An Insertion Algorithm and Leaders of Rooted Trees

Qing-Hu Hou

Center for Applied Mathematics
Tianjin University, Tianjin 300072, P. R. China
hou@nankai.edu.cn

Abstract

The notion of leaders of labeled rooted trees was introduced by Seo. A vertex in a labeled rooted tree is called a leader if it has no smaller descendants. We present an algorithm which leads to a bijection between labeled rooted trees and integer sequences $a_1 \cdots a_{n-1}$ with $a_i \in \{1, 2, \ldots, n\}$ such that the number of leaders is exactly one more than the number of anti-excedances, namely, the positions i for which $a_i \leq i$. Our bijection gives a refinement of an identity of Gessel and Seo which takes the degree of 2 into account. By taking the reverse complement of a sequence, we obtain a combinatorial interpretation of a symmetry property on the enumeration of forests by the number of leaders and the number of components. This question was raised by Gessel and Seo. Applying a theorem of Lyapunov, we show that the distribution of the number of leaders of a random rooted tree is asymptotically normal.

Keywords: labeled rooted tree, leader, Cayley's formula, Postnikov's hook length formula

AMS Classification: 05A15, 05C05, 05C30

1 Introduction

Throughout the paper, we consider labeled rooted trees in which each vertex is given a unique integer label. Especially, a tree on $[n] = \{1, 2, \ldots, n\}$ refers to a tree with n vertices labeled by $1, 2, \ldots, n$. A vertex u is called a descendant of a vertex v if v is in the path from u to the root. A vertex v of a tree T is called a leader (or proper vertex by Seo [19]) if v is smaller than all its descendants. Note that a leaf itself is regarded as a leader. The set of leaders, the number of leaders and the degree of a vertex v (the number of children of v) are denoted by Lead (T), lead (T), and $\deg_T(v)$, respectively.

The number of leaders turns out to be an important statistic for the enumeration of rooted trees. The well-known formula of Cayley [2] states that the number of rooted labeled trees with n vertices equals n^{n-1}. Gessel and Seo [6] discovered the following remarkable generalization of Cayley's formula:

$$\sum_T u^{\mathrm{lead}\,(T)} c^{\deg_T(1)} = cu^2 \prod_{i=1}^{n-2} (i + (n-1-i)u + cu), \tag{1.1}$$

where T ranges over all trees on $[n]$ with root 1. Callan [1] provided an involution to count the number of forests of unrooted trees on $[n]$ with respect to the number of leaders. Guo and Zeng [7] noticed that the generating function of plane trees with respect to the number of younger children and the number of elder children has a similar form to (1.1). Chen and Peng [3] gave a combinatorial interpretation.

The notion of leaders was introduced by Seo in his combinatorial proof of Postnikov's hook length formula [15]:

$$\frac{n!}{2^n}\sum_{T}\prod_{v\in V(T)}\left(1+\frac{1}{h(v)}\right)=(n+1)^{n-1}, \tag{1.2}$$

where T ranges over all unlabeled binary trees on n vertices, $V(T)$ is the vertex set, and $h(v)$ is the number of descendants of v under the convention that v is counted as a descendant of itself. Seo [19] found a proof of (1.2) by using the generating function (1.3) for labeled trees with respect to the number of leaders. For other combinatorial proofs of Postnikov's identity and its generalizations, see Seo [19], Chen and Yang [4], Du and Liu [5], Liu [13], Shin and Zeng [21], Han [9], and Kuba and Panholzer [11, 12].

Gessel and Seo [6] counted various kinds of labeled rooted trees and forests by the number of leaders. They obtained the corresponding generating functions by establishing and solving differential equations. Let $a_1a_2\cdots a_{n-1}$ be the reverse Prüfer code of a tree. Seo and Shin [20] showed that among the n choices of a_i, there are exactly i choices leading to a new leader. By this observation, they gave a combinatorial interpretation for (1.1) and for other generating functions. However, from a given Prüfer code or a given reverse Prüfer code, it appears that the number of leaders cannot be directly deduced without explicitly constructing the tree from the code.

The main objective of this paper is to present a bijection between rooted trees and sequences $a_1\cdots a_{n-1}$ with $a_i\in[n]$ such that the number of leaders of a tree can be easily determined from the corresponding sequence. For this purpose, we introduce an insertion algorithm on rooted trees. Given a sequence $a_1\cdots a_{n-1}$, the algorithm constructs a tree recursively starting from the rooted tree with only one vertex 1. In each step, we either add a new maximum leaf or insert a vertex above the maximum leaf. It is shown that the algorithm builds up a bijection between trees and sequences such that the number of leaders is exactly one more than that of anti-excedances, i.e., the positions i's satisfying $a_i\le i$. Moreover, the number of children of the vertex 1 coincides with the number of 1's appearing in the corresponding sequence. Thus the relation (1.1) follows immediately from our bijection. A further consideration leads us to a refinement of (1.1) by taking the degree of the vertex 2 into account, see Theorem 2.3. Moreover, by taking the reverse complement, namely, $a_i\to(n+1)-a_{n-i}$, we are led to a combinatorial interpretation for the symmetry relation due to Gessel and Seo that the number of trees with j non-leaders is equal to the number of trees with $j+1$ leaders.

The insertion algorithm can be easily extended to forests, k-ary trees, k-colored ordered trees and k-colored ordered forests. The corresponding bijections lead to combinatorial proofs for the formulas obtained by Gessel and Seo using generating functions (see [6]). In particular, the bijection for k-ary trees implies that

$$\sum_{U} u^{\mathrm{lead}\,(U)} = u \prod_{i=1}^{n-1} \big((n-i)k + (ki - i + 1)u\big), \tag{1.3}$$

where U ranges over all k-ary trees on $[n]$. Consequently, Postnikov's formula (1.2) can be deduced from (1.3) by taking $k = u = 2$. In fact, among the $n!$ ways of labeling an unlabeled binary tree, there are exactly

$$\frac{n!}{\prod_{v \in S} h(v)}$$

of them such that all the vertices in S are leaders. When $S = [n]$, the above formula reduces to the formula for the number of increasing labelings of a binary tree [18, Formula (4)]. Thus,

$$n! \sum_{T} \prod_{v \in V(T)} \left(1 + \frac{1}{h(v)}\right) = \sum_{T} \sum_{S \subseteq V(T)} \frac{n!}{\prod_{v \in S} h(v)} = \sum_{T'} \sum_{S \subseteq \mathrm{Lead}\,(T')} 1 = \sum_{T'} 2^{\mathrm{lead}\,(T')},$$

where T' ranges over all labeled binary trees on $[n]$.

Adopting the notation in [6], we denote by $\mathrm{tree}\,(F)$ the number of components (trees) in a forest F and define

$$P_n(a,b,c) = \sum_{F} a^{n-\mathrm{lead}\,(F)} b^{\mathrm{lead}\,(F)-\mathrm{tree}\,(F)} c^{\mathrm{tree}\,(F)},$$

where F ranges over all forests on $[n]$. Like the case of rooted trees, the operation of reverse complement leads to a simple bijective proof for the symmetry relation $P_n(a,b,c) = P_n(b,a,c)$ for forests. This question was raised by Gessel and Seo [6].

The last section deals with the distribution of the number of leaders in a random rooted trees on $[n]$. By a theorem of Lyapunov, we show that the limiting distribution is normal.

2 An Insertion Algorithm for Rooted Trees

In this section, we present an insertion algorithm which bijectively maps the sequences $a_1 a_2 \cdots a_{n-1}$ with $a_i \in [n]$ to the labeled trees on $[n]$. This bijection has many features compared with the classical Prüfer code. In Prüfer's bijection, one adds a leaf at every step in the construction of a tree from a sequence. While in our insertion algorithm, there are two kinds of operations. One may add a leaf or insert an inner vertex on a leaf edge, that is, an edge associated with a leaf. Prüfer's code records the labeling of the parent vertex and hence preserves the information of the degrees of vertices. Our bijection focuses on the ordering of the labels of vertices in order to keep track of the number of leaders, as described in the following theorem.

Theorem 2.1. *There is a bijection between trees on* $[n]$ *and integer sequences* $a_1 \cdots a_{n-1}$ *with* $a_i \in [n]$ *such that the number of leaders is exactly one more than that of* i*'s satisfying* $a_i \leq i$ *in the corresponding sequence. Further more, the degree of the vertex* 1 *in a tree equals the number of appearances of* 1*'s in the corresponding sequence.*

Proof. The bijective map ϕ from sequences to trees is given by the insertion algorithm described as follows.

Initially, we have a tree T_0 which consists of only one vertex, the root 1. Suppose that we have constructed the tree T_{k-1} with k vertices and label set $S_{k-1} \subseteq [n]$. Now let us construct T_k from T_{k-1} according to a_k, denoted by $(T_{k-1}, a_k) \to T_k$. There are two cases: either (a) $a_k \leq k$ or (b) $a_k > k$. For case (a), we add a child to the a_k-th vertex of T_{k-1} (vertices are ordered by their labels) and label the child by the smallest unused label, i.e., the smallest element in $[n] \setminus S_{k-1}$. For case (b), we insert a vertex above the largest leaf of T_{k-1} and label the new vertex by the $(a_k - k)$-th element in $[n] \setminus S_{k-1} = \{l_1 < l_2 < \cdots < l_{n-k}\}$. For examples, see Figure 2.1 where the new vertices are represented by empty circles. Repeating the above procedure, we finally get a tree

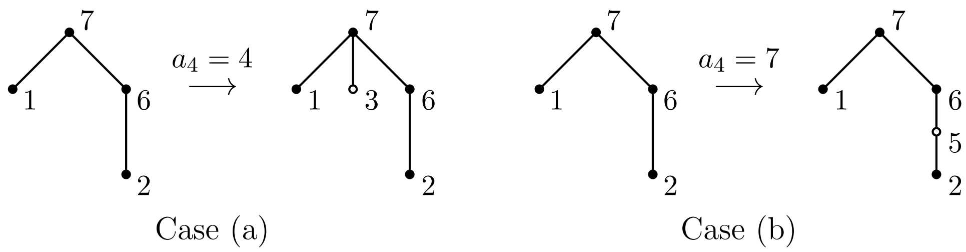

Figure 2.1: The two cases in the construction.

T_{n-1} with label set $[n]$. Now define $\phi(a_1 a_2 \cdots a_{n-1}) = T_{n-1}$.

Notice that in the insertion algorithm, we have used the greedy algorithm in the sense that at each stage when a leaf is added, we always give the minimum available label to the added leaf. Therefore the trees T_k $(0 \leq k \leq n-1)$ have the greedy property: If j is a leaf of T_k, then $[j] \subseteq S_k$, i.e., all labels less than or equal to j appear in T_k. We call such a tree a greedy tree. As we will see below, the greedy property enables us to count the number of leaders based on the corresponding sequence.

Let T be a greedy tree with label set $S \subseteq [n]$ and j its largest leaf. Unless the trivial case ($T = T_0$), the vertex j has always a parent j'. Now consider the tree T' obtained from T by removing the leaf j. Either T' is a greedy tree, or j' is a leaf of T' and $[j'] \not\subseteq S \setminus \{j\}$. In the former case, let a be the ordering of j' in the label set of T'. Then we have $(T', a) \to T$ as in case (a). In the latter case, let T'' be the tree obtained from T' by changing the label j' into j. Since $[j'] \not\subseteq S \setminus \{j\}$ but $[j] \subseteq S$, it holds that $j' > j$ and thus T'' is a greedy tree. Let S'' be the label set of T'' and a the number of vertices of T'' plus the ordering of j' in $[n] \setminus S''$. Then we have $(T'', a) \to T$ as in case (b). In summary, given a greedy tree T with at least two vertices, there always exist a greedy tree T' and an integer $a \in [n]$ such that $(T', a) \to T$. Notice further that all trees with label set $[n]$ are greedy trees and that the only greedy tree with one vertex is T_0. We thus derive that ϕ is surjective.

Let T be a greedy tree as above. Suppose that we have two ways to construct T: $(T', a') \to T$ and $(T'', a'') \to T$. If both of the constructions are of case (a), then $T' = T''$ is the unique tree obtained by removing the largest leaf from T and thus $a' = a''$. If both of the constructions are of case (b), then we also have $T' = T''$ is the unique tree obtained by removing the parent vertex of the largest leaf from T and thus $a' = a''$. Finally, suppose that one construction is of case (a) and the other is of case (b). Then the tree obtained by removing the largest leaf from T is both a greedy tree and a non-greedy tree, which is a contradiction. Therefore, the mapping ϕ is injective. Recall that we have proved that ϕ is surjective. Thus we know that ϕ is a bijection.

Now we will show that for $0 \leq k \leq n-1$, the number of leaders of T_k is exactly one more than that of i's satisfying $a_i \leq i$ in the sequence $a_1 \cdots a_k$. We prove it by induction on k. Clearly the statement holds for $k = 0$. Suppose that it holds for $k-1$.

If $a_k \leq k$, then T_k is obtained from T_{k-1} by adding one leaf. Therefore, if a vertex j is not a leader of T_{k-1}, it is also not a leader of T_k. Conversely, we will show that if a vertex $j \in T_{k-1}$ is a leader then it is also a leader of T_k. Suppose on the contrary that j is a leader of T_{k-1} but not a leader of T_k. This happens only when the label for the new leaf is less than j, i.e., the smallest integer in $[n] \setminus S_{k-1}$ is less than j. Let j' be any leaf in the subtree consisting of all the descendants of j in T_{k-1}. Since j is a leader of T_{k-1}, we have $j < j'$ and hence by the greedy property $[j] \subset [j'] \subseteq S_{k-1}$, which is a contradiction. Thus, T_k has exactly one more leader than T_{k-1}, i.e., the new leaf.

If $a_k > k$, then T_k is obtained from T_{k-1} by inserting a vertex above the largest leaf. Since the label of the new vertex is larger than the label of the largest leaf by the greedy property, the insertion does not change the number of leaders.

Summarizing, we prove that the assertion holds for k. Notice that $k = n-1$ means that the number of leaders of a tree is exactly one more than that of i's satisfying $a_i \leq i$ in the corresponding sequences.

Finally, since 1 is the minimal vertex in T_{k-1}, we add a child to the vertex 1 if and only if $a_k = 1$. Thus, the number of children of 1 equals the number of 1's appearing in the sequence $a_1 \cdots a_{n-1}$. ∎

Here is an example for $n = 7$ and the code $3, 5, 2, 4, 3, 7$. The corresponding tree is constructed as in Figure 2.2.

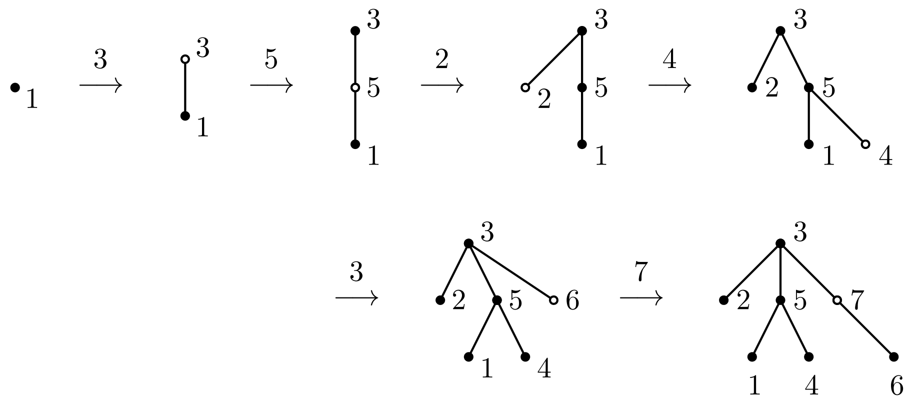

Figure 2.2: The tree corresponding to the sequence $(3, 5, 2, 4, 3, 7)$.

The bijection immediately leads to the generating function for the number of leaders:

Corollary 2.2. *We have*

$$\sum_T u^{\mathrm{lead}\,(T)} = u \prod_{i=1}^{n-1} (iu + (n-i)), \tag{2.1}$$

where T ranges over all trees on $[n]$.

It is well-known that the Prüfer code records the degrees of each vertex. Although our bijection does not, we can read out the degrees of vertex 1. In fact, we have the following generalization of the identity (1.1).

Theorem 2.3. *We have*

$$\sum_T u^{\mathrm{lead}\,(T)} x_1^{\deg_T(1)} x_2^{\deg_T(2)} = x_1 u^2 \prod_{i=1}^{n-2} (x_1 u + x_2 u + (i-1)u + (n-1-i)), \tag{2.2}$$

where T ranges over all trees on $[n]$ with root 1.

Proof. Notice that the labeled trees with root 1 are in one-to-one correspondce to the sequences $1a_2 \cdots a_{n-1}$ by the insertion algorithm. Furthermore, 2 is the second minimal vertex in T_i for $i \geq 1$. As in the proof of Theorem 2.1, the number of children of 2 equals the number of ocurrences of 2's in the sequence $1a_2 \cdots a_{n-1}$. Thus the bijection implies that

$$\begin{aligned}
&\sum_T u^{\mathrm{lead}\,(T)} x_1^{\deg_T(1)} x_2^{\deg_T(2)} \\
&\quad = x_1 u^2 \cdot (x_1 u + x_2 u + (n-2)) \cdot (x_1 u + x_2 u + u + (n-3)) \qquad (2.3)\\
&\qquad \cdots (x_1 u + x_2 u + (n-3)u + 1) \\
&\quad = x_1 u^2 \prod_{i=1}^{n-2} (x_1 u + x_2 u + (i-1)u + (n-1-i)), \qquad (2.4)
\end{aligned}$$

as desired. ∎

Now, the identity (1.1) follows immediately by setting $x_2 = 1$ in the identity (2.2).

When defining the sequence a_i' via $a_i' = (n+1) - a_{n-i}$, then it holds

$$a_{n-i} > n - i \quad \Longleftrightarrow \quad a_i' \leq i.$$

Therefore, by taking the reverse complement $a_i \to (n+1) - a_{n-i}$, we deduce the following symmetry property.

Corollary 2.4. *The number of trees on $[n]$ with j non-leaders is equal to that of trees with $j+1$ leaders.*

It should be noted that the above symmetry is also a consequence of Corollary 2.2. Let $[u^n]f(u)$ denote the coefficient of u^n in the expansion of $f(u)$. We see that

$$[u^{n-j}]u \prod_{i=1}^{n-1} (iu + (n-i)) = [u^j] u^n u^{-1} \prod_{i=1}^{n-1} (iu^{-1} + (n-i)) = [u^{j+1}] u \prod_{i=1}^{n-1} (iu + (n-i)).$$

The concept of leaders of a tree can be naturally generalized to forests. We say a vertex is a leader of a forest F if it is a leader of a certain tree (i.e., connected component) in F. Define the multi-variable generating function $P_n(a, b, c)$ for forests by

$$P_n(a, b, c) = \sum_F a^{n - \mathrm{lead}\,(F)} b^{\mathrm{lead}\,(F) - \mathrm{tree}\,(F)} c^{\mathrm{tree}\,(F)},$$

where F ranges over labeled forests on $[n]$, $\mathrm{lead}\,(F)$ is the number of leaders in F, and $\mathrm{tree}\,(F)$ denotes the number of trees of F. As pointed by Gessel and Seo [6], Corollary 2.4 implies a general symmetry $P_n(a, b, c) = P_n(b, a, c)$, which can be done by applying the operation of reverse complement to each tree in F.

3 Forests

Slight modifications of the insertion algorithm for trees lead to the bijections for other types of trees and forests. The principle remains the same, that is, adding a leaf whenever the code is no more than the total number of ways of adding a new leaf to a certain vertex and inserting an inner vertex above the largest leaf otherwise. The only difference lies in that for different types of trees and forests, we have different number of ways of adding a leaf and inserting an inner vertex.

We first consider the bijection for forests. They are in one-to-one correspondence to the sequences $a_1a_2\cdots a_{n-1}$ with $a_i \in \{0, 1, \ldots, n\}$ using the same algorithm as trees. We need only notice that for $a_i = 0$ we add a new tree with the root labeled by the minimal unused label. Then the number of trees and the number of leaders equals one more than that of 0's and anti-excedances in the sequence, respectively. Let $\mathrm{tree}\,(F)$ denotes the number of trees in F. The bijection implies the generating function formula

$$\sum_F b^{\mathrm{lead}\,(F)} c^{\mathrm{tree}\,(F)} = bc \prod_{i=1}^{n-1} ((n-i) + ib + bc),$$

which can be written in the following homogeneous form.

Theorem 3.1. *We have*

$$\sum_F a^{n-\mathrm{lead}\,(F)} b^{\mathrm{lead}\,(F)-\mathrm{tree}\,(F)} c^{\mathrm{tree}\,(F)} = c \prod_{i=1}^{n-1} (ia + (n-i)b + c), \tag{3.1}$$

where F ranges over all forests on $[n]$.

Equation (3.1) was proved by Gessel-Seo [6] and Moon-Yang [14] using algebraic methods. Here we give a combinatorial approach.

Recall that the left hand side of (3.1) is denoted by $P_n(a, b, c)$. Using the above coding for forests, we derive a simple bijective proof for the symmetry of $P_n(a, b, c) = P_n(b, a, c)$, which is a generalization of Corollary 2.4. In fact, let $a_1a_2\cdots a_{n-1}$ be the code of a forest F_a. Set

$$b_i = \begin{cases} 0, & \text{if } a_{n-i} = 0, \\ n + 1 - a_{n-i}, & \text{otherwise,} \end{cases}$$

and denote by F_b the forest corresponding to the sequence $b_1b_2\cdots b_{n-1}$. Then we have

$$\mathrm{tree}\,(F_b) = \mathrm{tree}\,(F_a), \quad \text{and} \quad n - \mathrm{lead}\,(F_b) = \mathrm{lead}\,(F_a) - \mathrm{tree}\,(F_a).$$

We have seen in Section 2 that the symmetry can also be deduced by applying the bijection for trees to each tree in a forest. Such a bijection preserves the labels in each tree. While the above bijection change the labels globally.

For example, the forests corresponding to $a = (4, 0, 3, 1, 2)$ and $b = (5, 6, 4, 0, 3)$ are shown in Figure 3.1. We see that both F_a and F_b consist of two trees. F_a has one non-leader and $\text{lead}(F_a) - \text{tree}(F_a) = 3$, while F_b has three non-leaders and $\text{lead}(F_b) - \text{tree}(F_b) = 1$.

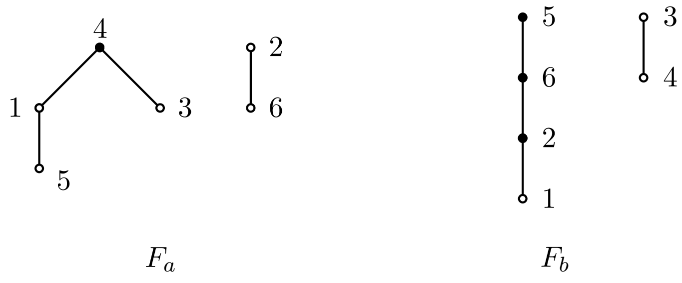

Figure 3.1: The forests F_a and F_b. Leaders are marked by empty circles.

4 k-Ary Trees

In this section, we consider the bijection between k-ary trees and sequences $a_1a_2 \cdots a_{n-1}$ with $a_i \in [nk - i + 1]$. Similar to the case of ordinary trees, the number of leaders is one more than that of i's satisfying $a_i \leq ki - i + 1$. This bijection leads to combinatorial proofs for Theorems 4.1 and 4.2.

A k-ary tree is a labeled tree such that each vertex has at most k children and each child of a vertex is designated as its first, second, ..., or k-th child. Clearly, for a vertex of degree d, we have $k - d$ ways to add a new child. Hence, for a k-ary tree with i vertices, we have $\sum(k - d) = ki - (i - 1)$ ways to add a new leaf. Moreover, we have k ways to insert a vertex with a fixed label above the largest leaf, i.e., designate the largest leaf as the first, the second, ..., or the k-th child of the new vertex. Thus, a k-ary tree is in one-to-one correspondence to the sequence $a_1a_2 \cdots a_{n-1}$ with $a_i \in [nk - i + 1]$ and the number of leaders is one more than that of i's satisfying $a_i \leq ki - i + 1$, which implies a result of Seo [19, Corollary 7] (see also [6, Theorem 7.1]).

Theorem 4.1. *We have*

$$\sum_U u^{\text{lead}\,(U)} = u \prod_{i=1}^{n-1} \big((n-i)k + (ki - i + 1)u\big), \tag{4.1}$$

where U ranges over all k-ary trees on $[n]$.

In [6], Gessel and Seo defined the decomposing path of a k-ary tree as follows. Starting from the root, the path goes always to the first child unless the vertex is a non-leader (or an improper vertex by Gessel and Seo) and its smallest descendant is a descendant of its first child. In the later case, the path goes to the second child. As in Section 2, we denote by $T_0, T_1, \ldots, T_{n-1}$ the successive trees in the construction. One can verify that the length of the decomposing path of T_i is the same as or one more than that of T_{i-1}. Moreover, among the $nk - i + 1$ possible values of a_i, there exists one and only one such that the length increases by one. Thus we recover Theorem 7.3 in [6].

Theorem 4.2. *For $k > 1$ we have*

$$\sum_{T} v^{n-\mathrm{lead}\,(T)} u^{\mathrm{lead}\,(T)-\mathrm{comp}\,(T)} w^{\mathrm{comp}\,(T)} = w \prod_{i=1}^{n-1} \big((n-i)kv + i(k-1)u + w\big),$$

where T ranges over all k-ary trees on $[n]$ and $\mathrm{comp}\,(T)$ *denotes the number of vertices on the decomposing path of T.*

5 k-Colored Ordered Trees and Forests

At last, we provide the insertion algorithms for k-colored ordered trees and forests. We will see that k-colored ordered trees are in one-to-one correspondence to sequences $a_1 a_2 \cdots a_{n-1}$ with $a_i \in [nk + i - 1]$ and the number of leaders is exactly one more than that of i's satisfying $a_i \leq ki + i - 1$. The bijection for k-colored ordered forests leads to Theorem 5.3, which is the analogue of Theorem 3.1.

An ordered tree is a labeled tree such that the children of each vertex are linearly ordered. A k-colored ordered tree is an ordered tree such that each edge is colored in colors $1, 2, \ldots, k$ and that among the children of any vertex, those joint with edges of color i precede those joint with edges of color j for $i < j$. We remark that k-colored ordered trees have also been studied under the notion of k-bundled ordered trees, see [10] for example. A k-colored ordered forest is a forest of k-colored ordered trees such that the roots are linearly ordered.

The generating functions for k-colored ordered trees and forests are given by the following two theorems. Moon and Yang [14, Theorem 2] gave an equivalent result on k-colored ordered trees based on a recurrence relation. We provide combinatorial proofs for these two theorems.

Suppose a vertex has r_j children joint with edges of color j. Then we have $r_j + 1$ ways to add a new child joint with an edge of color j. Therefore, for a vertex of degree d, we have $\sum (r_j + 1) = d + k$ different ways to add a new child. Hence, for a k-colored ordered tree with i vertices, we have $ki + i - 1$ ways to add a new leaf. Moreover, we have only k ways to insert a vertex with a fixed label above the largest leaf, i.e., to choose a color for the edge. Therefore, k-colored ordered trees are in one-to-one correspondence to sequences $a_1a_2\cdots a_{n-1}$ with $a_i \in [nk+i-1]$ and the number of leaders is exactly one more than that of i's satisfying $a_i \le ki+i-1$.

Thus we have the following theorem.

Theorem 5.1. *We have*

$$\sum_T u^{\mathrm{lead}\,(T)} = u\prod_{i=1}^{n-1}\big((n-i)k+(ki+i-1)u\big), \tag{5.1}$$

where T ranges over all k-colored ordered trees on $[n]$.

We have shown that for a k-colored ordered tree with t vertices, there are $kt+t-1$ different ways to add a new leaf. Hence for a k-colored ordered forest with i vertices, we have $\sum(tk+t-1)=(k+1)i-c$ different ways to add a new leaf, where c is the number of trees. Since the roots are linearly ordered, there are $c+1$ ways to insert a new root. Thus altogether we have $(k+1)i+1$ ways to add a new root or a new leaf. Moreover, we have k ways to insert a vertex with a fixed label above the maximal leaf. We are led to the following theorem.

Theorem 5.2. *We have*

$$\sum_F u^{\mathrm{lead}\,(F)} = u\prod_{i=1}^{n-1}\big((n-i)k+(ki+i+1)u\big), \tag{5.2}$$

where F ranges over all k-colored ordered forests on $[n]$.

Similar to the case of trees, the forests F_i have the greedy property. So the minimal unused label is greater than the minimal label in each tree of F_i. Let S_i be the sequence of the minimal element of each tree of F_i. Then the number of left-to-right minima of S_i is equal to or one more than that of S_{i-1}. Moreover, if and only if we insert a new tree in front of the other trees, the number increases by one. Thus we have given a combinatorial proof of Theorem 8.3 in [6].

Theorem 5.3. *We have*

$$\sum_F u^{n-\mathrm{lead}\,(F)}v^{\mathrm{lead}\,(F)-\mathrm{comp}\,(F)}w^{\mathrm{comp}\,(T)} = w\prod_{i=1}^{n-1}\big((n-i)ku+i(k+1)v+w\big),$$

where F ranges over all k-colored ordered forests on $[n]$ and comp (F) *is the number of left-to-right minima in the sequence consisting of the smallest element of each tree of F.*

6 The Limiting Distribution

In this section, we investigate the limiting distribution of the number of leaders under the equiprobable model. We assume that each of the n^{n-1} labeled rooted trees on $[n]$ occurs with probability $1/n^{n-1}$ and define the random variable ξ_n as the number of leaders in a tree. We will consider the limiting distribution of ξ_n as n tends to infinity.

The bijection given in Section 2 implies that

$$\xi_n = 1 + \sum_{i=1}^{n-1} \xi_{n,i} = \sum_{i=1}^{n} \xi_{n,i}, \tag{6.1}$$

where $\xi_{n,i}$ are independent Bernoulli distributions with $p_{n,i} = i/n$, i.e.,

$$\text{Prop}\,\{\xi_{n,i} = 1\} = i/n \quad \text{and} \quad \text{Prop}\,\{\xi_{n,i} = 0\} = (n-i)/n, \quad \forall\, 1 \le i \le n.$$

The random variables $\xi_{n,i}$ form a Poisson sequence [17, p. 23]. The mean and the variance of ξ_n can be easily computed:

$$\mathbf{E}\,\xi_n = \sum_{i=1}^{n} \mathbf{E}\,\xi_{n,i} = \sum_{i=1}^{n} i/n = (n+1)/2,$$

$$\mathbf{Var}\,\xi_n = \sum_{i=1}^{n} \mathbf{Var}\,\xi_{n,i} = \sum_{i=1}^{n} i(n-i)/n^2 = (n^2-1)/6n.$$

As a corollary of Lyapunov's theorem, we have [17, p. 23]

Lemma 6.1. *Let $X_{n,i}$ be a Poisson sequence with $P\{X_{n,i} = 1\} = p_{n,i}$ and*

$$B_n^2 = \sum_{i=1}^{n} p_{n,i}(1-p_{n,i}), \quad \eta_n = B_n^{-1} \sum_{i=1}^{n} (X_{n,i} - p_{n,i}).$$

If $B_n \to \infty$ as $n \to \infty$, then the sequence $\{\eta_n\}$ is asymptotically normal with parameters $(0,1)$.

Applying the lemma to $\xi_{n,i}$, we obtain the following limiting distribution of ξ_n.

Theorem 6.2. *Let ξ_n be the number of leaders of a random equiprobable labeled rooted tree on $[n]$. Then the distribution of the random variable*

$$\eta_n = \left(\xi_n - \frac{n+1}{2}\right) \bigg/ \sqrt{\frac{n^2-1}{6n}}$$

converges to the standard normal distribution as $n \to \infty$.

Acknowledgments. We would like to thank William Y.C. Chen, Seunghyun Seo, Heesung Shin and referees for valuable comments. This work was supported by the PCSIRT Project of the Ministry of Education and the National Science Foundation of China.

References

[1] D. Callan, A combinatorial derivation of the number of labeled forests, J. Integer Seq. **6** (2003) Article 03.4.7.

[2] A. Cayley, A theorem on trees, Quart. J. Math. **23** (1889) 376–378.

[3] W.Y.C. Chen and J.F.F. Peng, Disposition polynomials and plane trees, European J. Combin. **36** (2014) 122–129.

[4] W.Y.C. Chen and L.L.M. Yang, On Postnikov's hook length formula for binary trees, European J. Combin. **29** (2008) 1563–1565.

[5] R.R.X. Du and F. Liu, (k, m)-Catalan numbers and hook length polynomials for trees, European J. Combin. **28** (2007) 1312–1321.

[6] I.M. Gessel and S. Seo, A refinement of Cayley's formula for trees, Electron. J. Combin. **11(2)** (2004-2006) R27.

[7] V.J.W. Guo and J. Zeng, A generalization of the Ramanujan polynomials and plane trees, Adv. Appl. Math. **39(1)** (2007) 96–115.

[8] G.-N. Han, Discovering hook length formulas by an expansion technique, Electron. J. Combin. **15** (2008) R133.

[9] G.-N. Han, Yet another generalization of Postnikov's hook length formula for binary trees, SIAM J. Discrete Math. **23(2)** (2009) 661–664.

[10] M. Kuba and A. Panholzer, Enumeration results for alternating tree families, European J. Combin. **31** (2010) 1751–1780.

[11] M. Kuba and A. Panholzer, Bilabelled increasing trees and hook-length formulas, European J. Combin. **33** (2012) 248–258.

[12] M. Kuba and A. Panholzer, A unifying approach for proving hook-length formulas for weighted tree families, Graphs Combin. **29(6)** (2013) 1839–1865.

[13] F. Liu, Hook length polynomials for plane forests of a certain type, Ann. Combin. **13** (2009) 315–322.

[14] J.W. Moon and L.L.M. Yang, Postnikov identities and Seo's formula, Bull. Inst. Combin. Appl. **49** (2007) 21–31.

[15] A. Postnikov, Permutohedra, associahedra, and beyond, Retrospective in Combinatorics: Honoring Richard Stanley's 60th Birthday, MIT, Cambridge, 2004. Slides of the talk are available at http://www-math.mit.edu/~apost/talks/perm-slides.pdf.

[16] H. Prüfer, Neuer Beweis eines Satzes über Permutationen, Archiv der Math. **27(3)** (1918) 142–144.

[17] V. N. Sachkov, Probabilistic Methods in Combinatorial Analysis, Cambridge University Press, 1997.

[18] B.E. Sagan, Probabilistic proofs of hook length formulas involving trees, Séminaire Lotharingien de Combinatoire **61A** (2009) Article B61A.

[19] S. Seo, A combinatorial proof of Postnikov's identity and a generalized enumeration of labeled trees, Electron. J. Combin. **11(2)** (2004-2006) N3.

[20] S. Seo and H. Shin, A generalized enumeration of labeled trees and reverse Prüfer algorithm, J. Combin. Theory, Ser. A, **114** (2007) 1357–1361.

[21] H. Shin and J. Zeng, A further correspondence between (bc, b)-parking functions and (bc, b)-forests, DMTCS Proceedings (2009) 793–804.