Discrete Appl. Math.
155 (2007), no. 17, 2187-2201

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]Department of Mathematics, Howard University, Washington, DC
20059, USA

emails: [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 as well as some generalizations and variations follow quickly from the butterfly decomposition. We next obtain two involutions on free Dyck paths and free Schröder paths, leading to parity results and combinatorial identities. We also use the butterfly decomposition to give a combinatorial treatment of Klazar's generating function for the number of chains in plane trees. Finally we study the size of chains in plane trees with n edges and show that the average size of such chains tends asymptotically to $\frac{n+9}{6}$.

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

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

Suggested Running Title: The Butterfly Decomposition

Corresponding Author: Louis W. Shapiro, e-mail: lshapiro@howard.edu

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 and approaches still emerge. 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 giving a combinatorial link 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 the x-axis. (The Chung-Feller theorem was first proved by Major Percy A. MacMahon in 1909 [16, p.168] but named after its 1949 re-discoverers [4].) The previous combinatorial approaches to the Chung-Feller theorem are based on the cycle lemma or cyclic paths, see Raney [19], Narayana [18] and Dershowitz-Zaks [6]. There are other Chung-Feller type results and generalizations in [2, 3, 10, 11, 23, 26].

In the settings of free Dyck paths and free Schröder paths, we obtain two involutions which lead to combinatorial identities as well as a parity result on plane trees concerning the number of leaves at odd height vs the number of leaves at even height.

The butterfly decomposition also leads to the following results: a correspondence between leaf-colored doubly rooted plane trees and free Schröder paths, and a combinatorial interpretation of the generating function for the number of chains in plane trees obtained by Klazar [14]. We set up a one-to-one correspondence between chains in plane trees and tricolored plane trees.

In the last section of this paper we find the generating function for the size of chains in plane trees with n edges and this gives us the asymptotic value of $\frac{n+9}{6}$ for the size of an average chain.

We use the standard notations where

$$C = C(x) = \frac{1-\sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty} \frac{1}{n+1}\binom{2n}{n} x^n = 1 + xC^2 = \frac{1}{1-xC}$$

is the generating function for the Catalan numbers c_n [25, Exe. 6.19] and

$$B = B(x) = \frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty} \binom{2n}{n} x^n = 1 + 2xBC = \frac{1}{1-2xC}$$

is the generating function for the central binomial coefficients.

Recall that a *Dyck path* of length $2n$ is a lattice path from the origin to $(2n, 0)$ consisting of up steps $U = (1, 1)$ and down steps $D = (1, -1)$ 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 final 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 [11]. The reflection of a Dyck path with respect to the x-axis is called a *negative Dyck path*. A *negative elevated (irreducible) Dyck path* is defined in the same manner. 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 $\binom{2n}{n}$.

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. The main result of this section is a correspondence between doubly rooted plane trees and free Dyck paths, from which 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. What is more, the distinguished vertex is allowed to coincide with the 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, v_1, 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.

It is clear that the generating function for a single butterfly is xC^2, and the generating function for a sequence of k butterflies is $(xC^2)^k$. Define the *height* of a vertex as the number of edges on the unique path from the root to it. Note that the root is at height 0. Then the generating function for trees with a distinguished vertex at height

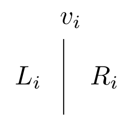

Figure 1: A butterfly

m is $(xC^2)^m C$, the m butterflies leading down to the distinguished vertex contribute $(xC^2)^m$ and the subtree rooted at the distinguished vertex contributes the factor C. 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:

$$B = C + C(xC^2) + C(xC^2)^2 + \cdots = \frac{C}{1 - xC^2}. \tag{2.2}$$

A natural question arises: is there a simple combinatorial argument that leads to this conclusion without resorting to the formula for the Catalan numbers? Several of the more obvious approaches seem to go nowhere but the butterfly decomposition sets up an easy bijection.

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 (or the worm crawling around the tree, see [25, pp. 33-34], [12, p.239], [15, pp. 21-27] and [5]) that takes plane trees to Dyck paths. The Dyck path corresponding to a plane tree is given by the sequence of steps when traversing the plane tree in preorder (visiting the root first, then traversing its subtrees from left to right). More precisely, we use U to denote the move from a vertex to a child and use D to denote the move from a vertex to its parent. This bijection is well known and Figure 2 illustrates it.

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 $v_1 v_2 \ldots v_k w$ 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.

We apply the glove bijection to each of the L_i $(1 \leq i \leq k)$ and T', and call the resulting Dyck paths P_i and P_{k+1}. For every R_i, we first create T_i by adding a new edge on the top of R_i, then use the glove bijection to produce an elevated Dyck path.

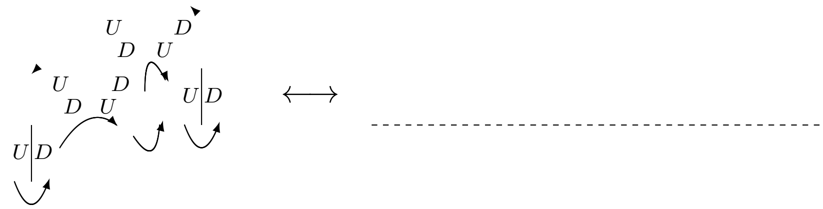

Figure 2: Glove bijection

We finish by reflecting this elevated path about the x-axis to give a negative elevated path Q_i. Now

$$P_1\, Q_1\, P_2\, Q_2\, \cdots\, P_k\, Q_k\, P_{k+1} \tag{2.3}$$

is a free Dyck path of length $2n$. Conversely, given a free Dyck path we may decompose it into segments where each return to the x-axis from below concludes a segment. 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 (2.3) 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 path allowed. Thus we have established the bijection. ∎

An example of the above bijection is shown in Figure 3.

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

$$\frac{C}{1-xC^2} = \frac{1}{1-2xC}, \tag{2.4}$$

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. If we want to think in terms of free Dyck paths we could call these two colors "up" and "down" instead. Note that here only the children of the root are colored. The relation (2.4) 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.

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 $2n$.*

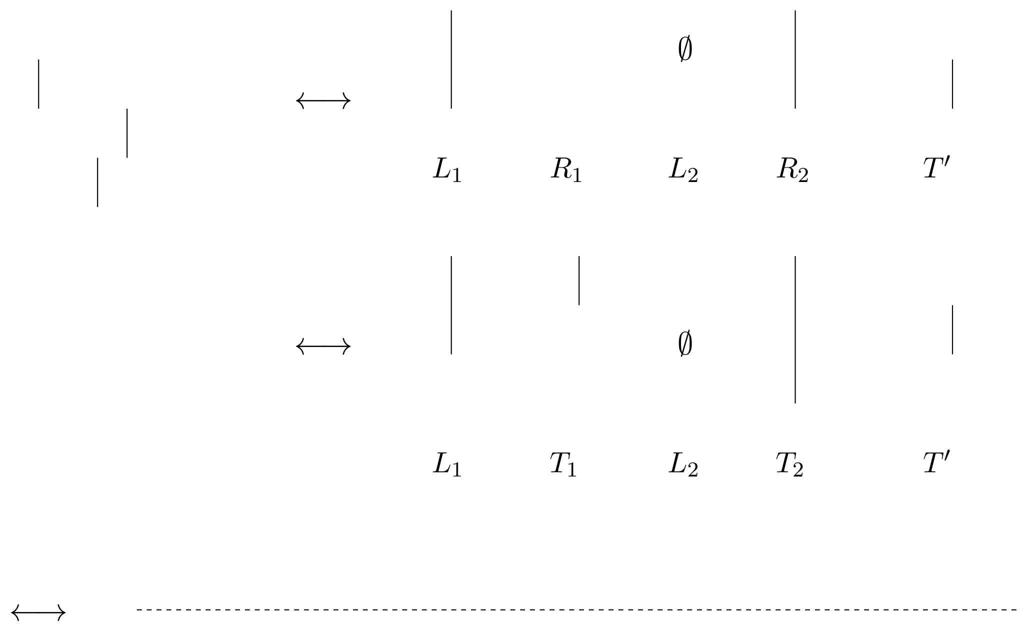

Figure 3: Doubly rooted plane trees and free Dyck paths

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 a 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 $P_1 P_2 \ldots P_k$. 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 direct bijection between doubly rooted plane trees and bicolored plane trees.

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. Let T be a doubly rooted plane tree. By the butterfly decomposition, we get subtrees L_i, R_i and T'. Then we create T_i by taking a new root with a single child that is the root of R_i. By coloring the children of the roots of L_i and T' white and coloring the child of the root of T_i black, and identifying their roots as the root of the corresponding bicolored plane tree, we have its subtrees listed from left to right as

$$(L_{1,1} \cdots L_{1,t_1})\, R_1\, (L_{2,1} \cdots L_{2,t_2})\, R_2\, \cdots\, (L_{k,1} \cdots L_{k,t_k})\, R_k\, (T'_1 \cdots T'_t),$$

for some $t_1, \ldots, t_k$ and t, where $L_{i,1}, \ldots, L_{i,t_i}$ are the subtrees of L_i and $T'_1, \ldots, T'_t$ are the subtrees of 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 [10] on Dyck paths and Schröder paths with flaws.

Theorem 3.1 (Chung-Feller) *For any $0 \le m \le n$, 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. Here we perform a *right-to-left preorder traversal* of a plane tree T on its vertices with numbers $0, 1, 2, \ldots, n$ in the order they are visited and the root getting label 0. Figure 4 gives the plane tree corresponding to the free Dyck path in Figure 3 and the labels show the right-to-left preorder traversal.

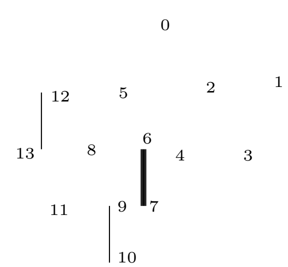

Figure 4: Labels for the Chung-Feller theorem

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.*

Theorem 3.2 holds by the bijection created in Theorem 2.1. By the butterfly decomposition, if w is a leaf, then the free Dyck path obtained under the bijection ends on an up step; and otherwise this path ends on a down step.

As a corollary, we note that half of all free Dyck paths end on a down step. Thus over all plane trees with n edges, half of the vertices are leaves, see Problem 10753 of the American Mathematical Monthly [13, 22] and Seo [21].

We can extend this as follows. Let a *crew cut vertex* be one which is at distance 1 from all the leaves that are below it. Consider the set $A_{n,m}$ of ordered pair (T, w), where T is a doubly rooted plane tree with n edges, w is the distinguished vertex of T, and moreover w is a crew cut vertex at height m. In the terminology of the butterfly decomposition, the generating function for $|A_{n,m}|$ is

$$\left(xC^2\right)^m \cdot \frac{x}{1-x},$$

where the factor $\frac{x}{1-x}$ is contributed by the terminal subtree T'. Summing $|A_{n,m}|$ over all m yields that the generating function for the number of crew cut vertices in all plane trees with n edges is

$$\frac{B+1}{2} \cdot \frac{x}{1-x}.$$

Continuing in this vein, we say that a vertex is *u-uniform* if all the leaves below it are at distance u. Thus leaves are 0-uniform and crew cut vertices are 1-uniform. The generating function for the number of u-uniform vertices in all plane trees with n edges is

$$\frac{B+1}{2} \cdot \frac{x^u}{1-x-x^2-x^3-\cdots-x^u},$$

where the factor $V_u(x) = \frac{x^u}{1-x-x^2-x^3-\cdots-x^u}$ comes from the terminal subtree T', which can be decomposed into a path of length u and an edge-disjoint union of paths of length less than or equal to u.

The following Figure 5 shows a plane tree with four crew cut vertices drawn with large solid dots and one 2-uniform vertex drawn with an open circle.

Figure 5: Crew cut vertices and a uniform vertex

The sequences of the numbers of 1-uniform vertices and 2-uniform vertices start with $0, 1, 2, 5, 15, 50, 176, 638, 2354, 8789, 33099, \cdots$ and $0, 0, 1, 2, 6, 18, 59,$ $203, 724, 2643, 9802, \cdots$ and are sequences A024718 and A121320 respectively in S-loane's EIS [24].

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

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

Let $A(x) = \frac{B+1}{2}$ and $B(x) = V_u(x)$. As u increases, the radius β of convergence of $B(x)$ decreases from 1 to $\frac{1}{2}$. Since the radius α of convergence of $A(x)$ is $\frac{1}{4}$, we can apply Bender's lemma and get that the n-th coefficient of $\frac{B+1}{2} \cdot V_u(x)$ equals $\frac{1}{2}\binom{2n}{n} \cdot V_u\left(\frac{1}{4}\right)$. It is a fact that the generating function for the total number of vertices in plane trees just equals B and the n-th coefficient is $\binom{2n}{n}$. Thus the proportion of u-uniform vertices approaches

$$\frac{1}{2} \cdot V_u\left(\frac{1}{4}\right) = \frac{3}{2} \cdot \frac{1}{2 \cdot 4^u + 1}, \tag{3.1}$$

albeit more slowly as u increases. As special cases, the average number of crew cut vertices approaches $\frac{1}{6}$, and the proportion of 2-uniform vertices approaches $\frac{1}{22}$ as the number of edges increases.

The above interpretation of the Chung-Feller theorem also implies some refinements and generalizations recently obtained by Eu, Fu and Yeh [10]. 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 [10]. 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 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 4 where the prefix edges are drawn with thick lines.

Using the preorder traversal of plane trees, we get the following generalization of the refined version of the Chung-Feller theorem [10].

Theorem 3.3 *For $0 \leq k \leq m \leq n$, there is a bijection between the set of free Dyck paths of length $2n$ with m flaws in k flaw blocks and the set of doubly rooted plane trees of n edges with stem length 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 length 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 length 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 [25, Sec. 5.4], we have

$$[x^n]C^k = \frac{k}{2n+k}\binom{2n+k}{n}. \tag{3.2}$$

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

$$\frac{k}{2n-k}\binom{2n-k}{n}, \tag{3.3}$$

see Engelberg [9, Cor. 3.2], Mohanty [17, Cor. 1 (iv)] and Deutsch [7, Sec. 6.6].

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 contains a leaf as an end vertex; 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.

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 goes below the x-axis. These paths are enumerated by the Schröder numbers r_n [25, Exe. 6.39]. An *elevated (irreducible) Schröder path*, a *free Schröder path* and a *negative Schröder path* are defined in the same manner as Dyck paths. 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.

There is a simple bijection that transforms Schröder paths into Dyck paths with bicolored peaks. Just change each H step into a red UD peak.

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 $2n$ by 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 a 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 a H step. It is easy to see that we obtain a Schröder path of length $2n$ and the above procedure is reversible. ∎

By 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 k leaves is given by the Narayana number (see [20] and [25, Exe. 6.36])

$$T_{n,k}=\frac{1}{n}\binom{n}{k}\binom{n}{k-1}.$$

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

$$\sum_{k=1}^{n}\left[(n+1-k)2^kT_{n,k}+k2^{k-1}T_{n,k}\right]=\sum_{k=1}^{n}(2n+2-k)2^{k-1}T_{n,k}.$$

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

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

Hence Theorem 3.6 yields the following identity:

$$\sum_{k=1}^{n} (2n+2-k) 2^{k-1} T_{n,k} = \sum_{k=0}^{n} \binom{2n-k}{k} \binom{2n-2k}{n-k}. \tag{3.4}$$

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 [10].

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 to get a plane tree with each leaf being colored red or blue, which is enumerated by the Schröder number r_n followed by Theorem 3.5. 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 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 a vertex labelled by m. Let T_w 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_w corresponds to a free Schröder path with m flaws.*

From the above theorem, we immediately get the following refinement.

Theorem 3.9 *For $0 \leq k \leq m \leq n$, there is a bijection between the set of free Schröder paths 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 length k and m prefix edges.*

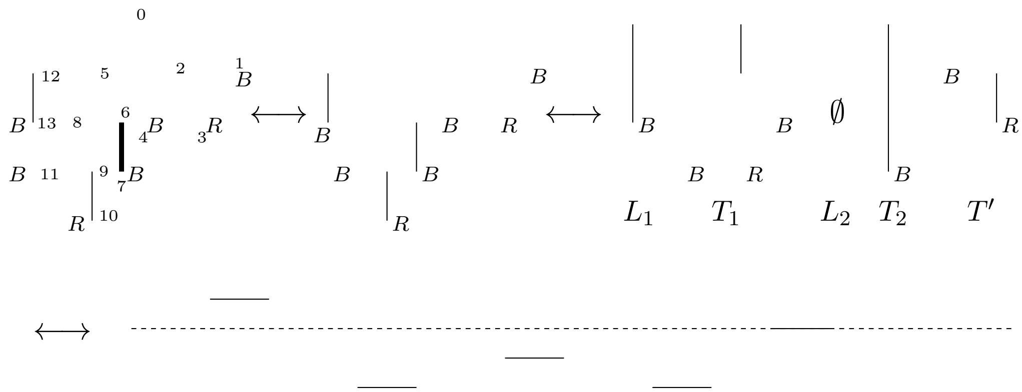

Figure 6: Leaf-colored plane trees and free Schröder paths

An example of the above bijection between leaf-colored doubly rooted plane trees and free Schröder paths is illustrated in Figure 6.

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 length 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 length 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

$$a(n,k) = [x^n]S^k. \tag{3.5}$$

Set $a(0,k) = 1$. When $n \geq 1$, using the Lagrange inversion formula [25, Sec. 5.4], we obtain that

$$a(n,k) = \frac{k}{n}\sum_{i=0}^{n-1} 2^{i+1}\binom{n+k-1}{i}\binom{n}{i+1}. \tag{3.6}$$

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:*

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

Proof. Let P be a free Dyck path of length $2n$. We construct ϕ by reflecting the last elevated subpath, be it positive or negative, about the x-axis. Clearly, ϕ is a parity reversing involution. By the butterfly decomposition, the number of free Dyck paths with k flaw blocks equals the number of doubly rooted plane trees with stem length k, that is, $[x^{n-k}]C^{2k+1}$. Hence the relation (4.1) follows from (3.2).

∎

If we consider the set of all plane trees with n edges where $n \geq 1$, we will get the following refinement of the fact that half of the vertices are leaves [13, 21, 22].

Corollary 4.2 *Over all plane trees with n edges, the number of leaves at height $m+1$ equals the number of internal vertices at height m.*

Proof. The involution ϕ switches the last elevated subpath from positive to negative or from negative to positive. If positive we have by the butterfly decomposition a tree with a distinguished vertex, say at height m, and T' is nonempty. Then T' can be written as $(L_{m+1})U(R_{m+1})D$ where L_{m+1} and R_{m+1} are possibly trivial trees by the glove bijection. The involution keeps all the L_i and R_i with $1 \leq i \leq m$ unchanged but transforms T' into a left subtree L_{m+1} and a right subtree R_{m+1}, with the new leaf being the distinguished vertex at height $m+1$ between them. ∎

Since the involution ϕ interchanges the distinguished vertex at height $m+1$ and m, we get the following property.

Corollary 4.3 *Over all plane trees with n edges, the total number of vertices at even height equals the total number of vertices at odd height.*

If we consider leaves instead of vertices then there are more leaves at odd height than at even height. The more precise result is as follows.

Corollary 4.4 *Consider the set of all plane trees with n edges. There are a Catalan number c_{n-1} more of leaves at odd height than at even height for $n \geq 1$. Dually there are c_{n-1} more of internal vertices at even height than at odd height.*

Proof. Let $P_o(n)$ ($P_e(n)$, respectively) denote the number of leaves at odd (even, respectively) height in all plane trees with n edges. We aim to establish a parity reversing involution Φ on the leaves in plane trees, where the parity of a leaf is defined as the parity of its height. Moreover, we define the sign of a leaf as 1 if it is at odd height, and as -1 if it is at even height.

If $n = 1$, it is easy to check that the corollary holds and Φ is trivial. We now assume that $n \geq 2$. Let r and v be the root and the leftmost child of the root in a given plane tree T respectively. Let T_v denote the subtree rooted at v. Let T_r denote the tree obtained from T by removing T_v and the edge rv. The desired involution $\Phi(T)$ is constructed by attaching T_r to v so that r is the first child of v. We use $\mathcal{A}_n$ to denote the set of plane trees with n edges such that both T_r and T_v are not trivial, and we use $\mathcal{B}_n$ ($\mathcal{C}_n$, respectively) to denote the set of plane trees with n edges such that T_r (T_v, respectively) is a single vertex. Then $\mathcal{A}_n \cup \mathcal{B}_n \cup \mathcal{C}_n$ is exactly the set of all plane trees with n edges. It is easy to see that Φ is a parity reversing involution on $\mathcal{A}_n$ and has no contribution to $P_o(n) - P_e(n)$. What is more, Φ is a parity reversing involution between $\mathcal{B}_n$ and $\mathcal{C}_n$ such that it contributes 1 to $P_o(n) - P_e(n)$ for each $T \in \mathcal{B}_n$, more precisely, this contribution 1 is for odd height 1. It is known that the number of planted plane trees with n edges, which is also $|\mathcal{B}_n|$, is c_{n-1}. Hence

$$P_o(n) - P_e(n) = c_{n-1},$$

and this completes the proof. ∎

Note that the involution Φ between $\mathcal{B}_n$ and $\mathcal{C}_n$ is ϕ by setting r and v as the distinguished vertices respectively by the butterfly decomposition. See [8] for similar results concerning vertices of even and odd degree.

We also have an involution on free Schröder paths.

Theorem 4.5 *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 (3.6):*

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

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 k flaw blocks equals $[x^{n-k}]S^{2k+1}$, that is, $a(n-k, 2k+1)$. 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 7, where those open circles stand for vertices in chains and solid dots stand for normal vertices. For instance, the last structure of Figure 7 has a chain of size 2.

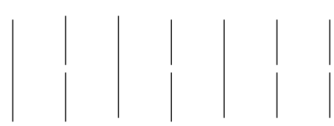

Figure 7: Chains in plane trees with 2 edges

The main result of this section is a combinatorial interpretation of the generating function for the number of chains in all plane trees obtained by Klazar [14]. 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 [14] derived the following generating function for the number of chains in plane trees with n edges:

$$\frac{C}{1-2xC^2} = \sum_{n=0}^{\infty}\sum_{k=0}^{n} 3^k \frac{k}{2n-k}\binom{2n-k}{n} x^n. \tag{5.1}$$

Note that here we use a slightly different formulation of the generating function C from that used by Klazar [14].

We now give a combinatorial proof of the fact that the generating function for the number of chains in all 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 $v_1v_2\ldots v_kw$ 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

$$\frac{C}{1-2xC^2} = \frac{1}{1-3xC}, \tag{5.2}$$

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 $v_1v_2\ldots v_kw$ 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 taking a new root with a single child that is the root of R_i. Color 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 from left to right as

$$(L_{1,1}\cdots L_{1,t_1})\, R_1\, (L_{2,1}\cdots L_{2,t_2})\, R_2\, \cdots\, (L_{k,1}\cdots L_{k,t_k})\, R_k\, (T'_1\cdots T'_t),$$

for some $t_1, \ldots, t_k$ and t, where $L_{i,1}, \ldots, L_{i,t_i}$ are the subtrees of L_i and $T'_1, \ldots, T'_t$ are the subtrees of T'. The reverse procedure is easy to construct. This completes the proof. ∎

An example of the above bijection is shown in Figure 8.

From the above bijection, we easily see that chains of size m 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 a plane tree is just a two colored path in the butterfly decomposition. Hence we can color the vertices in a chain with t colors and preserve

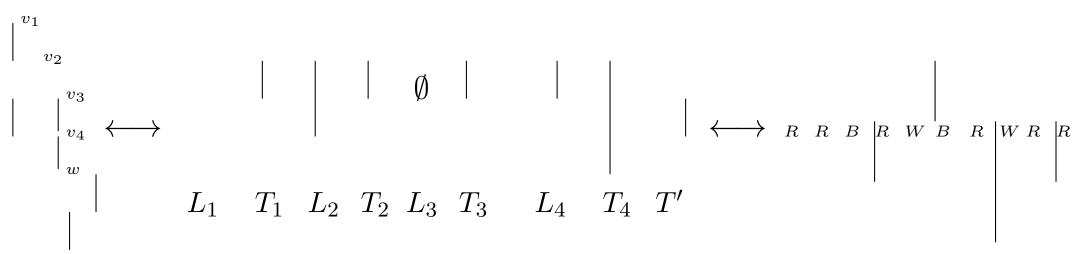

Figure 8: Chains and tricolored plane trees

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 butterfly decomposition, the generating function for the number of chains in plane trees with n edges that end with a leaf equals

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

It is a new combinatorial explanation for Sequence A067336 in [24].

6 Average Size of Chains

In this section, we use the generating function B of the central binomial coefficients 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 $\frac{n+9}{6}$.

Bear in mind that the generating function for the number of chains of size 1 in all plane trees with n edges equals B. We let L^* be the generating function for the number of plane trees with a distinguished leaf in all plane trees with n edges. 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 for the number 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 *For $1 \leq k \leq n+1$, let $r_{n,k}$ be the total number of chains of size k in all plane trees with n edges. Define $R(x, y) = \sum_{n\geq 0} \sum_{k=1}^{n+1} r_{n,k} y^k x^n$. Then*

$$R(x, y) = \frac{yB}{1 - \frac{y(B-1)}{2}}.$$

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 size k in 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 yL^*, 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 yL and the generating function for T' equals C. Hence the generating function for the total number of chains of size k in plane trees with n edges equals $yL^* \cdot (yL)^{k-1} \cdot C = yB \cdot \left(\frac{y(B-1)}{2}\right)^{k-1}$. We sum over k to get the required generating function. ∎

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, *A*008549, also counts [24]:

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

Taking the partial derivative by y of $R(x, y)$ and evaluating it at $(x, 1)$ gives the following generating function.

Theorem 6.2 *The generating function for the total size of all chains in plane trees with n edges equals $R_y(x,1) = \frac{4B}{(3-B)^2}$.*

Now we consider the asymptotic approximations. Let H_n be the total number of chains in plane trees with n edges. Klazar [14] has shown that

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

After some algebraic calculations we get

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

Here we recall the Bender's lemma [1, p.496] again as we used in Section 3. 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 $A_n = \frac{n+9}{2}\left(\frac{9}{2}\right)^{n-1}$. Thus we obtain the following asymptotic property.

Theorem 6.3 *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

$$\frac{R_n}{H_n} \sim \frac{n+9}{6}.$$

For example,

$$\frac{R_{50}}{H_{50}} = \frac{2\,250\,588\,247\,788\,344\,466\,951\,528\,963\,319\,620}{228\,878\,511\,199\,384\,804\,987\,952\,173\,176\,432} \approx 9.833\,1,$$

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

Acknowledgments. The authors thank David S. Hough, Martin Klazar, Christian Krattenthaler, Xavier Viennot and Peter Winkler for helpful suggestions. This work was supported by the 973 Project on Mathematical Mechanization, the National Science Foundation, the PCSIRT Project of 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, *Congr. Numer.*, **147** (2000), 129-143.

[4] K.L. Chung and W. 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, *European 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 survey of the Fine numbers, *Discrete Math.*, **241** (2001), 241-265.

[9] O. Engelberg, On some problems concerning a restricted random walk, *J. Appl. Probab.*, **2** (1965), 396-404.

[10] 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.

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

[12] M. Gardner, *Time Travel and Other Mathematical Bewilderments*, W.H. Freeman and Co., 1988.

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

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

[15] J.H. van Lint, Combinatorial Theory Seminar, Eindhoven University of Technology, *Lecture Notes in Mathematics, Vol. 382*, Springer, Berlin, 1974.

[16] 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.

[17] S.G. Mohanty, On some generalization of a restricted random walk, *Studia Sci. Math. Hungar.*, **3** (1968), 225-241.

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

[19] G.N. Raney, Functional composition patterns and power series reversion, *Trans. Amer. Math. Soc.*, **94** (1960), 441-451.

[20] J. Riordan, Enumeration of plane trees by branches and endpoints, *J. Combin. Theory Ser. A*, **19** (1975), 214-222.

[21] S. Seo, A pairing of the vertices of ordered trees, *Discrete Math.*, **241** (2001), 471-477.

[22] L.W. Shapiro, Problem 10753, *Amer. Math. Monthly*, **106** (1999), 777.

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

[24] N.J.A. Sloane, The On-Line Encyclopedia of Integer Sequences, *http://www.research.att.com/˜njas/sequences.*

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

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