#A35 INTEGERS 22 (2022)

A BIJECTION BETWEEN SYMMETRIC VALLEYS IN DYCK PATHS AND EVEN DEGREE INTERNAL VERTICES IN ORDERED TREES

Farzan Byramji
Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India
farzan@iitk.ac.in

Received: 10/16/21, Accepted: 3/2/22, Published: 3/14/22

Abstract

Symmetric valleys in Dyck paths were recently studied by Elizalde, who proved using generating functions that the total number of symmetric valleys in all Dyck paths of semilength n equals the total number of internal nodes of even outdegree in all rooted ordered trees with n edges. In this note, we give a bijection exhibiting this. This bijection has nice properties allowing us to refine the counting with respect to other statistics such as height of the symmetric valley and depth of the even degree vertex in an ordered tree.

1. Introduction

We start with the relevant definitions, following Elizalde [1] for the most part. A *Dyck path* of semilength n is a lattice path with steps $\mathtt{u} = (1, 1)$ and $\mathtt{d} = (1, -1)$ that starts at $(0, 0)$, ends at $(2n, 0)$, and never goes below the x-axis. Let $\mathcal{D}_n$ be the set of Dyck paths of semilength n, and let $\mathcal{D} = \bigcup_{n\geq 0} \mathcal{D}_n$ be the set of all Dyck paths. An *ordered tree* (or *plane tree*) is a rooted tree for which the subtrees of every vertex are linearly ordered from left to right. Let $\mathcal{T}_n$ be the set of ordered trees with n edges, and let $\mathcal{T} = \bigcup_{n\geq 0} \mathcal{T}_n$. It is well-known that $|\mathcal{D}_n| = |\mathcal{T}_n|$. The standard bijection G from ordered trees to Dyck paths is as follows. G maps the tree with no edges to the empty path. If the root r of a tree T has $T_1, T_2, \ldots, T_m$ as its subtrees, $G(T) = \mathtt{u}G(T_1)\mathtt{du}G(T_2)\mathtt{d}\ldots\mathtt{u}G(T_m)\mathtt{d}$. This is the same as traversing the tree in preorder, writing $\mathtt{u}$ when going down an edge and writing $\mathtt{d}$ when going up an edge.

A *valley* in a Dyck path is an occurrence of $\mathtt{du}$. We say that a valley $\mathtt{du}$ is *symmetric* if the maximal subsequence of the form $\mathtt{d}^i\mathtt{u}^j$ that contains it satisfies $i = j$; see Figure 1 (from [1]) for an example. For a path $D \in \mathcal{D}$, we denote its number of symmetric valleys by $\mathrm{sval}(D)$. The *weight* of a symmetric valley is the

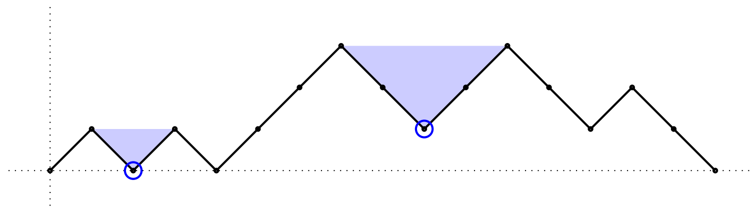


Figure 1: A Dyck path with two symmetric valleys, circled in blue, having weights 1 and 2.

largest i such that the valley is contained in a subsequence of the form $\mathtt{d}^i\mathtt{u}^i$. Let $\mathcal{D}'_n$ be the set of all Dyck paths of semilength n with a marked symmetric valley and $\mathcal{D}' = \bigcup_{n\geq 0} \mathcal{D}'_n$. Then $|\mathcal{D}'_n| = \sum_{D\in\mathcal{D}_n} \mathrm{sval}(D)$. In what follows, by a marked path we refer to such an element of $\mathcal{D}'$. The *height* of a valley is defined to be the y-coordinate of its lowest vertex. In Figure 1, the symmetric valley of weight 1 has height 0 while the other symmetric valley has height 1.

The *degree* of a vertex v (denoted $\deg(v)$) in an ordered tree is the number of children of that vertex v. A vertex is *internal* if it has non-zero degree. For a tree $T \in \mathcal{T}$, denote by $\mathrm{eval}(T)$ the number of internal nodes of T with even degree. Let $\mathcal{T}'_n$ be the set of all ordered trees with a marked internal node of even degree and let $\mathcal{T}' = \bigcup_{n\geq 0} \mathcal{T}'_n$. Then $|\mathcal{T}'_n| = \sum_{T\in\mathcal{T}_n} \mathrm{eval}(T)$. The *depth* of a vertex is the length of the path from the root to that vertex. In particular, the root has depth 0.

Elizalde [1] found that $|\mathcal{D}'_n| = |\mathcal{T}'_n|$ using the corresponding generating functions and asked for a combinatorial proof. Possibly this was known even earlier according to the comments for this sequence A014301 in the OEIS. We prove this here by giving a bijection F from $\mathcal{T}'$ to $\mathcal{D}'$ which is size-respecting. Moreover the bijection gives the following refinement. Let $\mathcal{D}'_{n,h}$ be the set of all Dyck paths of semilength n with a marked symmetric valley of height h. Let $\mathcal{T}'_{n,h}$ be the set of all ordered trees with a marked internal node of even degree that has depth h. Then F is a bijection from $\mathcal{T}'_{n,h}$ to $\mathcal{D}'_{n,h}$ which leads to the following theorem.

Theorem 1. *The number of symmetric valleys at height h in all Dyck paths in $\mathcal{D}_n$ is equal to the number of even degree internal nodes at depth h in all ordered trees in $\mathcal{T}_n$.*

More statistics are preserved by F but other than the weight of a valley, they are perhaps not as natural as those considered above, so we describe them in the last section.

Algorithm 1 The bijection F from $\mathcal{T}'$ to $\mathcal{D}'$

T is an ordered tree with some even degree internal node marked.
Let $c_1, c_2, \ldots, c_k$ be the children of the root and $T_1, T_2, \ldots, T_k$ the corresponding subtrees.
if the root is the marked node **then**
  $j \leftarrow k/2$ ▷ k is even since the root is marked
  $P_l \leftarrow$ the path from the root going to c_j and then repeatedly following the rightmost child
  $P_r \leftarrow$ the path from the root going to c_{j+1} and then repeatedly following the leftmost child
  $s \leftarrow \min(|P_l|, |P_r|)$
  $v_l \leftarrow$ the node on P_l at depth s
  $v_r \leftarrow$ the node on P_r at depth s
  $L_T \leftarrow$ the tree rooted at v_l in T
  $R_T \leftarrow$ the tree rooted at v_r in T
  $T'_j \leftarrow T_j$ with the subtree L_T (at v_l) replaced by a single node
  $T'_{j+1} \leftarrow T_{j+1}$ with the subtree R_T (at v_r) replaced by a single node
  $T' \leftarrow$ the tree with subtrees $T_1, T_2, \ldots, T_{j-1}, T'_j, T'_{j+1}, T_{j+2}, \ldots, T_k$
  $D' \leftarrow G(T')$ where the valley between the Dyck paths corresponding to the modified subtrees T'_j and T'_{j+1} is marked
  $D \leftarrow G(L_T)D'G(R_T)$
  return D
else
  Let m be such that T_m contains the marked node.
  $D \leftarrow \mathtt{u}G(T_1)\mathtt{du}G(T_2)\mathtt{d}\ldots\mathtt{u}G(T_{m-1})\mathtt{du}F(T_m)\mathtt{du}G(T_{m+1})\mathtt{d}\ldots\mathtt{u}G(T_{k-1})\mathtt{du}G(T_k)\mathtt{d}$
  return D
end if

2. The Bijection

The bijection is described in Algorithm 1. Some clarifying remarks are in order. We use $|P_l|$ to denote the number of edges in P_l. Equivalently $|P_l|$ is the depth (in T) of the rightmost leaf of T_j. Similarly $|P_r|$ is the depth (in T) of the leftmost leaf of T_{j+1}. By the choice of s, v_l and v_r must exist and at least one of v_l and v_r is a leaf. The tree T'_j is obtained from T_j by removing the children of v_l and similarly for T'_{j+1}. If both v_l and v_r are leaves, then T' is simply T. We shall continue to use the notation introduced in the algorithm in what follows.

An example is given in Figure 2. We explain next how F is applied to the ordered tree T in the figure. The root has three children of which the second is marked. Let T_1, T_2 and T_3 be the corresponding subtrees as in Algorithm 1. So T_1 is an ordered tree where the root has two leaves as its children while in T_3, the root has just one leaf as its child. Then $G(T_1)$ is `udud` and $G(T_2)$ is the empty path. So the

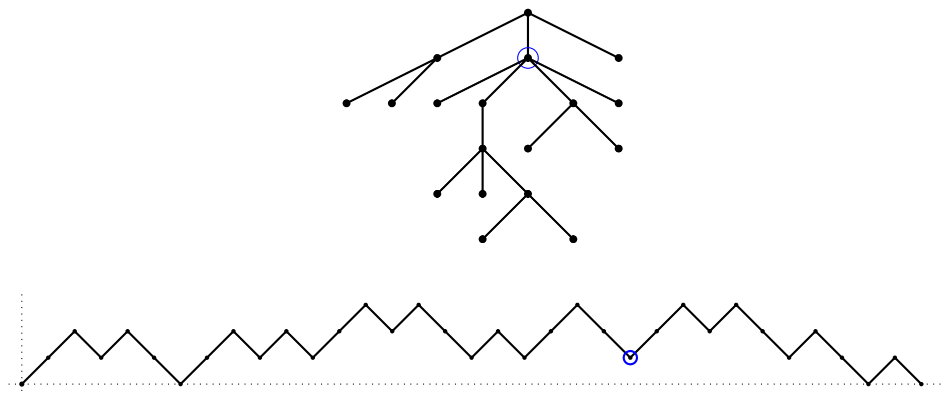


Figure 2: A marked ordered tree and the corresponding marked Dyck path for $n = 17$

marked Dyck path corresponding to T would finally be uududdu$F(T_2)$dud once we have $F(T_2)$.

The trees involved in the computation of $F(T_2)$ are shown in Figure 3. In the figure and what follows, T_2 will be denoted by T to be consistent with the notation in the algorithm. For this tree, $j = 4/2 = 2$ and $s = \min(4, 2) = 2$. The vertex circled in olive is v_l and the one in red is v_r. Now $G(L_T) =$ ududuududd, $G(T') =$ uduudduududdud and $G(R_T)$ is the empty path. So D' is $\texttt{uduud}\underline{\texttt{du}}\texttt{ududdud}$ where the underlined valley is marked. Finally D is $\texttt{ududuududduduud}\underline{\texttt{du}}\texttt{ududdud}$. Coming back to the tree of Figure 2, putting this marked Dyck path D for $F(T_2)$ in uududdu$F(T_2)$dud gives the path shown in Figure 2.

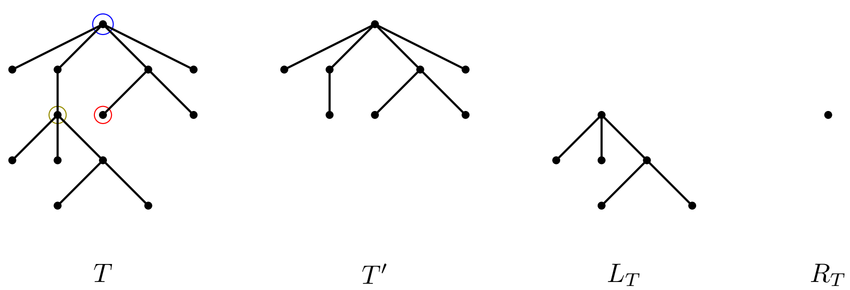


Figure 3: Trees required for obtaining $F(T_2)$

Table 1 gives the correspondence between all marked Dyck paths of semilength 4 and marked ordered trees with 4 edges.

Lemma 1. *For any marked ordered tree T with n edges, $F(T)$ is a Dyck path of semilength n with a marked symmetric valley. Moreover the height of the marked*

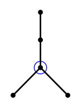
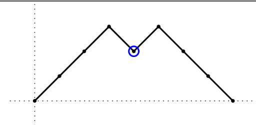
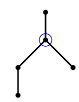
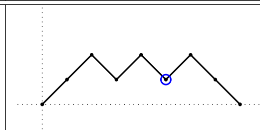
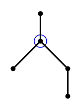
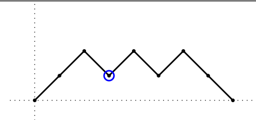
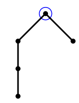
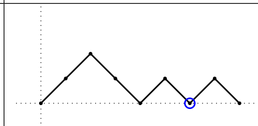
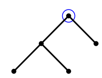
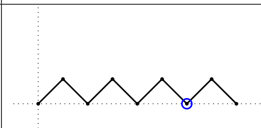
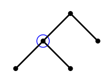
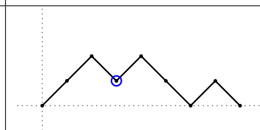
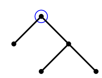
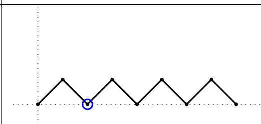
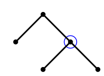
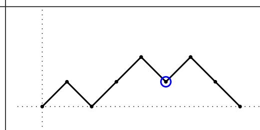
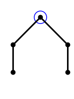
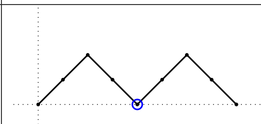
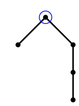
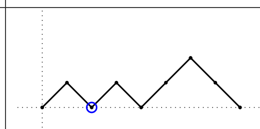
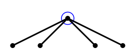
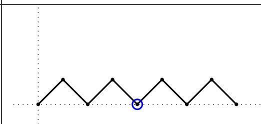
Marked ordered tree	Marked Dyck path	Marked ordered tree	Marked Dyck path
			
			
			
			
			
			

Table 1: All 11 marked ordered trees and corresponding marked Dyck paths for $n = 4$

valley in $F(T)$ is the same as the depth of the marked node in T.

Proof. We prove this by induction on the depth of the marked node.

If the root is marked (depth of the marked node is 0), then the algorithm first obtains T', L_T and R_T. The total number of edges among T', L_T and R_T is n. So D has semilength n. The marked valley between the paths corresponding to T'_j and T'_{j+1} is symmetric since the depth of the rightmost leaf of T'_j is the same as the depth of the leftmost leaf of T'_{j+1} by construction. It is also clear that the marked valley has height 0.

Now suppose the statement holds for all trees with marked node at depth d. Take any marked tree with the marked node at depth $d + 1$. The root of this tree is unmarked and T_m has the marked node for some m and the depth of the marked node in T_m is d. Let $n_1, n_2, \ldots, n_k$ be the number of edges in $T_1, T_2, \ldots, T_k$

respectively so that $\sum_{i=1}^{k} n_i + k = n$. By the induction hypothesis, we have that $F(T_m)$ is a marked Dyck path of semilength n_m and the marked valley has height d. Also $G(T_1), G(T_2), \ldots, G(T_k)$ are Dyck paths with semilengths $n_1, n_2, \ldots, n_k$ respectively. So D as defined in the algorithm is indeed a Dyck path of semilength n. Moreover the marked valley in D is symmetric since it was symmetric in $F(T_m)$ and it has height $d + 1$ as required. □

Algorithm 2 $H : \mathcal{D}' \to \mathcal{T}'$

D is a Dyck path with some symmetric valley marked
if the marked valley has height 0 **then**
  Let $\mathtt{u}D^l_p\mathtt{du}D^l_{p-1}\mathtt{d}\ldots\mathtt{u}D^l_1\mathtt{du}D^r_1\mathtt{du}D^r_2\mathtt{d}\ldots\mathtt{u}D^r_q\mathtt{d}$ be the unique decomposition of D where $D^l_p, D^l_{p-1}, \ldots, D^l_1, D^r_1, D^r_2, \ldots D^r_q \in \mathcal{D}$ and the valley between D^l_1 and D^r_1 is marked.
  $j \leftarrow \min(p, q)$
  $D' \leftarrow \mathtt{u}D^l_j\mathtt{du}D^l_{j-1}\mathtt{d}\ldots\mathtt{u}D^l_1\mathtt{du}D^r_1\mathtt{du}D^r_2\mathtt{d}\ldots\mathtt{u}D^r_j\mathtt{d}$
  $L_D \leftarrow \mathtt{u}D^l_p\mathtt{du}D^l_{p-1}\mathtt{d}\ldots\mathtt{u}D^l_{j+1}\mathtt{d}$
  $R_D \leftarrow \mathtt{u}D^r_{j+1}\mathtt{du}D^r_{j+2}\mathtt{d}\ldots\mathtt{u}D^l_q\mathtt{d}$
  $T' \leftarrow G^{-1}(D')$ and let $T'_1, T'_2, \ldots, T'_j, T'_{j+1}, \ldots, T'_{2j}$ be its subtrees
  $L_T \leftarrow G^{-1}(L_D)$
  $R_T \leftarrow G^{-1}(R_D)$
  $v_l \leftarrow$ the rightmost leaf of T'_j
  $v_r \leftarrow$ the leftmost leaf of T'_{j+1}
  $T \leftarrow T'$ with v_l replaced by L_T and v_r replaced by R_T
  return T with its root marked
else
  Let $\mathtt{u}D_1\mathtt{du}D_2\mathtt{d}\ldots\mathtt{u}D_k\mathtt{d}$ be the unique decomposition of D such that there is m $(1 \le m \le k)$ with $D_1, D_2, \ldots, D_{m-1}, D_{m+1}, \ldots, D_k \in \mathcal{D}$ and $D_m \in \mathcal{D}'$.
  Let T be the ordered tree with subtrees $G^{-1}(D_1), G^{-1}(D_2), \ldots, G^{-1}(D_{m-1}), H(D_m), G^{-1}(D_{m+1}), \ldots, G^{-1}(D_k)$.
  return T
end if

The inverse of F, call it H, is described in Algorithm 2. Here we note that at least one of p and q is equal to j, so at least one of the paths D_l and D_r must be empty.

Lemma 2. *For any marked Dyck path D of semilength n, $H(D)$ is an ordered tree with n edges and a marked internal node having even degree.*

Proof. The proof is by induction on the height of the marked valley in D.

Suppose the marked node has height 0. Let n', n_l and n_r be the semilengths of the paths D', L_D and R_D respectively, so that $n'+n_l+n_r = n$. Then $G^{-1}(D'), G^{-1}(L_D)$ and $G^{-1}(R_D)$ have n', n_l and n_r edges respectively. By construction, T then has

$n' + n_l + n_r = n$ edges. The root of T' has even degree $2j$ by the choice of D'. Since the degree of the root remains unaffected while transforming T' to T, the root of T also has even degree.

Now assume the statement holds for all marked paths with the marked valley having height d. Consider any marked path with the marked valley having height $d+1$. Then the path decomposes as $\mathtt{u}D_1\mathtt{du}D_2\mathtt{d}\ldots\mathtt{u}D_k\mathtt{d}$ such that there is m $(1 \le m \le k)$ with $D_1, D_2, \ldots, D_{m-1}, D_{m+1}, \ldots, D_k \in \mathcal{D}$ and $D_m \in \mathcal{D}'$. Let $n_1, n_2, \ldots, n_k$ be the semilengths of $D_1, D_2, \ldots, D_k$ respectively. Then $G^{-1}(D_1), G^{-1}(D_2), \ldots, G^{-1}(D_k)$ have $n_1, n_2, \ldots, n_k$ edges respectively. By induction, $H(D_m)$ has n_m edges and a marked internal node with even degree. So T has $\sum_{i=1}^{k} n_i + k = n$ edges and the marked node has even degree since it had even degree in $H(D_m)$. □

Now we show that H is indeed the inverse of F.

Lemma 3. *$H \circ F$ is the identity map on $\mathcal{T}'$.*

Proof. Again we use induction on the depth of the marked node. Let $T \in \mathcal{T}'$.

Suppose the marked node is the root. Consider the case $|P_l| \le |P_r|$ so that $s = |P_l|$. Then the tree L_T is trivial and so $G(L_T)$ is empty. The path D' is $\mathtt{u}G(T_1)\mathtt{du}G(T_2)\mathtt{d}\ldots\mathtt{u}G(T'_j)\mathtt{du}G(T'_{j+1})\mathtt{d}\ldots G(T_{2j})$ and the marked valley is between $G(T'_j)$ and $G(T'_{j+1})$. So when H is applied to $F(T)$, $p \le q$ since $G(L_T)$ is empty. Thus j is set to p, which is precisely the j of Algorithm 1. The paths D', $G(L_T)$ and $G(R_T)$ are recovered in this way, and the corresponding trees are also obtained by applying G^{-1}. Now v_r in T' is replaced by R_T to get T. Indeed this is the reverse of how T' was created from T in Algorithm 1. Therefore we have $(H \circ F)(T) = T$ in this case. The other case $|P_l| > |P_r|$ is similar and so we have proved the base case of the induction.

Suppose the statement is true for all trees with the marked node at depth d. Let $T \in \mathcal{T}'$ have the marked node at depth $d+1$. Let $T_1, T_2, \ldots, T_k$ be the subtrees and T_m contain the marked node. Then $F(T)$ is $\mathtt{u}G(T_1)\mathtt{du}G(T_2)\mathtt{d}\ldots\mathtt{u}G(T_{m-1})\mathtt{du}F(T_m)\mathtt{d}$ $\mathtt{u}G(T_{m+1})\mathtt{d}\ldots\mathtt{u}G(T_{k-1})\mathtt{du}G(T_k)\mathtt{d}$. Applying H to this marked path gives a tree with subtrees $G^{-1}(G(T_1)), G^{-1}(G(T_2)), \ldots, G^{-1}(G(T_{m-1})), H(F(T_m))$, $G^{-1}(G(T_{m+1})), \ldots, G^{-1}(G(T_k))$. By induction, we have $H(F(T_m)) = T_m$. Therefore $H(F(T))$ is the tree with subtrees $T_1, T_2, \ldots, T_m, \ldots, T_k$ and this tree is just T. □

Lemma 4. *$F \circ H$ is the identity map on $\mathcal{D}'$.*

The proof is similar to that of the previous lemma, so we skip it.

Combining these lemmas we obtain our main theorem.

Theorem. *The number of symmetric valleys at height h in all Dyck paths in $\mathcal{D}_n$ is equal to the number of even degree internal nodes at depth h in all ordered trees in $\mathcal{T}_n$.*

3. Some Statistics on Marked Dyck Paths and Marked Ordered Trees

In this section, we define some more statistics on marked Dyck paths and marked order trees, which behave nicely under the bijection F.

Recall that the weight (denoted wt) of a symmetric valley is the largest i such that the valley is contained in a subsequence of the form $\mathtt{d}^i\mathtt{u}^i$. The corresponding statistic s for an internal node v of even degree, say $2j$, is defined as follows. If the subtrees attached to v are $T_1, T_2, \ldots, T_j, T_{j+1}, \ldots, T_{2j}$, let v_l be the rightmost leaf of T_j and v_r be the leftmost leaf of T_{j+1}. Then s is defined to be the smaller of the distance from v to v_l and the distance from v to v_r. Note that this coincides with the definition of s in Algorithm 1. For the tree T in Figure 2, $\mathrm{s}(T) = 2$. Also for the corresponding Dyck path D in Figure 2, we have $\mathrm{wt}(D) = 2$.

Proposition 1. *For any* $T \in \mathcal{T}'$, $\mathrm{s}(T) = \mathrm{wt}(F(T))$.

Proof. In the case where the root is marked, this follows from the observation that the rightmost leaf of T'_j and the leftmost leaf of T'_{j+1} are at the same distance s from the root of T'. In the other case, use the fact that the weight of a marked Dyck path $\mathtt{u}D_1\mathtt{du}D_2\mathtt{d}\ldots\mathtt{u}D_m\mathtt{d}\ldots\mathtt{u}D_k\mathtt{d}$ (where D_m is a marked Dyck path and all $D_1, D_2, \ldots, D_{m-1}, D_{m+1}, \ldots, D_k$ are Dyck paths) is the same as the weight of D_m. Similarly the statistic s of a marked ordered tree T which has a marked tree T_i as a subtree is the same as s of T_i. □

The *difference* (denoted diffp) of a marked Dyck path is defined recursively. If the marked valley has height 0, then the difference is $p - q$ where p and q are as described in Algorithm 2. For a marked Dyck path $D = \mathtt{u}D_1\mathtt{du}D_2\mathtt{d}\ldots\mathtt{u}D_m\mathtt{d}\ldots\mathtt{u}D_k\mathtt{d}$, $\mathrm{diffp}(D) = \mathrm{diffp}(D_m)$. The marked Dyck path in Figure 2 has difference $3 - 0 = 3$. The *difference* (denoted difft) of a marked ordered tree is $\deg(v_l) - \deg(v_r)$, where v_l and v_r are defined above. Recall that at most one of $\deg(v_l)$ and $\deg(v_r)$ can be positive, so we can recover both $\deg(v_l)$ and $\deg(v_r)$ from this single statistic. The path in Figure 2 has difference $3 - 0 = 3$. Like the previous proposition, we have an easy correspondence between these statistics.

Proposition 2. *For any* $T \in \mathcal{T}'$, $\mathrm{difft}(T) = \mathrm{diffp}(F(T))$.

It would be interesting to know how many marked Dyck paths of semilength n have difference 0. Equivalently this is the number of marked Dyck paths on which the action of H is essentially the same as G^{-1}, the standard bijection from Dyck paths to ordered trees.

The *spread* (sp) of a marked Dyck path is also defined recursively. For a marked valley at height 0, it is defined as $\min(p, q)$. For a marked Dyck path $D = \mathtt{u}D_1\mathtt{du}D_2\mathtt{d}\ldots\mathtt{u}D_m\mathtt{d}\ldots\mathtt{u}D_k\mathtt{d}$, $\mathrm{sp}(D) = \mathrm{sp}(D_m)$. The path in Figure 2 has spread $\min(5, 2) = 2$. The *semiwidth* sw of a marked ordered tree is $\deg(v)/2$ by where v is the marked node. (This was denoted by j in the algorithms.) The tree in Figure 2 has semiwidth $4/2 = 2$.

Proposition 3. *For any* $T \in \mathcal{T}'$, $\mathrm{sw}(T) = \mathrm{sp}(F(T))$.

These propositions immediately imply further refinements of the main theorem.

Finally we note that the statistics difference and spread for valleys in Dyck paths make sense even when the valleys are allowed to be asymmetric. Enumerating all valleys in Dyck paths with respect to these statistics could be interesting.

Acknowledgement. The author thanks the anonymous referee for suggestions that helped improve this paper.

References

[1] S. Elizalde, Symmetric peaks and symmetric valleys in Dyck paths, *Discrete Math.* **344** (2021), 112364, 13pp.