THE HEIGHT OF RECORD-BIASED TREES

BENOÎT CORSINI

ABSTRACT. Given a permutation σ, its corresponding binary search tree is obtained by recursively inserting the values $\sigma(1), \ldots, \sigma(n)$ into a binary tree so that the label of each node is larger than the labels of its left subtree and smaller than the labels of its right subtree. In this paper, we study the height of binary search trees drawn from the record-biased model of permutations whose probability measure on the set of permutations is proportional to $\theta^{\mathrm{record}(\sigma)}$, where $\mathrm{record}(\sigma) = |\{i \in [n] : \forall j < i, \sigma(i) > \sigma(j)\}|$. We show that the height of a binary search tree built from a record-biased permutation of size n with parameter θ is of order $(1 + o_{\mathbb{P}}(1)) \max\{c^* \log n, \theta \log(1 + n/\theta)\}$, hence extending previous results of Devroye on the height or random binary search trees.

CONTENTS

1. INTRODUCTION

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of non-negative integers. For $n \in \mathbb{N}$, write $\mathcal{S}_n$ for the set of permutations of the set $[n] = \{1, \ldots, n\}$. In this article, we study properties of *record-biased permutations* [5], defined as follows. Let $n \in \mathbb{N}$ and $\theta \in [0, \infty)$. The record-biased distribution with parameters n and θ is the probability measure $w_{n,\theta}$ on $\mathcal{S}_n$ defined by

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

where $\mathrm{record}(\sigma) = |\{i \in [n] : \forall j \in [i-1], \sigma(i) > \sigma(j)\}|$ is the number of records of σ and $W_{n,\theta} = \sum_{\sigma \in \mathcal{S}_n} \theta^{\mathrm{record}(\sigma)}$ is a normalizing constant. This model of random permutations was introduced by Auger, Bouvel, Nicaud, and Pivoteau [5], who characterized some of its properties: the asymptotic distribution of the number of records, the number of descents, the first value $\sigma(1)$, and the number of inversions.

Record-biased permutations were originally defined as a more realistic model for representing real-life data than random uniform permutations, in particular when it came to *presortedness* [1]. It is therefor natural to consider the binary search trees built from these permutations, as they correspond to a popular method for storing the data contained in the permutation. In particular, the heights of these trees relate to the worst retrieval time for the entries of the permutation. This paper derives the first order asymptotic behaviour of the heights of the binary search trees associated to record-biased permutations, and in particular shows that it compares to the heights of the binary search trees associated to random uniform permutations when $\theta \in [0, \infty)$ is fixed.

[1]presortedness refers to the fact that most real-life data contains more pre-sorted subsequences than predicted by the uniform model.

Let $T_\infty = \{\varnothing\} \cup \bigcup_{k \geq 1} \{0,1\}^k$ be the complete infinite rooted binary tree, where nodes u at depth $|u| \geq 1$ are indexed by strings written as $u = u_1 \ldots u_{|u|} \in \{0,1\}^{|u|}$. This means that u has parent $u_1 \ldots u_{|u|-1}$ and children $u0$ and $u1$. For a set $V \subset T_\infty$ and a node $u \in T_\infty$, let $uV = \{uv, v \in V\}$.

Call a *subtree* of T_∞ (or just "tree", for short) a set $T \subset T_\infty$ which is connected when viewed as a subgraph of T_∞. For any subtree T of T_∞, its *root* is defined to be the unique node of T of minimum depth. Given a tree T and any node $u \in T_\infty$, let $T(u) = (uT_\infty) \cap T$ be the subtree of T rooted at u; note that if $\varnothing \in T$, then $T(u) = \emptyset$ if and only if $u \notin T$. Finally, for any subtree $T \subset T_\infty$, write $h(T) = \sup(|u|, u \in T) - \inf(|u|, u \in T)$ for its height, corresponding to the greatest distance between any node and the root of T.

Call a *labelled tree* any pair (T, τ) where T is a subtree of T_∞ and $\tau : T \to \mathbb{N}$ is an injective function. Furthermore say that (T, τ) is a *binary search tree* if, for any $u \in T$, and for any $v \in T(u0)$ (respectively $v \in T(u1)$), we have $\tau(u) > \tau(v)$ (respectively $\tau(u) < \tau(v)$); in other words, a binary search tree is a labelled tree where the label of each node is larger than the labels of its left subtree and smaller than the labels of its right subtree.

Given a binary search tree (T, τ) and an integer $k \in \mathbb{N}$ not in the image of τ, there exists a unique node $u \in T_\infty \setminus T$ where it is possible to *insert* k in (T, τ). In other words, there exists a unique $u \in T_\infty \setminus T$ such that the labelled tree (T', τ') defined by $T' = T \cup \{u\}$, $\tau'(v) = \tau(v)$ for $v \in T$, and $\tau'(u) = k$, is a binary search tree. This implies that, given an injective function $f : [n] \to \mathbb{N}$, there exists a unique binary search tree $(T\langle f \rangle, \tau\langle f \rangle)$ obtained by recursively inserting the values $f(1), f(2), \ldots, f(n)$ in that order. We call $(T\langle f \rangle, \tau\langle f \rangle)$ the *binary search tree of f*. An example of the construction of $(T\langle f \rangle, \tau\langle f \rangle)$ when $f = (2,4,1,6,3,5)$ can be found in Figure 1.
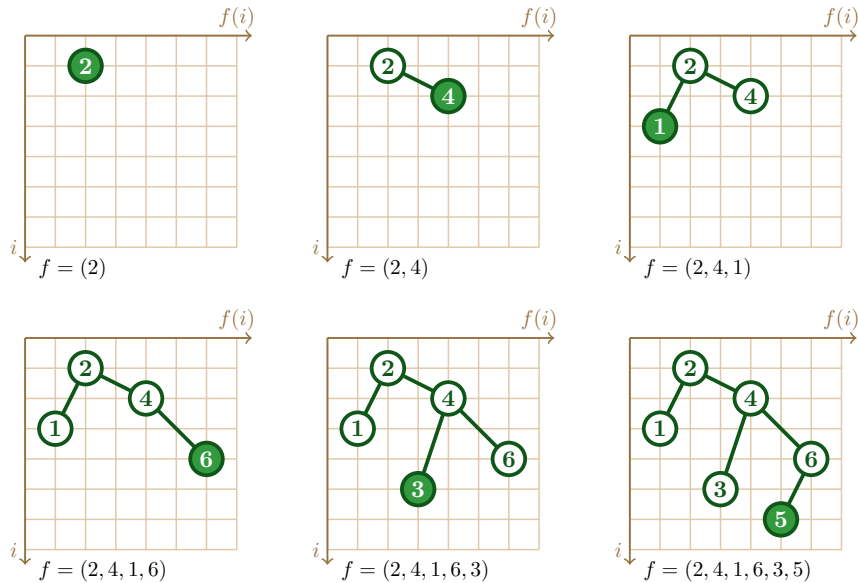

FIGURE 1. An example of construction of the binary search tree $(T\langle \sigma \rangle, \tau\langle \sigma \rangle)$ for the permutation $\sigma = (2,4,1,6,3,5)$. The nodes highlighted in green correspond to the most recently inserted elements, and the values on the nodes correspond to the values of the labelling function $\tau\langle \sigma \rangle$. This figure represents the recursive construction of binary search trees, where values are inserted one after the other at the unique location such that the label of each node is larger than the labels of its left subtree and smaller than the labels of its right subtree.

For the rest of the paper, write $T_{n,\theta}$ for a *record-biased tree*, defined to be distributed as $T\langle \sigma \rangle$ where σ is $w_{n,\theta}$-distributed. Let $c^* = 4.311\ldots$ be the unique solution of $c \log\left(\frac{2e}{c}\right) = 1$ with $c \geq 2$. The goal of this work is to prove the following result on the height of record-biased trees.

Theorem 1.1. *Let $(\theta_n)_{n\geq 0}$ be any sequence of non-negative numbers. Then, as n tends to infinity,*

$$\frac{h(T_{n,\theta_n})}{\max\left\{c^* \log n,\ \theta_n \log\left(1 + \frac{n}{\theta_n}\right)\right\}} \longrightarrow 1\,,$$

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

For any $n \in \mathbb{N}$ and $\theta \in [0, \infty)$, write

$$\mu(n, \theta) = \sum_{0 \leq i < n} \frac{\theta}{\theta + i}\,. \tag{1}$$

By comparison to the integral, we can see that

$$\left| \mu(n, \theta) - \left[1 + \theta \log\left(1 + \frac{n}{\theta}\right)\right] \right| \leq \theta \log\left(1 + \frac{1}{\theta}\right)\,,$$

which implies that, as n tends to infinity, we have $\frac{\mu(n,\theta_n)}{1 + \theta_n \log(1 + n/\theta_n)} \to 1$. It follows that, for any sequence $(\theta_n)_{n\geq 0}$ of non-negative numbers,

$$\max\left\{c^* \log n,\ \mu(n, \theta_n)\right\} = \left(1 + o(1)\right) \max\left\{c^* \log n,\ \theta_n \log\left(1 + \frac{n}{\theta_n}\right)\right\}\,.$$

This implies that the convergence in Theorem 1.1 is equivalent to the statement that

$$\frac{h(T_{n,\theta_n})}{\max\left\{c^* \log n,\ \mu(n, \theta_n)\right\}} \longrightarrow 1$$

in probability and in L^p for any $p > 0$. For the rest of the paper, we aim at proving this statement rather than the one of Theorem 1.1.

By taking subsequences, to prove Theorem 1.1, it suffices to consider two cases: if $\theta_n \equiv \theta \in [0, \infty)$ is constant, or if $\theta_n \to \infty$ as $n \to \infty$. In the case where $\theta_n = \theta$ for all $n \geq 0$, then $\mu(n, \theta_n) \sim \theta \log n$. This case of Theorem 1.1 can be rewritten as follows.

Theorem 1.2. *For any non-negative number $\theta \in [0, \infty)$, as n tends to infinity,*

$$\frac{h(T_{n,\theta})}{\log n} \longrightarrow \max\left\{c^*, \theta\right\}$$

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

When $\theta = 1$, the tree $T_{n,\theta}$ is a *random binary search tree*, the binary search tree of a uniformly random permutation, and this case of Theorem 1.2 corresponds to a well-known result of Devroye [9]; this case will also be used as input for the proof of the general result. Theorem 1.2 furthermore shows that there is a change of behaviour for the height of record-biased trees at the value $\theta = c^*$. Since previous results [5] on record-biased permutations proved that the number of records is of order $\theta \log n$, this change of behaviour corresponds to the moment where the first-order height of the tree becomes characterized by the length of its rightmost path.

Note that, Theorem 1.2 also states that the asymptotic behaviour of the height does not change around the value $\theta = 1$. This result might be unexpected since, for $\theta < 1$ the records of the permutations are penalized, whereas they are rewarded when $\theta > 1$. This strong change of behaviour for the permutation does not affect the height of the tree since, in the case where $\theta < c^*$, the height of $T_{n,\theta}$ is actually mainly characterized by the height of the left subtree of the root.

Theorem 1.2 covers the case where $(\theta_n)_{n\geq 0}$ is constant in Theorem 1.1. For the case where $(\theta_n)_{n\geq 0}$ diverges to infinity, first note that, for such sequences, $\mu(n, \theta_n) = \omega(\log n)$, where $a_n = \omega(b_n)$ means that $|a_n/b_n| \to \infty$. This case of Theorem 1.1 can thus be rewritten (and strengthened) as follows.

Theorem 1.3. *Let $(\theta_n)_{n\geq 0}$ be a sequence of non-negative numbers such that $\theta_n \to \infty$. Then, as n tends to infinity,*

$$\frac{h(T_{n,\theta_n})}{\mu(n, \theta_n)} \longrightarrow 1$$

in probability and in L^p for any $p > 0$. Moreover, for any $\epsilon > 0$ and $\lambda > 0$,

$$\mathbb{P}\left(\left|\frac{h(T_{n,\theta_n})}{\mu(n,\theta_n)} - 1\right| > \epsilon\right) = O\left(\frac{1}{n^\lambda}\right).$$

By using the first Borel-Cantelli lemma, the second bound of this result implies that, whenever the random variables of the sequence $(T_{n,\theta_n})_{n \geq 0}$ are defined on a common probability space, the first convergence also occurs almost-surely. Together, Theorems 1.2 and 1.3 establish Theorem 1.1.

1.1. **Overview of the proofs.** The main strategy to prove Theorems 1.2 and 1.3 is based on a similar method to that of [1]: controlling the size of the rightmost path and the behaviour of the left subtrees hanging from this rightmost path. We now provide an overview of the proof of Theorem 1.2 - a similar method is used to prove Theorem 1.3. In particular Proposition 1.4, Lemma 1.5, and Proposition 1.6 below are all non-asymptotic results, so can be applied in the setting of either theorem.

The main result that we use to understand the overall structure of the tree is given by the following proposition.

Proposition 1.4. *Let $n \in \mathbb{N}$ and $\theta \in [0, \infty)$. Let $T_{n,\theta}$ be a record-biased tree with parameters n and θ. Then, for any $k \in [n]$,*

$$\mathbb{P}\Big(\big|T_{n,\theta}(\mathsf{0})\big| = k - 1\Big) = \mathbb{P}\Big(\big|T_{n,\theta}(\mathsf{1})\big| = n - k\Big) = \frac{\theta}{\theta + n - k} \prod_{1 \leq i < k} \left(1 - \frac{\theta}{\theta + n - i}\right).$$

Moreover, conditionally given that $|T_{n,\theta}(\mathsf{0})| = k - 1$, $T_{n,\theta}(\mathsf{0})$ is distributed as a random binary search tree of size $k - 1$, $T_{n,\theta}(\mathsf{1})$ is distributed as a record-biased tree with parameters $n - k$ and θ, and $T_{n,\theta}(\mathsf{0})$ and $T_{n,\theta}(\mathsf{1})$ are independent of each other. Conversely, the preceding properties completely characterize record-biased binary search trees.

The proof of Proposition 1.4 can be found in Section 2.1. This result implies that record-biased trees can be generated inductively starting from the split between left and right subtree at the root. It also states that only the right subtree keeps the θ-record-biased distribution, since the left subtree is a random binary search tree.

Note that Proposition 1.4 implies that

$$\mathbb{P}\Big(\big|T_{n,\theta}(\mathsf{0})\big| \geq k\Big) = \prod_{1 \leq i \leq k} \left(1 - \frac{\theta}{\theta + n - i}\right),$$

a fact that will be used multiple times in what follows. In particular, for any fixed $\theta \in [0, \infty)$ and as n goes to infinity, this fact implies that

$$\mathbb{P}\left(\frac{|T_{n,\theta}(\mathsf{0})|}{n} \geq t\right) \longrightarrow (1 - t)^\theta.$$

This implies that the proportion of nodes to the right of the root is asymptotically distributed as a Beta$(\theta, 1)$ random variable. Since, conditionally given its size, the right subtree of a θ-record-biased tree is itself a θ-record-biased tree, the Beta distribution plays a key role in understanding the structure of record-biased trees and will be used to stochastically bound the number of nodes contained in a given subtree (see Proposition 1.6 below).

Since this work studies the binary search trees associated to record-biased permutations, it is important to understand the relation between the number of records of the permutation and the structure of the tree. Given a permutation σ, say that $\sigma(i)$ is a record of σ if, for all $j < i$, $\sigma(j) < \sigma(i)$. With this definition, note that $\sigma(i)$ is a record of σ if and only if the unique node $u \in T\langle\sigma\rangle$ labelled $\sigma(i)$ lies on the rightmost path of $T\langle\sigma\rangle$. Next, for any subtree T of T_∞, let

$$\text{record}(T) := \big|\{k : \mathsf{1}^k \in T\}\big| = 1 + \max\big\{k : \mathsf{1}^k \in T\big\} - \min\big\{k : \mathsf{1}^k \in T\big\};$$

by the above remark, if $T = T\langle\sigma\rangle$, then $\text{record}(T) = \text{record}(\sigma)$. From this definition, it is also useful to note that, for a subtree T of T_∞, we have $h(T) \geq \text{record}(T) - 1$.

An important input to the proofs, which is a fairly straightforward consequence of [5, Theorem 3], is the following lemma, stating bounds on the asymptotic behaviour of the number of records of record-biased permutations and trees.

Lemma 1.5. *Let $\epsilon > 0$. Then, there exists $\delta = \delta(\epsilon) > 0$ such that, for all $n \in \mathbb{N}$ and $\theta \in [0, \infty)$, for any $w_{n,\theta}$-distributed permutation σ, we have*

$$\mathbb{P}\left(\left|\frac{\text{record}(\sigma)}{\mu(n,\theta)} - 1\right| > \epsilon\right) = \mathbb{P}\left(\left|\frac{\text{record}(T_{n,\theta})}{\mu(n,\theta)} - 1\right| > \epsilon\right) \leq e^{-\delta\mu(n,\theta)},$$

where $\mu(n, \theta)$ is defined as in (1).

The proof of Lemma 1.5, whose first equality simply follows from the definition of $\text{record}(T)$, can be found in Section 2.2. Combining this results with the asymptotic behaviour of μ gives tight bounds on the asymptotic behaviour of the number of records of $T_{n,\theta}$. In particular, if θ is fixed, and as n tends to infinity, we have $\text{record}(T_{n,\theta}) = (\theta + o_{\mathbb{P}}(1)) \log n$.

An easy way to see the interest in bounding the number of records of record-biased trees is apparent when considering the following lower bound for their height:

$$h(T_{n,\theta}) \geq \max\left\{1 + h(T_{n,\theta}(\mathtt{0})), \text{record}(T_{n,\theta}) - 1\right\}.$$

Indeed, this bound, combined with Lemma 1.5, implies that $h(T_{n,\theta}) \geq (\theta + o_{\mathbb{P}}(1)) \log n$. Moreover, from Proposition 1.4, we can verify that

$$\log\left|T_{n,\theta}(\mathtt{0})\right| = (1 + o_{\mathbb{P}}(1)) \log n.$$

Since, conditioned on its size, $T_{n,\theta}(\mathtt{0})$ is a random binary search tree, the case $\theta = 1$ of Theorem 1.2 (which was already established in [9]) implies that

$$h(T_{n,\theta}(\mathtt{0})) = (c^* + o_{\mathbb{P}}(1)) \log n.$$

These two results together show that

$$h(T_{n,\theta}) \geq \max\left\{c^*, \theta\right\} \log n + o_{\mathbb{P}}(\log n),$$

corresponding to the lower bound of Theorem 1.2.

For the upper bound, since the left subtrees in $T_{n,\theta}$ are all binary search trees, for which the heights are already known to be concentrated around their expected height [9, 10, 21], we expect the height of the left subtrees to be well-behaved, conditioned on their sizes. This means that, in order to bound the height of record-biased trees from above, we essentially need to understand two quantities: the size of the rightmost path, and the sizes of the left subtrees hanging from that path.

Let $T_{n,\theta}$ be a record-biased tree with parameters n and θ. Since Lemma 1.5 already gives us strong bounds on the length of the rightmost path of $T_{n,\theta}$, it only remains to better understand the properties of the sizes of the left subtrees hanging from that path $(|T_{n,\theta}(\mathtt{1}^j\mathtt{0})|)_{j\geq 0}$. Using Proposition 1.4, it is easy to verify that

$$\mathbb{P}\left(\left|T_{n,\theta}(\mathtt{1}^j\mathtt{0})\right| \geq k \,\bigg|\, \left|T_{n,\theta}(\mathtt{0})\right|, \ldots, \left|T_{n,\theta}(\mathtt{1}^{j-1}\mathtt{0})\right|\right) = \prod_{1\leq i\leq k}\left(1 - \frac{\theta}{\theta + n - \sum_{0\leq\ell<j}(|T_{n,\theta}(\mathtt{1}^\ell\mathtt{0})| + 1) - i}\right);$$

this equality coming from the fact that, given $|T_{n,\theta}(\mathtt{0})|, \ldots, |T_{n,\theta}(\mathtt{1}^{j-1}\mathtt{0})|$, the tree $T_{n,\theta}(\mathtt{1}^j)$ is a record-biased tree of size $n - \sum_{0\leq\ell<j}(|T_{n,\theta}(\mathtt{1}^\ell\mathtt{0})| + 1)$.

Given two random variables X and Y, we say that X is stochastically smaller than Y, and write $X \preceq Y$, if and only if, for all $t \in \mathbb{R}$, we have $\mathbb{P}(X \geq t) \leq \mathbb{P}(Y \geq t)$. Using the above mentioned properties, we can prove that the sizes of the left subtrees $(|T_{n,\theta}(\mathtt{1}^j\mathtt{0})|)_{j\geq 0}$ are stochastically bounded as follows.

Proposition 1.6. *Let $\theta \in [0, \infty)$ and $(B_j)_{j\geq 0}$ be a sequence of independent and identically distributed $\text{Beta}(\theta, 1)$ random variables. Then, for any $n \in \mathbb{N}$ and $j \in \mathbb{N}$, we have*

$$\left|T_{n,\theta}(\mathtt{1}^j\mathtt{0})\right| \preceq j + n \prod_{0\leq i<j} B_i$$

The proof of Proposition 1.6 can be found in Section 2.3. Using the bound from Proposition 1.6, by the law of large numbers and the fact that $\mathbb{E}[\log B_0] = -\frac{1}{\theta}$, we expect to have

$$\left|T_{n,\theta}(\mathtt{1}^j\mathtt{0})\right| \leq j + ne^{-(1+o_{\mathbb{P}}(1))\frac{j}{\theta}}.$$

Moreover, since $|T_{n,\theta}(\mathbf{1}^j\mathbf{0})| = 0$ whenever $j > \mathrm{record}(T_{n,\theta})$ and, by Lemma 1.5, $\mathrm{record}(T_{n,\theta}) = (\theta + o_{\mathbb{P}}(1))\log n$, the j term in the upper bound will not notably contribute and the previous inequality can be strengthened as follows:

$$\left|T_{n,\theta}(\mathbf{1}^j\mathbf{0})\right| \leq ne^{-(1+o_{\mathbb{P}}(1))\frac{j}{\theta}}\,.$$

By the definition of the height of trees, for any subtree T of T_∞, we have that

$$h(T) = \max_{0 \leq j \leq \mathrm{record}(T)} \left\{j + h\left(T(\mathbf{1}^j\mathbf{0})\right)\right\}\,.$$

Replacing T with $T_{n,\theta}$ in this equality and using the previous upper bound on the size of the left subtrees and the fact that, conditioned on their sizes, the left subtrees are independent random binary search trees, we obtain that

$$h(T_{n,\theta}) \leq \max_{0 \leq j \leq \mathrm{record}(T_{n,\theta})} \left\{j + c^*\left(\log n - \left(1 + o_{\mathbb{P}}(1)\right)\frac{j}{\theta}\right)\right\}$$

$$= \max_{0 \leq j \leq \mathrm{record}(T_{n,\theta})} \left\{c^* \log n + j\left(1 - \frac{c^*}{\theta} + o_{\mathbb{P}}(1)\right)\right\}\,.$$

The $o_{\mathbb{P}}(1)$ term comes from the law of large numbers for the heights of binary search trees and these heights are sufficiently concentrated that this term can in fact be taken out of the maximum. This leads to the following upper bound for the height of record-biased trees

$$h(T_{n,\theta}) \leq c^* \log n + \max\left\{0,\, 1 - \frac{c^*}{\theta} + o_{\mathbb{P}}(1)\right\}\mathrm{record}(T_{n,\theta})\,.$$

Finally, by using that $\mathrm{record}(T_{n,\theta}) = (\theta + o_{\mathbb{P}}(1))\log n$ from Lemma 1.5, we obtain that

$$h(T_{n,\theta}) \leq \max\left\{c^*,\, \theta\right\}\log n + o_{\mathbb{P}}(\log n)\,,$$

which corresponds to the upper bound of Theorem 1.2.

The rest of the paper is organized in two sections. The first one, Section 2, uses a generative model for record-biased permutations introduced in [5] to deduce properties of record-biased trees. The second one, Section 3, combines the previously described properties to prove the theorems.

Before moving into the details of the proof, we define an important set of events for the study of record-biased trees. Given any $\mathbb{N}$-valued sequence $K = (k_j)_{j \geq 0}$, let $E_{n,\theta}(K)$ be the event that the left subtree sizes $(|T_{n,\theta}(\mathbf{1}^j\mathbf{0})|)_{j \geq 0}$ are given by the entries of K:

$$(2) \qquad\qquad E_{n,\theta}(K) := \left\{\left|T_{n,\theta}(\mathbf{1}^j\mathbf{0})\right| = k_j, \forall j \in \mathbb{N}\right\}\,.$$

Naturally, we will only be interested in K such that $\mathbb{P}(E_{n,\theta}(K)) > 0$. Note that, for any finite subtree T of T_∞, $\mathrm{record}(T)$ corresponds to the unique value $r \geq 1$ such that

$$r + \sum_{0 \leq j < r} \left|T(\mathbf{1}^j\mathbf{0})\right| = |T|\,.$$

This implies that, for any vector K with $\mathbb{P}(E_{n,\theta}(K)) > 0$, there is a unique non-negative integer $r = r(K)$ such that $r + \sum_{0 \leq j < r} k_j = n$ and on the event $E_{n,\theta}(K)$, necessarily $\mathrm{record}(T_{n,\theta}) = r(K)$. In particular, conditioning on $E_{n,\theta}(K)$ determines $\mathrm{record}(T_{n,\theta})$.

1.2. **Related work.** As mentioned before, record-biased permutations were just recently introduced [5]; and prior to the current work, to the best of our knowledge, [5] was the only paper studying the model. However, it is worth situating this model in a slightly larger literature on random permutations. An observation on this model made by the authors is that record-biased permutations can be bijectively mapped to Ewens permutations [13] using the Foata bijection [14]. This connection leads to interesting properties of both record-biased permutations and Ewens permutations related to their numbers of records and to their cycle structures. Moreover, although the record-biased model was originally connected with Ewens permutations [13], it is worth noting its strong resemblance to the Mallows model of permutations [18]; this resemblance was part of the inspiration for the current work.

Contrary to record-biased permutations, the literature on binary search trees and their height is vast and we only provide a glimpse here. The first order asymptotic behaviour of the height of random binary search trees was proven by Devroye [9], who showed that their height is of order $(c^* + o_{\mathbb{P}}(1))\log n$; this result was

built off a previous work of Pittel [20], stating that the height was of order $(\alpha + o_{\mathbb{P}}(1)) \log n$, while not being able to identify the constant α. Since then, these results have been extended to higher order asymptotic behaviour [10, 21], and to other models of increasing trees [7, 11]. Recently, similar results upon which this work is based on, proved the first order asymptotic behaviour of the height of Mallows trees, as well as distributional limits under some assumptions on the parameters of the model [1].

Finally, it is worth mentioning that the heights of binary search trees are often closely related to properties of extremes in *branching random walks*. An important example comes from the results of [7, 9], which strongly rely on the Hammersley-Kingman-Biggins theorem [6, 15, 17], providing a law of large numbers for the minimum of a branching random walk. Related results on the minimal position in branching random walks can be found in [2, 3, 8, 16].

2. Properties of record-biased trees

In this section we provide useful properties of record-biased trees and prove Proposition 1.4, Lemma 1.5 and Proposition 1.6. Moreover, we use Proposition 1.6 to obtain upper tail bounds on the sizes of the left subtrees $(|T_{n,\theta}(1^j 0)|)_{j \geq 0}$. Most results follow from the generative model of record-biased permutations developed in [5].

2.1. **Generating record-biased trees.** The following proposition was proven in [5] and gives an easy way to understand and generate record-biased permutations.

Proposition 2.1. *Let $n \in \mathbb{N}$ and $\theta \in [0, \infty)$. Let σ be a $w_{n,\theta}$-distributed permutation. Then, for all $k \in [n]$, we have*

$$\mathbb{P}\big(\sigma^{-1}(1) = k\big) = \begin{cases} \frac{\theta}{\theta+n-1} & \text{if } k = 1 \\ \frac{1}{\theta+n-1} & \text{otherwise} \end{cases} .$$

Moreover, by defining $A_i = [n] \setminus \{\sigma^{-1}(1), \ldots, \sigma^{-1}(i-1)\}$, we have

$$\mathbb{P}\big(\sigma^{-1}(i) = k \mid \sigma^{-1}(1), \ldots, \sigma^{-1}(i-1)\big) = \begin{cases} \frac{\theta}{\theta+n-i} & \text{if } k = \min(A_i) \\ \frac{1}{\theta+n-i} & \text{otherwise} \end{cases} .$$

The previous proposition fully describes the joint distribution of the values of $\sigma^{-1}(1), \ldots, \sigma^{-1}(n)$, so uniquely characterizes record-biased permutations. Note that $\sigma(k)$ is a record if and only if $k = \min(A_i)$, and that the probability for $\sigma(k)$ to be a record is independent of the previous values of $\sigma^{-1}(1), \ldots, \sigma^{-1}(i-1)$, which gives a useful way to compute the number of records of a record-biased permutation, as stated in the following corollary.

Corollary 2.2. *Let $n \in \mathbb{N}$ and $\theta \in [0, \infty)$. Let σ be a $w_{n,\theta}$-distributed permutation. Then, for any $t \in \mathbb{R}$, the moment generating function of $\mathrm{record}(\sigma)$ is given by*

$$\mathbb{E}\left[e^{t \cdot \mathrm{record}(\sigma)}\right] = \prod_{1 \leq i \leq n} \left(1 + (e^t - 1)\frac{\theta}{\theta + n - i}\right) .$$

Given a permutation $\sigma \in \mathcal{S}_n$ with $k = \sigma(1)$, write σ_- for the unique permutation of $[k-1]$ corresponding to the ordering of σ on the set $\sigma^{-1}([k-1])$, that is

$$\sigma_-^{-1}(i) < \sigma_-^{-1}(j) \iff \sigma^{-1}(i) < \sigma^{-1}(j)$$

Similarly, write σ_+ for the unique permutation of $[n-k]$ corresponding to the ordering of σ on the set $\sigma^{-1}([n] \setminus [k])$, that is

$$\sigma_+^{-1}(i) < \sigma_+^{-1}(j) \iff \sigma^{-1}(k+i) < \sigma^{-1}(k+j)$$

The following corollary gives an interesting characterization of $\sigma(1)$, σ_-, and σ_+ when σ is a record-biased permutation.

Corollary 2.3. *Let $n \in \mathbb{N}$ and $\theta \in [0, \infty)$. Let σ be a $w_{n,\theta}$-distributed permutation. Then, for any $k \in [n]$, we have*

$$\mathbb{P}\big(\sigma(1) = k\big) = \frac{\theta}{\theta + n - k} \prod_{1 \leq i < k} \left(1 - \frac{\theta}{\theta + n - i}\right) .$$

Moreover, given that $\sigma(1) = k$, σ_- is a uniformly random permutation of $\mathcal{S}_{k-1}$, σ_+ is a record-biased permutation with parameters $n - k$ and θ, and σ_- and σ_+ are independent of each other.

Proof. For the distribution of $\sigma(1)$, note that we have

$$\mathbb{P}\big(\sigma(1) = k \mid \sigma(1) > k - 1\big) = \mathbb{P}\big(\sigma^{-1}(k) = 1 \mid \sigma(1) > k - 1\big) = \frac{\theta}{\theta + n - k},$$

where the second equality follows from Proposition 2.1 and the fact that $1 \in A_k \Leftrightarrow 1 = \min(A_k)$. By induction, this proves the desired distribution for $\sigma(1)$.

For the distribution of σ_-, note that, if $\sigma(1) = k$, then $1 \notin \{\sigma^{-1}(1), \ldots, \sigma^{-1}(k-1)\}$. Combining this with Proposition 2.1, we have that, for any $i < k$, for any $j_1, \ldots, j_{i-1}$ all distinct in $[n] \setminus \{1\}$, and for any $j \in [n] \setminus \{1, j_1, \ldots, j_{i-1}\}$

$$\mathbb{P}\Big(\sigma^{-1}(i) = j \,\Big|\, \sigma(1) = k, \sigma^{-1}(1) = j_1, \ldots, \sigma^{-1}(i-1) = j_{i-1}\Big)$$

$$= \mathbb{P}\Big(\sigma^{-1}(i) = j \,\Big|\, \sigma^{-1}(i) \neq 1, \sigma^{-1}(1) = j_1, \ldots, \sigma^{-1}(i-1) = j_{i-1}\Big)$$

$$\propto \frac{1}{\theta + n - i}.$$

Since the last value does not depend on j, $\sigma^{-1}(i)$ is uniformly distributed over $j \in [n] \setminus \{1, j_1, \ldots, j_{i-1}\}$. Moreover, the definition of σ_- implies that, for any $\tau \in \mathcal{S}_{k-1}$, we have

$$\mathbb{P}\Big(\sigma_- = \tau \,\Big|\, \sigma(1) = k\Big) = \mathbb{P}\Big(\sigma^{-1}\big(\tau(1)\big) < \sigma^{-1}\big(\tau(2)\big) < \ldots < \sigma^{-1}\big(\tau(k-1)\big) \,\Big|\, \sigma(1) = k\Big) = \frac{1}{(k-1)!}$$

which proves that, conditionally given $\sigma(1) = k$, the random permutation σ_- is uniformly distributed. Finally, for the distribution of σ_+, by writing $A_{k+1} = [n] \setminus \{\sigma^{-1}(1), \ldots, \sigma^{-1}(k)\} = \{x_1 < \ldots < x_{n-k}\}$, Proposition 2.1 implies that, for any $\tau \in \mathcal{S}_{n-k}$, we have

$$\mathbb{P}\Big(\sigma^{-1}(k+1) = x_{\tau(1)}, \ldots, \sigma^{-1}(n) = x_{\tau(n-k)} \,\Big|\, \sigma^{-1}(1), \ldots, \sigma^{-1}(k-1), \sigma(1) = k\Big)$$

$$= \prod_{k < i \leq n} \mathbb{P}\Big(\sigma^{-1}(i) = x_{\tau(i-k)} \,\Big|\, \sigma^{-1}(1), \ldots, \sigma^{-1}(k-1), \sigma(1) = k, \sigma^{-1}(j) = x_{\tau(j-k)}, \forall k < j < i\Big)$$

$$= w_{n-k,\theta}(\tau).$$

By the definition of σ_+, this proves that

$$\mathbb{P}\Big(\sigma_+ = \tau \,\Big|\, \sigma^{-1}(1), \ldots, \sigma^{-1}(k-1), \sigma(1) = k\Big) = w_{n-k,\theta}(\tau),$$

implying that, conditionally given $\sigma(1) = k$, the random permutation σ_+ is $w_{n-k,\theta}$-distributed and independent of σ_-.   □

Proposition 1.4 now almost directly follows from Corollary 2.3.

Proof of Proposition 1.4. The definition of binary search trees implies that, for any permutation σ, we have

$$T\langle\sigma\rangle(0) = 0T\langle\sigma_-\rangle$$

and

$$T\langle\sigma\rangle(1) = 1T\langle\sigma_+\rangle.$$

Combining this with Corollary 2.3 proves that, conditionally given their sizes, $T_{n,\theta}(0)$ is a random binary search tree, $T_{n,\theta}(1)$ is a θ-record-biased tree, and they are independent of each other. The direct statement of Proposition 1.4 now simply follows from Corollary 2.3 and the fact that $|T\langle\sigma\rangle(0)| = \sigma(1) - 1$. For the converse, use that binary search trees are completely characterized by their left and right subtree distributions to see that these distributions completely characterize record-biased trees.   □

2.2. **Length of the rightmost path.** As explained in Section 1.1, to prove our results on the height of record-biased trees, we control the length of the rightmost path and bound the sizes of the left subtrees hanging from that path. We now prove Lemma 1.5, bounding the number of records of a record-biased permutation, hence bounding the length of the rightmost path of the corresponding binary search tree.

Proof of Lemma 1.5. Using Corollary 2.2, for a $w_{n,\theta}$-distributed permutation σ, for any $t \in \mathbb{R}$, we have

$$\mathbb{P}\left[e^{t \cdot \text{record}(\sigma)}\right] = \prod_{1 \leq i \leq n} \left(1 + (e^t - 1)\frac{\theta}{\theta + n - i}\right).$$

Now, using Chernoff's bound, it follows that

$$\mathbb{P}\left(\text{record}(\sigma) > (1 + \epsilon)\mu(n, \theta)\right) \leq e^{-t(1+\epsilon)\mu(n,\theta)} \prod_{1 \leq i \leq n} \left(1 + (e^t - 1)\frac{\theta}{\theta + n - i}\right).$$

Since $1 + x \leq e^x$, we have that

$$\mathbb{P}\left(\text{record}(\sigma) > (1 + \epsilon)\mu(n, \theta)\right) \leq \exp\left(-t(1 + \epsilon)\mu(n, \theta) + (e^t - 1)\mu(n, \theta)\right),$$

where $\mu(n, \theta)$ is defined in (1). For t small enough, $-t(1 + \epsilon) + (e^t - 1) < 0$, proving the upper bound of the lemma. Similarly for the lower bound,

$$\mathbb{P}\left(\text{record}(\sigma) < (1 - \epsilon)\mu(n, \theta)\right) \leq e^{t(1-\epsilon)\mu(n,\theta)} \prod_{1 \leq i \leq n} \left(1 + (e^{-t} - 1)\frac{\theta}{\theta + n - i}\right)$$

$$\leq \exp\left(t(1 - \epsilon)\mu(n, \theta) + (e^{-t} - 1)\mu(n, \theta)\right),$$

and once again, for t small enough, $t(1 - \epsilon) + (e^{-t} - 1) < 0$. □

2.3. **Stochastic bound on the left subtrees.** We conclude this section with results on the sizes of the left subtree of $T_{n,\theta}$. We start with a lemma useful to bound the size of the right subtree at the root of a record-biased tree.

Lemma 2.4. *Let $\theta \in [0, \infty)$ and let B be a $\text{Beta}(\theta, 1)$-distributed random variable. Then, for any $n \in \mathbb{N}$,*

$$\left|T_{n,\theta}(\mathbb{1})\right| \preceq nB + 1.$$

Proof. This statement holds when $n = 0$ since $0 \leq 1$, so we can now assume that $n \geq 1$. By Proposition 1.4, for any $0 \leq k \leq n - 1$, we have

$$\mathbb{P}\left(\left|T_{n,\theta}(\mathbb{1})\right| \leq n - k - 1\right) = \mathbb{P}\left(\left|T_{n,\theta}(\mathbb{0})\right| \geq k\right) = \prod_{1 \leq i \leq k} \left(1 - \frac{\theta}{\theta + n - i}\right).$$

By using that $\frac{1}{1+x} \geq e^{-x}$ along with the fact that $1 - \frac{\theta}{\theta+n-i} = \frac{1}{1 + \frac{\theta}{n-i}}$, this yields the lower bound

$$\mathbb{P}\left(\left|T_{n,\theta}(\mathbb{1})\right| \leq n - k - 1\right) \geq \exp\left(-\theta \sum_{1 \leq i \leq k} \frac{1}{n - i}\right).$$

By comparison to the integral, it follows that

$$\mathbb{P}\left(\left|T_{n,\theta}(\mathbb{1})\right| \leq n - k - 1\right) \geq \exp\left(-\theta \int_{n-k-1}^{n-1} \frac{1}{t} dt\right) \geq \left(\frac{n - k - 1}{n}\right)^\theta.$$

Using this bound, for any $t \in [0, n - 1]$, we have

$$\mathbb{P}\left(\left|T_{n,\theta}(\mathbb{1})\right| \leq t\right) = \mathbb{P}\left(\left|T_{n,\theta}(\mathbb{1})\right| \leq \lfloor t \rfloor\right) \geq \left(\frac{\lfloor t \rfloor}{n}\right)^\theta.$$

Finally, use that

$$\mathbb{P}\left(nB + 1 \leq t\right) = \left(\left(\frac{t - 1}{n}\right)_+\right)^\theta \leq \left(\frac{\lfloor t \rfloor}{n}\right)^\theta$$

to conclude the proof of the proposition. □

Note that there is a natural stochastic lower bound for $|T_{n,\theta}(1)|$, given by $nB - 1 - \theta$, whose proof essentially follows the same line of argument as the upper bound. However, this lower bound is not useful for our argument, so we omit it.

Using Lemma 2.4, we can now prove Proposition 1.6.

Proof of Proposition 1.6. By Lemma 2.4,

$$\mathbb{P}\big(\big|T_{n,\theta}(1)\big| \leq k\big) \geq \mathbb{P}\big(1 + nB_0 \leq k\big).$$

Now, by using properties of subtrees of a record-biased tree from Proposition 1.4, we have that

$$\mathbb{P}\Big(\big|T_{n,\theta}(1^j)\big| \leq k \,\Big|\, \big|T_{n,\theta}(1^{j-1})\big| = \ell\Big) = \mathbb{P}\Big(\big|T_{\ell,\theta}(1)\big| \leq k\Big) \geq \mathbb{P}\Big(1 + \ell B_{j-1} \leq k\Big).$$

It follows that

$$\big|T_{n,\theta}(1^j)\big| \preceq 1 + B_{j-1}\left(j - 1 + n \prod_{0 \leq i < j-1} B_i\right) \leq j + n \prod_{0 \leq i < j} B_i.$$

To conclude this proof, simply use that $|T_{n,\theta}(1^j 0)| \leq |T_{n,\theta}(1^j)|$. □

We can now conclude this section with a rather sharp upper tail bound on the sizes of the left subtrees in $T_{n,\theta}$.

Proposition 2.5. *Let $n \in \mathbb{N}$ and $\theta \in [0,\infty)$. Fix $\epsilon > 0$ such that $\epsilon\theta < 1$. Then, for any $M \in [0,\infty)$ and $k \in \mathbb{N}$ such that $ke^{(\frac{1}{\theta}-\epsilon)k} < ne^M$, we have*

$$\mathbb{P}\Big(\exists j \leq k, \big|T_{n,\theta}(1^j 0)\big| > ne^{-(\frac{1}{\theta}-\epsilon)+M}\Big) \leq Ce^{-\lambda M} \cdot \left(1 - \frac{ke^{(\frac{1}{\theta}-\epsilon)k}}{ne^M}\right)^{-\lambda}$$

where $C = \frac{1}{1-(1-\epsilon\theta)e^{\epsilon\theta}}$ and $\lambda = \frac{\epsilon\theta^2}{1-\epsilon\theta}$.

Proof. Fix $j \leq k$. Using the bound from Proposition 1.6, we have that

$$\mathbb{P}\Big(\big|T_{n,\theta}(1^j 0)\big| > ne^{-(\frac{1}{\theta}-\epsilon)+M}\Big) \leq \mathbb{P}\left(\prod_{0 \leq i < j} B_i > e^{-(\frac{1}{\theta}-\epsilon)j+M} - \frac{j}{n}\right).$$

Now, let $c \in (0,\infty)$ and $t \in (0,\infty)$. Using Markov's inequality, we have

$$\mathbb{P}\left(\prod_{0 \leq i < j} B_i > c\right) \leq \frac{1}{c^t}\mathbb{E}\left[\left(\prod_{0 \leq i < j} B_i\right)^t\right] \leq \frac{1}{c^t}\mathbb{E}\left[B_0^t\right]^j = \frac{1}{c^t}\left(\frac{\theta}{\theta+t}\right)^j.$$

Since $j \leq k$ and by assumption on the value of k, we know that $e^{-(\frac{1}{\theta}-\epsilon)j+M} - \frac{j}{n} \geq e^{-(\frac{1}{\theta}-\epsilon)k+M} - \frac{k}{n} > 0$. This implies that we can apply the previous bound with $c = e^{-(\frac{1}{\theta}-\epsilon)j+M} - \frac{j}{n}$ and obtain

$$\mathbb{P}\Big(\big|T_{n,\theta}(1^j 0)\big| > ne^{-(\frac{1}{\theta}-\epsilon)+M}\Big) \leq \left(e^{-(\frac{1}{\theta}-\epsilon)j+M} - \frac{j}{n}\right)^{-t}\left(\frac{\theta}{\theta+t}\right)^j$$

$$= \exp\left(-t\log\left(e^{-(\frac{1}{\theta}-\epsilon)j+M} - \frac{j}{n}\right) + j\log\left(\frac{\theta}{\theta+t}\right)\right).$$

Write $\Xi_j = \frac{je^{(\frac{1}{\theta}-\epsilon)j}}{ne^M} < 1$. Using that

$$\log\left(e^{-(\frac{1}{\theta}-\epsilon)j+M} - \frac{j}{n}\right) = -\left(\frac{1}{\theta} - \epsilon\right)j + M + \log(1 - \Xi_j)$$

and that $\Xi_j \leq \Xi_k$, we eventually have that

$$\mathbb{P}\Big(\big|T_{n,\theta}(1^j 0)\big| > ne^{-(\frac{1}{\theta}-\epsilon)+M}\Big) \leq \exp\left(\left(\frac{1}{\theta} - \epsilon\right)tj - t\big(M + \log(1 - \Xi_k)\big) + j\log\left(\frac{\theta}{\theta+t}\right)\right).$$

To conclude the proof, set $t = \lambda = \frac{\epsilon\theta^2}{1-\epsilon\theta}$ to obtain

$$\mathbb{P}\Big(\big|T_{n,\theta}(1^j 0)\big| > n e^{-\left(\frac{1}{\theta}-\epsilon\right)+M}\Big) \leq \exp\Big(\big(\epsilon\theta + \log(1-\epsilon\theta)\big)j - \lambda\big(M + \log(1-\Xi_k)\big)\Big).$$

The desired result follows from using a union bound and summing the right hand side over $j \geq 0$. $\square$

3. THE HEIGHT OF RECORD-BIASED TREES

In this section, the characteristics of record-biased trees developed in Section 2 are used to obtain bounds on their height and eventually prove Theorem 1.2 and 1.3.

3.1. **Lower bounds on the height.** We start by proving the following result, corresponding to the lower bound of Theorem 1.2.

Proposition 3.1. *Let $\theta \in [0,\infty)$ and $\epsilon > 0$. Then, as n tends to infinity*

$$\mathbb{P}\Big(h(T_{n,\theta}) < \big(\max\{c^*,\theta\} - \epsilon\big)\log n\Big) \longrightarrow 0.$$

Proof. Using Lemma 1.5 and the fact that $h(T_{n,\theta}) \geq \operatorname{record}(T_{n,\theta}) - 1$ along with the asymptotic behaviour of $\mu(n,\theta)$ from (1) when θ is fixed, we first have that

$$\mathbb{P}\Big(h(T_{n,\theta}) < \big(\theta - \epsilon\big)\log n\Big) \longrightarrow 0,$$

Note that, in the case $\theta = 0$ the previous result trivially holds since $h(T_{n,\theta}) > 0$. It now only remains to prove that

$$\mathbb{P}\Big(h(T_{n,\theta}) < \big(c^* - \epsilon\big)\log n\Big) \longrightarrow 0.$$

By the definition of the height of trees, we know that

$$h(T_{n,\theta}) \geq 1 + h\big(T_{n,\theta}(0)\big).$$

Hence, it suffices to show that

$$\mathbb{P}\Big(h\big(T_{n,\theta}(0)\big) < \big(c^* - \epsilon\big)\log n\Big) \longrightarrow 0.$$

Using Proposition 1.4, we know that

$$\mathbb{P}\Big(\big|T_{n,\theta}(0)\big| \geq k\Big) = \prod_{1\leq i\leq k}\left(1 - \frac{\theta}{\theta+n-i}\right),$$

from which it follows that

$$\mathbb{P}\left(\big|T_{n,\theta}(0)\big| \geq \left\lfloor\frac{n}{\log n}\right\rfloor\right) = \prod_{1\leq i\leq \frac{n}{\log n}}\left(1 - \frac{\theta}{\theta+n-i}\right) \geq \left(1 - \frac{\theta}{\theta+n}\right)^{\frac{n}{\log n}} = 1 - o(1).$$

Recall now from Proposition 1.4 that, given $|T_{n,\theta}(0)| = k$, $T_{n,\theta}(0)$ is distributed as a random binary search tree of size k. Using the previous lower bound on the size of $T_{n,\theta}(0)$, and the law of large numbers for the height of a random binary search tree [9], it follows that

$$\mathbb{P}\Big(h\big(T_{n,\theta}(0)\big) < (c^* - \epsilon)\log n\Big) \longrightarrow 0.$$

This establishes the second desired lower bound for the height of $T_{n,\theta}$ and concludes the proof. $\square$

Proving the lower bound of Theorem 1.3 is similar, but easier.

Proposition 3.2. *Let $(\theta_n)_{n\geq 0}$ be a sequence of non-negative numbers, such that $\theta_n \to \infty$. Then, for all $\epsilon > 0$ and $\lambda > 0$, we have*

$$\mathbb{P}\Big(h(T_{n,\theta_n}) < (1-\epsilon)\mu(n,\theta_n)\Big) = O\left(\frac{1}{n^\lambda}\right).$$

Proof. Using Lemma 1.5 along with the fact that $h(T_{n,\theta}) \geq \text{record}(T_{n,\theta}) - 1$, we have that

$$\mathbb{P}\Big(h(T_{n,\theta}) \leq (1-\epsilon)\mu(n,\theta_n)\Big) \leq e^{-\delta(\epsilon)\mu(n,\theta_n)}\,.$$

Moreover, since $\mu(n,\theta_n) = \omega(\log n)$ whenever $\theta_n \to \infty$, we know that $e^{-\delta(\epsilon)\mu(n,\theta_n)} = O(n^{-\lambda})$, which concludes the proof of the proposition. □

3.2. **Upper bounds on the height.** The first result of this section proves a very useful upper tail bound on the height of a record-biased tree, conditionally given the sizes of its left subtrees. Before stating this result, recall the definition of $E_{n,\theta}(K)$ when $K = (k_j)_{j \geq 0}$ from (2):

$$E_{n,\theta}(K) = \Big\{ \big|T_{n,\theta}(1^j 0)\big| = k_j,\ \forall j \in \mathbb{N} \Big\}\,.$$

Moreover, recall from the end of Section 1.1 that, conditionally given $E_{n,\theta}(K)$ and given that $\mathbb{P}(E_{n,\theta}(K)) > 0$, there exists a unique $r = r(K)$ such that $\text{record}(T_{n,\theta}) = r(K)$.

Proposition 3.3. *Let $n \in \mathbb{N}$ and $\theta \in [0,\infty)$. Consider $K = (k_j)_{j \geq 0}$ such that $\mathbb{P}(E_{n,\theta}(K)) > 0$, where $E_{n,\theta}(K)$ is defined in (2) and let $r = r(K)$ be the unique value such that $r + \sum_{0 \leq j < k} k_j = n$. Then, for any $\eta \in \mathbb{N}$ and $t > 0$*

$$\mathbb{P}\Big(h(T_{n,\theta}) \geq \eta \,\Big|\, E_{n,\theta}(K)\Big) \leq \sum_{0 \leq j \leq r} (2e^{-t})^{\eta-j} \cdot (k_j+1)^{e^t - 1}$$

Proof. Thanks to Proposition 1.4, for any vector $K = (k_j)_{j \geq 0}$ such that $\mathbb{P}(E_{n,\theta}(K)) > 0$, conditionally given $E_{n,\theta}(K)$, the trees $(T_{n,\theta}(1^j 0))_{j \geq 0}$ are independent random binary search trees of respective sizes $K = (k_j)_{j \geq 0}$. Using this fact together with the identity

$$h(T_{n,\theta}) = \max_{0 \leq j \leq \text{record}(T_{n,\theta})} \Big\{ j + h\big(T_{n,\theta}(1^j 0)\big) \Big\}\,.$$

it follows that

$$\mathbb{P}\Big(h(T_{n,\theta}) \geq \eta \,\Big|\, E_{n,\theta}(K)\Big) = \mathbb{P}\left(\max_{0 \leq j \leq \text{record}(T_{n,\theta})} \Big\{ j + h\big(T_{n,\theta}(1^j 0)\big) \Big\} \geq \eta \,\Big|\, E_{n,\theta}(K) \right)$$

$$= \mathbb{P}\left(\max_{0 \leq j \leq r} \Big\{ j + h\big(\mathcal{T}_j\big) \Big\} \geq \eta \,\Big|\, E_{n,\theta}(K) \right),$$

where $(\mathcal{T}_j)_{j \geq 0}$ are independent random binary search trees of respective sizes k_j, also independent of $E_{n,\theta}(K)$. Using a union bound, we obtain that

$$(3) \qquad \mathbb{P}\Big(h(T_{n,\theta}) \geq \eta \,\Big|\, E_{n,\theta}(K)\Big) \leq \sum_{0 \leq j \leq r} \mathbb{P}\Big(j + h\big(\mathcal{T}_j\big) \geq \eta\Big)\,.$$

To bound the right-hand side, use a union bound over the $2^{\eta-j}$ paths of length $\eta - j$ from the root of T_∞ to obtain

$$(4) \qquad \mathbb{P}\Big(j + h\big(\mathcal{T}_j\big) \geq \eta\Big) \leq \sum_{v \in T_\infty : |v| = \eta - j} \mathbb{P}\big(v \in \mathcal{T}_j\big) = 2^{\eta-j}\mathbb{P}\big(1^{\eta-j} \in \mathcal{T}_j\big)\,.$$

Using the moment generating function from Corollary 2.2 and the fact that a binary search tree is simply a record-biased tree with parameter $\theta = 1$, we have that

$$\mathbb{P}\big(1^{\eta-j} \in \mathcal{T}_j\big) = \mathbb{P}\Big(\text{record}(\mathcal{T}_j) \geq \eta - j + 1\Big) \leq e^{-t(\eta-j+1)} \prod_{1 \leq \ell \leq k_j} \left(1 + (e^t - 1)\frac{1}{k_j - \ell + 1}\right)\,.$$

Now, using that

$$\prod_{1 \leq i \leq n} \left(1 + (e^t - 1)\frac{1}{n - i + 1}\right) \leq \exp\left((e^t - 1)\sum_{1 \leq \ell \leq k_j}\frac{1}{\ell}\right) \leq \exp\Big((e^t - 1)\log(k_j + 1)\Big)$$

and that $e^{-t} \leq 1$, we obtain that

$$(5) \qquad \mathbb{P}\big(1^{\eta-j} \in \mathcal{T}_j\big) \leq \exp\Big(-t(\eta - j) + (e^t - 1)\log(k_j + 1)\Big) = e^{-t(\eta-j)} \cdot (k_j + 1)^{e^t - 1}\,.$$

Combining (3), (4), and (5), we obtain the desired bound. □

We now use Proposition 3.3 to prove the following result, which allows us to extend our convergence in probability results to the convergence in L^p. Recall the definition of $\mu(n, \theta)$ from (1).

Proposition 3.4. *For any sequence $(\theta_n)_{n \geq 0}$ and any $p > 0$, the family of random variables*

$$\left(\left(\frac{h(T_{n, \theta_n})}{\max\{c^* \log n, \mu(n, \theta_n)\}} \right)^p \right)_{n \geq 0}$$

is uniformly integrable.

Proof. First, write $\nu_n = \max\{c^* \log n, \mu(n, \theta_n)\}$ and fix $a > 0$. We have that

$$\mathbb{P}\left(\left(\frac{h(T_{n, \theta_n})}{\nu_n} \right)^p \geq a \right) = \mathbb{P}\left(h(T_{n, \theta_n}) \geq a^{\frac{1}{p}} \nu_n \right).$$

Using Proposition 3.3 with $\eta = \lceil a^{\frac{1}{p}} \nu_n \rceil$ and the fact that $k_j \leq n$, it follows that

$$\mathbb{P}\left(h(T_{n, \theta_n}) \geq \eta \mid E_{n, \theta_n}(K) \right) \leq \sum_{0 \leq j \leq r(K)} (2e^{-t})^{\eta - j} (n+1)^{e^t - 1}.$$

Moreover, by using Lemma 1.5, we know that there exists a universal constant $\delta > 0$ such that

$$\mathbb{P}\left(\mathrm{record}(T_{n, \theta_n}) > 2\mu(n, \theta_n) \right) \leq e^{-\delta \mu(n, \theta_n)}.$$

In the case where $\mu(n, \theta_n)$ is bounded, the previous tail bound might not suffice for our purpose. However, in this case, we would have that $\mathbb{E}[\mathrm{record}(T_{n, \theta_n})] = \mu(n, \theta_n) = O(1)$, from which a simple application of Markov's inequality gives us that

$$\mathbb{P}\left(\mathrm{record}(T_{n, \theta_n}) > 2\mu(n, \theta_n) + \log(n+1) \right) = O\left(\frac{1}{\log n} \right) = o(1).$$

Combining both bounds, we have that

$$\mathbb{P}\left(\mathrm{record}(T_{n, \theta_n}) > 2\mu(n, \theta_n) + \log(n+1) \right) = o(1).$$

Combine the previous tail bounds to obtain that

$$\mathbb{P}\left(h(T_{n, \theta_n}) \geq \eta \right) \leq \sum_{K : r(K) \leq 2\mu(n, \theta_n) + \log(n+1)} \mathbb{P}\left(h(T_{n, \theta_n}) \geq \eta \mid E_{n, \theta_n}(K) \right) \mathbb{P}\left(E_{n, \theta_n}(K) \right) + o(1)$$

$$\leq \sum_{0 \leq j \leq 2\mu(n, \theta_n) + \log(n+1)} (2e^{-t})^{\eta - j} (n+1)^{e^t - 1} + o(1).$$

Let $t = \log(2e)$. The previous equation can then be rewritten as

$$\mathbb{P}\left(h(T_{n, \theta_n}) \geq \eta \right) \leq \sum_{0 \leq j \leq 2\mu(n, \theta_n) + \log(n+1)} \frac{1}{e^{\eta - j}} (n+1)^{1+e} + o(1)$$

$$= \frac{1}{e-1} \exp\left(2\mu(n, \theta_n) + (2+e)\log(n+1) - \eta \right) + o(1).$$

Recall now that $\eta = \lceil a^{\frac{1}{p}} \nu_n \rceil$ and that $\nu_n = \max\{c^* \log n, \mu(n, \theta_n)\}$. This implies that, whenever $a > \left(2 + \frac{2+e}{c^*} \right)^p$, we have

$$\eta - 2\mu(n, \theta_n) - (2+e)\log(n+1) > \left(a^{\frac{1}{p}} - 2 - \frac{2+e}{c^*} + o(1) \right) \max\left\{ c^* \log n, \mu(n, \theta_n) \right\} \longrightarrow \infty.$$

It follows that, for such a,

$$\lim_{n \to \infty} \mathbb{P}\left(\left(\frac{h(T_{n, \theta_n})}{\nu_n} \right)^p \geq a \right) = 0,$$

establishing the claimed uniform integrability. $\square$

3.3. **Convergence of the height.** In this section, we build on the results of the previous sections to finally prove Theorem 1.2 and 1.3.

Proof of Theorem 1.2. Thanks to the uniform integrability proved in Proposition 3.4, we only need to establish the convergence in probability in order to prove the theorem. Moreover, Proposition 3.1 already proved the requisite lower bound, so it only remains to prove the upper bound.

Assume first that $\theta = 0$. In this case, $\mathrm{record}(T_{n,\theta}) = 1$ and we have that

$$T_{n,\theta} = \{\varnothing\} \cup \big(0 T_{n,\theta}(0)\big),$$

which implies that

$$h(T_{n,\theta}) = 1 + h\big(T_{n,\theta}(0)\big).$$

Using that, in this case, $T_{n,\theta}(0)$ is a random binary search tree of size $n - 1$, along with the law of large numbers for the height of random binary search trees [9], it follows that

$$\mathbb{P}\Big(h(T_{n,\theta}) > (c^* + \epsilon)\log n\Big) \longrightarrow 0,$$

which is the desired upper bound. Assume for the rest of the proof that $\theta > 0$.

For any $a > 0$, by using the bound from Proposition 3.3, we have

$$\mathbb{P}\Big(h(T_{n,\theta}) \geq a\log n \,\Big|\, E_{n,\theta}(K)\Big) \leq \sum_{0 \leq j \leq r(K)} (2e^{-t})^{\lfloor a\log n \rfloor - j} \cdot (k_j + 1)^{e^t - 1}.$$

Moreover, from Lemma 1.5, we know that, for any $\alpha > 0$, as n tends to infinity

$$\mathbb{P}\Big(\mathrm{record}(T_{n,\theta}) > (\theta + \alpha)\log n\Big) \longrightarrow 0.$$

Finally, Proposition 2.5 tells us that, for any fixed $\beta > 0$ such that $\beta\theta < 1$, by setting $k = \left\lfloor \left(\frac{1}{\theta} - \beta\right)^{-1}\log n \right\rfloor$ and $M = 2\log\log n$, we have

$$\mathbb{P}\Big(\exists 0 \leq j \leq k, |T_{n,\theta}(1^j 0)| > n e^{-\left(\frac{1}{\theta} - \beta\right)j + M}\Big) \leq C e^{-\lambda M} \cdot \left(1 - \frac{k}{e^M}\right)^{-\lambda} \longrightarrow 0.$$

Now fix $\alpha, \beta > 0$ such that $\left(\frac{1}{\theta} - \beta\right)^{-1} > (\theta + \alpha)$, for example we could take $0 < \beta < \frac{1}{\theta}$ and set $\alpha = \frac{\beta\theta}{2}\left(\frac{1}{\theta} - \beta\right)^{-1}$. Combining the previous results, we obtain

$$\mathbb{P}\Big(h(T_{n,\theta}) \geq a\log n\Big) \leq \sum_{K:\left\{ \begin{smallmatrix} r(K) \leq (\theta+\alpha)\log n \\ k_j \leq n \cdot \exp\left(-\left(\frac{1}{\theta} - \beta\right)j + M\right) \end{smallmatrix} \right.} \mathbb{P}\Big(h(T_{n,\theta}) \geq a\log n \,\Big|\, E_{n,\theta}(K)\Big)\mathbb{P}\Big(E_{n,\theta}(K)\Big) + o(1)$$

$$\leq \sum_{0 \leq j \leq (\theta+\alpha)\log n} (2e^{-t})^{\lfloor a\log n \rfloor - j} \cdot \left(n e^{-\left(\frac{1}{\theta} - \beta\right)j + M} + 1\right)^{e^t - 1} + o(1),$$

where the bounds on k_j hold since we assumed that $\left(\frac{1}{\theta} - \beta\right)^{-1} > (\theta + \alpha)$, implying that $j \leq (\theta + \alpha)\log n \leq k = \left\lfloor \left(\frac{1}{\theta} - \beta\right)^{-1}\log n \right\rfloor$ for n large enough. Moreover, since $\left(\frac{1}{\theta} - \beta\right)j \leq \left(\frac{1}{\theta} - \beta\right)(\theta + \alpha)\log n < \log n$, it follows that $n e^{-\left(\frac{1}{\theta} - \beta\right)j + M} + 1 \leq e^{\log n - \left(\frac{1}{\theta} - \beta\right)j + o(\log n)}$. Similarly, $(2e^{-t})^{\lfloor a\log n \rfloor - j} = e^{(\log 2 - t)(a\log n - j) + o(\log n)}$. Combining these formulas with the previous upper tail bound on the height, we obtain that, for any $0 < \beta < \frac{1}{\theta}$, and for any $0 < \alpha < \beta\theta\left(\frac{1}{\theta} - \beta\right)^{-1}$,

$$\mathbb{P}\Big(h(T_{n,\theta}) \geq a\log n\Big) \leq \sum_{0 \leq j \leq (\theta+\alpha)\log n} e^{(\log 2 - t)(a\log n - j) + o(\log n)} \cdot \left(e^{\log n - \left(\frac{1}{\theta} - \beta\right)j + o(\log n)}\right)^{e^t - 1} + o(1)$$

$$= e^{[a(\log 2 - t) + (e^t - 1) + o(1)]\log n} \sum_{0 \leq j \leq (\theta+\alpha)\log n} \left(e^{-\log 2 + t - (e^t - 1)\left(\frac{1}{\theta} - \beta\right)}\right)^j + o(1).$$

Set now $t = \log c^*$ and recall that $c^* \log \left(\frac{2e}{c^*}\right) = 1$ to eventually rewrite the previous bound into

(6) $\quad \mathbb{P}\left(h(T_{n,\theta}) \geq a \log n\right)$

$$\leq \exp\left(-(c^* - 1)\left[\frac{a}{c^*} - 1 + o(1)\right]\log n\right) \sum_{0 \leq j \leq (\theta+\alpha)\log n} \exp\left((c^* - 1)\left[\frac{1}{c^*} - \frac{1}{\theta} + \beta\right]j\right) + o(1).$$

We now bound the right-hand side from above in two cases, according to whether $\theta < c^*$ or $\theta \geq c^*$.

Case 1: $\theta < c^*$. In this case, for β small enough, the sum in j in (6) can be bounded by a constant depending only on c^* and θ. Moreover, for any $\epsilon > 0$, taking $a = c^* + \epsilon$, the term before the sum converges to 0. This proves that

$$\mathbb{P}\left(h(T_{n,\theta}) \geq (c^* + \epsilon)\log n\right) \longrightarrow 0.$$

Since we assumed that $\theta < c^*$, this corresponds to the desired upper bound in this case.

Case 2: $\theta \geq c^*$. For any $\epsilon > 0$, taking $a = \theta + \epsilon$ and using (6), we have that

(7) $\mathbb{P}\left(h(T_{n,\theta}) \geq (\theta + \epsilon)\log n\right)$

$$\leq \exp\left(-(c^* - 1)\left[\frac{\theta + \epsilon}{c^*} - 1 + o(1)\right]\log n\right) \sum_{0 \leq j \leq (\theta+\alpha)\log n} \exp\left((c^* - 1)\left[\frac{1}{c^*} - \frac{1}{\theta} + \beta\right]j\right) + o(1)$$

$$\leq \exp\left(-(c^* - 1)\left[\frac{\theta + \epsilon}{c^*} - 1 + o(1)\right]\log n\right)\exp\left(-(c^* - 1)\left[\frac{1}{c^*} - \frac{1}{\theta} + \beta\right](\theta + \alpha)\log n + O(1)\right) + o(1)$$

$$= \exp\left(-(c^* - 1)\left[\frac{\theta + \epsilon}{c^*} - 1 - \left(\frac{1}{c^*} - \frac{1}{\theta} + \beta\right)(\theta + \alpha)\right]\log n + o(\log n)\right) + o(1),$$

In order to conclude the proof in this case, note that

$$\frac{\theta + \epsilon}{c^*} - 1 - \left(\frac{1}{c^*} - \frac{1}{\theta} + \beta\right)(\theta + \alpha) = \frac{\epsilon}{c^*} - \alpha\left(\frac{1}{c^*} - \frac{1}{\theta}\right) - \beta\theta - \alpha\beta.$$

Moreover, the upper tail bound of (7) holds for any $\epsilon > 0$, $0 < \beta < \frac{1}{\theta}$, and $0 < \alpha < \beta\theta\left(\frac{1}{\theta} - \beta\right)^{-1}$. Hence, for any fixed $\epsilon > 0$, we can choose α and β small enough, so that

$$\frac{\theta + \epsilon}{c^*} - 1 - \left(\frac{1}{c^*} - \frac{1}{\theta} + \beta\right)(\theta + \alpha) > 0.$$

This implies that

$$\mathbb{P}\left(h(T_{n,\theta}) \geq (\theta + \epsilon)\log n\right) \longrightarrow 0,$$

which proves the desired upper bound in the case $\theta \geq c^*$. This concludes the proof of Theorem 1.2. □

We conclude this section with the proof of Theorem 1.3.

Proof of Theorem 1.3. Let $(\theta_n)_{n \geq 0}$ be such that $\theta_n \to \infty$ and fix $\epsilon > 0$. We start by proving the desired upper tail bound.

Using Proposition 3.3, we know that

$$\mathbb{P}\left(h(T_{n,\theta_n}) \geq \eta \,\Big|\, E_{n,\theta}(K)\right) \leq \sum_{0 \leq j \leq r(K)} (2e^{-t})^{\eta-j} \cdot (k_j + 1)^{e^t - 1}.$$

Moreover, we know that $k_j \leq n$. Using this upper bound for k_j and letting $t = \log(2e)$ and $\eta = \lceil(1 + 2\epsilon)\mu(n, \theta_n)\rceil$, we obtain

$$\mathbb{P}\left(h(T_{n,\theta_n}) \geq (1 + 2\epsilon)\mu(n, \theta_n)\,\Big|\, E_{n,\theta}(K)\right) \leq \sum_{0 \leq j \leq r(K)} e^{j-\eta} \cdot (n+1)^{e+1}$$

$$\leq \frac{1}{e-1} \cdot (n+1)^{e+1} \cdot e^{r(K)-\eta}.$$

Finally, using Lemma 1.5, we know that

$$\mathbb{P}\left(\text{record}(T_{n,\theta_n}) > (1 + \epsilon)\mu(n, \theta_n)\right) \leq e^{-\delta(\epsilon)\mu(n, \theta_n)},$$

from which we obtain that

$$\mathbb{P}\Big(h(T_{n,\theta_n}) \geq (1+2\epsilon)\mu(n,\theta_n)\Big)$$

$$\leq \sum_{K:r(K)\leq(1+\epsilon)\mu(n,\theta_n)} \mathbb{P}\Big(h(T_{n,\theta_n}) \geq (1+2\epsilon)\mu(n,\theta_n) \,\Big|\, E_{n,\theta}(K)\Big)\mathbb{P}\Big(E_{n,\theta_n}(K)\Big) + e^{-\delta(\epsilon)\mu(n,\theta_n)}$$

$$\leq \frac{1}{e-1} \cdot (n+1)^{e+1} \cdot e^{(1+\epsilon)\mu(n,\theta_n)-\eta} + e^{-\delta(\epsilon)\mu(n,\theta_n)} \,.$$

From the definition of η, we know that $(1+\epsilon)\mu(n,\theta_n) - \eta \leq -\epsilon\mu(n,\theta_n)$. Moreover, in the case where $\theta_n \to \infty$, we know that $\mu(n,\theta_n) = \omega(\log n)$, from which we obtain that

$$\mathbb{P}\Big(h(T_{n,\theta_n}) \geq (1+2\epsilon)\mu(n,\theta_n)\Big) \leq e^{-\epsilon\mu(n,\theta_n)+o(\mu(n,\theta_n))} + e^{-\delta(\epsilon)\mu(n,\theta_n)} = O\left(\frac{1}{n^\lambda}\right) \,.$$

This corresponds to the desired upper tail bound. Combining this result with the lower tail bound from Proposition 3.2, we proved both the tail bound of the theorem and the convergence in probability. To conclude this proof, use Proposition 3.4 to extend the convergence in probability to the convergence in L^p for any $p > 0$. □

4. Future work and open questions

- This work describes the first order asymptotic behaviour of the height of record-biased trees, and it is natural to next consider lower-order fluctuations. In the regime where $\theta > c^*$, the height is principally controlled by the number of records, and the latter quantity satisfies a central limit theorem; this is a consequence of the central limit theorem for the number of cycles of Ewens permutations [4, Section 5] along with the Foata bijection [14]. It is is therefore natural to suspect that the height also satisfies a central limit theorem. On the other hand, when $\theta < c^*$, the height of record-biased permutations and the results of [10, 21] suggest that, in this case, $h(T_{n,\theta})$ has asymptotically bounded variance as $n \to \infty$. In the case that $\theta = c^*$, or more generally that $\theta_n \to c^*$, it is less clear to us what second-order behaviour to expect.
- The similarities between this work and [1] raise the question about more general frameworks under which these two models and results fall. Possible developments might come from using a more general model of random trees [12], or a more general model of random permutations [19] upon which the binary search trees are built. Considering the height of these generalized models of trees might lead to a unifying framework for Mallows and record-biased trees, as well as other models.

References

[1] L. Addario-Berry and B. Corsini, *The height of Mallows trees*, The Annals of Probability, 49 (2021), pp. 2220–2271.

[2] L. Addario-Berry and B. Reed, *Minima in branching random walks*, The Annals of Probability, 37 (2009), pp. 1044–1079.

[3] E. Aïdékon, *Convergence in law of the minimum of a branching random walk*, The Annals of Probability, 41 (2013), pp. 1362–1426.

[4] R. Arratia, A. D. Barbour, and S. Tavaré, *Logarithmic combinatorial structures: a probabilistic approach*, vol. 1, European Mathematical Society, 2003.

[5] N. Auger, M. Bouvel, C. Nicaud, and C. Pivoteau, *Analysis of algorithms for permutations biased by their number of records*, arXiv preprint arXiv:1605.02905, (2016).

[6] J. D. Biggins, *The first-and last-birth problems for a multitype age-dependent branching process*, Advances in Applied Probability, 8 (1976), pp. 446–459.

[7] N. Broutin, L. Devroye, E. McLeish, and M. de la Salle, *The height of increasing trees*, Random Structures & Algorithms, 32 (2008), pp. 494–518.

[8] F. M. Dekking and B. Host, *Limit distributions for minimal displacement of branching random walks*, Probability theory and related fields, 90 (1991), pp. 403–426.

[9] L. Devroye, *A note on the height of binary search trees*, Journal of the ACM (JACM), 33 (1986), pp. 489–498.

[10] M. DRMOTA, *An analytic approach to the height of binary search trees II*, Journal of the ACM (JACM), 50 (2003), pp. 333–374.

[11] ———, *The height of increasing trees*, Annals of Combinatorics, 12 (2009), pp. 373–402.

[12] S. EVANS, R. GRÜBEL, AND A. WAKOLBINGER, *Trickle-down processes and their boundaries*, Electronic Journal of Probability, 17 (2012), pp. 1–58.

[13] W. J. EWENS, *The sampling theory of selectively neutral alleles*, Theoretical population biology, 3 (1972), pp. 87–112.

[14] D. FOATA, *On the netto inversion number of a sequence*, Proceedings of the American Mathematical Society, 19 (1968), pp. 236–240.

[15] J. M. HAMMERSLEY, *Postulates for subadditive processes*, The Annals of Probability, 2 (1974), pp. 652–680.

[16] Y. HU AND Z. SHI, *Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees*, The Annals of Probability, 37 (2009), pp. 742–789.

[17] J. F. C. KINGMAN, *The first birth problem for an age-dependent branching process*, The Annals of Probability, (1975), pp. 790–801.

[18] C. L. MALLOWS, *Non-null ranking models. i*, Biometrika, 44 (1957), pp. 114–130.

[19] J. PITMAN AND W. TANG, *Regenerative random permutations of integers*, The Annals of Probability, 47 (2019), pp. 1378–1416.

[20] B. PITTEL, *On growing random binary trees*, Journal of Mathematical Analysis and Applications, 103 (1984), pp. 461–480.

[21] B. REED, *The height of a random binary search tree*, Journal of the ACM (JACM), 50 (2003), pp. 306–332.