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 $T_\infty$ be the infinite binary tree, identified with words on $\{0, 1\}$. A subtree of $T_\infty$ is a connected subset of $T_\infty$ 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 $\tau$ such that $\tau(v) > \tau(u)$ (respectively $\tau(v) < \tau(u)$) for all $u \in T$ such that $v$ is a prefix of $u$. For a permutation $\sigma \in S_n$ let $T_\sigma$ be the only binary search tree such that, for all $1 \leq i \leq n$, $(\tau^{-1}(\sigma(1)), \ldots, \tau^{-1}(\sigma(i)))$ is a subtree of $T_\infty$ rooted at $\varnothing$.

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

Mallows permutations
For a permutation $\sigma \in S_n$, let $\text{Inv}(\sigma) = \{i < j : \sigma(i) > \sigma(j)\}$ be its inversion number. We consider the following model of random permutation.

Definition [Mallows permutations [3]]. For $n \geq 1$ and $q \in [0, 1)$, the Mallows distribution with parameters $n$ and $q$ is the probability measure $\pi_{n,q}$ on $S_n$ given by

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

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 $S_n$. Note that, if $\sigma \sim \pi_{n,1}$, then $\sigma' = n + 1 - \sigma$ is distributed as $\pi_{n,1}/q$. Moreover, $T_{\sigma'}$ corresponds to the mirror tree of $T_{\sigma}$, where the role of the left and right subtrees are swapped. See Figure 2 for a depiction of $T_T$ and $T_{\sigma'}$.

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

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

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

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

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

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 $(q_n)_{n \geq 0}$, further results can be proven for the height of Mallows trees, stating distributional variation.

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

$$\frac{h(T_{n,q_n}) - n(1-q_n) - c^* \log n}{\sqrt{n(1-q_n)q_n}} \to \text{Normal}(0, 1),$$

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

$$n - 1 - h(T_{n,q_n}) \sim \text{Poisson}((1 \to \text{Poisson}(\lambda)).$$

both convergences in distribution occurring in distribution.

References
