The Subtree Size Profile of Plane-oriented Recursive Trees

Michael FUCHS*
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, 300, Taiwan
Email: mfuchs@math.nctu.edu.tw

Abstract

In this extended abstract, we outline how to derive limit theorems for the number of subtrees of size k on the fringe of random plane-oriented recursive trees. Our proofs are based on the method of moments, where a complex-analytic approach is used for constant k and an elementary approach for k which varies with n. Our approach is of some generality and can be applied to other simple classes of increasing trees as well.

1 Introduction

Several recent studies have been concerned with the *node profile* of rooted random trees, where the node profile is defined as the number of nodes at distance k from the root; for random binary search trees and recursive trees see Chauvin et al. [4], Chauvin et al. [5], Drmota and Hwang [11], [12], Fuchs et al. [22]; for random plane-oriented recursive trees see Hwang [23]; for other types of random trees see Drmota and Gittenberger [10], Drmota et al. [13], Drmota and Szpankowski [14], Park et al. [25].

Here, we are going to investigate another kind of profile which is defined as the number of subtrees of size k. This profile is called *subtree size profile* and has so far only been investigated for random binary search trees, random recursive trees and random Catalan trees; see Chang and Fuchs [3], Dennert and Grübel [7], Feng et al. [16], Feng et al. [17], Feng et al. [18], Fuchs [21].

Similar to the node profile, the subtree size profile is an important tree characteristic carrying a lot of information on the shape of a tree. For instance, total path length (sum of distances of all nodes to the root) and Wiener index (sum of distances between all nodes) can be easily computed from the subtree size profile. Also, as we will explain in more details below, results about the subtree size profile will in turn entail results about the occurrence of pattern sizes. Studying patterns in random trees is an important issue with many applications in computer science (for instance in the context of compressing; see Devroye [8] and Flajolet et al. [19]) and mathematical biology (see [3] and Rosenberg [26]).

In this paper, we will consider the subtree size profile for random plane-oriented recursive trees (or random PORTs for short). Random PORTs have surfaced in several recent applications sometimes under different names such as heap-ordered trees or scale-free random trees. They are for instance used as one of the most simplest model of random networks; see Barabási and Albert [1] and the thorough discussion in [23].

We will start by defining random PORTs. First, PORTs of size n are rooted plane trees with n nodes that are labeled such that the sequence of labels from the root to any node is increasing. Alternatively, PORTs can be defined via a tree evolution process: start with the root; inductively assume that a node of out-degree d has $d+1$ free places (before the first child, between the first and second child, etc.); attach the next node to a free place. Stop when you have attached n nodes. It is easy to see that this gives the same sets. Also, from the second definition, we can easily count the number τ_n of PORTs of size n. Therefore, observe that the sum over all out-degrees in any tree of size i equals $i-1$. Hence, the number of free places after inserting i nodes equals $2i-1$. Finally, τ_n is the product over all free places after inserting i nodes with $1 \le i \le n-1$. Thus,

$$\tau_n = 1 \cdot 3 \cdots (2n-3) = 2^{1-n} n! C_n,$$

where $C_n = \binom{2n-2}{n-1}/n$ are the (shifted) Catalan numbers. Note that the exponential generating function $P(z)$ of τ_n is given by $1-\sqrt{1-2z}$ and satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}z} P(z) = \frac{1}{1-P(z)} = 1 + P(z) + P(z)^2 + \cdots \qquad (1.1)$$

with initial condition $P(0)=0$. The latter equation can also be obtained by symbolic combinatorics: a PORT without the root (this is the left-hand-side) is the ordered sequence of PORTs tangling from the root (this is the right-hand-side).

A random PORT of size n is now defined as a PORT which is chosen from all PORTs of size n with probability

*Supported by National Science Council under the grant NSC-99-2115-M-009-007-MY2.

$1/\tau_n$. Again, an alternative definition can be given via the above tree evolution process: attach the next node uniformly to one of the free places. Note that under this tree evolution process, a node with high out-degree is more likely to attract the next node. This preferential attachment rule is one of the reasons of the importance of PORTs.

In the sequel, we will be interested in the subtree size profile of random PORTs which is a double-indexed random variable denoted by $X_{n,k}$. We will derive limit laws for fixed k and for k tending to infinity as n tends to infinity. More precisely, we will prove the following result.

THEOREM 1.1. *(i) (Normal range) Let $k = k_n$ such that $1 \leq k = o(\sqrt{n})$. Then,*

$$\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} \xrightarrow{d} N(0,1),$$

where $\mu_{n,k} = (2n-1)/(4k^2-1)$ and, as $n \to \infty$,

$$\sigma_{n,k}^2 \sim \left(\frac{8k^2-4k-8}{(4k^2-1)^2} - \frac{(2k-3)!!^2}{(k-1)!^2 4^{k-1} k(2k+1)}\right) n.$$

(ii) (Poisson range) Let $k = k_n$ such that $k \sim c\sqrt{n}$ as $n \to \infty$. Then,

$$X_{n,k} \xrightarrow{d} \text{Poisson}(2^{-1}c^{-2}).$$

(iii) (Degenerate range) Let $k = k_n$ such that $k < n$ and $\sqrt{n} = o(k)$ as $n \to \infty$. Then,

$$X_{n,k} \xrightarrow{L_1} 0.$$

We will explain some consequences of this result for the occurrences of pattern sizes (the following interpretation of our result was pointed out to us by one of the anonymous referees). Therefore, observe that the number of patterns of size k is given by C_k (the number of rooted plane trees). From Stirling's formula, we have

$$C_k \sim \pi^{-1/2} k^{-3/2} 4^{k-1} \qquad (k \to \infty). \tag{1.2}$$

Consequently, all patterns can only occur up to a size which is $\mathcal{O}(\log n)$. Beyond this, certain patterns cease to exist. Our result on the other hand shows that despite of this, all pattern sizes do still occur up to $o(\sqrt{n})$. Pattern sizes of order $\sqrt{n}$ then only exist sporadically and are Poisson. Finally, all pattern sizes beyond the order of $\sqrt{n}$ are highly unlikely.

The non-existence of patterns beyond $\sqrt{n}$ is consistent with other well-known properties of PORTs. For instance, if we denote by T_n the total path length, then the order of T_n is known to be $O(n \log n)$. Moreover, as mentioned above, T_n can easily be computed from $X_{n,k}$ as

$$T_n = \sum_{k=0}^{n-1} k X_{n,k}.$$

Hence, there cannot be patterns with sizes much beyond $\sqrt{n \log n}$.

Our result should be compared with the corresponding results for random binary search trees and recursive trees in [16] where the same phenomena was observed (with a slightly different expression for the mean and a more simpler expression for the variance). The proof method from [16], however, cannot be applied here since it rested on an exact expression for all centered moments which does not seem to be available here. So, we will have to devise a new method of proof. Our new method will use an inductive procedure called ”moment pumping” that was used in several recent studies in the analysis of algorithms (see Chern et al. [6] and references therein). In comparison to [16], our method will incorporate the cancelations from [16] already in the induction step making the resulting proof much simpler. Another advantage of our method is that it can be applied to other classes of random trees as well. This will be briefly indicated in a final section, where we are going to apply our method to other simple classes of random increasing trees, thereby showing that the above phenomena is universal for these tree families. We mention in passing that the subtree size profile for simple classes of random trees (for a definition see Drmota [9]) was investigated as well. For instance, Catalan trees were treated in [3] and a similar result as above was proved with the critical point $\sqrt{n}$ moving up to $n^{2/3}$. Similar to simple families of random increasing trees, the same phenomena is again expected to hold universally for many other simple families of random trees.

Before presenting details, we will give a short sketch of the paper. In the next section, we will look at the case of constant k and derive mean value and variance. Therefore, we will use complex-analytic tools, more precisely, singularity analysis with its closure properties. This method combined with induction can be extended to derive the first order asymptotics of all higher centered moments, too. By the Fréchet-Shohat Theorem, this then implies Theorem 1.1, part (i) for constant k. As for varying k, we will use an elementary approach (in the number-theoretic sense, i.e., no complex analysis is used) again based on moments and induction. Finally, in the last section, we will briefly indicate how our tools can be used to derive similar results for other simple classes of random increasing trees, too.

2 Constant Subtree Size - An Analytic Approach

Here, we will sketch the proof of Theorem 1.1, part (i) for constant k. Therefore, we will work with moments.

The first step will be to derive the mean value. Next, we will shift-the-mean and show that generating functions of centered moments satisfy a recursive relation (see Lemma 2.1 below). This combined with singularity analysis with its closure properties (see Chapter VI in Flajolet and Sedgewick [20]) will allow us to obtain the singularity expansion of the generating function of the variance from properties of the mean. Then, an asymptotic expansion for the variance will follow from the transfer theorems in [20]. Moreover, this procedure can be generalized to all higher centered moments as well. Finally, our result will follow from the Fréchet-Shohat Theorem (see Lemma 1.43 Elliott in [15]).

We mention in passing that one could alternatively try an approach based on the bivariate generating function (2.4) and Hwang's quasi-power theorem (see [20]). Such an approach has the advantage that one does not have to treat all moments, but is likely to be restricted to constant k only. Our approach via moments, however, also works for varying k as we will demonstrate in Section 3.

The starting point of our proof is the following (trivial) observation: the number of subtrees in a random PORT of size n is the sum of the number of subtrees in all subtrees of the root which are again random PORTs. This observation translates into the following distributional recurrence

$$X_{n,k} \stackrel{d}{=} \sum_{i=1}^{N} X_{I_i,k}^{(i)} \qquad (n > k) \tag{2.3}$$

with initial conditions $X_{k,k} = 1, X_{n,k} = 0$ for $n < k$ and $X_{n,k}^{(i)} \stackrel{d}{=} X_{n,k}$. Moreover, $X_{n,k}, X_{n,k}^{(i)}, (N, I_1, I_2, \dots)$ are independent random variables, where N is the out-degree of the root and $I_1, \dots, I_N$ are the sizes of the subtrees of the root. Due to the uniform probability model, we have for the joint distribution of $(N, I_1, I_2, \dots)$

$$\begin{aligned}\pi_{n,r,i_1,\dots,i_r} &:= P(N = r, I_1 = i_1, \dots, I_r = i_r) \\ &= \binom{n-1}{i_1,\dots,i_r}\frac{\tau_{i_1}\cdots\tau_{i_r}}{\tau_n},\end{aligned}$$

where $i_1, \dots, i_r \geq 1$ and $i_1 + \cdots + i_r = n - 1$. Since we are interested in moments of $X_{n,k}$, we define the bivariate generating function

$$P_k(z,y) = \sum_{n\geq 1} \tau_n \mathbb{E}\left(\exp(X_{n,k}y)\right)\frac{z^n}{n!}. \tag{2.4}$$

Then, (2.3) translates into

$$\frac{\partial}{\partial z}P_k(z,y) = \frac{1}{1-P_k(z,y)} + (e^y - 1)2^{1-k}kC_k z^{k-1} \tag{2.5}$$

with initial condition $P_k(0, y) = 0$. Note that $P_k(z, 0) = P(z)$ for all $k \geq 1$.

As (1.1), the above differential equation can also be alternatively obtained by symbolic combinatorics: a PORT without the root is the sequence of PORTs tangling from the root. Moreover, the number of subtrees of size k in a PORT is the sum of the number of subtrees of size k in the PORTs tangling from the root (this is the first term of the right-hand-side in (2.5)) except in the case where the PORT itself has size k where we have to add one (this is the correction term on the right-hand-side of (2.5)).

Next, we compute the mean. Therefore, set

$$M_k(z) = \sum_{n\geq 1} \tau_n \mathbb{E}(X_{n,k})\frac{z^n}{n!}.$$

Then, by differentiating the above differential equation with respect to y and setting $y = 0$, we obtain the following differential equation

$$\frac{\mathrm{d}}{\mathrm{d}z}M_k(z) = \frac{M_k(z)}{1-2z} + 2^{1-k}kC_k z^{k-1}$$

with initial condition $M_k(0) = 0$.

The solution of this differential equation is easily obtained as

$$M_k(z) = \frac{2^{1-k}kC_k}{\sqrt{1-2z}}\int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t. \tag{2.6}$$

From this we can derive the mean value.

PROPOSITION 2.1. *We have, for $n > k$,*

$$\mathbb{E}\left(X_{n,k}\right) = \frac{2n-1}{4k^2-1}.$$

Proof. From (2.6), we obtain

$$\mathbb{E}(X_{n,k}) = \frac{n!2^{1-k}kC_k}{\tau_n}[z^n]\frac{1}{\sqrt{1-2z}}\int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t.$$

Next,

$$\begin{aligned}&\frac{1}{\sqrt{1-2z}}\int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t \\ &= \frac{1}{\sqrt{1-2z}}\int_0^{1/2} t^{k-1}\sqrt{1-2t}\mathrm{d}t \\ &\quad + \frac{1}{\sqrt{1-2z}}\int_{1/2}^z t^{k-1}\sqrt{1-2t}\mathrm{d}t \\ &= \frac{2^{-k}B(k,3/2)}{\sqrt{1-2z}} \\ &\quad + \frac{2^{1-k}}{\sqrt{1-2z}}\int_{1/2}^z\sum_{l=0}^{k-1}\binom{k-1}{l}(-1)^l(1-2t)^{l+1/2}\mathrm{d}t \\ &= \frac{(k-1)!}{(2k+1)!!}\cdot\frac{1}{\sqrt{1-2z}} \\ &\quad + 2^{1-k}\sum_{l=0}^{k-1}\binom{k-1}{l}\frac{(-1)^{l+1}}{2l+3}(1-2z)^{l+1},\end{aligned} \tag{2.7}$$

where $B(x,y)$ is the beta function. Plugging this into the above expression gives for $n > k$

$$\mathbb{E}(X_{n,k}) = \frac{n!2^{1-k}k!C_k}{\tau_n(2k+1)!!}[z^n]\frac{1}{\sqrt{1-2z}} = \frac{2n-1}{4k^2-1},$$

where the last line follows by a simple computation.

REMARK 2.1. *Note that Theorem 1.1, part (iii) follows from the above exact expression of the mean value.*

Next, we consider the variance and higher centered moments. Therefore, we shift-the-mean. Set $\mu := 2/(4k^2-1)$ and

$$\begin{aligned}\bar{P}_k(z,y) &= \sum_{n\geq 1}\tau_n\mathbb{E}(\exp((X_{n,k}-\mu n)y))\frac{z^n}{n!}\\ &= P_k(ze^{-\mu y}, y).\end{aligned}$$

Then, (2.5) becomes

$$\frac{\partial}{\partial z}\bar{P}_k(z,y) = \frac{e^{-\mu y}}{1-\bar{P}_k(z,y)} + (e^y-1)e^{-k\mu y}2^{1-k}kC_kz^{k-1}$$

with initial condition $\bar{P}_k(0,y) = 0$. Next, set

$$\bar{A}_k^{[m]}(z) = \sum_{n\geq 1}\tau_n\mathbb{E}(X_{n,k}-\mu n)^m\frac{z^n}{n!}.$$

Then, again by differentiation, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}z}\bar{A}_k^{[m]}(z) = \frac{\bar{A}_k^{[m]}(z)}{1-2z} + \bar{B}_k^{[m]}(z)$$

with initial condition $\bar{A}_k^{[m]}(0) = 0$ and

$$\begin{aligned}&\bar{B}_k^{[m]}(z)\\ &\quad= \left((-k\mu+1)^m - (-k\mu)^m\right)2^{1-k}kC_kz^{k-1}\\ &\qquad+ \sum_{i=0}^{m-1}\binom{m}{i}(-\mu)^{m-i}\frac{\partial^i}{\partial y^i}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}\\ &\qquad+ \sum_{\substack{i_1+i_2+i_3=m-1\\ i_1<m-1}}\binom{m-1}{i_1,i_2,i_3}\bar{A}_k^{[i_1+1]}(z)\\ &\qquad\cdot\frac{\partial^{i_2}}{\partial y^{i_2}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}\frac{\partial^{i_3}}{\partial y^{i_3}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0},\end{aligned}$$

where

$$\begin{aligned}&\frac{\partial^i}{\partial y^i}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}\\ &\qquad= \sum_{l=0}^{i}\binom{i}{l}\mu^l\frac{\mathrm{d}}{\mathrm{d}z}\bar{A}_k^{[l]}(z)\\ &\qquad\quad- \left((1-k\mu)^i - (-k\mu)^i\right)2^{1-k}kC_kz^{k-1}.\end{aligned}$$

Note that $\bar{B}_k^{[m]}(z)$ is a function of $\bar{A}_k^{[i]}(z)$ with $i < m$.

The above differential equation is easily solved.

LEMMA 2.1. *We have,*

$$\bar{A}_k^{[m]}(z) = \frac{1}{\sqrt{1-2z}}\int_0^z\bar{B}_k^{[m]}(t)\sqrt{1-2t}\,\mathrm{d}t.$$

This is a recurrence for the generating functions of all centered moments. Hence, we can derive singularity expansions of these generating functions by the closure properties in Chapter VI in [20] and induction. Asymptotic expansions for the centered moments will then follow from the transfer theorems of Chapter VI in [20].

We will first demonstrate this procedure by treating the variance. Since we want to use the results from Chapter VI in [20], we need to show that the generating functions are analytic in a domain of the form: for $R > 1$ and $0 < \phi < \pi/2$,

$$\Delta(R,\phi) = \{z\ :\ |z| < R, z \neq 1/2, |\arg(z-1/2)| > \phi\}.$$

We call such a domain a Δ-*domain*. Moreover, a function is called *analytic in a* Δ-*domain* if it is analytic in a Δ-domain for some R and ϕ.

Now, we prove the following result.

PROPOSITION 2.2. *$\bar{A}_k^{[2]}(z)$ is analytic in a Δ-domain. Moreover, we have the singularity expansion*

$$\bar{A}_k^{[2]}(z) = \frac{\sigma^2}{2\sqrt{1-2z}} + \mathcal{O}(1)\qquad(z\to 1/2, z\in\Delta),$$

where

$$\sigma^2 = \frac{8k^2-4k-8}{(4k^2-1)^2} - \frac{(2k-3)!!^2}{(k-1)!^2 4^{k-1}k(2k+1)}.$$

Proof. First, note that $\bar{B}_k^{[2]}(z)$ is a function of

$$\bar{A}_k^{[1]}(z) = M_k(z) - \frac{z\mu}{\sqrt{1-2z}}.$$

Hence, from the expression for $M_k(z)$ derived in the proof of Proposition 2.1, we obtain that $\bar{B}_k^{[2]}(z)$ is Δ-analytic and $\bar{B}^{[2]}(z) = \mathcal{O}(1/(1-2z))$ as $z \to 1/2$ and $z \in \Delta$. Consequently, by Lemma 2.1 and Theorem VI.9 in [20], $\bar{A}^{[2]}(z)$ is also Δ-analytic and

$$\bar{A}^{[2]}(z) = \frac{c}{\sqrt{1-2z}} + \mathcal{O}(1)$$

as $z \to 1/2$ and $z \in \Delta$, where

$$c = \int_0^{1/2}\bar{B}_k^{[2]}(t)\sqrt{1-2t}\,\mathrm{d}t.$$

Finally, the claimed expression for c is obtained from the exact expression of $\bar{B}_k^{[2]}(z)$ (by plugging in the exact expression of $M_k(z)$) and a straightforward computation.

Applying the transfer theorems in [20] gives now the following consequence.

COROLLARY 2.1. *As $n \to \infty$,*

$$\mathrm{Var}\,(X_{n,k}) = \sigma^2 n + \mathcal{O}\left(n^{1/2}\right).$$

REMARK 2.2. *The above method can be used to derive longer expansions of the variance, too.*

As for higher centered moments, we use the same procedure together with induction (details will be given in the journal paper).

PROPOSITION 2.3. *$\bar{A}_k^{[m]}(z)$ is analytic in a Δ-domain for all $m \geq 3$. Moreover, we have the singularity expansions*

$$\bar{A}_k^{[2m-1]}(z) = \mathcal{O}\left((1-2z)^{3/2-m}\right) \quad (z \to 1/2, z \in \Delta)$$

and

$$\bar{A}_k^{[2m]}(z) = \frac{(2m)!(2m-3)!!\sigma^{2m}}{4^m m!}(1-2z)^{1/2-m} + \mathcal{O}\left((1-2z)^{1-m}\right) \quad (z \to 1/2, z \in \Delta).$$

Theorem 1.1, part (i) for constant k follows from this by the transfer theorems in [20] and the Fréchet-Shohat Theorem.

3 Variable Subtree Size - An Elementary Approach

Now, we will sketch the proof of the remaining cases of Theorem 1.1. Therefore, assume that $k = k_n$ with $k \to \infty$ as $n \to \infty$. We will concentrate on part (i), part (ii) being proved similarly.

As in the previous section, our proof is based on moments and induction. However, here we will work directly with the underlying sequences instead of their generating functions. Moreover, we will use a slightly different starting point. Therefore, observe that a random PORT can be decomposed into two random PORTs, one being the leftist subtree and the other the remaining tree. This gives the following distribution recurrence

$$X_{n,k} \stackrel{d}{=} X_{I_n,k} + X^*_{n-I_n,k} - \mathbf{1}_{\{n-I_n=k\}} \qquad (n > k) \tag{3.8}$$

with initial conditions $X_{k,k} = 1, X_{n,k} = 0$ for $n < k$ and $X^*_{n,k} \stackrel{d}{=} X_{n,k}$. Here, I_n is the size of the leftist subtree. Due to the uniform random model, we have

$$\pi_{n,j} := P(I_n = j) = \frac{2(n-j)C_j C_{n-j}}{nC_n} \quad (1 \leq j < n).$$

Since the mean value was already derived in the previous section, we can immediately concentrate on centered moments. Therefore, set

$$\mu_{n,k} = \begin{cases} 0, & \text{if } n < k, \\ 1, & \text{if } n = k; \\ 2n/(4k^2-1), & \text{if } n > k; \end{cases}$$

and $\bar{P}_{n,k}(z) = \mathbb{E}(\exp((X_{n,k} - \mu_{n,k})z))$. Then, from (3.8)

$$\bar{P}_{n,k}(z) = \sum_{1 \leq j < n} \pi_{n,j} \bar{P}_{j,k}(z) \bar{P}_{n-j,k}(z) e^{\Delta_{n,j,k} z} \ (n > k),$$

where $\bar{P}_{n,k}(z) = 1$ for $n \leq k$ and

$$\Delta_{n,j,k} = \mu_{j,k} + \mu_{n-j,k} - \mu_{n,k} - \mathbf{1}_{\{j=n-k\}}.$$

Next, set $\bar{A}_{n,k}^{[m]} = \mathbb{E}(X_{n,k} - \mu_{n,k})^m$. Differentiating yields the following recurrence

$$\bar{A}_{n,k}^{[m]} = \sum_{1 \leq j < n} \pi_{n,j} \left(\bar{A}_{j,k}^{[m]} + \bar{A}_{n-j,k}^{[m]}\right) + \bar{B}_{n,k}^{[m]} \ (n > k), \tag{3.9}$$

where $\bar{A}_{n,k}^{[m]} = 0$ for $n \leq k$ and

$$\bar{B}_{n,k}^{[m]} = \sum_{\substack{i_1+i_2+i_3=m \\ 0 \leq i_1, i_2 < m}} \binom{m}{i_1, i_2, i_3} \sum_{1 \leq j < n} \pi_{n,j} \bar{A}_{j,k}^{[i_1]} \bar{A}_{n-j,k}^{[i_2]} \Delta_{n,j,k}^{i_3}. \tag{3.10}$$

The above recurrence can be easily solved.

LEMMA 3.1. *For $n > k$,*

$$\bar{A}_{n,k}^{[m]} = \sum_{k+1 \leq j \leq n} \frac{C_j(n+1-j)}{C_n} \bar{B}_{j,k}^{[m]}. \tag{3.11}$$

Proof. This is a straightforward justification.

Consequently, since $\bar{B}_{n,k}^{[m]}$ is a function of $\bar{A}_{n,k}^{[i]}$ with $i < m$, we again have a recursive relation for $\bar{A}_{n,k}^{[m]}$. Therefore, we can again use induction to obtain first order asymptotics of all higher centered moments. However, to estimate error terms, we will need a uniform bound as well which is derived by another induction (see [22] where the same approach was used in a different context).

We first treat the variance.

PROPOSITION 3.1. *(i) For $k, n \geq 1$,*

$$\mathrm{Var}(X_{n,k}) = \mathcal{O}\left(\frac{n}{k^2}\right).$$

(ii) Let $k = k_n$ such that $k \to \infty$ as $n \to \infty$. Then, as $n \to \infty$,

$$\mathrm{Var}(X_{n,k}) \sim \frac{n}{2k^2}.$$

Proof. Here, (3.10) becomes

$$\bar{B}_{n,k}^{[2]} = \sum_{1 \leq j < n} \pi_{n,j} \Delta_{n,j,k}^2.$$

Next, note that
(3.12)
$$\bar{B}^{[2]}_{n,k} = \mathcal{O}\left(\frac{1}{k^2} + \pi_{n,k}\right) = \mathcal{O}\left(\frac{1}{k^2} + \frac{(n-k)C_kC_{n-k}}{nC_n}\right).$$

We will plug this into Lemma 3.1 and treat the resulting two sums separately. First, for the first sum,

$$\begin{aligned}
&\frac{1}{k^2}\sum_{k+1\le j\le n}\frac{C_j(n+1-j)C_{n+1-j}}{C_n}\\
&\quad= \mathcal{O}\left(\frac{n^{3/2}}{k^2}\sum_{k+1\le j\le n} j^{-3/2}(n+1-j)^{-1/2}\right)\\
&\quad= \mathcal{O}\left(\frac{\sqrt{n}}{k^2}\int_{k/n}^{1} x^{-3/2}(1-x)^{-1/2}dx\right)\\
&\quad= \mathcal{O}\left(\frac{n}{k^{5/2}}\right),
\end{aligned}$$

where we have used (1.2). Next, for the second sum

$$\begin{aligned}
&\frac{C_k}{C_n}\sum_{k+1\le j\le n}\frac{(j-k)C_{j-k}(n+1-j)C_{n+1-j}}{j}\\
&\quad= \frac{C_k}{C_n}\sum_{1\le j\le n-k} C_j(n-k+1-j)C_{n-k+1-j}\\
&\qquad - \frac{kC_k}{C_n}\sum_{k+1\le j\le n}\frac{C_{j-k}(n+1-j)C_{n+1-j}}{j}\\
&\quad= \frac{C_k(n-k+1)C_{n-k+1}}{2C_n}\\
&\qquad - \frac{kC_k}{C_n}\sum_{k+1\le j\le n}\frac{C_{j-k}(n+1-j)C_{n+1-j}}{j}\\
&\quad= \frac{1}{2}\mathbb{E}(X_{n,k}) = \frac{2n-1}{2(2k+1)(2k-1)},
\end{aligned}$$

where the first claim in the last line is proved by showing that the expression solves the recurrence for the mean. This proves part (i).

As for part (ii), observe that (3.12) can be refined to

$$\bar{B}^{[2]}_{n,k} = \pi_{n,k}\Delta^2_{n,k,k} + \mathcal{O}\left(1/k^2\right).$$

Now, we again use Lemma 3.1. The error term is treated as above. As for the main term, observe that $\Delta^2_{n,k,k} \sim 1$ as $n \to \infty$. Consequently,

$$\begin{aligned}
&\sum_{k+1\le j\le n}\frac{C_j(n+1-j)C_{n+1-j}}{C_n}\pi_{j,k}\Delta^2_{j,k,k}\\
&\quad\sim \frac{2C_k}{C_n}\sum_{k+1\le j\le n}\frac{(j-k)C_{j-k}(n+1-j)C_{n+1-j}}{j}\\
&\quad\sim \frac{n}{2k^2},
\end{aligned}$$

where the last line follows as above. This proves part (ii).

The next step is to use induction to generalize the previous result to all higher centered moments. The proof is long and cumbersome and we will not give details here.

PROPOSITION 3.2. *(i) For $k, n \ge 1$ and $m \ge 2$,*

$$\bar{A}^{[m]}_{n,k} = \mathcal{O}\left(\max\left\{\frac{n}{k^2}, \left(\frac{n}{k^2}\right)^{m/2}\right\}\right).$$

(ii) Let $k = k_n$ such that $k = o(\sqrt{n})$ and $k \to \infty$ as $n \to \infty$. Then, as $n \to \infty$,

$$\bar{A}^{[2m-1]}_{n,k} = o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\right);$$

$$\bar{A}^{[2m]}_{n,k} \sim g_m\left(\frac{n}{2k^2}\right)^m,$$

for $m \ge 1$, where $g_m = (2m)!/(2^m m!)$.

From this, our claimed result follows by the Fréchet-Shohat Theorem.

4 Other Simple Classes of Increasing Trees

In this final section, we indicate some extensions of our result to other simple classes of increasing trees. We will be rather brief and postpone further details to the journal version of this paper.

Simple classes of increasing trees were introduced in Bergeron et al. [2] and are defined as follows: first, an *increasing tree* is a rooted, plane, node-labeled tree with the sequence of labels from the root to any node being increasing. *A simple class of increasing trees* is then defined as the class of increasing trees together with a *weight sequence* ϕ_r with $\phi_0 > 0$ and $\phi_r > 0$ for some $r \ge 2$. The ordinary generating function of the weight sequence, i.e.,

$$\phi(\omega) = \sum_{r\ge 0}\phi_r\omega^r$$

is called *weight function*. Using the weight sequence, a weight is attached to every increasing tree T as follows

$$\omega(T) = \prod_{v\in T}\phi_{d(v)},$$

where the product runs over all nodes of T and $d(v)$ is the out-degree of v.

Next, we equip a simple class of increasing trees with a probability model, where the probability of T is proportional to its weight, i.e., if

$$\tau_n = \sum_{T \text{ has } n \text{ nodes}}\omega(T),$$

then the probability of a tree T with n nodes is $\omega(T)/\tau_n$. Note that if $P(z)$ denotes the exponential generating function of τ_n then

$$\frac{\mathrm{d}}{\mathrm{d}z}P(z) = \phi(P(z)) = \phi_0 + \phi_1 P(z) + \cdots$$

with initial condition $P(0) = 0$. As (1.1), this differential equation can again be derived by symbolic combinatorics: an increasing tree without the root (this is the left-hand-side) is the sequence of increasing trees tangling from the root where we have to multiply with ϕ_d if the root has out-degree d (this is the right-hand-side).

It is easy to see that random binary search trees, random recursive trees and random PORTs are all simple classes of increasing trees. Note that all of them can be generated by a tree evolution process. It is interesting to ask which other simple classes of increasing trees admit such a construction (via a natural tree evolution process). This question was solved in Panholzer and Prodinger [24], where it was shown that up to scaling such a construction exists if and only if the class belongs to following three types.

- Random d-ary trees: $\phi(\omega) = (1 + t)^d$, where $d \in \{2, 3, 4, \ldots\}$.
- Random recursive trees: $\phi(\omega) = e^t$.
- Generalized random PORTs: $\phi(\omega) = (1 - t)^{-r+1}$, where $r > 1$.

Using our previous approach, a similar result as Theorem 1.1 can be proved for those classes as well (for instance, again by symbolic combinatorics, one easily derives a differential equation similar to (2.5) for the bivariate generating function). We just state our result for generalized random PORTs.

First, the mean value of generalized random PORTs has again a simple expression.

THEOREM 4.1. *We have for $n > k$*

$$\mu_{n,k} := \mathbb{E}(X_{n,k}) = \frac{(r-1)(rn-1)}{(rk+r-1)(rk-1)}.$$

Moreover, results for the variance and limit laws can be given as well. We just state the result for varying k.

THEOREM 4.2. *(i) (Normal range) Let $k = k_n$ such that $k \to \infty$ as $n \to \infty$. Then,*

$$\frac{X_{n,k} - \mu_{n,k}}{\sigma_{n,k}} \xrightarrow{d} N(0, 1),$$

where, as $n \to \infty$,

$$\sigma^2_{n,k} \sim \frac{r-1}{r} \cdot \frac{n}{k^2}.$$

(ii) (Poisson range) Let $k = k_n$ such that $k \sim c\sqrt{n}$ as $n \to \infty$. Then,

$$X_{n,k} \xrightarrow{d} \text{Poisson}((r-1)r^{-1}c^{-2}).$$

(iii) (Degenerate range) Let $k = k_n$ such that $k < n$ and $\sqrt{n} = o(k)$ as $n \to \infty$. Then,

$$X_{n,k} \xrightarrow{L_1} 0.$$

Acknowledgments

We thank Markus Kuba and Alois Panholzer for bringing the current topic to our attention and for some insightful discussions. Moreover, we thank Brigitte Vallée and the anonymous referees for many helpful comments.

References

[1] A.-L. Barabási and R. Albert (1999). Emergence of scaling in random networks, *Science*, **286**, 509-512.

[2] F. Bergeron, P. Flajolet, and B. Salvy (1992). Varieties of increasing trees, *Lectures Notes in Computer Science*, **581**, 24-48.

[3] H. Chang and M. Fuchs (2010). Limit theorems for pattern in phylogenetic trees, *Journal of Mathematical Biology*, **60**, 481-512.

[4] B. Chauvin, M. Drmota, and J. Jabbour-Hattab (2001). The profile of binary search trees, *Annals of Applied Probability*, **11**, 1042-1062.

[5] B. Chauvin, T. Klein, J.-F. Marckert, and A. Rouault (2005). Martingales and profile of binary search trees, *Electronic Journal of Probability*, **10**, 420-435.

[6] H.-H. Chern, M. Fuchs, and H.-K. Hwang (2007). Phase changes in random point quadtrees, *ACM Transactions on Algorithms*, **3**, 51 pages.

[7] F. Dennert and R. Grübel (2010). On the subtree size profile of binary search trees, *Combinatorics, Probability and Computing*, **19**, 561-578.

[8] L. Devroye (1998). On the richness of the collection of subtrees in random binary search trees, *Information Processing Letters*, **65**, 195-199.

[9] M. Drmota. *Random Trees*. SpringerWienNewYork, 2009.

[10] M. Drmota and B. Gittenberger (1997). On the profile of random trees, *Random Structures and Algorithms*, **10**, 421-451.

[11] M. Drmota and H.-K. Hwang (2005). Bimodality and phase transitions in the profile variance of random binary search trees, *SIAM Journal on Discrete Mathematics*, **19**, 19-45.

[12] M. Drmota and H.-K. Hwang (2005). Profile of random trees: correlation and width of random recursive trees and binary search trees, *Advances in Applied Probability*, **37**, 321-341.

[13] M. Drmota, S. Janson, and R. Neininger (2008). A functional limit theorem for the profile of search trees, *The Annals of Applied Probability*, **18**, 288-333.

[14] M. Drmota and W. Szpankowski (2010). The expected behavior of digital search tree profile, submitted.

[15] P. D. T. A. Elliott. *Probabilistic Number Theory I. Central Limit Theorems*. Springer, New-York, 1979.

[16] Q. Feng, H. Mahmoud, and A. Panholzer (2008). Phase changes in subtree varieties in random recursive trees and binary search trees, *SIAM Journal on Discrete Mathematics*, **22**, 160-184.

[17] Q. Feng, H. Mahmoud, and C. Su (2007). On the variety of subtrees in a random recursive tree, Technical report, The George Washington University, Washington, DC.

[18] Q. Feng, B. Miao, and C. Su (2006). On the subtrees of

binary search trees, *Chinese Journal of Applied Probability and Statistics*, **22**, 304-310.

[19] P. Flajolet, X. Gourdon, C. Martinez (1997). Patterns in random binary search trees, *Random Structures and Algorithms*, **11**, 223-244.

[20] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[21] M. Fuchs (2008). Subtree Sizes in Recursive Trees and Binary Search Trees: Berry-Esseen Bounds and Poisson Approximations, *Combinatorics, Probability and Computing*, **17**, 661-680.

[22] M. Fuchs, H.-K. Hwang, and R. Neininger (2006). Profiles of random trees: Limit theorems for random recursive trees and binary search trees, *Algorithmica*, **46**, 367-407.

[23] H.-K. Hwang (2007). Profiles of random trees: plane-oriented recursive trees, *Random Structures and Algorithms*, **30**, 380-413.

[24] A. Panholzer and H. Prodinger (2007). The level of nodes in increasing trees revisited, *Random Structures and Algorithms*, **31**, 203-226.

[25] G. Park, H.-K. Hwang, P. Nicodeme, and W. Szpankowski (2009). Profiles of tries, *SIAM Journal on Computing*, **38**, 1821-1880.

[26] N. A. Rosenberg (2006). The mean and variance of the number of r-pronged nodes and r-caterpillars in Yule-generated genealogical trees, *Annals of Combinatorics*, **10**, 129-146.