the left-subtree has size $\sigma(1) - 1$ and the right-subtree has size $n - \sigma(1)$.

II. Universality of the height

Devroye '86

If $\sigma^n \sim \text{Unif}(\mathfrak{S}_n)$ then, as $n \rightarrow \infty$:

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

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

Idea:

The root is labeled by $\sigma(1) \sim \text{Unif}([n])$; the left-subtree has size $\sigma(1) - 1$ and the right-subtree has size $n - \sigma(1)$.

At each node, the split between the left- and right-subtree is uniform \rightarrow the height is related to products of U_i or $1 - U_i$, where the U_i 's are i.i.d. uniform variables in $[0, 1]$.

II. Universality of the height

Devroye '86

If $\sigma^n \sim \text{Unif}(\mathfrak{S}_n)$ then, as $n \rightarrow \infty$:

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

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:

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

II. Universality of the height

Devroye '86

If $\sigma^n \sim \text{Unif}(\mathfrak{S}_n)$ then, as $n \rightarrow \infty$:

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

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:

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

Addario-Berry – Corsini '21

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

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

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

II. Universality of the height

Devroye '86

If $\sigma^n \sim \text{Unif}(\mathfrak{S}_n)$ then, as $n \rightarrow \infty$:

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

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

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

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

Corsini '23

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

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

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

II. Universality of the height

Assumption (A1):

μ has a bounded density ρ on $[0, 1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0, 1]$.

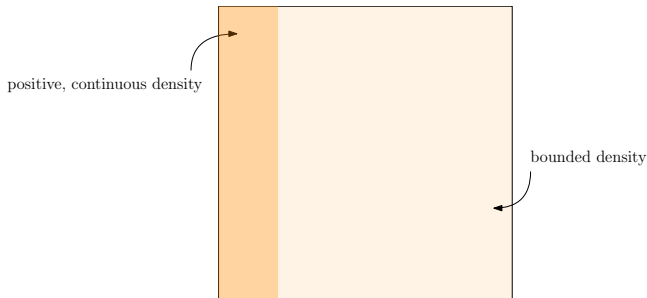
Corsini – D. – Féray '24

Let μ be a permuton satisfying (A1).

Then:

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

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



II. Universality of the height

Assumption (A1):

μ has a bounded density ρ on $[0, 1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0, 1]$.

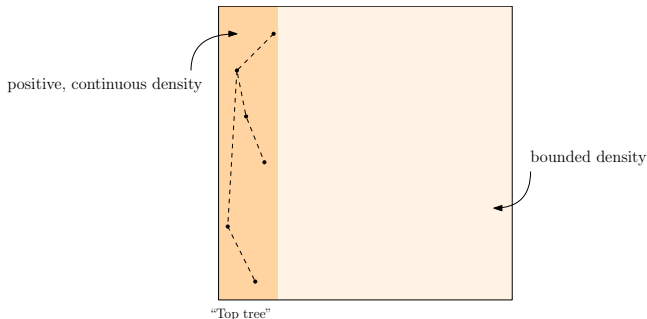
Corsini – D. – Féray '24

Let μ be a permuton satisfying (A1).

Then:

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

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



II. Universality of the height

Assumption (A1):

μ has a bounded density ρ on $[0, 1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0, 1]$.

Corsini – D. – Féray '24

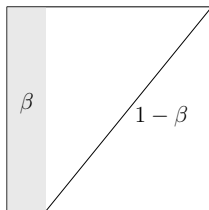
Let μ be a permuton satisfying (A1).

Then:

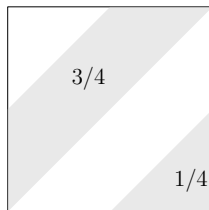
$$\frac{h(\mathcal{T}\langle\sigma_\mu^n\rangle)}{c^* \log n} \rightarrow 1$$

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 $\geq c\sqrt{n}$

II. Universality of the height

Assumption (A1):

μ has a bounded density ρ on $[0, 1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0, 1]$.

Corsini – D. – Féray '24

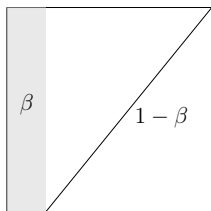
Let μ be a permuton satisfying (A1).

Then:

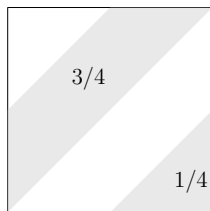
$$\frac{h(\mathcal{T}\langle\sigma_\mu^n\rangle)}{c^* \log n} \rightarrow 1$$

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

Question: Is there a permuton μ for which $\liminf \frac{h(\mathcal{T}\langle\sigma_\mu^n\rangle)}{\log n} < c^*$?

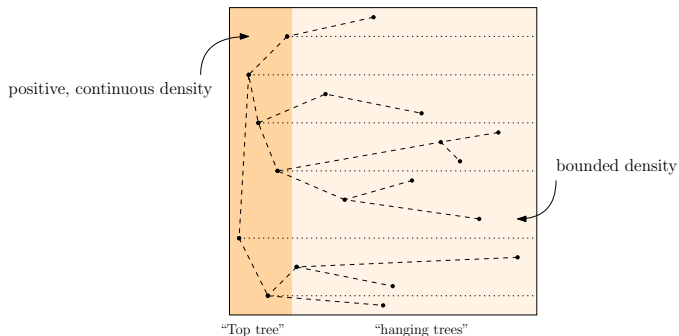


Height $\sim c_\beta \log n$ for some $c_\beta > c^*$



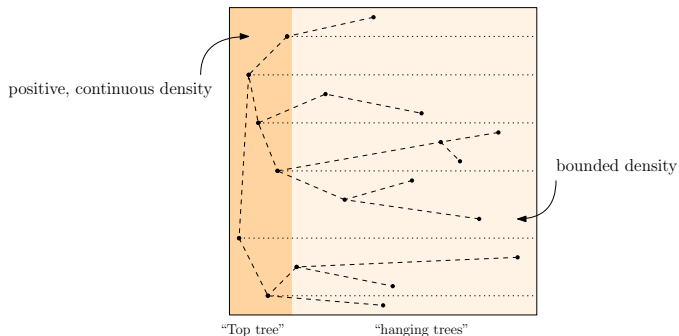
Height $\geq c\sqrt{n}$

III. Proof ideas



$$h(\mathcal{T}_{\text{top}}) \leq h(\mathcal{T}(\sigma_{\mu}^n)) \leq h(\mathcal{T}_{\text{top}}) + 1 + \max_k h(\mathcal{T}_{\text{hanging}}^k)$$

III. Proof ideas



$$h(\mathcal{T}_{\text{top}}) \leq h(\mathcal{T}(\sigma_{\mu}^n)) \leq h(\mathcal{T}_{\text{top}}) + 1 + \max_k h(\mathcal{T}_{\text{hanging}}^k)$$

Two steps:

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

III. Proof ideas

Two steps:

1 $h(\mathcal{T}_{\text{top}}) \sim c^* \log n;$

2 $\max_k h(\mathcal{T}_{\text{hanging}}^k) = o(\log n).$

This is mainly done via coupling techniques and deviation estimates.

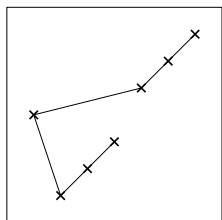
III. Proof ideas

Two steps:

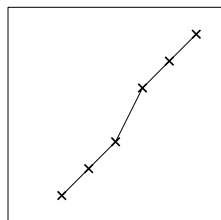
- 1 $h(\mathcal{T}_{\text{top}}) \sim c^* \log n$;
- 2 $\max_k h(\mathcal{T}_{\text{hanging}}^k) = o(\log n)$.

This is mainly done via coupling techniques and deviation estimates.

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



height = 4



height = 6

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 C is a chain of maximal size in $\mathcal{T}\langle\mathcal{P}_+\rangle$ then:

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

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 $\mathcal{T}(\mathcal{P}_+)$, then $C \cap \mathcal{P}_-$ is a chain in $\mathcal{T}(\mathcal{P}_-)$.

If C is a chain of maximal size in $\mathcal{T}(\mathcal{P}_+)$ then:

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

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:

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

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

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 $n\rho$;

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 $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$
→ 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 $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$
→ 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 $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$
→ (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 $h(\mathcal{T}\langle\sigma_\rho^n\rangle) \sim c^* \log n$.

III. Proof ideas

Using Devroye's result:

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

III. Proof ideas

Using Devroye's result:

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

→ 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

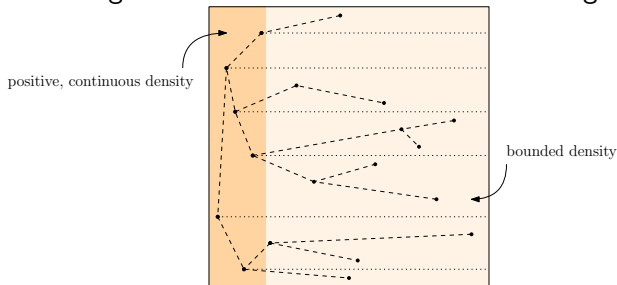
Using Devroye's result:

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

→ 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.

Top tree:
height $\sim c^* \log n$

Hanging trees:
height $\ll \log n$?



III. Proof ideas

Lemma

Let $\{Y_i, 1 \leq i \leq n\}$ be i.i.d. random variables with law μ_0 on $[0, 1]$, such that μ_0 has a positive density. Let $Y_{(1)} < \dots < Y_{(n)}$ be their increasing reordering, and $\zeta_k := Y_{(k+1)} - Y_{(k)}$ be the k -th spacing. Then:

$$\max_k \zeta_k = \mathcal{O}_{\mathbb{P}} \left(\frac{\log n}{n} \right).$$

III. Proof ideas

Lemma

Let $\{Y_i, 1 \leq i \leq n\}$ be i.i.d. random variables with law μ_0 on $[0, 1]$, such that μ_0 has a positive density. Let $Y_{(1)} < \dots < Y_{(n)}$ be their increasing reordering, and $\zeta_k := Y_{(k+1)} - Y_{(k)}$ be the k -th spacing. Then:

$$\max_k \zeta_k = \mathcal{O}_{\mathbb{P}} \left(\frac{\log n}{n} \right).$$

Hanging trees: the largest vertical gap of the points on the left is $\mathcal{O}(\log(n)/n)$, so the maximum number of points in a hanging tree is $\mathcal{O}(\log n)$.

III. Proof ideas

Lemma

Let $\{Y_i, 1 \leq i \leq n\}$ be i.i.d. random variables with law μ_0 on $[0, 1]$, such that μ_0 has a positive density. Let $Y_{(1)} < \dots < Y_{(n)}$ be their increasing reordering, and $\zeta_k := Y_{(k+1)} - Y_{(k)}$ be the k -th spacing. Then:

$$\max_k \zeta_k = \mathcal{O}_{\mathbb{P}} \left(\frac{\log n}{n} \right).$$

Hanging trees: the largest vertical gap of the points on the left is $\mathcal{O}(\log(n)/n)$, so the maximum number of points in a hanging tree is $\mathcal{O}(\log n)$.

A *typical* hanging tree should have height $\mathcal{O}(\log \log n)$. Can we argue that they *all* have height $o(\log n)$?

III. Proof ideas

Lemma

Let $\{Y_i, 1 \leq i \leq n\}$ be i.i.d. random variables with law μ_0 on $[0, 1]$, such that μ_0 has a positive density. Let $Y_{(1)} < \dots < Y_{(n)}$ be their increasing reordering, and $\zeta_k := Y_{(k+1)} - Y_{(k)}$ be the k -th spacing. Then:

$$\max_k \zeta_k = \mathcal{O}_{\mathbb{P}} \left(\frac{\log n}{n} \right).$$

Hanging trees: the largest vertical gap of the points on the left is $\mathcal{O}(\log(n)/n)$, so the maximum number of points in a hanging tree is $\mathcal{O}(\log n)$.

A *typical* hanging tree should have height $\mathcal{O}(\log \log n)$. Can we argue that they *all* have height $o(\log n)$?

→ comparison arguments along with deviation estimates.

III. Proof ideas

Comparison with monotone subsequences

$$h(\mathcal{T}\langle\sigma\rangle) \leq \text{LIS}(\sigma) + \text{LDS}(\sigma) .$$

III. Proof ideas

Comparison with monotone subsequences

$$h(\mathcal{T}\langle\sigma\rangle) \leq \text{LIS}(\sigma) + \text{LDS}(\sigma) .$$

Extreme deviations for monotone subsequences

If σ^n is a uniformly random permutation of $[n]$ then:

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

III. Proof ideas

Comparison with monotone subsequences

$$h(\mathcal{T}\langle\sigma\rangle) \leq \text{LIS}(\sigma) + \text{LDS}(\sigma) .$$

Extreme deviations for monotone subsequences

If σ^n is a uniformly random permutation of $[n]$ then:

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

→ Since $\rho \leq M$, we can write $\mathcal{P} \subseteq \mathcal{P}_+$ (homogeneous PPP)

III. Proof ideas

Comparison with monotone subsequences

$$h(\mathcal{T}\langle\sigma\rangle) \leq \text{LIS}(\sigma) + \text{LDS}(\sigma) .$$

Extreme deviations for monotone subsequences

If σ^n is a uniformly random permutation of $[n]$ then:

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

→ Since $\rho \leq M$, we can write $\mathcal{P} \subseteq \mathcal{P}_+$ (homogeneous PPP)

→ Deviation inequality for $\text{LIS}(\mathcal{P}_{\text{hanging}}^k) \leq \text{LIS}(\mathcal{P}_+^k)$ on each horizontal band

III. Proof ideas

Comparison with monotone subsequences

$$h(\mathcal{T}\langle\sigma\rangle) \leq \text{LIS}(\sigma) + \text{LDS}(\sigma) .$$

Extreme deviations for monotone subsequences

If σ^n is a uniformly random permutation of $[n]$ then:

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

→ Since $\rho \leq M$, we can write $\mathcal{P} \subseteq \mathcal{P}_+$ (homogeneous PPP)

→ Deviation inequality for $\text{LIS}(\mathcal{P}_{\text{hanging}}^k) \leq \text{LIS}(\mathcal{P}_+^k)$ on each horizontal band

→ All hanging trees have height $\ll \log n$.

IV. Subtree sizes

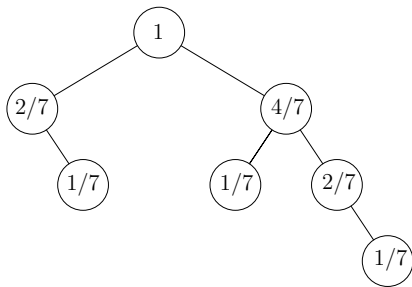
Even when the height is the same, can we detect the difference between two permutation BSTs?

IV. Subtree sizes

Even when the height is the same, can we detect the difference between two permutation BSTs?

Subtree sizes

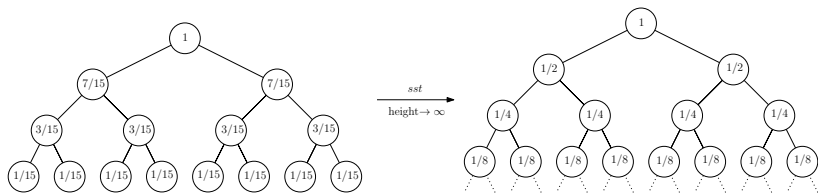
Let \mathcal{T} be a finite rooted tree. If $v \in \mathcal{T}$, we let $t(\mathcal{T}, v)$ be the proportion of nodes that are descendants of v (including itself).



IV. Subtree sizes

Subtree size convergence [Grübel '23]

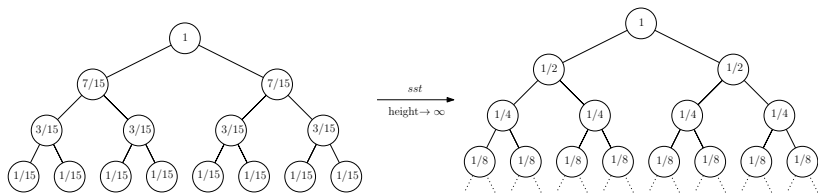
Let (\mathcal{T}_n) be a sequence of binary trees. We say that it converges for the *subtree size topology* if $t(\mathcal{T}_n, v)$ converges for all fixed v . The decoration of the infinite binary tree with these limits is the *subtree size limit*.



IV. Subtree sizes

Subtree size convergence [Grübel '23]

Let (\mathcal{T}_n) be a sequence of binary trees. We say that it converges for the *subtree size topology* if $t(\mathcal{T}_n, v)$ converges for all fixed v . The decoration of the infinite binary tree with these limits is the *subtree size limit*.



Rem: The limit object can be thought of as a probability measure on infinite paths.

IV. Subtree sizes

Subtree size limit sampled by a measure

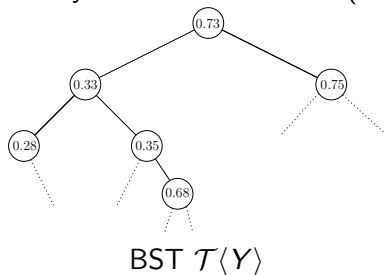
Let m be a probability measure on $[0, 1]$ with no atom. Let (Y_1, Y_2, \dots) be i.i.d. under m , and let $\mathcal{T}\langle Y \rangle$ be their infinite BST. For each node v , let $\psi_m(v)$ be the difference between the labels of the leftmost ancestor to the right of v , and the rightmost ancestor to the left of v . Thus: ψ_m is a *random* subtree size limit.

IV. Subtree sizes

Subtree size limit sampled by a measure

Let m be a probability measure on $[0, 1]$ with no atom. Let (Y_1, Y_2, \dots) be i.i.d. under m , and let $\mathcal{T}\langle Y \rangle$ be their infinite BST. For each node v , let $\psi_m(v)$ be the difference between the labels of the leftmost ancestor to the right of v , and the rightmost ancestor to the left of v . Thus: ψ_m is a *random* subtree size limit.

Ex: say $m = 2xdx$ and $Y = (0.73, 0.33, 0.75, 0.35, 0.68, 0.28, \dots)$.

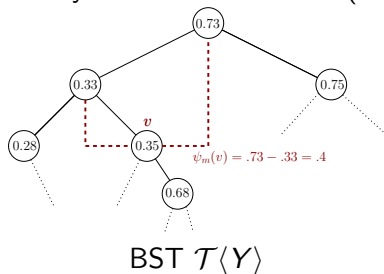


IV. Subtree sizes

Subtree size limit sampled by a measure

Let m be a probability measure on $[0,1]$ with no atom. Let (Y_1, Y_2, \dots) be i.i.d. under m , and let $\mathcal{T}\langle Y \rangle$ be their infinite BST. For each node v , let $\psi_m(v)$ be the difference between the labels of the leftmost ancestor to the right of v , and the rightmost ancestor to the left of v . Thus: ψ_m is a *random* subtree size limit.

Ex: say $m = 2xdx$ and $Y = (0.73, 0.33, 0.75, 0.35, 0.68, 0.28, \dots)$.

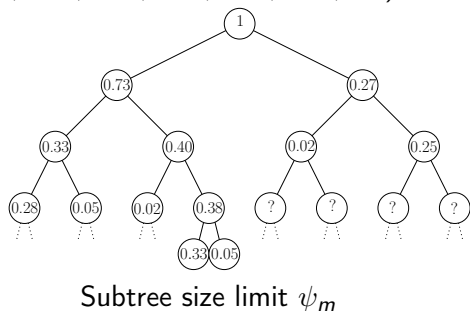
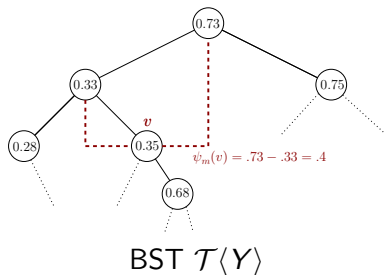


IV. Subtree sizes

Subtree size limit sampled by a measure

Let m be a probability measure on $[0,1]$ with no atom. Let (Y_1, Y_2, \dots) be i.i.d. under m , and let $\mathcal{T}\langle Y \rangle$ be their infinite BST. For each node v , let $\psi_m(v)$ be the difference between the labels of the leftmost ancestor to the right of v , and the rightmost ancestor to the left of v . Thus: ψ_m is a *random* subtree size limit.

Ex: say $m = 2xdx$ and $Y = (0.73, 0.33, 0.75, 0.35, 0.68, 0.28, \dots)$.



IV. Subtree sizes

Assumption (A2):

There exists a measure μ_0 on $[0, 1]$ with no atoms such that

$$\frac{1}{x} \mu([0, x] \times \cdot) \xrightarrow{x \rightarrow 0^+} \mu_0. \quad (1)$$

(A2) is weaker than (A1).

IV. Subtree sizes

Assumption (A2):

There exists a measure μ_0 on $[0, 1]$ with no atoms such that

$$\frac{1}{x} \mu([0, x] \times \cdot) \xrightarrow{x \rightarrow 0^+} \mu_0. \quad (1)$$

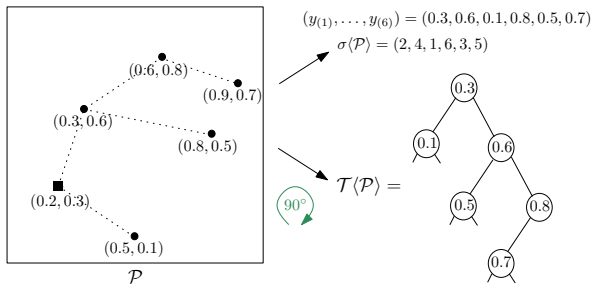
(A2) is weaker than (A1).

Corsini – D. – Féray '24

- If μ satisfies (A2) then, in distribution:

$$\mathcal{T}\langle \mathcal{P}_\mu^n \rangle \xrightarrow{\text{SSC}} \psi_{\mu_0}.$$

- If $\mathcal{T}\langle \mathcal{P}_\mu^n \rangle$ converges in distribution for the subtree size topology then (1) is satisfied, but μ_0 may have atoms.



Thank you for your attention!

