The height of Mallows trees

Benoît Corsini
McGill University

Overview

This poster presents a work with Louigi Addario-Berry [1] on the height of Mallows trees. Mallows trees are binary search trees drawn from Mallows permutations. We prove asymptotics for their height, extending results on the height of random binary search trees.

Binary search trees

Let T_∞ be the infinite binary tree, identified with words on $\{\overline{0}, \overline{1}\}$. A subtree of T_∞ is a connected subset of T_∞ and its root is the only node which is a prefix of all its other nodes. If v is the root of T, then T is said to be rooted at v.

A *binary search tree* is a tree T, with labelling function τ such that $\tau(v) > \tau(u)$ (respectively $\tau(v) < \tau(u)$) for all $u \in T$ such that $v\overline{0}$ (respectively $v\overline{1}$) is a prefix of u. For a permutation $\sigma \in \mathcal{S}_n$, let T_σ be the only binary search tree such that, for all $1 \le i \le n$, $\{\tau^{-1}(\sigma(1)), \ldots, \tau^{-1}(\sigma(i))\}$ is a subtree of T_∞ rooted at $\varnothing$.

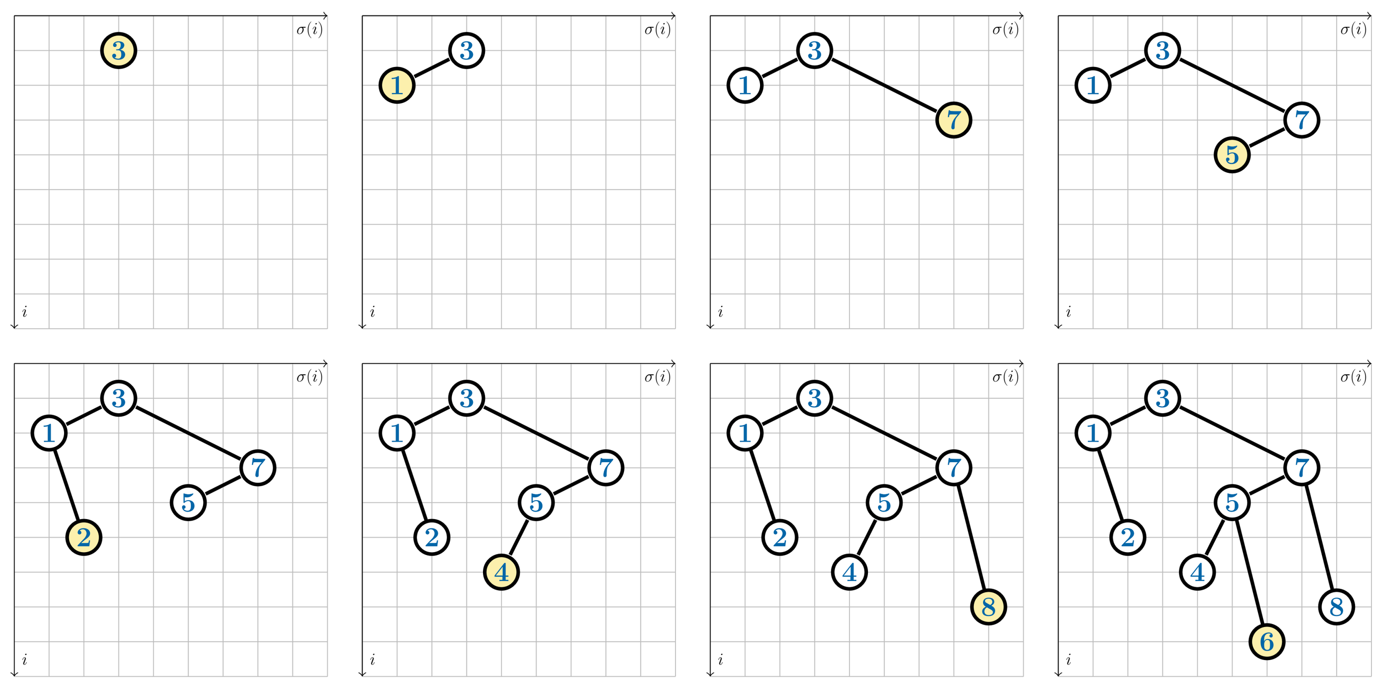

Fig. 1: Constructing T_σ for $\sigma = (3, 1, 7, 5, 2, 4, 8, 6)$. Labels are represented in blue.

Let $h(T)$ be the height of a subtree T. This first theorem states the asymptotic height of a random binary search tree.

Theorem [Devroye [2], 1986]. *Let σ_n be a uniformly random permutation of size n. Then, as $n \to \infty$,*

$$\frac{h(T_{\sigma_n})}{c^* \log n} \longrightarrow 1$$

in probability and in L^p for all $p > 0$. Here, c^* is the unique solution to $c \log(2e/c) = 1$, with $c \ge 2$.

Mallows permutations

For a permutation $\sigma \in \mathcal{S}_n$, let $\mathrm{Inv}(\sigma) = \big|\{i < j : \sigma(i) > \sigma(j)\}\big|$ be its inversion number. We consider the following model of random permutation.

Definition [Mallows permutations [3]]. *For $n \ge 1$ and $q \in [0, \infty)$, the Mallows distribution with parameters n and q is the probability measure $\pi_{n,q}$ on $\mathcal{S}_n$ given by*

$$\pi_{n,q}(\sigma) := Z_{n,q}^{-1} \cdot q^{\mathrm{Inv}(\sigma)},$$

where $Z_{n,q}$ is a normalizing constant.

From the definition, $\pi_{n,0}$ only gives weight 1 to the identity, and $\pi_{n,1}$ is the uniform distribution on $\mathcal{S}_n$. Note that, if $\sigma \sim \pi_{n,q}$, then $\sigma' = n + 1 - \sigma$ is distributed as $\pi_{n,1/q}$. Moreover, $T_{\sigma'}$ corresponds to the mirror tree of T_σ, where the role of the left and right subtrees are swapped. See Figure 2 for a depiction of T_σ and $T_{\sigma'}$.

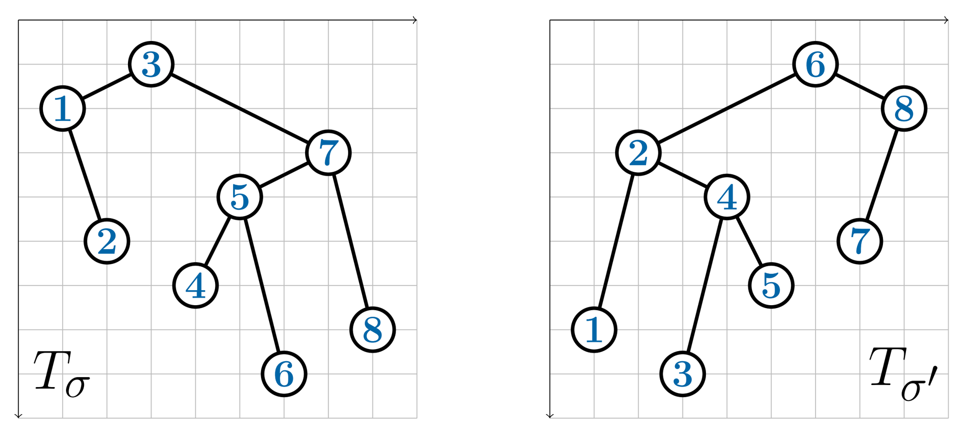

Fig. 2: T_σ and $T_{\sigma'}$ for $\sigma = (3, 1, 7, 5, 2, 4, 8, 6)$ and $\sigma' = (6, 8, 2, 4, 7, 5, 1, 3)$.

For $n \ge 1$ and $q \in [0, \infty)$, write $\textsc{Mallows}(n, q)$ for the distribution of T_σ, with $\sigma \sim \pi_{n,q}$, and call T_σ a Mallows tree. From the previous symmetries, it follows that the height of a $\textsc{Mallows}(n, q)$-distributed tree has the same distribution as the height of a $\textsc{Mallows}(n, 1/q)$-distributed tree. This remark allows us to restrict our study to $q \in [0, 1]$.

Converging results

For the following results, we consider a subsequence $(q_n)_{n\ge0}$ taking values in $[0, 1]$ and let $(T_{n,q_n})_{n\ge1}$ be a sequence of trees such that $T_{n,q_n} \sim \textsc{Mallows}(n, q_n)$. c^* refers to the only solution to $c \log(2e/c) = 1$ with $c \ge 2$, as in [2].

Convergence of the height

The first results on the height of Mallows trees proves an asymptotic behaviour.

Theorem [Addario-Berry & C., 2020]. *For any sequence $(q_n)_{n\ge0}$ taking values in $[0, 1]$, we have*

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

in probability and in L^p for all $p > 0$.

This extends Devroye's result [2], which corresponds to $q_n = 1$ for all n.

Convergence in distribution

When stronger assumptions are made on $(q_n)_{n\ge0}$, further results can be proven for the height of Mallows trees, stating distributional variation.

Theorem [Addario-Berry & C., 2020]. *Let $(q_n)_{n\ge0}$ be taking values in $[0, 1]$ such that $n(1 - q_n)/\log n \to \infty$ Then, if $nq_n \to \infty$, we have*

$$\frac{h(T_{n,q_n}) - n(1 - q_n) - c^* \log\left((1 - q_n)^{-1}\right)}{\sqrt{n(1 - q_n)q_n}} \longrightarrow \textsc{Normal}(0, 1),$$

and if $nq_n \to \lambda \in \mathbb{R}^+$, we have

$$n - 1 - h(T_{n,q_n}) \longrightarrow \textsc{Poisson}(\lambda),$$

both convergence occurring in distribution.

References

[1] Louigi Addario-Berry and Benoît Corsini. "The height of Mallows trees". In: *arXiv preprint arXiv:2007.13728* (2020).

[2] Luc Devroye. "A note on the height of binary search trees". In: *Journal of the ACM (JACM)* 33.3 (1986), pp. 489–498.

[3] Colin L Mallows. "Non-null ranking models. I". In: *Biometrika* 44.1/2 (1957), pp. 114–130.