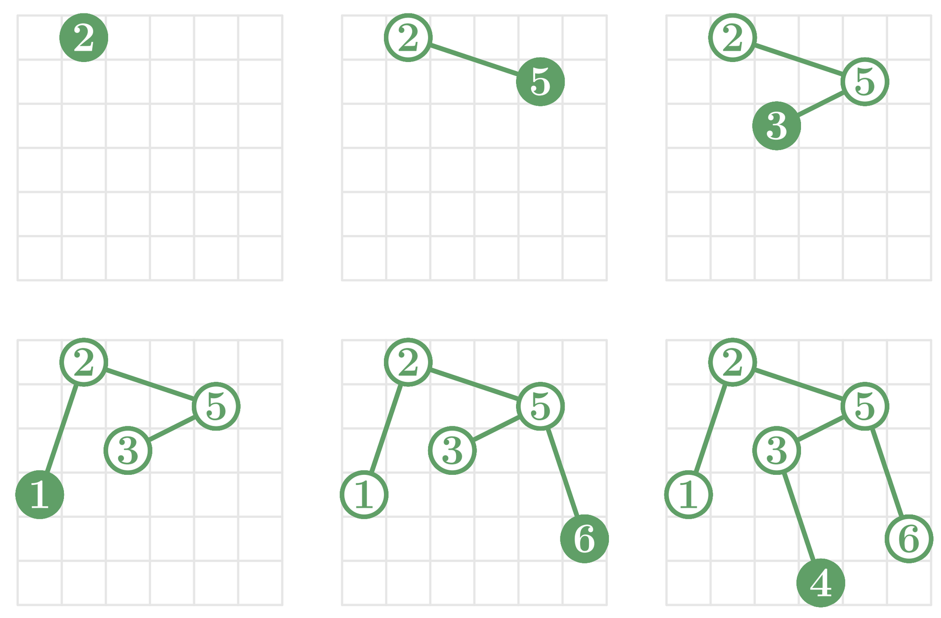
Random models of binary search trees

presented by Benoît Corsini

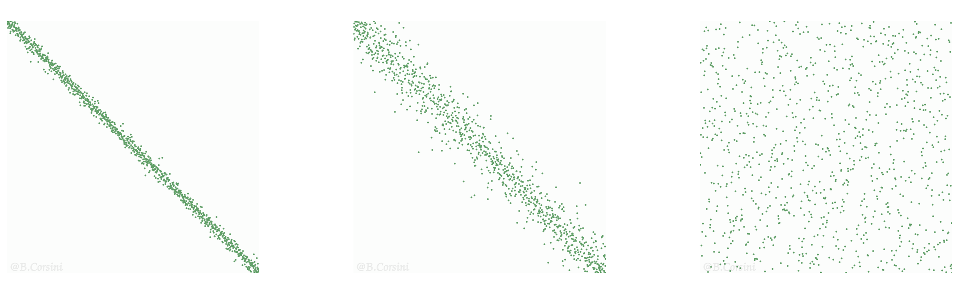
Binary search trees

Binary search trees are labelled binary tree such that the label of each node is larger than the labels of each node of its left subtree and smaller than the labels of each node of its right subtree. In particular, given a sequence of numbers, there exists a unique binary search tree obtained by inserting the values one after the other.

The binary search tree of the sequence $(2, 5, 3, 1, 6, 4)$ is provided on the right as an example.

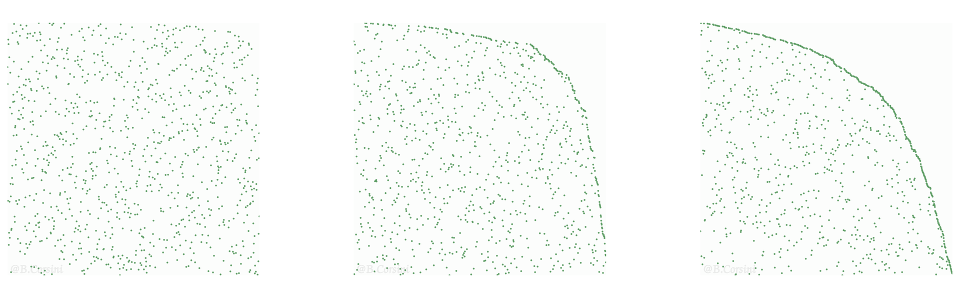


Mallows permutations ($n = 1000$)



$q = 0.9$ $q = 0.97$ $q = 1$

record-biased permutations ($n = 1000$)



$\theta = 5$ $\theta = 50$ $\theta = 200$

Random models of permutations

Definition (Mallows permutation)
For any $n \in \mathbb{N}$ and $q \in [0, \infty)$, the Mallows distribution is defined by

$$\pi_{n,q}(\sigma) = \frac{q^{\mathrm{Inv}(\sigma)}}{Z_{n,q}},$$

where $\mathrm{Inv}(\sigma) = |\{i < j : \sigma(i) > \sigma(j)\}|$ is the number of inversions of σ and $Z_{n,q}$ is a normalizing constant.

Definition (Record-biased permutation)
For any $n \in \mathbb{N}$ and $\theta \in [0, \infty)$, the record-biased distribution is defined by

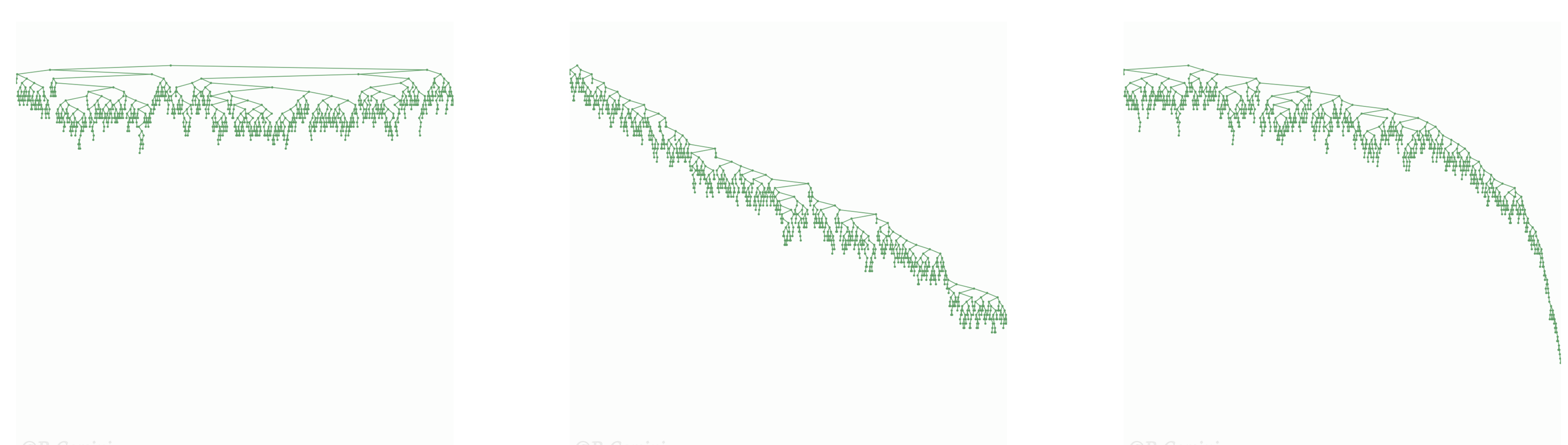
$$w_{n,\theta}(\sigma) = \frac{\theta^{\mathrm{Rec}(\sigma)}}{W_{n,q}},$$

where $\mathrm{Rec}(\sigma) = |\{i : \forall j < i, \sigma(i) > \sigma(j)\}|$ is the number of records of σ and $W_{n,q}$ is a normalizing constant.

Models of binary search trees

Consider the following random models of binary search trees.

- T_n for a random binary search tree (uniform permutation).
- $T^M_{n,q}$ for a Mallows tree with parameters n and q.
- $T^R_{n,\theta}$ for a record-biased tree with parameters n and θ.



random binary search tree — Mallows tree — record-biased tree

Results

We aimed at extending the following result on the asymptotic height of random binary search trees.

Theorem (Devroye, 1986)
Let $c^* \simeq 4.311\ldots$ be the unique solution to $c\log(2e/c) = 1$ with $c \geq 2$. Then

$$\frac{h(T_n)}{c^* \log n} \longrightarrow 1,$$

where the convergence occurs in probability and L_p for all $p > 0$.

We proved similar results to the case of Mallows trees and record-biased trees, thus extending this theorem.

In the case of Mallows trees, by mirroring $T^M_{n,q}$ we obtain a tree distributed as $T^M_{n,1/q}$, so $h(T^M_{n,q})$ is distributed as $h(T^M_{n,1/q})$ and we reduce q to $[0, 1]$ instead of $[0, \infty)$.

Theorem (Addario-Berry & , 2021)
For any $(q_n)_{n\geq 1}$ in $[0, 1]$, we have

$$\frac{h(T^M_{n,q_n})}{n(1 - q_n) + c^* \log n} \longrightarrow 1$$

where the convergence occurs in probability and L_p for all $p > 0$.

Theorem (, 2023)
For any $(\theta_n)_{n\geq 1}$ in $[0, \infty)$, we have

$$\frac{h(T^R_{n,\theta_n})}{\max\{c^* \log n, \theta_n \log(1 + n/\theta_n)\}} \longrightarrow 1$$

where the convergence occurs in probability and L_p for all $p > 0$.

In the case of Mallows trees, we also further prove second order behaviour in the case where $n(1 - q_n)/\log n \to \infty$.

Upcoming work

We are currently applying these results to sets of points obtained from a permuton. These results would once again extend the case of random binary search trees and cover some cases of Mallows trees and record-biased trees.

References

- **Addario-Berry, L., & Corsini, B. (2021).** The height of Mallows trees. *The Annals of Probability, 49(5), 2220-2271.*
- **Auger, N., Bouvel, M., Nicaud, C., & Pivoteau, C. (2016).** Analysis of algorithms for permutations biased by their number of records. *In 27th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithm 2016.*
- **Corsini, B. (2023).** The height of record-biased trees. *Random Structures & Algorithms, 62(3), 623-644.*
- **Devroye, L. (1986).** A note on the height of binary search trees. *Journal of the ACM (JACM), 33(3), 489-498.*
- **Mallows, C. L. (1957).** Non-null ranking models. I. *Biometrika, 44(1/2), 114-130.*

https://www.benoitcorsini.com
Eindhoven University of Technology