Limit Theorems for Subtree Size Profiles of Increasing Trees

Michael FUCHS
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, 300
Taiwan

October 31, 2011

Abstract

Simple families of increasing trees have been introduced by Bergeron, Flajolet and Salvy. They include random binary search trees, random recursive trees and random plane-oriented recursive trees (PORTs) as important special cases. In this paper, we investigate the number of subtrees of size k on the fringe of some classes of increasing trees, namely generalized PORTs and d-ary increasing trees. We use a complex-analytic method to derive precise expansions of mean value and variance as well as a central limit theorem for fixed k. Moreover, we propose an elementary approach to derive limit laws when k is growing with n. Our results have consequences for the occurrence of pattern sizes on the fringe of increasing trees.

1 Introduction

Several recent studies have been concerned with the *profile* of rooted random trees, where a couple of different notions of the profile have been proposed. The oldest and most widely-used notion counts the number of nodes at a fixed distance k from the root. This kind of profile which is called *node profile* has been extensively studied for many different families of trees; for random binary search trees and recursive trees see Chauvin, Drmota, and Jabbour-Hattab [4], Chauvin, Klein, Marckert, and Rouault [5], Fuchs, Neininger, and Hwang [22], Drmota and Hwang [10], [11]; for random plane-oriented recursive trees see Hwang [23]; for other types of random trees see Drmota and Gittenberger [9], Drmota, Janson, and Neininger [12], Drmota and Szpankowski [13], Park, Hwang, Nicodeme, and Szpankowski [26].

In this paper, we are interested in another notion of profile defined as the number of subtrees of size k on the fringe of rooted random trees. This profile which is called the *subtree size profile* has so far only been investigated for random binary search trees and recursive trees. More specifically, limit theorems have been derived in Feng, Mahmoud, and Panholzer [15], Feng, Mahmoud, and Su [16], Feng, Miao, and Su [17]; Berry-Esseen bounds, local limit theorems and Poisson approximation results have been discussed in Fuchs [21]; and functional limit laws have been proved by Dennert and Grübel [7]. Similar to the node profile, the subtree size profile is a fine tree characteristic carrying a lot of information about the shape of

Key words: subtree size profile, increasing trees, plane-oriented recursive trees, limit laws, singularity analysis.

a tree. For instance, the total path length (sum of distances of all nodes to the root) and the Wiener index (sum of distances between any two nodes) can be easily computed from the subtree size profile. Moreover, results about the subtree size profile in turn entail results about the occurrence of pattern sizes (for a more thorough discussion see below). Studying pattern occurrence in random trees is an important issue in Computer Science (for instance in the context of compression; see Devroye [8] and Flajolet, Gourdon, and Martinez [19]) as well as in Phylogenetics (see Chang and Fuchs [3] and Rosenberg [27]).

Here, we are going to investigate the subtree size profile for other families of rooted random trees. More precisely, we will propose a method which is applicable to several classes of simple families of random increasing trees as defined in Bergeron, Flajolet, and Salvy [2]. These families have numerous applications (see [2]) and contain random binary search trees, random recursive trees and random plane-oriented recursive trees (PORTs) as special cases. We will explain our approach and work out all details for random PORTs in the next two sections (random binary search trees and recursive trees have been treated in [16]). Generalized PORTs and d-ary increasing trees will then be briefly treated in a final section.

We will start by giving some more details on random PORTs. Random PORTs have surfaced in a couple of different applications sometimes under different names such as random 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]. As for the definition of random PORTs, first a PORT is a rooted, plane tree together with a labeling of the vertices, where labels along any path from the root to a leaf form an increasing sequence. If we fix the number of nodes to be n, then an easy counting argument shows that the number τ_n of PORTs with n nodes is given by

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

where $C_n = \binom{2n-2}{n-1}/n$ are the (shifted) Catalan numbers. A random PORT is then obtained by uniformly picking a PORT of size n.

There is an equivalent definition of a random PORT of size n via a tree evolution process: start from the root and recursively attach new nodes, where an existing node with d children is supposed to have $d+1$ free places (in front of the first child, between the first child and the second child, etc.) and the new incoming node is attached to a place that is chosen uniformly from all those free places. Stop when you have attached n nodes and the resulting tree is again a random PORT. Note that in this tree evolution process, nodes with a large number of children are more likely to attract the new incoming node. This preferential attachment rule is the reason for the importance of PORTs as simple network models.

Now, for a random PORT of size n, we denote by $X_{n,k}$ the number of subtrees of size k (the same notation for the subtree size profile will also be used for other types of random trees). In this work, we are interested in the limiting properties of $X_{n,k}$ both for fixed k and for k tending to infinity as n tends to infinity. More precisely, we will prove the following result.

Theorem 1. *(i) (Normal range) Let $k = k(n)$ such that $1 \le k = o(\sqrt{n})$. 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,

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

For fixed k, this result will follow by standard tools. Hence, our main contribution will be the treatment of varying k. Here, the result entails some interesting consequences concerning the occurrence of patterns (by which we mean the subtrees rooted at the nodes of a tree). In order to explain these consequences, first observe that the number of patterns of size k is equal to the number of rooted, plane trees of size k which is given by

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

Hence, *all pattern* can just occur about to size $k = \mathcal{O}(\log n)$. Beyond this order, some pattern will cease to exist. Our result on the other hand shows that *all pattern sizes* are still present up to $k = o(\sqrt{n})$. Moreover, patterns sizes of order $\sqrt{n}$ exist only sporadically and are Poisson. Finally, pattern sizes of order beyond $\sqrt{n}$ are highly unlikely. Note that the latter non-existence is consistent with the stochastic behavior of other shape parameters. For instance, it is well-known that the total path length T_n of random PORTs is of order $\mathcal{O}(n \log n)$; see Smythe and Mahmoud [28]. Now notice that we have the following connection between T_n and the subtree size profile

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

Consequently, if all pattern sizes exist up to index k_0, then

$$\Theta(k_0^2) = \sum_{k=0}^{k_0} k \le \sum_{k=0}^{n-1} k X_{n,k} = T_n = \mathcal{O}(n \log n).$$

Thus, pattern sizes beyond $\sqrt{n \log n}$ are very unlikely.

Next, we are going to discuss the method of proof of the above theorem. Therefore, note that our result is almost identical to the results for random binary search trees and recursive trees in [15] (only the variance has a more complicated shape). The proof in the latter paper, however, rested on a precise expression for all moments of $X_{n,k}$ and such an expression is not available in the current situation. Consequently, a new method of proof has to be devised. Our new approach will again work with moments, but in difference to [15] we will use induction to derive the first order asymptotics of all moments (such an approach was nicknamed "moment-pumping" in several recent papers; see Chern, Fuchs, and Hwang [6] and references therein). Moreover, we will directly work with central moments. This will incorporate the tedious cancelations from [15] in the induction step, making the resulting proof much easier. Another advantage is that this approach will be applicable to other families of random trees as well, thereby showing that the above phenomena hold more generally for many families of random trees.

We will conclude the introduction by giving a more detailed sketch of our approach. In Section 2 we will consider the case of fixed k. As already mentioned before, this case is standard and a variety of approaches could be used (e.g. bivariate generating functions combined with Hwang's quasi-power theorem or contraction method; see Flajolet and Sedgewick [20] for the former and [7] for the latter). The approach we choose will work with moments and use complex-analytic tools. To give some more details, our starting point is the easy observation that the number of subtrees of size k in a random PORT is obtained as the sum of the numbers of subtrees of size k in all subtrees of the root which again are random PORTs. This yields the following distributional recurrence

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

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, $X_{n,k}, X_{n,k}^{(i)}, (N, I_1, I_2, \ldots)$ are independent random variables, where N is the out-degree of the root and $I_1, \ldots, I_N$ are the sizes of the subtrees of the root. Due to the uniform probability model, we have

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

where $i_1, \ldots, i_r \geq 1$ and $i_1+\cdots+i_r = n-1$. Next, we consider the following (scaled) exponential generating function of the m-th moment

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

Straightforward computation then reveals that all these generating functions satisfy the following type of differential equation

$$A'(z) = \frac{A(z)}{1-2z} + B(z), \qquad A(0)=0,$$

where $B(z)$ is a function of generating functions of moments of smaller order. Generating functions of centered moments also satisfy the same differential equation. This differential equation is easily solved

$$A(z) = \frac{1}{\sqrt{1-2z}} \int_0^z B(t)\sqrt{1-2t}\mathrm{d}t. \tag{2}$$

Thus, we have a recursive scheme. From this, the above result for fixed k is obtained as follows: we first derive an exact expression for the mean value and an asymptotic expansion for the variance. We then shift-the-mean and use induction to deduce the first order asymptotics of all higher centered moments. Here, we will use singularity analysis with its closure properties (see Chapter VI in [20]). Finally, our result will follow from the Fréchet-Shohat Theorem (see Lemma 1.43 in [14]). A similar strategy was used in a recent paper of Fill and Kapur [18] for studying additive functionals in Catalan trees.

The above approach using generating functions and singularity analysis has the drawback that the dependency of the error terms on k is not clear. Hence, in Chapter 3, we will devise another approach for the more complicated case of variable k. Here, we will not work with generating functions, but directly with the underlying sequences. Even though one could read off a recurrence relation for the m-th moment from the above differential equation, it is easier to use a slightly different starting point. Therefore, observe that a random PORT can be decomposed into two random PORTs, one being the leftist subtree of the root and the other the remaining tree. This yields the following distributional recurrence for $X_{n,k}$

$$X_{n,k} \stackrel{d}{=} X_{I_n,k} + X^*_{n-I_n,k} - \mathbf{1}_{\{n-I_n=k\}} \qquad (n>k) \tag{3}$$

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. From our random model, it is easy to see that

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

Now, by taking expectations, one observes that all (centered or non-centered) moments satisfy a recurrence of the form

$$a_{n,k} = 2\sum_{1\leq j<n} \frac{C_jC_{n-j}}{C_n} a_{j,k} + b_{n,k} \qquad (n>k) \tag{4}$$

with $a_{k,k}$ given and $b_{n,k}$ a function of moments of smaller order. This recurrence is easily solved

$$a_{n,k} = \sum_{k+1\le j\le n} \frac{C_j(n+1-j)}{C_n} b_{j,k} + \frac{C_k(n+1-k)C_{n+1-k}}{C_n} a_{k,k} \qquad (n > k). \tag{5}$$

Thus one again has a recursive scheme. Now, one can apply the approach from [22] which works as follows: first we re-derive the mean. Then, we shift-the-mean and use induction to derive a uniform bound for all centered moments. Next, we use this uniform bound and another induction to derive the first order asymptotics of all centered moments in the normal range of Theorem 1. For the Poisson range, a similar approach is used, the only difference being that we will work with factorial moments. The final step is then again the Fréchet-Shohat Theorem. The same approach was already applied in [3] for random binary search trees, but in the current situation the technicalities are much more demanding.

Our two approaches above are of some generality and can be applied to other families of random trees as well. In particular, our approaches work for certain classes of simple families of increasing trees. We will recall the definition of these tree families in the final section and shortly outline how to deduce similar results for those families, too.

2 Constant Subtree Size - An Analytic Approach

Here, we will prove Theorem 1 for constant k. Therefore, we will follow the approach sketched in the introduction. First consider the double generating function

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

Then, (1) translates into the following differential equation.

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

with the initial condition $P_k(0,y) = 0$.

Next, we recall that $A_k^{[m]}(z)$ is the m-th derivative of $P_k(z,y)$ with respect to y at $y = 0$. Taking derivatives in the above differential equation yields

$$\frac{\mathrm{d}}{\mathrm{d}z} A_k^{[m]}(z) = \frac{A_k^{[m]}(z)}{1-2z} + B_k^{[m]}(z), \tag{6}$$

where $B_k^{[m]}(z)$ is a suitable function and $A_k^{[m]}(0) = 0$ (note that this verifies the claim from the introduction). We list two instances for $B_k^{[m]}(z)$ which will be needed below

$$B_k^{[1]}(z) = 2^{1-k}kC_k z^{k-1}, \qquad B_k^{[2]}(z) = \frac{2A_k^{[1]}(z)^2}{(1-2z)^{3/2}} + 2^{1-k}kC_k z^{k-1}. \tag{7}$$

More generally, $B_k^{[m]}(z)$ is a function of $A_k^{[i]}(z)$ with $i < m$.

As already mentioned in the introduction, we will use singularity analysis to obtain asymptotic expansions of moments. First, we need a definition. For some $R > 1$ and $0 < \phi < \pi/2$ set

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

which will be a called a Δ-*domain*. Moreover, we say that a function $f(z)$ is *defined* (or *analytic*, etc.) *on a* Δ-*domain* if $f(z)$ is analytic on $\Delta(R,\phi)$ for some R and ϕ.

Now, we can start by deriving the mean value and the variance of $X_{n,k}$.

Proposition 2. *For $n > k$,*

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

and, as $n \to \infty$,

$$\begin{aligned}\operatorname{Var}\left(X_{n,k}\right) = &\left(\frac{8k^2-4k-8}{\left(4k^2-1\right)^2} - \frac{(2k-3)!!^2}{(k-1)!^2 4^{k-1} k(2k+1)}\right) n - \frac{4k^2-2k-2}{\left(4k^2-1\right)^2} \\ &+ \frac{(2k-3)!!^2}{(k-1)!^2 4^{k-1} k(2k+1)} + \mathcal{O}\left(n^{-5/2}\right).\end{aligned}$$

Proof. We start with the mean value. Therefore, observe that by substituting (7) into the solution (2) of the differential equation (6), we obtain

$$A_k^{[1]}(z) = \frac{2^{1-k} k C_k}{\sqrt{1-2z}} \int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t.$$

Consequently,

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

Next,

$$\begin{aligned}\frac{1}{\sqrt{1-2z}} &\int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t \\ &= \frac{1}{\sqrt{1-2z}} \int_0^{1/2} t^{k-1}\sqrt{1-2t}\mathrm{d}t + \frac{1}{\sqrt{1-2z}} \int_{1/2}^z t^{k-1}\sqrt{1-2t}\mathrm{d}t \\ &= \frac{2^{-k} B(k,3/2)}{\sqrt{1-2z}} + \frac{2^{1-k}}{\sqrt{1-2z}} \int_{1/2}^z \sum_{l=0}^{k-1} \binom{k-1}{l} (-1)^l (1-2t)^{l+1/2}\mathrm{d}t \\ &= \frac{(k-1)!}{(2k+1)!!} \cdot \frac{1}{\sqrt{1-2z}} + 2^{1-k} \sum_{l=0}^{k-1} \binom{k-1}{l} \frac{(-1)^{l+1}}{2l+3} (1-2z)^{l+1}, \end{aligned} \tag{8}$$

where $B(x,y)$ denotes the beta function. Substituting this into the above expression yields 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 straightforward simplifications. Hence, the first claim of the result is proved.

In order to prove the second claim, again by (7) and (2),

$$\begin{aligned}A_k^{[2]}(z) &= \frac{2}{\sqrt{1-2z}} \int_0^z \frac{A_k^{[1]}(t)^2}{1-2t}\mathrm{d}t + \frac{2^{1-k} k C_k}{\sqrt{1-2z}} \int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t \\ &= \frac{2^{3-2k} k^2 C_k^2}{\sqrt{1-2z}} \int_0^z \frac{1}{(1-2t)^2} \left(\int_0^t u^{k-1}\sqrt{1-2u}\mathrm{d}u\right)^2 \mathrm{d}t + \frac{2^{1-k} k C_k}{\sqrt{1-2z}} \int_0^z t^{k-1}\sqrt{1-2t}\mathrm{d}t. \end{aligned} \tag{9}$$

The second term is the same as above. Thus, we only have to concentrate on the first term. Note that according to the computation above,

$$\frac{1}{(1-2t)^2} \left(\int_0^t u^{k-1}\sqrt{1-2u}\mathrm{d}u\right)^2 = \frac{(k-1)!^2}{(2k+1)!!^2} \cdot \frac{1}{(1-2t)^2} + \mathcal{O}\left(\frac{1}{\sqrt{1-2t}}\right)$$

as $t \to 1/2$ in a suitable Δ-domain. Consequently, by Theorem VI.9 in [20],

$$\frac{1}{\sqrt{1-2z}}\int_0^z \frac{1}{(1-2t)^2}\left(\int_0^t u^{k-1}\sqrt{1-2u}\mathrm{d}u\right)^2 \mathrm{d}t$$
$$= \frac{(k-1)!^2}{2(2k+1)!!^2}\cdot\frac{1}{(1-2z)^{3/2}} + \frac{c}{\sqrt{1-2z}} + \mathcal{O}\left(\sqrt{1-2z}\right)$$

as $z \to 1/2$ in a suitable Δ-domain. Here,

$$c = -\frac{(k-1)!^2}{2(2k+1)!!^2} + \int_0^{1/2}\left(\frac{1}{(1-2t)^2}\left(\int_0^t u^{k-1}\sqrt{1-2u}\mathrm{d}u\right)^2 - \frac{(k-1)!^2}{(2k+1)!!^2}\frac{1}{(1-2t)^2}\right)\mathrm{d}t.$$

Before simplifying this constant, we substitute what we have so far into (9) and use the transfer theorems from Chapter VI in [20]. This yields

$$\mathbb{E}(X_{n,k}^2) = \frac{n!2^{2-2k}k!^2C_k^2}{\tau_n(2k+1)!!^2}[z^n]\frac{1}{(1-2z)^{3/2}} + \frac{cn!2^{3-2k}k^2C_k^2}{\tau_n}[z^n]\frac{1}{\sqrt{1-2z}} + \frac{2n-1}{4k^2-1} + \mathcal{O}\left(\frac{n!}{\tau_n}2^n n^{-3/2}\right)$$

as $n \to \infty$. Now, recall the well-known asymptotics for the n-th Catalan number

$$C_n \sim \pi^{-1/2}n^{-3/2}4^{n-1}, \qquad (n \to \infty). \tag{10}$$

Consequently, the error term above is $\mathcal{O}(1)$. Further simplifying the other terms yields

$$\mathbb{E}(X_{n,k}^2) = \frac{4n^2}{(4k^2-1)^2} + \left(\frac{4c(2k-3)!!^2}{(k-1)!^2} + \frac{2}{4k^2-1}\right)n + \mathcal{O}(1)$$

as $n \to \infty$ and thus

$$\mathrm{Var}(X_{n,k}) = \left(\frac{4c(2k-3)!!^2}{(k-1)!^2} + \frac{2}{4k^2-1} - \frac{4}{(4k^2-1)^2}\right)n + \mathcal{O}(1)$$

as $n \to \infty$.

So, what is left is to find a simple expression for c. Therefore, we substitute (8) into the integral in the above expression for c. This together with some computation yields

$$c = -\frac{(k-1)!^2}{2(2k+1)!!^2} + \frac{2^{2-k}(k-1)!}{(2k+1)!!}\sum_{l=0}^{k-1}\binom{k-1}{l}\frac{(-1)^{l+1}}{(2l+1)(2l+3)}$$
$$+ 2^{1-2k}\sum_{l=0}^{k-1}\sum_{i=0}^{k-1}\binom{k-1}{l}\binom{k-1}{i}\frac{(-1)^{l+i}}{(2l+3)(2i+3)(l+i+2)}.$$

Now, either by standard simplifications or using Maple,

$$\sum_{l=0}^{k-1}\binom{k-1}{l}\frac{(-1)^{l+1}}{(2l+1)(2l+3)} = -\frac{2^{k-1}k!}{(2k+1)!!}$$

$$\sum_{l=0}^{k-1}\sum_{i=0}^{k-1}\binom{k-1}{l}\binom{k-1}{i}\frac{(-1)^{l+i}}{(2l+3)(2i+3)(l+i+2)} = -\frac{2}{k(2k+1)} + \frac{2^{2k-1}k!(k-1)!}{(2k+1)!!^2}.$$

Substituting this into the expression above and substituting the expression for c in turn into the above expression for the variance together with straightforward computations yields the claimed result.

Note that our method above just yields the main term in the asymptotic expansion of the variance. However, it is straightforward to extend our approach to obtain arbitrary long expansions of the variance, too. ∎

Remark 1. Note that Theorem 1, part (iii) immediately follows from the above explicit expression for the mean value.

Next, we are going to generalize the previous method to obtain an asymptotic expansion of all higher centered moments. Therefore, we first have to shift-the-mean. Thus, set $\bar{X}_{n,k} = X_{n,k} - \mu n$, where $\mu = 2/(4k^2 - 1)$. Moreover, set

$$\bar{P}_k(z,y) = \sum_{n\geq 1} \tau_n \mathbb{E}\left(\exp(\bar{X}_{n,k}y)\right)\frac{z^n}{n!} = P_k\left(y, ze^{-\mu y}\right).$$

Then, our original differential equation can be replaced by the following one

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

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

Now, denote by $\bar{A}_k^{[m]}(z)$ the m-th derivative of $\bar{P}_k(z,y)$ with respect to y at $y = 0$. These functions again satisfy our fundamental differential equation

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

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

$$\begin{aligned}\bar{B}_k^{[m]}(z) &= \left((-k\mu+1)^m - (-k\mu)^m\right)2^{1-k}kC_k z^{k-1} + \sum_{i=0}^{m-1}\binom{m}{i}(-\mu)^{m-i}\frac{\partial^i}{\partial y^i}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} \\ &\quad + \sum_{\substack{i_1+i_2+i_3=m-1\\ i_1<m-1}}\binom{m-1}{i_1,i_2,i_3}\bar{A}_k^{[i_1+1]}(z)\frac{\partial^{i_2}}{\partial y^{i_2}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}\frac{\partial^{i_3}}{\partial y^{i_3}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}. \qquad (11)\end{aligned}$$

Our next aim is to prove the following result.

Proposition 3. *$\bar{A}_k^{[m]}(z)$ is analytic in a Δ-domain for all $m \geq 1$. Moreover, we have the singularity expansions*

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

and

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

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. We will use induction, where apart from the above claim, we will also prove the following one

$$\frac{\partial^{2m-1}}{\partial y^{2m-1}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} = \mathcal{O}\left((1-2z)^{1/2-m}\right)$$

as $z \longrightarrow 1/2$ and $z \in \Delta$, and

$$\frac{\partial^{2m}}{\partial y^{2m}} \frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} = \frac{(2m)!(2m-1)!!\sigma^{2m}}{4^m m!}(1-2z)^{-1/2-m} + \mathcal{O}\left((1-2z)^{-m}\right) \tag{12}$$

as $z \longrightarrow 1/2$ and $z \in \Delta$.

Now, the claims are easily verified for $m = 1$. Thus, we may assume that the claims hold for all $m' < m$. We want to prove them for m. We will only concentrate on the even case, the odd case being similar.

First, we will investigate the terms in (11). Therefore, observe that by the induction hypothesis

$$\sum_{i=0}^{2m-1} \binom{2m}{i} (-\mu)^{2m-i} \frac{\partial^i}{\partial y^i} \frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} = \mathcal{O}\left((1-2z)^{1/2-m}\right)$$

as $z \longrightarrow 1/2$ and $z \in \Delta$. Next, we consider the last term in (11). Here, again from the induction hypothesis

$$\begin{aligned}
&\sum_{\substack{i_1+i_2+i_3=2m-1\\ i_1<2m-1}} \binom{2m-1}{i_1,i_2,i_3} \bar{A}_k^{[i_1+1]}(z) \frac{\partial^{i_2}}{\partial y^{i_2}} \frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} \frac{\partial^{i_3}}{\partial y^{i_3}} \frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} \\
&= \sum_{\substack{i_1+i_2+i_3=2m-1\\ i_1 \text{ odd}; i_1<2m-1; i_2,i_3 \text{ even}}} \binom{2m-1}{i_1,i_2,i_3} \bar{A}_k^{[i_1+1]}(z) \frac{\partial^{i_2}}{\partial y^{i_2}} \frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} \frac{\partial^{i_3}}{\partial y^{i_3}} \frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0} \\
&\qquad + \mathcal{O}\left((1-2z)^{-m}\right) \\
&= \frac{c(2m-1)!\sigma^{2m}}{4^m}(1-2z)^{-1/2-m} + \mathcal{O}\left((1-2z)^{-m}\right).
\end{aligned}$$

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

$$\begin{aligned}
c &= \sum_{\substack{i_1+i_2+i_3=2m-1\\ i_1 \text{ odd}; i_1<2m-1; i_2,i_3 \text{ even}}} (i_1+1) \frac{(i_1-2)!!}{((i_1+1)/2)!} \frac{(i_2-1)!!}{(i_2/2)!} \frac{(i_3-1)!!}{(i_3/2)!} \\
&= 2^{2-m} \sum_{\substack{i_1+i_2+i_3=2m-1\\ i_1 \text{ odd}; i_1<2m-1; i_2,i_3 \text{ even}}} \binom{i_1-1}{(i_1-1)/2} \binom{i_2}{i_2/2} \binom{i_3}{i_3/2} \\
&= 2^{2-m} \sum_{i=0}^{m-2} \binom{2i}{i} \sum_{j=0}^{m-i-1} \binom{2j}{j} \binom{2m-2i-2-2j}{m-i-1-j} = 2^{3-m}(m-1)\binom{2m-2}{m-1}.
\end{aligned}$$

By substituting this into the expression above and collecting all terms, we obtain for (11)

$$\bar{B}_k^{[2m]}(z) = \frac{(m-1)(2m)!(2m-3)!!\sigma^{2m}}{2^{2m-1}m!}(1-2z)^{-1/2-m} + \mathcal{O}\left((1-2z)^{-m}\right)$$

as $z \longrightarrow 1/2$ and $z \in \Delta$. Using (2), we have

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

and the claim follows from Theorem VI.9 in [20].

What is left is to show (12). Here, observe that

$$\frac{\partial^{2m}}{\partial y^{2m}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}=\frac{\bar{A}_k^{[2m]}(z)}{1-2z}$$
$$+\sum_{\substack{i_1+i_2+i_3=2m-1\\ i_1<2m-1}}\binom{2m-1}{i_1,i_2,i_3}\bar{A}_k^{[i_1+1]}(z)\frac{\partial^{i_2}}{\partial y^{i_2}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}\frac{\partial^{i_3}}{\partial y^{i_3}}\frac{1}{1-\bar{P}_k(z,y)}\bigg|_{y=0}.$$

Hence, the claim follows from the expansion for $\bar{A}_k^{[2m]}(z)$ and the computations above. This concludes the proof of the result. ∎

Now, we can conclude the proof of Theorem 1 for fixed k.

Proof of Theorem 1, (i) for fixed k. Applying the transform theorems of Chapter VI in [20] yields

$$\mathbb{E}\left(X_{n,k}-\mu n\right)^{2m-1}=\mathcal{O}\left(\frac{n!}{\tau_n}2^n n^{m-5/2}\right)$$

as $n\to\infty$, and

$$\mathbb{E}\left(X_{n,k}-\mu n\right)^{2m}=\frac{n!(2m)!(2m-3)!!\sigma^{2m}}{\tau_n 4^m m!}[z^n](1-2z)^{1/2-m}+\mathcal{O}\left(\frac{n!}{\tau_n}2^n n^{m-2}\right)$$

as $n\to\infty$. Using (10) and standard computations in turn gives

$$\mathbb{E}\left(X_{n,k}-\mu n\right)^{2m-1}=\mathcal{O}\left(n^{m-1}\right)$$

and

$$\mathbb{E}\left(X_{n,k}-\mu n\right)^{2m}=g_m\sigma^{2m}n^m+\mathcal{O}\left(n^{m-1/2}\right),$$

where $g_m=(2m)!/(2^m m!)$. The claimed result follows from this by the Fréchet-Shohat Theorem. ∎

Remark 2. It might be possible that with the approach just presented the case of variable k can be treated as well. However, in order to do so, one needs error terms which are uniform in both n and k. Such error terms seem to be easier to derive when one considers recurrences instead of generating function and avoids using complex analysis. This we are going to do next.

3 Variable Subtree Size - An Elementary Approach

In this section, we will concentrate on the remaining cases of Theorem 1, namely, all cases where $k=k(n)$ and $k\to\infty$ as $n\to\infty$. As already mentioned, we will propose a different approach which will work with sequences and is elementary in the sense that complex analysis is avoided.

We will follow the approach outlined in the introduction. Therefore, set $P_{n,k}(z)=\mathbb{E}(\exp(X_{n,k}z))$. Then, we obtain

$$P_{n,k}(z)=\sum_{1\le j<n}\pi_{n,j}P_{j,k}(z)P_{n-j,k}(z),\qquad (n>k),$$

where $P_{n,k}(z)=1$ for $n<k$ and $P_{k,k}(z)=e^z$. By taking derivatives and evaluating at $z=0$, we observe that all moments satisfy the recurrence

$$a_{n,k}=\sum_{1\le j<n}\pi_{n,j}(a_{j,k}+a_{n-j,k})+b_{n,k},\qquad (n>k),\tag{13}$$

where $b_{n,k}$ is a given sequence that involves moments of lower order. Here, $a_{n,k} = 0$ for $n < k$, $a_{k,k}$ is given and $b_{n,k} = 0$ for $n \leq k$. Note that the above recurrence can be rewritten to (4).

Now, in order to prove (5), we set $A(z) = \sum_{n=1}^{\infty} C_n a_{n,k} z^n$ and $B(z) = \sum_{n=1}^{\infty} C_n b_{n,k} z^n$. Then, (4) becomes

$$A(z) - C_k a_{k,k} z^k = 2A(z) \sum_{n=1}^{\infty} C_n z^n + B(z) = \left(1 - \sqrt{1-4z}\right) A(z) + B(z).$$

Solving for $A(z)$ yields

$$A(z) = \frac{1}{\sqrt{1-4z}} \left(B(z) + C_k a_{k,k} z^k\right).$$

Reading off coefficients immediately gives (5).

We first demonstrate how to re-derive the expression for the mean value of the previous section from (5). Therefore, observe that for the mean we have (13) with

$$b_{n,k} = -\mathbb{P}(I_n = n-k) = -\frac{2kC_k C_{n-k}}{nC_n}.$$

By substituting this into (5) (with $a_{k,k} = 1$), we obtain

$$\begin{aligned}
\mathbb{E}X_{n,k} &= -\frac{2kC_k}{C_n} \sum_{j=k+1}^{n} \frac{C_{j-k}(n+1-j)C_{n+1-j}}{j} + \frac{C_k(n+1-k)C_{n+1-k}}{C_n} \\
&= -\frac{2kC_k}{C_n}[z^{n-k+1}] \sum_{i=1}^{\infty} iC_i z^i \sum_{j=1}^{\infty} \frac{C_j}{j+k} z^j + \frac{C_k(n+1-k)C_{n+1-k}}{C_n} \\
&= -\frac{kC_k}{C_n}[z^n] \frac{1}{\sqrt{1-4z}} \int_0^z t^{k-1} \left(1 - \sqrt{1-4t}\right) \mathrm{d}t + \frac{C_k(n+1-k)C_{n+1-k}}{C_n} \\
&= \frac{kC_k}{C_n}[z^n] \frac{1}{\sqrt{1-4z}} \int_0^z t^{k-1} \sqrt{1-4t}\mathrm{d}t \\
&= \frac{2^n kC_k}{2^k C_n}[u^n] \frac{1}{\sqrt{1-2u}} \int_0^z t^{k-1} \sqrt{1-2t}\mathrm{d}t.
\end{aligned}$$

The remaining derivation is as in the previous section.

Next, we will look at the variance. In difference to the previous section, we will already now shift-the-mean. Therefore, set

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

Moreover, set $\bar{X}_{n,k} = X_{n,k} - \mu_{n,k}$, and $\bar{P}_{n,k}(z) = \mathbb{E}e^{\bar{X}_{n,k} z}$. Then, (3) becomes

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

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

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

Put $\bar{A}^{[m]}_{n,k} = \mathbb{E}(X_{n,k} - \mu_{n,k})^m = \bar{P}^{(m)}_{n,k}(0)$. By taking m-th derivatives and evaluating at $z = 0$, we obtain again a recurrence of type (13)

$$\bar{A}^{[m]}_{n,k} = \sum_{1 \le j < n} \pi_{n,j} \left(\bar{A}^{[m]}_{j,k} + \bar{A}^{[m]}_{n-j,k} \right) + \bar{B}^{[m]}_{n,k}, \qquad (n > k), \tag{14}$$

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

$$\bar{B}^{[m]}_{n,k} = \sum_{\substack{i_1+i_2+i_3=m \\ 0 \le i_1, i_2 < m}} \binom{m}{i_1, i_2, i_3} \sum_{1 \le j < n} \pi_{n,j} \bar{A}^{[i_1]}_{j,k} \bar{A}^{[i_2]}_{n-j,k} \Delta^{i_3}_{n,j,k}. \tag{15}$$

Now, we first derive a uniform bound for the variance. Therefore, we will use the above recurrence together with (5).

Lemma 4. *For $k, n \ge 1$,*

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

Proof. Setting $m = 2$ in (14), we get our usual recurrence with

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

An easy observation shows that

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

We substitute now the latter into (5) and treat the resulting two sums differently. First, for the first sum,

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

where we have used (10). Using the same arguments as in our above derivation of the mean value, we can evaluate the second sum

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

Overall, the claimed bound follows. ∎

Next, we refine the above result for varying k.

Proposition 5. *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. Observe that in the proof of the last lemma, we more precisely have

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

Moreover, we found that the contribution of the $\mathcal{O}$-term is negligible compared with the claimed order of the variance. Hence, we just have to concentrate on the first term.

First, direct computation shows that $\Delta_{n,k,k}^2 \sim 1$ as $n \to \infty$. Consequently,

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

where the last line follows as in the proof of the previous lemma. Hence, we get the claimed result. ∎

Next, we use induction to extend the uniform bound for the variance to all higher centered moments. From a technical point of view, this is the most demanding part of the proof.

Lemma 6. *For $k, n \ge 1$ and $m \ge 2$,*

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

Proof. First note that Lemma 4 implies the validity of the assertion for $m = 2$. Next, assume that the assertion holds for all $m' < m$. We are going to show that it holds for m as well.

Therefore, we first bound (15). In order to do so, we start by considering the range where $n > 2k^2$ and break the double sum into three parts

$$\bar{B}_{n,k}^{[m]} = \sum_{i_1,i_2,i_3}\sum_{j\le k^2} + \sum_{i_1,i_2,i_3}\sum_{k^2<j<n-k^2} + \sum_{i_1,i_2,i_3}\sum_{n-k^2\le j} =: \Sigma_1 + \Sigma_2 + \Sigma_3.$$

We will treat each sum separately. We start with the second sum

$$\begin{aligned}\Sigma_2 &= \sum_{\substack{i_1+i_2+i_3=m\\0\le i_1,i_2<m}}\binom{m}{i_1,i_2,i_3}\sum_{k^2<j<n-k^2}\pi_{n,j}\bar{A}_{j,k}^{[i_1]}\bar{A}_{n-j,k}^{[i_2]}\Delta_{n,j,k}^{i_3}\\ &= \sum_{i=1}^{m-1}\binom{m}{i}\sum_{k^2<j<n-k^2}\pi_{n,j}\bar{A}_{j,k}^{[i]}\bar{A}_{n-j,k}^{[m-i]}\\ &= m\sum_{k^2<j<n-k^2}\pi_{n,j}\bar{A}_{j,k}^{[1]}\bar{A}_{n-j,k}^{[m-1]} + \sum_{i=2}^{m-1}\binom{m}{i}\sum_{k^2<j<n-k^2}\pi_{n,j}\bar{A}_{j,k}^{[i]}\bar{A}_{n-j,k}^{[m-i]} =: \Sigma_{2,1} + \Sigma_{2,2}.\end{aligned}$$

The above two parts can be bounded as follows

$$\begin{aligned}\Sigma_{2,1} &= \mathcal{O}\left(\frac{1}{k^2}\sum_{k\le j<n}\pi_{n,j}\left(\frac{n-j}{k^2}\right)^{(m-1)/2}\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{(m-1)/2}\frac{1}{k^2}\sum_{k\le j<n}j^{-3/2}\left(1-\frac{j}{n}\right)^{(m-2)/2}\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{(m-1)/2}\frac{1}{k^{5/2}}\right) = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\frac{1}{\sqrt{n}k^{3/2}}\right)\end{aligned}$$

and

$$\begin{aligned}\Sigma_{2,2} &= \mathcal{O}\left(\sum_{i=2}^{m-1}\binom{m}{i}\sum_{1\le j<n}\pi_{n,j}\left(\frac{j}{k^2}\right)^{i/2}\left(\frac{n-j}{k^2}\right)^{(m-i)/2}\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\frac{1}{\sqrt{n}}\sum_{i=2}^{m-1}\binom{m}{i}\frac{1}{n}\sum_{1\le j<n}\left(\frac{j}{n}\right)^{(i-3)/2}\left(1-\frac{j}{n}\right)^{(m-i-1)/2}\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where we used (10) and the induction hypothesis.

Next, we estimate the third sum

$$\begin{aligned}\Sigma_3 &= \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m}}\binom{m}{i_1,i_2,i_3}\sum_{n-k^2\le j<n}\pi_{n,j}\bar{A}_{j,k}^{[i_1]}\bar{A}_{n-j,k}^{[i_2]}\Delta_{n,j,k}^{i_3}\\ &= \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m,i_1\ge 2}}\sum_{n-k^2\le j<n}+\sum_{i_2+i_3=m-1}\sum_{n-k^2\le j<n}+\sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}}\sum_{n-k^2\le j<n}\\ &=: \Sigma_{3,1}+\Sigma_{3,2}+\Sigma_{3,3}.\end{aligned}$$

Again, we bound the last three parts separately. First, we treat the first part

$$\begin{aligned}\Sigma_{3,1} &= \mathcal{O}\left(\sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m,i_1\ge 2}}\sum_{n-k^2\le j<n}\pi_{n,j}\left(\frac{j}{k^2}\right)^{i_1/2}\right)\\ &= \mathcal{O}\left(\frac{1}{\sqrt{n}}\sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m,i_1\ge 2}}\left(\frac{n}{k^2}\right)^{i_1/2}\frac{1}{n}\sum_{1\le j<n}\left(\frac{j}{n}\right)^{(i_1-3)/2}\left(1-\frac{j}{n}\right)^{-1/2}\right)\\ &= \mathcal{O}\left(\frac{1}{\sqrt{n}}\sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m,i_1\ge 2}}\left(\frac{n}{k^2}\right)^{i_1/2}\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\frac{k}{n}\right).\end{aligned}$$

Next, we bound the second sum

$$\Sigma_{3,2} = \mathcal{O}\left(\frac{1}{k^2}\sum_{i_2+i_3=m-1}\sum_{n-k^2\le j<n}\pi_{n,j}\right) = \mathcal{O}\left(\frac{1}{k^2}\right).$$

Finally, we bound the third sum

$$\begin{aligned}\Sigma_{3,3} &= \mathcal{O}\left(\frac{1}{k}\sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}}\sum_{n-k^2\le j<n}\pi_{n,j}\right) = \mathcal{O}\left(\frac{1}{k}\sum_{k^2\le j<n}j^{-3/2}\left(1-\frac{j}{n}\right)^{-1/2}\right)\\ &= \mathcal{O}\left(\frac{1}{k}\left(\sum_{k^2\le j<n/2}j^{-3/2}+\frac{1}{\sqrt{n}}\cdot\frac{1}{n}\sum_{n/2\le j<n}\left(\frac{j}{n}\right)^{-3/2}\left(1-\frac{j}{n}\right)^{-1/2}\right)\right) = \mathcal{O}\left(\frac{1}{k^2}\right).\end{aligned}$$

What is still left is to estimate Σ_1. Again, we break the sum into three parts

$$\begin{aligned}
\Sigma_1 &= \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m}} \binom{m}{i_1,i_2,i_3} \sum_{j\le k^2} \pi_{n,j} \bar{A}_{j,k}^{[i_1]} \bar{A}_{n-j,k}^{[i_2]} \Delta_{n,j,k}^{i_3} \\
&= \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m, i_1\ge 2}} \sum_{j\le k^2} + \sum_{i_2+i_3=m-1} \sum_{j\le k^2} + \sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}} \sum_{j\le k^2} \\
&=: \Sigma_{1,1} + \Sigma_{1,2} + \Sigma_{1,3}
\end{aligned}$$

and bound all three parts individually. For the first sum, we have

$$\begin{aligned}
\Sigma_{1,1} &= \mathcal{O}\left(\frac{1}{k^2} \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m, i_1\ge 2}} \sum_{j\le k^2} \pi_{n,j} j \left(\frac{n-j}{k^2}\right)^{i_2/2} \right) \\
&= \mathcal{O}\left(\frac{1}{k^2} \sum_{\substack{i_1+i_2+i_3=m\\ 0\le i_1,i_2<m, i_1\ge 2}} \left(\frac{n}{k^2}\right)^{i_2/2} \sum_{j\le k^2} j^{-1/2} \left(1-\frac{j}{n}\right)^{(i_2-1)/2} \right) \\
&= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{(m-2)/2} \frac{1}{k} \right) = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2} \frac{k}{n} \right).
\end{aligned}$$

Next, we obtain the following bound for the second sum

$$\begin{aligned}
\Sigma_{1,2} &= \mathcal{O}\left(\frac{1}{k^2} \sum_{i_2+i_3=m-1} \sum_{j\le k^2} \pi_{n,j} \left(\frac{n-j}{k^2}\right)^{i_2/2} \right) \\
&= \mathcal{O}\left(\frac{1}{k^2} \sum_{i_2+i_3=m-1} \left(\frac{n}{k^2}\right)^{i_2/2} \sum_{j\le k^2} j^{-3/2} \left(1-\frac{j}{n}\right)^{(i_2-1)/2} \right) \\
&= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{(m-1)/2} \frac{1}{k^2} \right) = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2} \frac{1}{\sqrt{n}k} \right).
\end{aligned}$$

For the final bound, we have to work slightly harder

$$\begin{aligned}
\Sigma_{1,3} &= \mathcal{O}\left(\sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}} \sum_{j\le k^2} \pi_{n,j} \left(\frac{n-j}{k^2}\right)^{i_2/2} \Delta_{n,j,k}^{i_3} \right) \\
&= \mathcal{O}\left(\frac{1}{k^2} \sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}} \sum_{j<k} \pi_{n,j} j \left(\frac{n-j}{k^2}\right)^{i_2/2} + \pi_{n,k} \sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}} \left(\frac{n-k}{k^2}\right)^{i_2/2} \right) \\
&= \mathcal{O}\left(\frac{1}{k^2} \sum_{\substack{i_2+i_3=m\\ 0\le i_2<m}} \left(\frac{n}{k^2}\right)^{i_2/2} \sum_{j<k} j^{-1/2} \left(1-\frac{j}{n}\right)^{(i_2-1)/2} + k^{-3/2} \left(\frac{n}{k^2}\right)^{(m-1)/2} \left(1-\frac{k}{n}\right)^{-1/2} \right) \\
&= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{(m-1)/2} \frac{1}{k^{3/2}} \right) = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2} \frac{1}{\sqrt{n}k} \right).
\end{aligned}$$

So, overall we have proved the following estimate for $\bar{B}_{n,k}^{[m]}$

$$\bar{B}_{n,k}^{[m]} = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\frac{1}{\sqrt{n}}\right) \tag{16}$$

for $n > 2k^2$. Here, we should note that actually slightly more than the above bound was proved. The reason for working harder than it is in fact necessary at this stage will become apparent later on.

Next, we have to find a suitable bound for the remaining range $n \leq 2k^2$. Therefore, we first observe

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

Then, we break the first sum into four parts

$$\sum_{\substack{i_1+i_2+i_3=m\\ 0\leq i_1,i_2<m}} = \sum_{\substack{i_1+i_2=m\\ 0\leq i_1,i_2<m}} + \sum_{\substack{i_1+i_2+i_3=m\\ 0\leq i_1,i_2<m, i_1,i_3\geq 1}} + \sum_{\substack{i_2+i_3=m\\ 1\leq i_2<m}} + \sum_{i_3=m} = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4.$$

We will carefully estimate every part. We start with the first one

$$\begin{aligned}\Sigma_1 &= \mathcal{O}\left(\sum_{i=1}^{m-1}\sum_{1\leq j<n, j\neq k}\pi_{n,j}\left(\frac{j}{k^2}\right)\left(\frac{n-j}{k^2}\right)\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^2\frac{1}{\sqrt{n}}\sum_{i=1}^{m-1}\frac{1}{n}\sum_{1\leq j<n}\left(\frac{j}{n}\right)^{-1/2}\left(1-\frac{j}{n}\right)^{1/2}\right) = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^2\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Next, we treat the second part

$$\begin{aligned}\Sigma_2 &= \mathcal{O}\left(\sum_{\substack{i_1+i_2+i_3=m\\ 0\leq i_1,i_2<m, i_1,i_3\geq 1}}\sum_{1\leq j<n, j\neq k}\pi_{n,j}\left(\frac{j}{k^2}\right)\frac{1}{k^{i_3}}\right)\\ &= \mathcal{O}\left(\frac{n}{k^2}\frac{1}{\sqrt{n}k}\sum_{\substack{i_1+i_2+i_3=m\\ 0\leq i_1,i_2<m, i_1,i_3\geq 1}}\frac{1}{n}\sum_{1\leq j<n}\left(\frac{j}{n}\right)^{-1/2}\left(1-\frac{j}{n}\right)^{-1/2}\right) = \mathcal{O}\left(\frac{n}{k^3}\frac{1}{\sqrt{n}}\right).\end{aligned}$$

For the third part, we have

$$\begin{aligned}\Sigma_3 &= \mathcal{O}\left(\sum_{i=1}^{m-1}\sum_{1\leq j<n, j\neq k}\pi_{n,j}\left(\frac{n-j}{k^2}\right)\frac{1}{k^{m-i}}\right)\\ &= \mathcal{O}\left(\frac{n}{k^2}\frac{1}{k}\sum_{i=1}^{m-1}\sum_{1\leq j<n} j^{-3/2}\left(1-\frac{j}{n}\right)^{1/2}\right) = \mathcal{O}\left(\frac{n}{k^3}\right).\end{aligned}$$

Finally, the fourth part we bound crudely by

$$\Sigma_4 = \mathcal{O}\left(\sum_{1\leq j<n, j\neq k}\pi_{n,j}\frac{1}{k^m}\right) = \mathcal{O}\left(\frac{1}{k^m}\right).$$

Overall, we obtain the following bound

$$\bar{B}_{n,k}^{[m]} = \mathcal{O}\left(\frac{n}{k^3} + \pi_{n,k}\right) \tag{17}$$

for $n \le 2k^2$.

Now, we substitute what we have proved so far into the solution (5) of (14). Therefore, we will break the solution into two parts

$$\bar{A}_{n,k}^{[m]} = \sum_{k+1\le j\le n} \frac{C_j(n+1-j)C_{n+1-j}}{C_n} \bar{B}_{j,k}^{(m)} = \sum_{k+1\le j\le 2k^2} + \sum_{2k^2<j\le n} = \Sigma_1 + \Sigma_2.$$

In order to bound the second sum, we use (16) and obtain

$$\begin{aligned}\Sigma_2 &= \mathcal{O}\left(\sum_{2k^2<j\le n} \frac{C_j(n+1-j)C_{n+1-j}}{C_n}\left(\frac{j}{k^2}\right)^{m/2}\frac{1}{\sqrt{j}}\right)\\ &= \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\frac{1}{n}\sum_{1\le j\le n}\left(\frac{j}{n}\right)^{(m-4)/2}\left(1-\frac{j-1}{n}\right)^{-1/2}\right) = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m/2}\right).\end{aligned}$$

For the first part, we use (17). Consequently,

$$\begin{aligned}\Sigma_1 &= \mathcal{O}\left(\sum_{k+1\le j\le 2k^2} \frac{C_j(n+1-j)C_{n+1-j}}{C_n}\frac{j}{k^3}\right) + \mathcal{O}\left(\sum_{k+1\le j\le 2k^2}\frac{C_j(n+1-j)C_{n+1-j}}{C_n}\pi_{n,k}\right)\\ &= \Sigma_{1,1} + \Sigma_{1,2}.\end{aligned}$$

The second sum was already estimated in the proof of Lemma 4 where we obtained the bound $\mathcal{O}(n/k^2)$. For the first sum, we break our considerations into two cases. First, assume that $n > 3k^2$. Then,

$$\Sigma_{1,1} = \mathcal{O}\left(\frac{n}{k^3}\sum_{k+1\le j\le 2k^2} j^{-1/2}\left(1-\frac{j-1}{n}\right)^{-1/2}\right) = \mathcal{O}\left(\frac{n}{k^2}\right).$$

Next, we assume that $n \le 3k^2$. Then,

$$\Sigma_{1,1} = \mathcal{O}\left(\frac{n^{3/2}}{k^3}\frac{1}{n}\sum_{1\le j\le n}\left(\frac{j}{n}\right)^{-1/2}\left(1-\frac{j-1}{n}\right)^{-1/2}\right) = \mathcal{O}\left(\frac{n}{k^2}\right).$$

Collecting the above estimates, we obtain

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

which concludes the induction proof. ∎

The next step is to refine the previous bound in the normal range of Theorem 1.

Proposition 7. *Let $k = k(n)$ such that $k = o(\sqrt{n})$ and $k \to \infty$ as $n \to \infty$. Then, as $n \to \infty$,*

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

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

Proof. We will once more use induction on m. Due to Proposition 5, the assertion holds for $m = 1$. Now, assume that the assertion holds for all $m' < m$. We will prove that it holds for m as well.

First, we concentrate on $\bar{B}^{[l]}_{n,k}$. Note that we can assume that $n > 2k^2$. We break $\bar{B}^{[l]}_{n,k}$ into the following three parts

$$\bar{B}^{[l]}_{n,k} = \sum_{i_1,i_2,i_3}\sum_{j<\epsilon n} + \sum_{i_1,i_2,i_3}\sum_{\epsilon n<j<(1-\epsilon)n} + \sum_{i_1,i_2,i_3}\sum_{(1-\epsilon)n<j} =: \Sigma_1 + \Sigma_2 + \Sigma_3,$$

where $\epsilon > 0$ is a fixed constant.

First, we consider the case where l is odd, i.e., $l = 2m-1$. We use Lemma 6 to bound Σ_1 and Σ_2, where it is actually enough to use the bounds we have already deduced in the proof of the lemma. This yields

$$\Sigma_1 = o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{n}}\sum_{i=2}^{2m-2}\int_0^{\epsilon} x^{(i-3)/2}(1-x)^{m-1-i/2}\mathrm{d}x.\right).$$

Now, we let $\epsilon \longrightarrow 0$ and obtain

$$\Sigma_1 = o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{n}}\right).$$

Similarly, we deduce the same bound for Σ_3. As to Σ_2, we first break it into two parts

$$\Sigma_2 = \sum_{\substack{i_1+i_2=2m-1\\0\le i_1,i_2<2m-1,i_1\ge 2}}\sum_{\epsilon n<j<(1-\epsilon)n} + \text{remaining terms},$$

where the remaining terms can be bounded as in the proof of Lemma 6 yielding

$$\text{remaining terms} = o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{n}}\right).$$

For the other part, we use the induction hypothesis. Therefore, note that either i_1 or i_2 must be odd. Hence,

$$\begin{aligned}
\sum_{\substack{i_1+i_2=2m-1\\0\le i_1,i_2<2m-1,i_2\ge 1}}\sum_{\epsilon n<j<(1-\epsilon)n} &= o\left(\sum_{i=2}^{2m-1}\sum_{\epsilon n<j<(1-\epsilon)n}\pi_{n,j}\left(\frac{n}{k^2}\right)^{i/2}\left(\frac{n-j}{k^2}\right)^{m-1/2-i/2}\right)\\
&= o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{n}}\sum_{i=2}^{2m-1}\frac{1}{n}\sum_{1\le j<n}\left(\frac{j}{n}\right)^{(i-3)/2}\left(1-\frac{j}{n}\right)^{m-1-i/2}\right)\\
&= o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

So, overall we have proved that

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

for $k = o(\sqrt{n})$ and $k \to \infty$ as $n \to \infty$.

Next, we consider the case where l is even, i.e., $l = 2m$. Here, Σ_1 and Σ_3 can be treated as above and we obtain the bound $o((n/k^2)^m 1/\sqrt{n})$. As to Σ_2, we divide it as above and again obtain the previous bound for the remaining parts. So, what is left is to consider

$$\sum_{\substack{i_1+i_2=2m\\0\le i_1,i_2<2m,i_1\ge 2}}\sum_{\epsilon n<j<(1-\epsilon)n}.$$

Note that either i_1 and i_2 are both odd or both even. The first case is treated as above and we obtain once more the bound $o((n/k^2)^m 1/\sqrt{n})$. For the second case, we have

$$\begin{aligned}
\sum_{\substack{i_1+i_2=2m\\ 0\le i_1,i_2<2m, i_1 \text{ is even}}} \sum_{\epsilon n<j<(1-\epsilon)n} &\sim \sum_{i=1}^{m-1}\binom{2m}{2i} g_i g_{m-i} \sum_{\epsilon n<j<(1-\epsilon)n} \pi_{n,j}\left(\frac{j}{2k^2}\right)^i \left(\frac{n-j}{2k^2}\right)^{m-i}\\
&\sim \left(\frac{n}{2k^2}\right)^m \frac{1}{\sqrt{n}} \sum_{i=1}^{m-1}\binom{2m}{2i} g_i g_{m-i} \frac{1}{2\sqrt{\pi}}\frac{1}{n} \sum_{\epsilon n<j<(1-\epsilon)n} \left(\frac{j}{n}\right)^{i-3/2}\left(1-\frac{j}{n}\right)^{m-i-1/2}\\
&\sim \left(\frac{n}{2k^2}\right)^m \frac{1}{\sqrt{n}} \sum_{i=1}^{m-1}\binom{2m}{2i} g_i g_{m-i} \frac{1}{2\sqrt{\pi}} \int_{\epsilon}^{1-\epsilon} x^{i-3/2}(1-x)^{m-i-1/2}\mathrm{d}x,
\end{aligned}$$

where we used the induction hypothesis in the first step and (10) in the second step. Collecting all contributions and letting $\epsilon \longrightarrow 0$ yields

$$\bar{B}_{n,k}^{[2m]} \sim \bar{g}_m \left(\frac{n}{2k^2}\right)^m \frac{1}{\sqrt{n}}.$$

for $k = o(\sqrt{n})$ and $k \longrightarrow \infty$ as $n \longrightarrow \infty$, where ($\Gamma(x)$ denotes the Γ-function)

$$\begin{aligned}
\bar{g}_m &= \sum_{i=1}^{m-1}\binom{2m}{2i} g_i g_{m-i} \frac{1}{2\sqrt{\pi}} \int_0^1 x^{i-3/2}(1-x)^{m-i-1/2}\mathrm{d}x.\\
&= \frac{(2m)!}{2^{m+1}\sqrt{\pi}} \sum_{i=1}^{m-1} \frac{1}{i!(m-i)!} B(i-1/2, m-i+1/2)\\
&= \frac{(2m)!}{2^{m+1}\sqrt{\pi}} \sum_{i=1}^{m-1} \frac{1}{i!(m-i)!} \frac{\Gamma(i-1/2)\Gamma(m-i+1/2)}{\Gamma(m)}\\
&= \frac{(2m)!}{2^{m+1}(m-1)!\sqrt{\pi}} \sum_{i=1}^{m-1} \frac{1}{i!(m-i)!} \frac{(2i-2)!}{(i-1)!4^{i-1}}\sqrt{\pi} \frac{(2m-2i)!}{(m-i)!4^{m-i}}\sqrt{\pi}\\
&= \frac{2\sqrt{\pi}(2m)!}{8^m(m-1)!} \sum_{i=1}^{m-1} \frac{1}{i}\binom{2m-2i}{m-i}\binom{2i-2}{i-1}\\
&= \frac{4\sqrt{\pi}(2m)!(2m-2)!}{8^m m!(m-1)!(m-2)!}.
\end{aligned}$$

Next, we substitute what we have proved so far into the solution of (14) which is given by (5). We break the solution into two parts

$$\bar{A}_{n,k}^{[l]} = \sum_{k+1\le j<\epsilon n} + \sum_{\epsilon n\le j\le n} = \Sigma_1 + \Sigma_2,$$

where $\epsilon > 0$ is a constant.

Again let us first consider the case where l is odd, i.e. $l = 2m-1$. Then, Σ_1 can be bounded as in the proof of Lemma 6 and we obtain

$$\Sigma_1 = \mathcal{O}\left(\left(\frac{n}{k^2}\right)^{m-1/2} \int_0^{\epsilon} x^{m-5/2}(1-x)^{-1/2}\mathrm{d}x\right).$$

Upon letting $\epsilon \to 0$, we obtain $o((n/k^2)^{m-1/2})$. For bounding Σ_2, we use the induction hypothesis

$$\Sigma_2 = o\left(\sum_{\epsilon n\le j\le n}\frac{C_j(n+1-j)C_{n+1-j}}{C_n}\left(\frac{j}{k^2}\right)^{m-1/2}\frac{1}{\sqrt{j}}\right)$$
$$= o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\frac{1}{n}\sum_{1\le j\le n}\left(\frac{j}{n}\right)^{m-5/2}\left(1-\frac{j-1}{n}\right)^{-1/2}\right) = o\left(\left(\frac{n}{k^2}\right)^{m-1/2}\right).$$

This proves the result in the case where l is odd.

Finally, we consider the case where l is even, i.e. $l = 2m$. Here, Σ_1 can be treated as above and we obtain $o((n/k^2)^m)$. So, what is left is to deduce the asymptotics of Σ_1. Therefore, observe that

$$\Sigma_1 \sim \bar{g}_m \sum_{\epsilon n\le j\le n}\frac{C_j(n+1-j)C_{n+1-j}}{C_n}\left(\frac{j}{2k^2}\right)^{m}\frac{1}{\sqrt{j}}$$
$$\sim \left(\frac{n}{2k^2}\right)^m \bar{g}_m\frac{1}{\sqrt{\pi}}\frac{1}{n}\sum_{\epsilon n\le j\le n}\left(\frac{j}{n}\right)^{m-2}\left(1-\frac{j-1}{n}\right)^{-1/2}$$
$$\sim \left(\frac{n}{2k^2}\right)^m \bar{g}_m\frac{1}{\sqrt{\pi}}\int_\epsilon^1 x^{m-2}(1-x)^{-1/2}\mathrm{d}x,$$

where we used the induction hypothesis in the first step and (10) in the second step. Letting $\epsilon \to 0$ and some simplification yields

$$\Sigma_1 \sim \left(\frac{n}{2k^2}\right)^m \bar{g}_m\frac{1}{\sqrt{\pi}}\int_0^1 x^{m-2}(1-x)^{-1/2}\mathrm{d}dx$$
$$= \left(\frac{n}{2k^2}\right)^m \frac{4(2m)!(2m-2)!}{8^m m!(m-1)!(m-2)!}B(m-1,1/2)$$
$$= \left(\frac{n}{2k^2}\right)^m \frac{4(2m)!(2m-2)!}{8^m m!(m-1)!(m-2)!}\cdot\frac{\Gamma(m-1)\Gamma(1/2)}{\Gamma(m-1/2)}$$
$$= \left(\frac{n}{2k^2}\right)^m \frac{4(2m)!(2m-2)!}{8^m m!(m-1)!(m-2)!}\cdot\frac{4^{m-1}(m-2)!(m-1)!}{(2m-2)!}$$
$$= \left(\frac{n}{2k^2}\right)^m \frac{(2m)!}{2^m m!} = \left(\frac{n}{2k^2}\right)^m g_m.$$

This concludes the induction proof and consequently also the proof of the proposition. ∎

The last proposition together with the Fréchet-Shohat Theorem proves Theorem 1, part (i) for $k \to \infty$.

With a similar method of proof, the Poisson range can be handled as well. We just give a rough sketch of the proof. The reader should have no problems in filling in the missing details.

Proof of Theorem 1, part (ii). In view of the claimed result, it is better to work here with factorial moments instead of central moments. Therefore, set $Q_{n,k}(\gamma) = \mathbb{E}\left(\gamma^{X_{n,k}}\right)$. Then, (3) translates into

$$Q_{n,k}(\gamma) = \sum_{1\le j<n}\pi_{n,j}Q_{j,k}(\gamma)Q_{n-j,k}(\gamma)\gamma^{-\mathbf{1}_{\{j=n-k\}}},$$
$$= \sum_{1\le j<n}\pi_{n,j}Q_{j,k}(\gamma)Q_{n-j,k}(\gamma) - (\gamma-1)\pi_{n,n-k}Q_{n-k,k}(\gamma), \qquad (n>k),$$

where $Q_{n,k}(\gamma) = 1$ for $n < k$ and $Q_{k,k}(\gamma) = \gamma$.

Next, we denote by $\bar{A}^{[m]}_{n,k}$ the m-th derivative of $Q_{n,k}(\gamma)$ evaluated at $\gamma = 1$. Then, for $m \geq 2$, the above recurrence in turn yields

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

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

$$\bar{B}^{[m]}_{n,k} = \sum_{i=1}^{m-1} \binom{m}{i} \sum_{1\leq j<n} \pi_{n,j} \bar{A}^{[i]}_{j,k} \bar{A}^{[m-i]}_{n-j,k} - m\pi_{n,n-k} \bar{A}^{(m-1)}_{n-k,k}. \tag{18}$$

As before, the next step is to obtain a uniform estimate. A careful analysis reveals that

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

for all $n, k \geq 1$ and $m \geq 2$. Note that the bound here is more simpler than in the previous analysis. This is due to the above simplified form of the "toll" sequence $\bar{B}^{[m]}_{n,k}$.

The latter estimate is then used to prove the following asymptotic expansion for $k = k(n)$ and $k \sim c\sqrt{n}$, as $n \to \infty$,

$$\bar{A}^{[m]}_{n,k} \sim \frac{1}{(2c^2)^m}, \qquad \text{for } m \geq 1.$$

Therefore, we proceed by induction. The case $m = 1$ follows from the explicit expression of the mean value. Next, assume that the claim holds for all integers $< m$. Then, by substituting the induction assumption into (18) and using the uniform estimate, we obtain, as $n \to \infty$,

$$\begin{aligned}
\bar{B}^{[m]}_{n,k} &\sim \sum_{i=1}^{m-1} \binom{m}{i} \sum_{\epsilon n<j<(1-\epsilon)n} \pi_{n,j} \bar{A}^{[i]}_{j,k} \bar{A}^{[m-i]}_{j,k} \\
&\sim \frac{1}{(2c^2)^m} \frac{1}{\sqrt{n}} \sum_{i=1}^{m-1} \binom{m}{i} \frac{1}{2\sqrt{\pi}} \frac{1}{n} \sum_{\epsilon n<j<(1-\epsilon)n} \left(\frac{j}{n}\right)^{i-3/2} \left(1-\frac{j}{n}\right)^{m-i-1/2} \\
&\sim \frac{1}{(2c^2)^m} \frac{1}{\sqrt{n}} \sum_{i=1}^{m-1} \binom{m}{i} \frac{1}{2\sqrt{\pi}} \int_{\epsilon}^{1-\epsilon} x^{i-3/2}(1-x)^{m-i-1/2} \mathrm{dx},
\end{aligned}$$

where $\epsilon > 0$. Letting $\epsilon \to 0$ and using similar computations as in the proof of the last proposition, we obtain, as $n \to \infty$,

$$\bar{B}^{[m]}_{n,k} \sim \frac{\sqrt{\pi} m(2m-2)!}{4^{m-1}(m-2)!} \frac{1}{(2c^2)^m} \frac{1}{\sqrt{n}}.$$

Now, we substitute this into (5) and again use the uniform bound. This implies, as $n \to \infty$,

$$\bar{A}^{[m]}_{n,k} \sim \sum_{\epsilon n\leq j\leq n} \frac{C_j(n+1-j)C_{n+1-j}}{C_n} \bar{B}^{[m]}_{j,k} \sim \frac{m(2m-2)!}{4^{m-1}(m-2)!} \frac{1}{(2c^2)^m} \int_{\epsilon}^{1} x^{m-2}(1-x)^{-1/2} \mathrm{d}x,$$

where $\epsilon > 0$. Letting $\epsilon \to \infty$ and some straightforward computations establishes the claimed results. Hence, the induction proof is completed. ∎

4 Other Simple Families of Increasing Trees

Random binary search trees, random recursive trees and random PORTs are all special cases of simple families of increasing trees. In this final section, we will briefly outline how our results can be extended to other simple families of increasing trees. We will only focus on the (more complicated) case of varying k.

We will start by recalling the definition of simple families of increasing trees from [2]. First, an *increasing tree* is a rooted, plane, node-labeled tree with labels along any path from the root to a leaf forming an increasing sequence. A *simple family of increasing trees* is then defined as increasing trees together with a sequence $(\phi_r)_{r\geq 0}$ of non-negative weights with $\phi_0 > 0$ and $\phi_r > 0$ for some $r \geq 2$. For a given increasing tree T, we define its weight as

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

where the product runs over all nodes v of T and $d(v)$ is the out-degree of node v. Moreover, we denote by $\#T$ the size of T and set

$$\tau_n = \sum_{\#T=n} \omega(T), \qquad \tau(z) = \sum_{n\geq 1} \tau_n \frac{z^n}{n!}.$$

Finally, we set

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

which is called the *weight function*. It was proved in [2] that

$$\tau'(z) = \phi(\tau(z)), \qquad \tau(0) = 0. \tag{19}$$

Now, we equip a simple class of increasing trees with a probability model, where the probability of a tree T is proportional to its weight. More precisely, if T has size n then its probability equals $\omega(T)/\tau_n$. The resulting class is called *simple class of random increasing trees*.

By specializing $(\phi_r)_{r\geq 0}$ (or equivalently $\phi(\omega)$) one recovers random binary search trees, random recursive trees and random PORTs as special cases.

- Random binary search trees (or equivalently random binary trees): $\phi_0 = 1, \phi_1 = 2, \phi_2 = 1$ and $\phi_r = 0$ for all $r \geq 3$, or equivalently $\phi(\omega) = (1+\omega)^2$.
- Random recursive trees: $\phi_r = 1/r!$ for all $r \geq 0$, or equivalently $\phi(\omega) = e^{\omega}$.
- Random PORTs: $\phi_r = 1$ for all $r \geq 0$, or equivalently $\phi(\omega) = (1-\omega)^{-1}$.

Subsequently, we will fix a simple class of random increasing trees. Then, the subtree size profile is a double-indexed random variable which we again denote by $X_{n,k}$. Arguing as in the introduction, we have

$$X_{n,k} \stackrel{d}{=} \sum_{i=1}^{N} X_{I_i,k}^{(i)} \qquad (n > k) \tag{20}$$

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, $X_{n,k}, X_{n,k}^{(i)}, (N, I_1, I_2, \ldots)$ are independent and the joint distribution of the latter random variable is given by

$$\pi_{n,r,i_1,\ldots,i_r} = P(N = r, I_1 = i_1, \ldots, I_r = i_r) = [\omega^r]\phi(\omega)\binom{n-1}{i_1,\ldots,i_r}\frac{\tau_{i_1}\cdots\tau_{i_r}}{\tau_n},$$

where $i_1, \ldots, i_r \geq 1$ and $i_1 + \cdots + i_r = n-1$. As in Section 2, we set

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

Then, we have

$$\frac{\partial}{\partial z} P_k(z, y) = \phi(P_k(z, y)) + (e^y - 1)\tau_k \frac{z^k}{k!}.$$

with initial condition $P_k(0, y) = 0$. Next, consider the (scaled) exponential generating function of the m-th moment

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

Taking derivatives shows that all these functions satisfy a differential equation of the type

$$A'(z) = \phi'(\tau(z))A(z) + B(z)$$

where $B(z)$ is a function of generating functions of moments of smaller order. From this, we can read off the following recurrence relation for the moments

$$a_{n,k} = \sum_{1 \leq j < n} \pi_{n,j} a_{j,k} + b_{n,k} \qquad (n > k)$$

with a suitable sequence $b_{n,k}$ (involving moments of lower order) and $a_{k,k}$ given, $b_{k,k} = 0$, and $a_{n,k} = b_{n,k} = 0$ for $n < k$. Moreover,

$$\pi_{n,j} := \frac{(n-1)!\tau_j}{j!\tau_n} [z^{n-1-j}] \phi'(\tau(z)).$$

As before, we need a general solution of this recurrence. Therefore, set

$$A(z) = \sum_{n \geq 1} \tau_n a_{n,k} \frac{z^n}{n!}, \qquad B(z) = \sum_{n \geq 1} \tau_n b_{n,k} \frac{z^n}{n!}.$$

Then, the recurrence translates into the differential equation

$$A'(z) = \phi'(\tau(z))A(z) + B'(z) + \tau_k a_{k,k} \frac{z^{k-1}}{(k-1)!}$$

with solution

$$A(z) = \tau'(z) \int_0^z \left(B'(t) + \tau_k a_{k,k} \frac{t^{k-1}}{(k-1)!} \right) (\tau'(t))^{-1} \mathrm{d}t.$$

By reading off coefficients, we obtain

$$a_{n,k} = \sum_{j \geq k+1} b_{j,k} \frac{n!\tau_j}{(j-1)!\tau_n} [z^n] \tau'(z) \int_0^z \frac{t^{j-1}}{\tau'(t)} \mathrm{d}t + a_{k,k} \frac{n!\tau_k}{(k-1)!\tau_n} [z^n] \tau'(z) \int_0^z \frac{t^{k-1}}{\tau'(t)} \mathrm{d}t. \tag{21}$$

Using this, one can in principle use the method from the preceding section to derive similar results as for PORTs for other simple families of increasing trees. We will state such results for special families of increasing trees which will be defined below.

Grown Simple Families of Increasing Trees. As mentioned in the introduction, random PORTs can be alternatively defined via a tree evolution process. It is an interesting question to ask which other simple families of random increasing trees admit such a construction (via a natural tree evolution process). This question was completely answered in Panholzer and Prodinger [25], where it was shown that such a construction is possible if and only if the class belongs to the following list.

- Random d-ary trees: $\phi(\omega) = \phi_0(1 + ct/\phi_0)^d$ where $\phi_0 > 0, c > 0$ and $d \in \{2, 3, 4, \dots\}$.
- Random recursive trees: $\phi(\omega) = \phi_0 e^{ct/\phi_0}$, where $\phi_0 > 0$ and $c > 0$.
- Generalized random PORTs: $\phi(\omega) = \phi_0(1 - ct/\phi_0)^{-r+1}$, where $\phi_0 > 0, c > 0$ and $r > 1$.

Moreover, as explained in Kuba and Panholzer [24], for stochastic properties it is sufficient to consider the following special cases.

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

We will state results similar to the one for random PORTs for these three simple families of random increasing trees (where random recursive trees are already covered by previous work).

Mean Value. Here, we show how to compute the mean value of the subtree size profile of random d-ary trees and generalized random PORTs. We will see that in all cases, the mean value admits a simple exact expression.

We start with generalized random PORTs. Therefore, observe that by solving (19) one obtains

$$\tau(z) = 1 - (1 - rz)^{1/r}.$$

This in particular gives

$$\tau_n = r^{n-1}(n-1)!\binom{n-1-1/r}{n-1}.$$

Note that the latter formula implies that $\tau_{n+1}/\tau_n = rn - 1$.

Now, we use (21) with $b_{n,k} = 0$ for all n, k and $a_{k,k} = 1$. This yields

$$\mathbb{E}(X_{n,k}) = \frac{n!\tau_k}{(k-1)!\tau_n}[z^n]\tau'(z)\int_0^z \frac{t^{k-1}}{\tau'(t)}\mathrm{d}t.$$

Next, consider for $n > k$

$$\begin{aligned}
[z^n]\tau'(z)\int_0^z \frac{t^{k-1}}{\tau'(t)}\mathrm{d}t &= [z^n]\tau'(z)\int_0^z t^{k-1}(1-rt)^{1-1/r}\mathrm{d}t \\
&= [z^n]\tau'(z)r^{-k+1}\sum_{l=0}^{k-1}\binom{k-1}{l}(-1)^l\int_0^z (1-rt)^{l+1-1/r}\mathrm{d}t \\
&= [z^n]\tau'(z)r^{-k+1}\sum_{l=1}^{k-1}\binom{k-1}{l}(-1)^l\frac{1}{rl+2r-1} \\
&= [z^n]\tau'(z)r^{-k}\int_0^1 t^{k-1}(1-t)^{1-1/r}\mathrm{d}t. \\
&= \frac{\tau_{n+1}}{n!}r^{-k}\frac{(k-1)!\Gamma(2-1/r)}{\Gamma(k+2-1/r)}.
\end{aligned}$$

Substituting this into the expression above yields for $n > k$

$$\mathbb{E}(X_{n,k}) = \frac{\tau_{n+1}}{\tau_n} r^{-1}(k-1)!\binom{k-1-1/r}{k-1}\frac{\Gamma(2-1/r)}{\Gamma(k+2-1/r)} = \frac{(r-1)(rn-1)}{(rk+r-1)(rk-1)}.$$

A similar computation gives the mean value of the subtree size profile for random d-ary trees. We collect our result in the following theorem.

Theorem 8. *(i) For generalized random PORTs, we have for $n > k$*

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

(ii) For random d-ary trees, we have for $n > k$

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

Limit Laws for Varying k. Here, we state limit laws for the subtree size profile of generalized random PORTs and random d-ary trees. These results follow from (21) with the method of proof from Section 3. Details will be left to the reader.

First, we state the result for generalized random PORTs.

Theorem 9. *(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 $n \to \infty$,

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

(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,

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

Similar, we have for random d-ary trees.

Theorem 10. *(i) (Normal range) Let $k = k(n)$ such $k \to \infty$ as $n \to \infty$. Then,*

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

where, as $n \to \infty$,

$$\sigma_{n,k}^2 \sim \frac{d}{d-1} \cdot \frac{n}{k^2}.$$

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

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

(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 acknowledge partial support by National Science Council under the grant NSC-99-2115-M-009-007-MY2. Finally, helpful comments of the reviewer and editor are appreciated as well.

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

[10] 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.

[11] 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.

[12] 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.

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

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

[15] 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.

[16] 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.

[17] 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.

[18] J. Fill and N. Kapur (2004). Limiting distributions for additive functionals on Catalan trees, *Theoretical Computer Science*, **326**, 69-102.

[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 (2009). *Analytic Combinatorics*. Cambridge University Press.

[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] M. Kuba and A. Panholzer (2007). On the degree distribution of the nodes in increasing trees, *Journal of Combinatorial Theory, Series A*, **114**, 597-618.

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

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

[27] 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.

[28] R. T. Smythe and H. Mahmoud (1995). A survey of recursive trees, *Theory of Probability and Mathematical Statistics*, **51**, 1-27.