Binary search trees of permutohedral samples

Benoît Corsini, Victor Dubach, Valentin Féray

AofA 2024, Bath

Full version at arXiv:2403.03151
I. BSTs and permutons

II. Universality of the height

III. Proof ideas
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 \( T\langle y \rangle \) by successively adding leaves with those labels.

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

Rem: same shape as \( T\langle \sigma \rangle \) with \( \sigma = (4, 2, 3, 5, 1) \).
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 \( T\langle y \rangle \) by successively adding leaves with those labels.

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

Rem: same shape as \( T\langle \sigma \rangle \) with \( \sigma = (4, 2, 3, 5, 1) \).
I. BSTs and permutons

Permutons

A *permuton* (or *copula*) is a probability measure $\mu$ 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 $\sigma = \sigma(\mathcal{P})$ such that $(y(1), \ldots, y(n))$ and $(\sigma(1), \ldots, \sigma(n))$ are in the same relative order.
1. BSTs and permutons

### Permuton samples

Let \( P = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2 \) 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 \( \sigma = \sigma(P) \) such that \((y(1), \ldots, y(n))\) and \((\sigma(1), \ldots, \sigma(n))\) are in the same relative order.

\[
(y(1), \ldots, y(6)) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7)
\]

\[
\sigma(P) = (2, 4, 1, 6, 3, 5)
\]
I. BSTs and permutons

**Permuton samples**

If $\mathcal{P}$ is random i.i.d. under a permuton $\mu$, this permutation $\sigma^n_\mu$ is a **permuton sample**.

Rem: if $\mu = \text{Leb}_{[0,1]^2}$ then $\sigma^n_\mu$ is uniformly random.
I. BSTs and permutons

Permuton samples

If $\mathcal{P}$ is random i.i.d. under a permuton $\mu$, this permutation $\sigma^n_{\mu}$ is a permuton sample.

Rem: if $\mu = \text{Leb}_{[0,1]^2}$ then $\sigma^n_{\mu}$ 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 $\mu$, this permutation $\sigma^n_\mu$ is a *permuton sample*.

\[
\begin{array}{c}
(0.2, 0.3) \\
(0.3, 0.6) \\
(0.5, 0.1) \\
(0.6, 0.8) \\
(0.8, 0.5) \\
(0.9, 0.7)
\end{array}
\]

\[\begin{align*}
(\mathcal{P}) & = (2, 4, 1, 6, 3, 5) \\
(\mathcal{P}) & = (0.2, 0.3, 0.6, 0.1, 0.8, 0.5, 0.7)
\end{align*}\]

- Rem: if $\mu = \text{Leb}_{[0,1]^2}$ then $\sigma^n_\mu$ 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 $\mu$, this permutation $\sigma^\mu_n$ is a permuton sample.

Rem: if $\mu = \text{Leb}_{[0,1]^2}$ then $\sigma^\mu_n$ 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
### BST of a point process

Let \( P = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2 \) 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 \( P \) is defined as \( T\langle P \rangle = T\langle y_1, \ldots, y_n \rangle \).
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 \( P = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^2 \) 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 \( P \) is defined as \( \mathcal{T}\langle P\rangle = \mathcal{T}\langle y(1), \ldots, y(n)\rangle \).

Rem: \( \mathcal{T}\langle P\rangle \) has the same shape as \( \mathcal{T}\langle \sigma\langle P\rangle\rangle \).
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 \( \mathcal{T}(\mathcal{P}) = \mathcal{T}(y(1), \ldots, y(n)) \).

\[ (y(1), \ldots, y(6)) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7) \]

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

\[ \mathcal{T}(\mathcal{P}) = \]

\[ \mathcal{T}(\mathcal{P}) \] has the same shape as \( \mathcal{T}(\sigma(\mathcal{P})) \).
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 $\sigma$ be a (random) permutation. What is the height of $T(\sigma)$?
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 $\sigma$ be a (random) permutation. What is the height of $T(\sigma)$?

In general:

$$n \geq h(T(\sigma(1), \ldots, \sigma(n))) \geq \log_2(n + 1)$$

\[ \text{Diagram of a long, thin tree} \quad \text{Diagram of a bushy tree} \]
The height of a BST determines the complexity of operations such as lookup, addition or removal of data.

Q: Let $\sigma$ be a (random) permutation. What is the height of $T<\sigma>$?

In general:

$$n \geq h(T<\sigma(1), \ldots, \sigma(n)>) \geq \log_2(n + 1)$$

Depending on the law of $\sigma$, asymptotic results for $h(T<\sigma>)$?
## II. Universality of the height

### Devroye ’86

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

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

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:

$$\forall \sigma \in \mathcal{S}_n, M_n, q(\sigma) \propto q^{\text{inv}(\sigma)}.$$
II. Universality of the height

Devroye ’86

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

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

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 $n(1 - q_n)/\log n \to 0$, then:

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

in probability and in $L^p$ for all $p \geq 1$. 
II. Universality of the height

Devroye ’86

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

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

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:

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

Corsini ’23

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

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

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

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

as $n \rightarrow \infty$, in probability and in $L^p$ for all $p \geq 1$. 
II. Universality of the height

Let $\mu$ 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:

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

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

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

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

---

**Diagram notes:**
- Positive, continuous density
- Bounded density
- "Top tree" and "hanging trees"
II. Universality of the height

Let $\mu$ 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:

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

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

Necessity of assumptions:

$\beta$

$1 - \beta$

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

$3/4$

$1/4$

Height $\geq c\sqrt{n}$
III. Proof ideas

\[ h(T_{\text{top}}) \leq h(T^{n}) \leq h(T_{\text{top}}) + 1 + \max_{k} h(T_{\text{hanging}}^{k}) \]
III. Proof ideas

Two steps:

1. $h(T_{\text{top}}) \sim c^* \log n$;

2. $\max_k h\left(T^k_{\text{hanging}}\right) = o(\log n)$. 

\[
 h(T_{\text{top}}) \leq h(T\langle \sigma^n_{\mu} \rangle) \leq h(T_{\text{top}}) + 1 + \max_k h\left(T^k_{\text{hanging}}\right)
\]
III. Proof ideas

Two steps:

1. \( h(T_{\text{top}}) \sim c^* \log n; \)

2. \( \max_k h(T_{\text{hanging}}^k) = o(\log n). \)

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. \( \max_{k} h\left(\mathcal{T}_{\text{hanging}}^{k}\right) = o(\log n). \)

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 $T$. 

---

**Lemma**

Let $P^- \subseteq P^+$ be two point sets. If $C \subseteq P^+$ is a chain in $T(P^+)$, then $C \cap P^-$ is a chain in $T(P^-)$. If $C$ is a chain of maximal size in $T(P^+)$ then:

$$h(T(P^-)) \geq h(T(P^+)) - |C \cap (P^+ \setminus P^-)|$$

If $P^- \subseteq P \subseteq P^+$, we can apply this lemma twice: with $P^- \subseteq P$, and with $P \subseteq P^+$. Thus:

$$h(T(P^+)) - |C^+ \cap (P^+ \setminus P^-)| \text{ negligible?} \leq h(T(P)) \leq h(T(P^-)) + |C \cap (P \setminus P^-)| \text{ negligible?}$$
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 $\mathcal{T}(\mathcal{P}_-)$. 
### 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 $\mathcal{T}\langle \mathcal{P}_+ \rangle$, 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:

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

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:

$$h(\mathcal{T}\langle \mathcal{P}_+ \rangle) - |C_+ \cap (\mathcal{P}_+ \setminus \mathcal{P})| \leq h(\mathcal{T}\langle \mathcal{P} \rangle) \leq h(\mathcal{T}\langle \mathcal{P}_- \rangle) + |C \cap (\mathcal{P} \setminus \mathcal{P}_-)| \text{ negligible?}$$
Hyp: $\mu$ has a density $\rho$, bounded above and below on $[0, \beta] \times [0, 1]$.

Goal: height of the top tree.
Hyp: $\mu$ has a density $\rho$, 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: $\mu$ has a density $\rho$, 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 $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$

$\rightarrow$ estimate on $h(T\langle\mathcal{P}\rangle)$;
III. Proof ideas

Hyp: $\mu$ has a density $\rho$, 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 $\mathcal{P}_- \subseteq \mathcal{P} \subseteq \mathcal{P}_+$
   $\rightarrow$ estimate on $h(\mathcal{T}(\mathcal{P}))$;
3. “dePoissonization” techniques.
III. Proof ideas

**Hyp:** $\mu$ has a density $\rho$, 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 $\sigma^n_\rho$ is uniform and $h(\mathcal{T}_\langle \sigma^n_\rho \rangle) \sim c^* \log n$. 
III. Proof ideas

Using Devroye’s result:

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

Hanging trees: the largest vertical gap of the points on the left is $O(\log(n)/n)$, then deviation estimates for $h(T)$ hanging 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.
III. Proof ideas

Using Devroye’s result:

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

$\longrightarrow$ if $\rho$ 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 $\sigma^n_\rho$ is uniform and $h\left(\mathcal{T}\langle \sigma^n_\rho \rangle\right) \sim c^* \log n$.

$\rightarrow$ if $\rho$ 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 $\mathcal{O}\left(\log(n)/n\right)$, 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 $\rho(x, y) = f(x) \cdot g(y)$, then $\sigma^n_\rho$ is uniform and $h(\mathcal{T}(\sigma^n_\rho)) \sim c^* \log n$.

$\rightarrow$ if $\rho$ 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 $O(\log(n)/n)$, then deviation estimates for $h\left(\mathcal{T}^k_{\text{hanging}}\right)$ 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.
\[ (y_1, \ldots, y_6) = (0.3, 0.6, 0.1, 0.8, 0.5, 0.7) \]
\[ \sigma(P) = (2, 4, 1, 6, 3, 5) \]
\[ T(P) = \]

Thank you for your attention!