The Butterfly Decomposition of Plane Trees

William Y.C. Chen[1], Nelson Y. Li[2], and Louis W. Shapiro[3]

[1,2]Center for Combinatorics, LPMC
Nankai University, Tianjin 300071, P.R. China

[3]Howard University, Washington, DC 20059, USA

[1]chen@nankai.edu.cn, [2]nelsonli@eyou.com, [3]lshapiro@howard.edu

Abstract. We introduce the notion of doubly rooted plane trees and give a decomposition of these trees, called the butterfly decomposition which turns out to have many applications. From the butterfly decomposition we obtain a one-to-one correspondence between doubly rooted plane trees and free Dyck paths, which implies a simple derivation of a relation between the Catalan numbers and the central binomial coefficients. We also establish a one-to-one correspondence between leaf-colored doubly rooted plane trees and free Schröder paths. The classical Chung-Feller theorem on free Dyck paths and some generalizations and variations with respect to Dyck paths and Schröder paths with flaws turn out to be immediate consequences of the butterfly decomposition and the preorder traversal of plane trees. We obtain two involutions on free Dyck paths and free Schröder paths, leading to two combinatorial identities. We also use the butterfly decomposition to give a combinatorial treatment of the generating function for the number of chains in plane trees due to Klazar. We further study the average size of chains in plane trees with n edges and show that this number asymptotically tends to $\frac{n+9}{6}$.

Keywords: Plane tree, doubly rooted plane tree, chains in plane trees, k-colored plane tree, butterfly decomposition, Dyck path, Schröder path.

AMS Classifications: 05A15, 05A19, 05C05.

1 Introduction

This paper is concerned with the enumeration of plane trees and the number of chains in plane trees with n edges. Although this subject has been very well studied over many decades, it seems that interesting problems still emerge. As we shall see, the enumeration of chains in plane trees leads us to discover a fundamental property of doubly rooted plane trees which has many applications. We call this the butterfly decomposition.

From the butterfly decomposition, we can establish a correspondence between doubly rooted plane trees and free Dyck paths. So we immediately get the relation between the Catalan numbers and the central binomial coefficients. The butterfly decomposition also implies the classical Chung-Feller theorem on free

Dyck paths with a given number of steps under x-axis. The Chung-Feller theorem was first proved by Major Percy A. MacMahon in 1909 [14, p.168] but named after its 1949 re-discoverers [4]. MacMahon proved it using formal series of words on an alphabet; Chung and Feller used generating functions. The previous combinatorial approaches to the Chung-Feller theorem are based on the cycle lemma or cyclic paths, see Dershowitz-Zaks [6] and Narayana [15]. There are other Chung-Feller type results and generalizations in [2, 3, 9, 10, 18, 22]. In some sense, the butterfly decomposition can be regarded as labelled variation of the cycle lemma.

The butterfly decomposition also leads to the following results: a correspondence between leaf-colored doubly rooted plane trees and free Schröder paths, a simple bijection between leaf-colored plane trees and Schröder paths, and a combinatorial interpretation of the generating function for the number of chains in plane trees obtained by Klazar [13]. We show that there is a one-to-one correspondence between chains in plane trees and tricolored plane trees (the definition is given in Section 5).

We obtain two involutions on free Dyck paths and free Schröder paths, which lead to two combinatorial identities. The last section of this paper gives an asymptotic formula for the average size of chains in plane trees with n edges.

2 The Butterfly Decomposition

In this section, we introduce the notion of doubly rooted plane trees and their butterfly decomposition. This decomposition seems to be fundamental for the enumeration of plane trees. It also implies the generating function for the number of chains in plane trees obtained by Klazar [13]. We will study the enumeration of chains in Section 5. The main result of this section is a correspondence between doubly rooted plane trees and free Dyck paths, from which it follows a combinatorial interpretation of the relation

$$(n+1)c_n = \binom{2n}{n}. \tag{2.1}$$

We will also establish a correspondence between free Dyck paths and 2-colored plane trees.

A (rooted) plane tree T with a distinguished vertex w is called a *doubly rooted plane tree*, where the distinguished vertex is regarded as the second root. The *butterfly decomposition* of a doubly rooted plane tree T with a distinguished vertex w is described as follows. Let $P = v_1 v_2 \ldots, v_k w$ be the path from the root of T to w. Let $L_1, L_2, \ldots, L_k$ be the subtrees such that L_i consists of the vertex v_i and its descendants on the left hand side of the path P. Similarly, we can define the subtrees $R_1, R_2, \ldots, R_k$ as the subtrees rooted at $v_1, v_2, \ldots, v_k$ consisting of the descendants on the right hand side of P. Moreover, the subtree of T rooted at w is denoted by T'. Therefore, a plane tree T with a distinguished vertex w can be decomposed into smaller structures $(U_1, U_2, \ldots, U_k; T')$, where U_i is called a *butterfly* consisting of L_i and R_i and the edge in the middle, as shown in Figure 1.

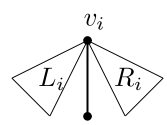

Figure 1: A Butterfly

Let B and C be the generating functions of the central binomial coefficients and Catalan numbers respectively:

$$B = \frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty} \binom{2n}{n} x^n, \tag{2.2}$$

$$C = \frac{1-\sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} x^n, \tag{2.3}$$

where $\binom{2n}{n}$ is called the *central binomial coefficient*, and $c_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th *Catalan number* [21]. It is clear that the generating function for a butterfly with n edges equals xC^2, and the generating function for a sequence of k butterflies with a total number of n edges equals $(xC^2)^k$. Note that the number of doubly rooted plane trees with n edges equals $n+1$ times the Catalan number, that is, the central binomial coefficient $\binom{2n}{n}$. Thus, we arrive at the following generating function relation:

$$B = C + C(xC^2) + C(xC^2)^2 + \cdots = \frac{C}{1-xC^2}. \tag{2.4}$$

A natural question arises: is there a simple combinatorial argument that leads to this conclusion without resorting to the formula for the Catalan numbers? The answer is affirmative, this leads to a quite simple derivation of the relation (2.1).

Recall that a *Dyck path* of length $2n$ is a lattice path from the origin to $(2n, 0)$ consisting of up steps $(1, 1)$ and down steps $(1, -1)$ that does not go below the x-axis. An *elevated Dyck path* or an irreducible Dyck path is defined as a Dyck path that does not touch the x-axis except for the origin and the destination. A lattice path from the origin to $(2n, 0)$ using the steps $(1, 1)$ and $(1, -1)$ without additional restrictions is called *a free Dyck path*, a free Dyck path is also called a Dyck path with flaws in the sense that the segments below the x-axis are regarded as flaws, see Eu-Liu-Yeh [10]. The reflection of a Dyck path with respect to the x-axis is called a negative Dyck path. An *elevated (irreducible) negative Dyck path* is defined in the same manner. As we shall see, free Dyck paths can be regarded as a labelled version of Dyck paths. Clearly, the set of free Dyck paths of length $2n$ is just the set of sequences consisting of n up steps and n down steps, as counted by the central binomial coefficient $\binom{2n}{n}$.

Theorem 2.1 *There is a bijection between the set of doubly rooted plane trees with n edges and the set of free Dyck paths of length $2n$.*

First we give a combinatorial setting for the proof of the above theorem. We recall the classical glove bijection between plane trees and Dyck paths [5]. This correspondence is also referred to as the preorder traversal of a plane tree. For the purpose of this paper, we may view the glove bijection as a recursive procedure. Recall that a *planted plane tree* is a plane tree whose root has only one child. Then the glove bijection gives a correspondence between the set of planted plane trees with n edges and the set of elevated Dyck paths of length $2n$. A planted plane tree with one edge corresponds to the elevated Dyck path of length two. Let T be a planted plane tree, and let $T_1, T_2, \ldots, T_k$ be the subtrees of the only one child of the root of T. Let $P_1, P_2, \ldots, P_k$ be elevated Dyck paths corresponding to $T_1, T_2, \ldots, T_k$ respectively. Then $UP_1P_2 \cdots P_kD$ is an elevated Dyck path of length $2n$, where U stands for an up step and D stands for a down step.

We are now ready to give a proof of Theorem 2.1.

Proof. Let T be a doubly rooted plane tree with n edges. Let w be the distinguished vertex of T and let $v_1v_2 \cdots v_kw$ be the path from the root to w. Suppose that $(L_1, R_1; L_2, R_2; \ldots; L_k, R_k; T')$ is the butterfly decomposition of T.

For the L_i $(1 \leq i \leq k)$ and T', we use the glove bijection to them and call the resulting Dyck paths P_i and P_{k+1}. For every R_i, we first add an edge at each root to form a planted plane tree T_i, then use the glove bijection to produce a negative elevated Dyck path Q_i. Now

$$P_1\, Q_1\, P_2\, Q_2\, \cdots\, P_k\, Q_k\, P_{k+1} \tag{2.5}$$

is a free Dyck path of length $2n$. Conversely, given a free Dyck path we may decompose it into elevated (irreducible) segments like the *first return decomposition* of a Dyck path [7], and we may reverse the above procedure to construct a doubly rooted plane tree because any free Dyck path P has a unique decomposition in the form (2.5) such that $Q_1, Q_2, \ldots, Q_k$ are negative elevated Dyck paths and $P_1, P_2, \ldots, P_{k+1}$ are the usual Dyck paths with the empty paths allowed. Thus we have established the bijection. ∎

An example of the above bijection is shown in Figure 2.

We next give another interpretation of the generating function for the number of bicolored plane trees. Guided by the following generating function identity

$$\frac{C}{1 - xC^2} = \frac{1}{1 - 2xC}, \tag{2.6}$$

we are led to introduce the notion of bicolored plane trees and k-colored plane trees, in general. A *k-colored plane tree* is a plane tree in which the children of the root are colored with k colors. A 2-colored plane tree is called a *bicolored plane tree*, and a 3-colored plane tree is called a *tricolored plane tree*. For bicolored plane trees, we assume that the two colors are black and white. Note that this terminology is somewhat misleading because in our context only the children of the root are colored. The relation (2.6) indicates that the set of bicolored plane trees are in one-to-one correspondence with doubly rooted plane trees. We next establish such a correspondence by making a connection between bicolored plane trees and free Dyck paths.

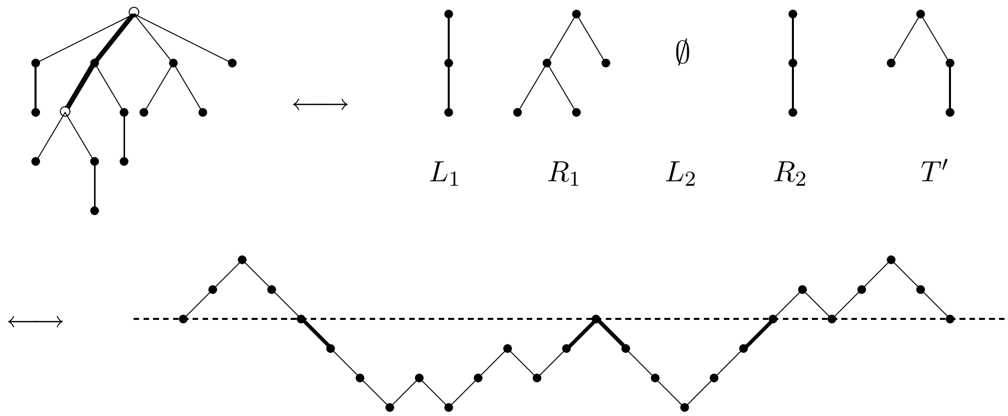

Figure 2: Doubly rooted plane trees and free Dyck paths

Theorem 2.2 *There is a one-to-one correspondence between the set of bicolored plane trees with n edges and the set of free Dyck paths of length $2n$.*

Proof. Let T be a bicolored plane tree, and let $T_1, T_2, \ldots, T_k$ be the planted subtrees of the root of T, listed from left to right. If T_i inherits the black color, then we construct an negative elevated Dyck path P_i from T_i; otherwise we construct an elevated Dyck path P_i above the x-axis. So we get a free Dyck path $P_1P_2 \ldots P_k$. Conversely, given a free Dyck path we may construct a bicolored plane tree. Hence we obtain the desired bijection. ∎

The bijections in Theorems 2.1 and 2.2 lead to a bijection between doubly rooted plane trees and bicolored plane trees. In fact, we may establish a direct correspondence without resorting to free Dyck paths.

Theorem 2.3 *There is a bijection between the set of doubly rooted plane trees with n edges and the set of bicolored plane trees with n edges.*

Proof. By the butterfly decomposition in Theorem 2.1, we get subtrees L_i, T_i and T'. By coloring L_i and T' black while T_i white, and identifying their roots as the root of the corresponding bicolored plane tree, we have its subtrees listed from left to right as

$$L_1\, T_1\, L_2\, T_2\, \cdots\, L_k\, T_k\, T'.$$

The reverse procedure is easy to construct. Thus we have established the bijection. ∎

3 The Chung-Feller Theorem

We begin this section by pointing out that the classical Chung-Feller theorem on Dyck paths is an immediate consequence of our bijection between doubly rooted plane trees and free Dyck paths. To see this connection, one only needs a simple observation on the preorder traversal of a plane tree. We also use this idea to derive some refinements and generalizations of the Chung-Feller theorem, including some recent results of Eu, Fu and Yeh [9] on Dyck paths and Schroöder paths with flaws.

Theorem 3.1 (Chung-Feller) *For any $0 \leq m \leq n$, the number of free Dyck paths of length $2n$ that contain exactly $2m$ steps below the x-axis is independent of m, and is equal to the n-th Catalan number c_n.*

Using the butterfly decomposition, we may transform the Chung-Feller theorem to an equivalent form on plane trees, which turns out to be a simple property of the preorder traversal. To be precise, we define the *right-to-left preorder traversal* of a plane tree T as a recursive procedure. First, visit the root of T. Let $T_1, T_2, \ldots, T_k$ be the subtrees of the root of T listed from left to right. Then traverse the subtrees in the order of $T_k, T_{k-1}, \ldots, T_1$. From the above traversal procedure, we may label the vertices of T with the numbers $0, 1, 2, \ldots, n$ in the order that they are visited. Figure 3 gives the plane tree corresponding to the free Dyck path in Figure 2 and the labels by the right-to-left preorder traversal.

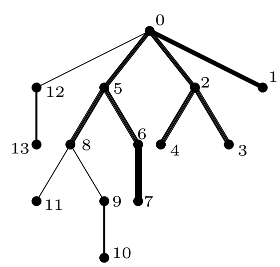

Figure 3: Labels for the Chung-Feller theorem

The following property immediately implies the Chung-Feller theorem since any plane tree can be regarded as a doubly rooted plane tree in which the distinguished vertex is chosen as the vertex with a given label m with respect to the right-to-left preorder traversal.

Theorem 3.2 *Let T be a plane tree with n edges. Assume that the vertices of T are labelled by $0, 1, 2, \ldots, n$ according to the right-to-left preorder traversal. Let w be the vertex labelled by m, where m is a given number not exceeding n. Then the doubly rooted plane tree T with w being the distinguished vertex corresponds to a free Dyck path with m down steps (up steps) below the x-axis.*

As another corollary, we note that half of all free Dyck paths end with an up step. Thus over all plane trees with n edges, half of the vertices are leafs, see Problem 10753 of the American Mathematical Monthly [12, 16].

The above interpretation of the Chung-Feller theorem also implies some refinements and generalizations recently obtained by Eu, Fu and Yeh [9]. Let us define some terminology. We say that a free Dyck path has m *flaws* if it contains m up (or down) steps below the x-axis. We note that a negative elevated (irreducible) Dyck path is called a *flaw block* by Eu, Fu and Yeh [9]. We define the *stem* of a doubly rooted plane tree as the path from the root to the distinguished vertex. Let T be a doubly rooted plane tree with a distinguished vertex w. An edge of T is said to be a *prefix edge* if it is either on on the stem of T or to the right of the stem. In other words, a prefix edge is an edge with labels not exceeding the label of the distinguished vertex with respect to the right-to-left preorder traversal. An example is shown in Figure 3 where the prefix edges are drawn with thick edges.

Using the preorder traversal of plane trees, we get the following generalization of the refined version of the Chung-Feller theorem [9].

Theorem 3.3 *For $0 \leq k \leq m \leq n$, there is a bijection between the set of free Dyck path of length $2n$ with m flaws in k flaw blocks and the set of doubly rooted plane trees of n edges with stem size k and m prefix edges.*

Proof. From the butterfly decomposition and the correspondence in Theorem 2.1, we see that the number of flaw blocks in a free Dyck path equals the stem size of the corresponding doubly rooted plane tree, and the number of flaws in a free Dyck path equals the number of prefix edges in the plane tree. This completes the proof. ∎

By the butterfly decomposition, one sees that the generating function for doubly rooted plane trees with stem size k equals $x^k C^k \cdot C^{k+1}$. It follows that the number of such trees with n edges and m prefix edges equals $[x^m]x^k C^k \cdot [x^{n-m}]C^{k+1}$, where $[x^n]C^k$ is the usual notation for the coefficient of x^n in the expansion of C^k. By the Lagrange inversion formula [21], we have

$$[x^n]C^k = \frac{k}{2n+k}\binom{2n+k}{n}. \tag{3.1}$$

Thus, we obtain the following expression.

Corollary 3.4 *For $0 < k \leq m \leq n$, the number of free Dyck paths of length $2n$ with m flaws and k flaw blocks equals*

$$\frac{k}{2m-k}\binom{2m-k}{m}\frac{k+1}{2n-2m+k+1}\binom{2n-2m+k+1}{n-m}.$$

Setting $m = n$ in the above corollary, one gets the number of Dyck paths of length $2n$ with k returns obtained by Deutsch [7]:

$$\frac{k}{2n-k}\binom{2n-k}{n}. \tag{3.2}$$

We next consider the enumeration of Schröder paths with flaws. For this purpose, we need to introduce the notion of *leaf-colored doubly rooted plane trees* which are defined as doubly rooted plane trees whose leaves are colored with two colors red (R) and blue (B) under the convention that the distinguished vertex receives no color even if it is a leaf. An edge of a plane tree is called an *external edge* if it's end vertex is a leaf; otherwise it is called an *internal edge*. As we shall see, such leaf-colored doubly rooted plane trees are in one-to-one correspondence with free Schröder paths. We note that there is a bijection between Schröder paths and plane trees with every leaf being colored red or blue, see Gouyou-Beauchamps and Vauquelin [11], here we create a new one.

Recall that a *Schröder path* of length $2n$ is a lattice path in the plane from $(0,0)$ to $(2n,0)$ with up steps $U=(1,1)$, horizontal steps $H=(2,0)$, and down steps $D=(1,-1)$, that never go below the x-axis. These paths are enumerated by the Schröder numbers r_n [21]. An *elevated (irreducible) Schröder path* and a *negative Schröder path* are defined in the same manner as with Dyck paths. A lattice path from $(0,0)$ to $(2n,0)$ with steps $U=(1,1)$, $H=(2,0)$, and $D=(1,-1)$ without additional restrictions is called a *free Schröder path*. We say that a free Schröder path has m *flaws* if the number of U steps and H steps under the x-axis equals m. A *flaw block* of a Schröder path is defined as a negative elevated Schröder path.

By the preorder traversal, we obtain the following correspondence.

Theorem 3.5 *There is a one-to-one correspondence between the set of plane trees with n edges in which each leaf is colored red or blue and the set of Schröder paths of length $2n$.*

Proof. Let T be a plane tree with n edges in which each leaf is colored red or blue. We proceed to construct a Schröder path of length $2n$ from the (left-to-right) preorder traversal. In the preorder traversal of the vertices of T, each edge is visited twice. Note that when an external edge $e=(u,v)$ (v is a leaf) is traversed, one always visits the vertex u, then the leaf v, and then immediately goes back to the vertex u. Now we may generate a sequence of U, D, and H steps by the following rule: (1) When an internal edge is visited for the first time, we get an U step. (2) When an internal edge is visited for the second time, we get a D step. (3) When an external edge with a red leaf is traversed, we get two steps UD. (4) When an external edge with a blue leaf is traversed, we get an H step. It is easy to see that we obtain a Schröder path of length $2n$ and the above procedure is reversible. ∎

By using the butterfly decomposition, we obtain the following correspondence.

Theorem 3.6 *There is a bijection between the set of leaf-colored doubly rooted plane trees with n edges and the set of free Schröder paths of length $2n$.*

Proof. Similar to that of Theorem 2.1.

Recall that the number of plane trees with n edges and i leaves is given by the Narayana number [21]

$$N_{n,i}=\frac{1}{n}\binom{n}{i}\binom{n}{i-1}.$$

It follows that that the number of leaf-colored doubly rooted plane trees equals

$$\sum_{i=1}^{n}\left[(n+1-i)2^iN_{n,i}+i2^{i-1}N_{n,i}\right]=\sum_{i=1}^{n}(2n+2-i)2^{i-1}N_{n,i}.$$

On the other hand, it is easy to see that the number of free Schröder paths of length $2n$ is given by the summation

$$\sum_{i=0}^{n}\binom{2n-i}{i}\binom{2n-2i}{n-i}.$$

Hence Theorem 3.6 yields the following identity:

$$\sum_{i=1}^{n}(2n+2-i)2^{i-1}N_{n,i}=\sum_{i=0}^{n}\binom{2n-i}{i}\binom{2n-2i}{n-i}. \tag{3.3}$$

By the right-to-left preorder traversal and the above correspondence, one may determine a distinguished vertex of a plane tree whose leaves are colored red and blue. This fact can be restated as a Schröder path analogue of the Chung-Feller theorem obtained by Eu, Fu and Yeh [9].

Theorem 3.7 *For each Schröder path P from $(0,0)$ to $(2n,0)$, assign weight 2 to P if P ends with a U step; otherwise P is assigned weight 1. Let m be a given number not exceeding n. Then the total weight of the set of free Schröder paths of length $2n$ with m flaws is always the Schröder number r_n.*

If a free Schröder path ends with an up step, then the corresponding subtree T' is empty and we have that the distinguished vertex is a leaf. There are now two possible ways to color it, hence we assign weight 2 to this kind of Schröder paths.

Using plane trees, we may reinterpret the above theorem as follows.

Theorem 3.8 *Let T a plane tree with n edges. Assume that the vertices of T are labelled by $0,1,2,\ldots,n$ according to the right-to-left preorder traversal. Let w be a vertex labelled by m. Let T' be a leaf-colored doubly rooted plane tree T with w being the distinguished vertex. Then by the correspondence between leaf-colored doubly rooted plane trees and free Schröder paths, T' corresponds to a free Schröder path with m flaws.*

From the above theorem, we immediately get the following refinement of Eu, Fu and Yeh [9].

Theorem 3.9 *For $0\le k\le m\le n$, there is a bijection between the set of free Schröder path of length $2n$ with m flaws in k flaw blocks and the set of leaf-colored doubly rooted plane trees of n edges with stem size k and m prefix edges.*

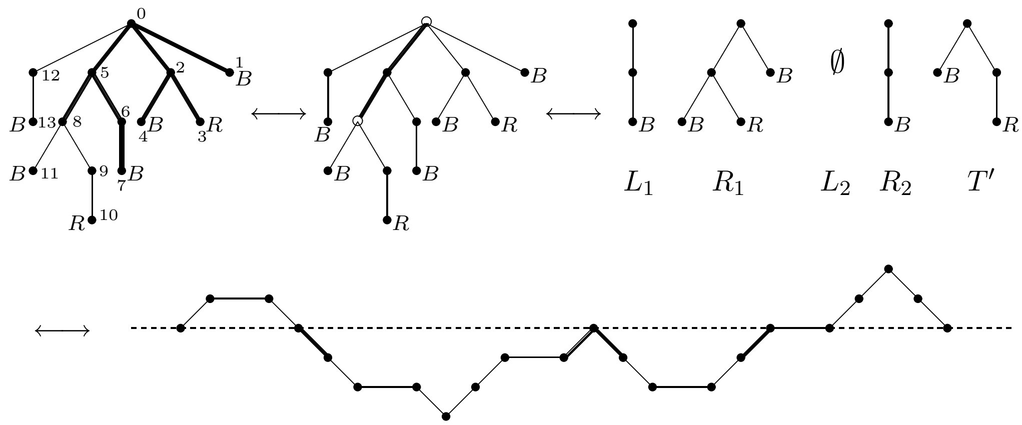

Figure 4: Leaf-colored plane trees and free Schröder paths

An example of the above bijection between leaf-colored doubly rooted plane trees and free Schröder paths is illustrated in Figure 4.

To conclude this section, we use the butterfly decomposition to obtain a formula for the total weight of leaf-colored doubly rooted plane trees of n edges with stem size k and m prefix edges. Let S be the generating function of the Schröder numbers as given by the equation $S = 1 + xS + xS^2$. Then the total weight of leaf-colored doubly rooted plane trees of n edges with stem size k and m prefix edges equals

$$2 \cdot [x^m]x^k S^k \cdot [x^{n-m}]S^k + [x^m]x^k S^k \cdot [x^{n-m}]S^k(S-1)$$

which can be rewritten as $[x^{m-k}]S^k[x^{n-m}](S^{k+1} + S^k)$. Let

$$a(n,k) = [x^n]S^k. \tag{3.4}$$

Set $a(0,k) = 1$. When $n \geq 1$, using the Lagrange inversion formula [21], we obtain that

$$a(n,k) = \frac{k}{n} \sum_{i=0}^{n-1} 2^{i+1} \binom{n+k-1}{i} \binom{n}{i+1}. \tag{3.5}$$

Note that $a(n,1)$ reduces to the Schröder number r_n.

Corollary 3.10 *For $0 < k \leq m \leq n$, the total weight of free Schröder paths of length $2n$ with m flaws and k flaw blocks equals*

$$a(m-k,k) \cdot [a(n-m,k+1) + a(n-m,k)].$$

4 Two Involutions

In this section, we present two parity reversing involutions on free Dyck paths and free Schröder paths, where the parity is defined as the parity of the number of its flaw blocks. We also derive two identities based on the computation via the butterfly decomposition.

Theorem 4.1 *For $n \geq 1$, there is a parity reversing involution on the set of free Dyck paths of length $2n$, which leads to the following identity*

$$\sum_{i=0}^{n}(-1)^i \frac{2i+1}{2n+1}\binom{2n+1}{n-i} = 0. \tag{4.1}$$

Proof. Let P be a free Dyck path of length $2n$. Let ϕ be the desired involution. If P ends with an elevated Dyck path, then we construct $\phi(P)$ by reflecting the last elevated Dyck path with respect to the x-axis. If P ends with a negative elevated Dyck path, then $\phi(P)$ is obtained by reflecting the last negative elevated Dyck path with respect to the x-axis. Clearly, ϕ is a parity reversing involution. By the butterfly decomposition, the number of free Dyck paths with i flaw blocks equals the number of doubly rooted plane trees with stem size i, that is, $[x^{n-i}]C^{2i+1}$. Hence the relation (4.1) follows from (3.1). ∎

We also have an involution on free Schröder paths.

Theorem 4.2 *For $n \geq 1$, there is a parity reversing involution on the set of free Schröder paths of length $2n$ containing at least one up step. So we have the following identity on $a(n,k)$ as defined by (3.5):*

$$\sum_{i=0}^{n}(-1)^i\, a(n-i, 2i+1) = 1. \tag{4.2}$$

Proof. Let P be a free Schröder path which contains at least one up step. Let Q be the last segment of P which is an elevated Schröder path or a negative elevated Schröder path. Note that Q may be followed by some horizontal steps in P. We reflect Q with respect to the x-axis to get a free Schröder path. Clearly, the resulting path contains at least one up step. It is easy to see that this construction is reversible and parity reversing. By the correspondence given in Theorem 3.9, the number of free Schröder paths of length $2n$ with i flaw blocks equals $[x^{n-i}]S^{2i+1}$, that is, $a(n-i, 2i+1)$. Therefore, identity (4.2) follows from the involution and the fact that the only Schröder path not affected by the involution is the path consisting of only horizontal steps. ∎

5 Chains in Plane Trees

Let us recall that a *chain* of a plane tree is a selection of vertices on a path from the root to a leaf. The *size* of a chain is defined as the number of vertices in the chain. Let Q_n be the number of nonempty chains in all plane trees with n edges. A tree with n edges may have as many as $2^{n+1} - 1$ non-empty chains and as few as $2n + 1$. The twelve chains in plane trees with 2 edges are illustrated in

Figure 5, where those empty circles stand for vertices in chains and black circles stand for normal vertices. For instance, the last structure of Figure 5 has a chain of size 2.

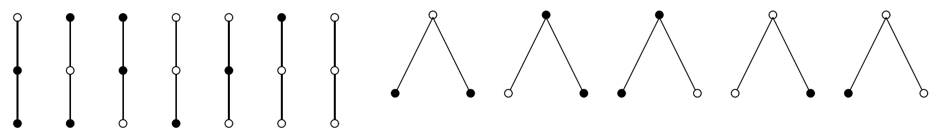

Figure 5: Chains in plane trees with 2 edges

The main result of this section is a combinatorial interpretation of the generating function for the number of chains in plane trees obtained by Klazar [13]. We also obtain a one-to-one correspondence between the set of chains in plane trees with n edges and the set of tricolored plane trees with n edges. Klazar [13] derived the following generating function for the number of chains in plane trees with n edges:

$$\frac{C}{1-2xC^2} = 1 + 3x + 12x^2 + 51x^3 + 222x^4 + 978x^5 + \cdots . \qquad (5.1)$$

Note that here we use a slightly different formulation of the generating function C from that used by Klazar [13].

We now give a combinatorial proof of the fact that the generating function of chains in plane trees with n edges equals $\frac{C}{1-2xC^2}$. Let T be a plane tree and Q be a chain of T. Suppose w is the vertex in Q such that the path $v_1v_2\cdots v_kw$ from the root of T to w contains all the vertices in Q. Moreover, we color the vertex v_i with the white color if it belongs to Q; otherwise, we color v_i with the black color. Such a coloring scheme leads to the following bijection.

Theorem 5.1 *There is a one-to-one correspondence between the set of chains in plane trees with n edges and the set of doubly rooted plane trees in which the vertices on the path from the root to the distinguished vertex (but not including the distinguished vertex) are colored with two colors.*

Using the above theorem and the butterfly decomposition of doubly rooted plane trees, we obtain the generating function of Klazar.

Motivated by the following relation

$$\frac{C}{1-2xC^2} = \frac{1}{1-3xC}, \qquad (5.2)$$

we are led to establish the following bijection.

Theorem 5.2 *There is a one-to-one correspondence between chains in plane trees and tricolored plane trees.*

Proof. Let T be a plane tree and Q be a chain of T. Let $v_1 v_2 \cdots v_k w$ be the path from the root to the vertex w, where w is the last vertex in the chain. Suppose that $(L_1, R_1; L_2, R_2; \ldots; L_k, R_k; T')$ is the butterfly decomposition of T. Let T_i be the planted plane tree obtained from R_i by adding a root. Coloring L_i and T' red, and color T_i white if the vertex v_i contained in T_i is a chain vertex, otherwise color T_i black. Identify their roots as the root of the corresponding tricolored plane tree, and set the subtrees of the root as

$$L_1\, T_1\, L_2\, T_2\, \cdots\, L_k\, T_k\, T'.$$

The reverse procedure is easy to construct. This completes the proof. ∎

An example of the above bijection is shown in Figure 6.

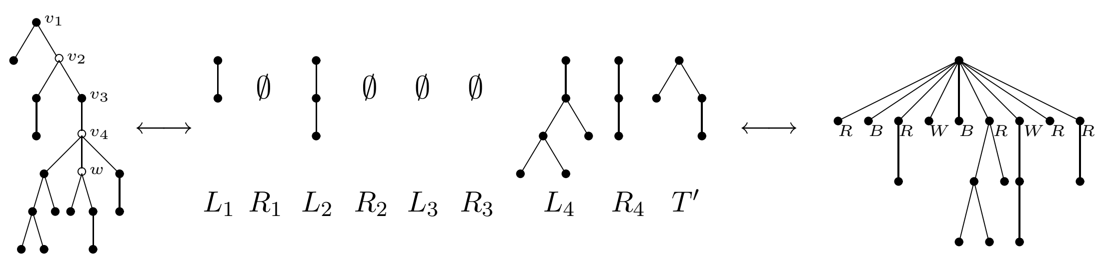

Figure 6: Chains and tricolored plane trees

From the above bijection, we easily see that chains with m vertices correspond to tricolored trees with $m-1$ white subtrees. Hence as a special case of Theorem 5.2, we obtain Theorem 2.3.

Notice that a chain in plane trees is just a two colored path in the butterfly decomposition. Hence we can color the vertices in chain with t colors and preserve these colors in the above bijection. Precisely speaking, a chain is called t-colored if its elements are t-colored. We have the following bijection.

Theorem 5.3 *There is a one-to-one correspondence between the set of $(k-2)$-colored chains in plane trees with n edges and the set of k-colored plane trees with n edges.*

The above bijection is a reflection of the following Catalan type identity

$$\frac{C}{1-(k-1)xC^2} = \frac{1}{1-kxC}.$$

Remark. By the path decomposition, the generating function for the number of chains with n edges that end with a leaf equals

$$\frac{1}{1-2xC^2} = \frac{1+\sqrt{1-4x}}{3\sqrt{1-4x}-1} = 1+2x+8x^2+34x^3+148x^4+652x^5+\cdots.$$

It is a new combinatorial explanation for Sequence A067336 in [19].

6 Average Size of Chains

In this section, we use the generating function B of central binomial coefficients as given by (2.2) to study the total size and average size of chains in plane trees with n edges. It turns out that by a decomposition of chains we may rewrite $\frac{C}{1-2xC^2}$ in order to give an asymptotic formula. We show that the average size of chains in plane trees with n edges asymptotically tends to $\frac{n+9}{6}$.

Bearing in mind that the generating function for the number of chains of size 1 equals the generating function B of the central binomial coefficients. We let L^* be the generating function for the number of plane trees with a distinguished leaf. Any tree with a distinguished vertex can be decomposed into a tree with a distinguished leaf and a subtree rooted at the distinguished vertex. Thus we have $B = L^*C$ and $L^* = B/C$.

We now consider plane trees with at least two vertices in which there is a distinguished leaf. Let L be the generating function of such plane trees with n edges. It is easy to obtain the following relations

$$L = L^* - 1 = \frac{B - C}{C} = \frac{B - 1}{2}.$$

Property 6.1 *The generating function for the total number of chains of size k in plane trees with n edges equals $B \cdot \left(\frac{B-1}{2}\right)^{k-1}$.*

Proof. The required generating function follows from a decomposition procedure for a plane tree with a given chain. Let T be a plane tree and Q be a chain of T. Let $w_1, w_2, \ldots, w_k$ be the chain vertices on the path from the root to the last vertex w_k. Then T can be decomposed into $k+1$ plane trees $T_1, T_2, \ldots, T_k$, and T', where T_1 is constructed from T by cutting off the subtrees of w_1, T_2 is obtained from the subtree of T rooted at w_1 by cutting off the subtrees of w_2, and so on, finally T' is the subtree of T rooted at w_k. The vertices $w_1, w_2, \ldots, w_k$ serve as distinguished vertices in $T_1, T_2, \ldots, T_k$. The generating function for the structure of T_1 equals $L^* = B/C$, since the distinguished vertex is allowed to coincide with the root in T_1. The generating function for other T_i $(2 \leq i \leq k)$ equals L and the generating function for T' equals C. Hence the required generating function equals $L^* \cdot L^{k-1} \cdot C = B \cdot \left(\frac{B-1}{2}\right)^{k-1}$. ∎

An interesting case arises if we look at chains of size 3 that include both the root and a leaf. In this case we have L^2 as our generating function. It is easily shown that $L^2 = x^2 + 6x^3 + 29x^4 + 130x^5 + \cdots$. This ubiquitous sequence, $A008549$, also counts [19]:

- The area under all Dyck paths of length $2n - 2$.
- The number of points at height one over all binomial paths of length $2n-2$.
- The number of inversions among all 321-avoiding permutations in S_n.

From the above theorem, we have the following generating function for the total number of chains in all plane trees with n edges.

Theorem 6.2 *The generating function for the total number of nonempty chains in all plane trees with n edges equals $\frac{2B}{3-B}$.*

Proof. We sum over k to get $B \cdot \sum_{k\geq 0} \left(\frac{B-1}{2}\right)^k = B \cdot \left(1 - \frac{B-1}{2}\right)^{-1} = \frac{2B}{3-B}$. ∎

Now we consider the asymptotic approximations. Let H_n be the total number of chains in plane trees with n edges. Klazar [13] has shown that

$$H_n \sim \frac{1}{2} \cdot \left(\frac{9}{2}\right)^n. \tag{6.1}$$

Now we use the language of Riordan arrays [17, 20] to compute the generating function for the total size of chains in all plane trees with n edges. The idea of Riordan arrays is represented as follows. Given two generating functions $g(x) = 1 + g_1x + g_2x^2 + \cdots$ and $f(x) = f_1x + f_2x^2 + \cdots$ with $f_1 \neq 0$, let $M = (m_{i,j})_{i,j\geq 0}$ be the infinite lower triangular matrix with nonzero entries on the main diagonal, where $m_{i,j} = [x^i](g(x)f^j(x))$, namely, $m_{i,j}$ equals the coefficient of x^i in the expansion of the series $g(x)f^j(x)$. If an infinite lower triangular matrix M can be constructed in this way from two generating functions $g(x)$ and $f(x)$, then it is called a *Riordan array* and is denoted by $M = (g(x), f(x)) = (g, f)$.

If we multiply the matrix $M = (g, f)$ by a column vector $(a_0, a_1, \cdots)^T$ to get a column vector $(b_0, b_1, \cdots)^T$, then the generating functions $A(x)$ and $R(x)$ of the sequences $(a_0, a_1, \cdots)$ and $(b_0, b_1, \cdots)$ satisfy the following relation

$$R(x) = g(x)A(f(x)).$$

We now have the following generating function for the total size of chains in plane trees with n edges.

Theorem 6.3 *The generating function for the total size of all chains in plane trees with n edges equals $\frac{4B}{(3-B)^2}$.*

Proof. Let $g(x) = B$ be the generating function for chains of size 1, and $f(x) = L = \frac{B-1}{2}$ be the generating function for plane trees with at least two vertices and a distinguished leaf. Consider the Riordan matrix (B, L). The generating function of the j-th ($j \geq 1$) column is $BL^{(j-1)}$, which is the generating function for the number of chains of size j. Since the generating function of $(1, 2, 3, 4 \cdots)^T$ is $A(x) = \frac{1}{(1-x)^2}$, it follows that the multiplication of the Riordan matrix (B, L) and the column vector $(1, 2, 3, 4 \cdots)^T$ gives the sequence of the total size of chains in plane trees with n edges. It follows that

$$R(x) = g(x)A(f(x)) = \frac{B}{(1-L)^2} = \frac{4B}{(3-B)^2}$$

is the generating function for the total size of chains. The matrix identity can be stated as

$$\begin{bmatrix} 1 & & & & & \\ 2 & 1 & & & & \\ 6 & 5 & 1 & & & \\ 20 & 22 & 8 & 1 & & \\ 70 & 93 & 47 & 11 & 1 & \\ & & \cdots & & & \ddots \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 19 \\ 92 \\ 446 \\ \vdots \end{bmatrix}.$$

This completes the proof. ∎

After some algebraic calculations we get

$$R = \frac{\frac{5-18x}{\sqrt{1-4x}} + 3}{8} \cdot \frac{1-4x}{\left(1 - \frac{9}{2}x\right)^2}.$$

Recall that Bender's lemma [1, p.496] basically says that if $C(x) = A(x) B(x)$ and the radii of convergence for $A(x)$ and $B(x)$ are α and β with $\alpha < \beta$, then

$$C_n \sim A_n B(\alpha).$$

Let $A(x) = \frac{1-4x}{\left(1-\frac{9}{2}x\right)^2}$ and $B(x) = \frac{1}{8} \cdot \left(\frac{5-18x}{\sqrt{1-4x}} + 3\right)$. We have $\alpha = 2/9 < \beta = 1/4$ for Bender's lemma. So we have $B\left(\frac{2}{9}\right) = \frac{3}{4}$ while $A_n = \frac{n+9}{2}\left(\frac{9}{2}\right)^{n-1}$. So we obtain the following asymptotic property.

Theorem 6.4 *Let R_n be the total size of chains in all plane trees with n edges. Then we have*

$$R_n \sim \frac{n+9}{12}\left(\frac{9}{2}\right)^n. \tag{6.2}$$

From Klazar's formula (6.1) and the above formula (6.2) it follows that the average size of chains in planes trees with n edges approaches

$$\lim_{n\to\infty} \frac{R_n}{H_n} = \frac{n+9}{6}.$$

For example,

$$\frac{R_{50}}{H_{50}} = \frac{2250\,588\,247\,788\,344\,466\,951\,528\,963\,319\,620}{228\,878\,511\,199\,384\,804\,987\,952\,173\,176\,432} \approx 9.833\,1,$$

while $\frac{50+9}{6} \approx 9.833\,3$.

Acknowledgments. The authors thank David S. Hough, Martin Klazar and Peter Winkler for helpful suggestions. This work was supported by the 973 Project on Mathematical Mechanization, the National Science Foundation, the Ministry of Education, and the Ministry of Science and Technology of China. The third author is partially supported by NSF grant HRD 0401697.

References

[1] E. Bender, Asymptotic methods in enumeration, *SIAM Rev.*, **16** (1974), 485-515; errata, ibid., **18** (1976), 292.

[2] D. Callan, Pair them up!: A visual approach to the Chung-Feller Theorem, *College Math. J.*, **26** (1995), 196-198.

[3] N. Cameron, The combinatorics of even trees, *Congressus Numeratium*, **147** (2000), 129-143.

[4] K.L. Chung and K. Feller, On fluctuations in coin-tossing, *Proc. Nat. Acad. Sci.* USA, **35** (1949), 605-608.

[5] N. Dershowitz and S. Zaks, Enumeration of ordered trees, *Discrete Math.*, **31** (1980), 9-28.

[6] N. Dershowitz and S. Zaks, The cycle lemma and some applications, *Europ. J. Combin.*, **11** (1990), 35-40.

[7] E. Deutsch, Dyck path enumeration, *Discrete Math.*, **204** (1999), 167-202.

[8] E. Deutsch and L. Shapiro, A bijection between ordered trees and 2-Motzkin paths and its many consequences, *Discrete Math.*, **256** (2002), 655-670.

[9] S.P. Eu, T.S. Fu and Y.N. Yeh, Refined Chung-Feller theorems for lattice paths, *J. Combin. Theory Ser. A*, **112** (2005), 143-162.

[10] S.P. Eu, S.C. Liu and Y.N. Yeh, Taylor expansions for Catalan and Motzkin numbers, *Adv. Appl. Math.*, **29** (2002), 345-357.

[11] D. Gouyou-Beauchamps and D. Vanquelin, Deux propriétés combinatoires des nombres de Schröder, *RAIRO Inform. Théor. Appl.*, **22** (1988) , 361-388.

[12] N. Hungerbühler, A simple proof of Shapiro's Theorem, *Amer. Math. Monthly*, **108** (2001), 873-874.

[13] M. Klazar, Twelve countings with rooted plane trees, *Europ. J. Combin.*, **18** (1997), 195-210.

[14] P.A. MacMahon, Memoir on the theory of the partitions of numbers, *Phil. Trans. Royal Soc. London (A)*, **209** (1909), 153-175, Also *Collected Works Vol.* 1, (G.E. Andrews, ed.), MIT Press, Cambridge, Mass 1978, 1292-1314.

[15] T.V. Narayana, Cyclic permutation of lattice paths and the Chung-Feller theorem, *Skandinavisk Aktuarietidskrift*, **50** (1967), 23-30.

[16] L. W. Shapiro, Problem 10753, *A*mer. Math. Monthly, **106** (1999), 777.

[17] L. W. Shapiro, S. Getu, Wen-Jin Woan and L.C. Woodson, The Riordan group, *Discrete Appl. Math.*, **34** (1991), 229-239.

[18] L. W. Shapiro and W.J. Woan, Some generating proofs of old and new results in probability, *Congressus Numerantium*, **143** (2000), 193-205.

[19] N.J.A. Sloane and S. Plouffe, The Online Encyclopedia of Integer Sequences, *www.research.att.com/˜njas/sequences/*.

[20] R. Sprugnoli, Riordan arrays and combinatorial sums, *Discrete Math.*, **132** (1994), 267-290.

[21] R.P. Stanley, *Enumerative Combinatorics, Vol. 2*, Cambridge University Press, Cambridge, 1999.

[22] W. Woan, Uniform partitions of lattice paths and Chung-Feller generalizations, *Amer. Math. Monthly*, **108** (2001), 556-559.