

General results on the enumeration of strings in Dyck paths

K. Manes, A. Sapounakis, I. Tasoulas and P. Tsikouras

Department of Informatics
University of Piraeus, Piraeus, Greece

{kmanes, arissap, jtas, pgtsik}@unipi.gr

Submitted: Dec 18, 2010; Accepted: Mar 23, 2011; Published: Mar 2011

Mathematics Subject Classifications: 05A15, 05A19

Abstract

Let τ be a fixed lattice path (called in this context string) on the integer plane, consisting of two kinds of steps. The Dyck path statistic “number of occurrences of τ ” has been studied by many authors, for particular strings only. In this paper, arbitrary strings are considered. The associated generating function is evaluated when τ is a Dyck prefix (or a Dyck suffix). Furthermore, the case when τ is neither a Dyck prefix nor a Dyck suffix is considered, giving some partial results. Finally, the statistic “number of occurrences of τ at height at least j ” is considered, evaluating the corresponding generating function when τ is either a Dyck prefix or a Dyck suffix.

1 Introduction

Throughout this paper, a *path* is considered to be a lattice path on the integer plane, consisting of steps $u = (1, 1)$ (called *rises*) and $d = (1, -1)$ (called *falls*). Since the sequence of steps of a path is encoded by a word in $\{u, d\}^*$, we will make no distinction between these two notions. The *length* $|\alpha|$ of a path α is the number of its steps. The *height of a point* of a path is its y -coordinate.

A *Dyck path* is a path that starts and ends at the same height and lies weakly above this height. It is convenient to consider that the starting point of a Dyck path is the origin of a pair of axes; (see Fig. 1).

The set of Dyck paths of semilength n is denoted by \mathcal{D}_n , and we set $\mathcal{D} = \bigcup_{n \geq 0} \mathcal{D}_n$, where $\mathcal{D}_0 = \{\varepsilon\}$ and ε is the empty path. It is well known that $|\mathcal{D}_n| = C_n$, where $C_n = \frac{1}{n+1} \binom{2n}{n}$ is the n -th Catalan number; (see sequence A000108 in [23]).

Every non-empty Dyck path α can be uniquely decomposed in the form $\alpha = u\beta d\gamma$, where $\beta, \gamma \in \mathcal{D}$. This is the so called *first return decomposition*. If $\gamma = \varepsilon$, then α is a *prime Dyck path*.

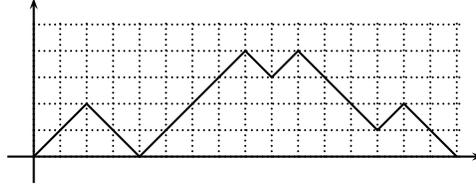


Figure 1: The Dyck path uudduuuudddudd.

A path which is a prefix (resp. a suffix) of a Dyck path, is called *Dyck prefix* (resp. *Dyck suffix*). For example, the path uudduu (resp. udddudd) consisting of the first six (resp. last seven) steps of the Dyck path of Fig. 1 is a Dyck prefix (resp. Dyck suffix). In the literature, Dyck prefixes are also called ballot paths.

We define the *depth* (resp. *height*) of a path α to be the difference between the height of the first (resp. last) point and the height of a lowest point of α . A path having depth δ and height h is referred as a (δ, h) -path. For example, the path udduuuud which lies between the second and the tenth point of the Dyck path of Fig. 1 is a $(1, 3)$ -path. Clearly, every Dyck prefix (resp. Dyck suffix) is a $(0, h)$ -path (resp. $(\delta, 0)$ -path), whereas a Dyck path is a $(0, 0)$ -path.

Every (δ, h) -path α , with $\delta, h > 0$, can be uniquely decomposed in the form $\alpha = \alpha_1\alpha_2$, where α_1 is a prime Dyck suffix (i.e., a suffix of a prime Dyck path) of depth δ and α_2 is a Dyck prefix of height h ; (see Fig. 2, where the semicircles represent Dyck paths). We call this the *leftmost lowest point decomposition* of α .

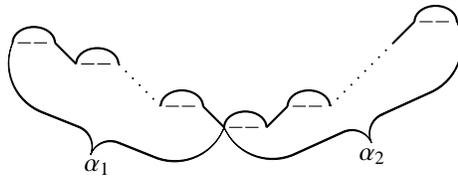


Figure 2: The leftmost lowest point decomposition of $\alpha = \alpha_1\alpha_2$.

A path $\tau \in \{u, d\}^*$, called in this context *string*, *occurs* in a path α if $\alpha = \beta\tau\gamma$, for some $\beta, \gamma \in \{u, d\}^*$. The number of occurrences of the string τ in α , is denoted by $|\alpha|_\tau$.

For the study of the Dyck paths statistic N_τ : “number of occurrences of τ ”, (with respect to the semilength) we consider the bivariate generating function

$$F = F(x, y) = \sum_{\alpha \in \mathcal{D}} x^{|\alpha|_u} y^{|\alpha|_\tau}.$$

We will also need the generating function A_p (resp. B_s) of the set of all Dyck paths having prefix p (resp. suffix s), as well as the generating function $\Gamma_{p,s}$ of the set of all Dyck paths having prefix p and suffix s at the same time. We denote, for simplicity, the generating functions A_{u^j} , B_{d^i} and Γ_{u^j, d^i} by A_j , B_i and $\Gamma_{j,i}$ respectively.

Given a string τ , the symmetric string of τ with respect to a vertical axis is called the *mirror string* of τ and it is denoted by $\bar{\tau}$. Clearly, the statistics N_τ and $N_{\bar{\tau}}$ are equidistributed.

Many articles dealing with the occurrence of strings in Dyck paths have appeared in the literature (e.g. see [1, 3, 5, 8, 12, 13, 14, 19, 20, 21, 24]). In particular, it has been proved (see [8]) that the statistic N_τ follows the Narayana distribution (A001263 of [23]), for every string τ of length 2, the statistic N_{udu} follows the Donaghey distribution (see [24]) and the statistic N_{duu} follows the Touchard distribution (see [8]). A systematic study of all strings with length up to 4 has been presented in [19], whereas some strings of arbitrary length have been studied in [13, 14]. Strings in k -colored Motzkin paths have been studied in [22], whereas strings in ballot paths have been studied in [15, 16].

So far, all results that appear in the literature involve particular strings. In this paper, we consider arbitrary strings, obtaining general results on this subject, which yield all known results as special cases. We will see that the statistic N_τ depends on some basic characteristics of the string τ , namely its number of rises, height, depth and periodicity. The importance of the notion of periodicity in words is well known, and it has been used extensively in various string enumeration problems.

In Section 2, we summarize some general results on the periodicity of words, which are used in the next sections.

In Section 3, we evaluate the generating function F when τ is a Dyck prefix (or equivalently a Dyck suffix) and we give several applications of the above result.

The same problem is studied in Section 4 for an arbitrary string which is neither a Dyck prefix nor a Dyck suffix. We give a complete answer for the case where the string is non-periodic. We also examine the class of strings of the form $d^\delta p$, where $\delta \in \mathbb{N}^*$ and p is a Dyck prefix.

In Section 5, we classify the occurrences of τ according to their height and we evaluate the associated generating functions.

Finally, in Section 6, we unify the main results of Sections 3, 4 and 5.

We note that some of the results of this paper have been announced in the 7th International Conference on Lattice Paths Combinatorics and Applications [11].

2 Periodic words

A non-empty word $w = a_1 a_2 \cdots a_n$ of length $|w| = n$, is called *periodic* if there exists a positive integer $\rho < |w|$, such that $a_{i+\rho} = a_i$, for all $i \in [n - \rho]$. The number ρ is called a *period* of w .

Equivalently, w is periodic iff there exist words λ, μ , with $\lambda \neq \varepsilon$, such that $w = (\lambda\mu)^k \lambda$, for some $k \in \mathbb{N}^*$. In this expression, the period $\rho = |\lambda\mu|$ uniquely determines λ, μ, k .

A non-empty word v that is both a proper prefix and suffix of w , is called a *border* of w . A word w is periodic iff it contains a border. More precisely, if ρ is a period of w , then the prefix v of length $|w| - \rho$ (i.e. $v = (\lambda\mu)^{k-1} \lambda$) is a border of w . Conversely, if v is a border of w , then $|w| - |v|$ is a period of w , as it follows immediately from the next result, which can be easily proved using induction.

Lemma 1. *Let w be a word and v any border of w . If k is the least positive integer such that $k|w| \geq (k+1)|v|$, then there exist unique words λ, μ , with $\lambda \neq \varepsilon$, such that*

$$w = (\lambda\mu)^k\lambda \quad \text{and} \quad v = (\lambda\mu)^{k-1}\lambda.$$

The borders of w are ordered with respect to their length. Clearly, the greatest border of w corresponds to the smallest period of w .

If v is a border of w and v' is a non-empty word with $|v'| < |v|$, then v' is a border of w iff v' is a border of v .

If λ is the least border of w , then $|w| \geq 2|\lambda|$, so that w can be written in the form $w = \lambda\mu\lambda$, where μ is a (possibly empty) word.

We also have the following result, the proof of which is easy and it is omitted.

Proposition 2. *Let w be a periodic word and let ν be the greatest positive integer such that there exist words λ, μ , with $\lambda \neq \varepsilon$, and $w = (\lambda\mu)^\nu\lambda$. Then, for every border v of $\lambda\mu\lambda$, we have that $|v| \leq |\lambda\mu|$.*

From the above Proposition, it follows easily that, for $\nu \geq 2$, the words λ, μ in the expression of w are unique. This expression is called the *canonical form* of w .

However, for $\nu = 1$, the expression $w = \lambda\mu\lambda$ is not unique. For example, the word $w = u^2du^2 = u(udu)u$ has two different expressions. Since in this case $w = \lambda\mu\lambda$, where λ is the greatest border of w , the canonical form can be also extended in the case $\nu = 1$, assuming that λ is the greatest border of w .

In the sequel, we determine the set \mathcal{V} of all borders of a periodic word. For this, we need the following two Lemmas.

Lemma 3. *For every periodic word w , words λ, μ , with $\lambda \neq \varepsilon$ and $\nu \in \mathbb{N}^*$, we have that $w = (\lambda\mu)^\nu\lambda$ is the canonical form of w iff $(\lambda\mu)^{\nu-1}\lambda$ is the greatest border of w (i.e., $|\lambda\mu|$ is the smallest period of w).*

Lemma 4. *For any positive integers $\nu, k \geq 2$ and any two words λ, μ , we have that $(\lambda\mu)^{\nu-1}\lambda$ is the greatest border of $(\lambda\mu)^\nu\lambda$ iff $(\lambda\mu)^{k-1}\lambda$ is the greatest border of $(\lambda\mu)^k\lambda$.*

Lemma 3 is an immediate consequence of Lemma 1, whereas the proof of Lemma 4 is based on the observation that it is enough to show that $(\lambda\mu)^{\nu-1}\lambda$ is the greatest border of $(\lambda\mu)^\nu\lambda$ iff $\lambda\mu\lambda$ is the greatest border of $\lambda\mu\lambda\mu\lambda$, for $\nu \geq 3$.

Proposition 5. *If $w = (\lambda\mu)^\nu\lambda$ is the canonical form of the periodic word w , then v is a border of w iff it is either a border of $\lambda\mu\lambda$ or of the form $v_k = (\lambda\mu)^k\lambda$, $k = 0, 1, \dots, \nu - 1$.*

Proof. Clearly, it is enough to show that for $\nu \geq 2$ and for every border v of w with $|v| \geq |v_1|$, there exists $k \in [\nu - 1]$, such that $v = v_k$.

Let k be the greatest element of $[\nu - 1]$ such that $|v_k| \leq |v|$. Then $|v| < |v_{k+1}|$, so that v is a border of v_{k+1} . Since, by Lemmas 3 and 4, v_k is the greatest border of v_{k+1} , we deduce that $v = v_k$. \square

For every border v of a periodic word w , we denote by $r(v)$ the complementary to v suffix of w , i.e., $w = vr(v)$.

Proposition 6. *Let $w = (\lambda\mu)^\nu\lambda$ be the canonical form of the periodic word w . Then we have that*

- i) for every border v of w , $r(v)$ starts with $\mu\lambda$ iff $v = v_k$, for some $k \in \{0, 1, \dots, \nu - 1\}$,*
- ii) for every two borders v, v' of $\lambda\mu\lambda$ with $|v| < |v'|$, $r(v)$ does not start with $r(v')$.*

Proof. *i)* Clearly, $r(v_k) = (\mu\lambda)^{\nu-k}$ starts with $\mu\lambda$ for every $k \in \{0, 1, \dots, \nu - 1\}$. For the converse, in view of Proposition 5, it is enough to show that if $r(v)$ starts with $\mu\lambda$ and v is a border of $\lambda\mu\lambda$, then $v = \lambda$. Indeed, we can easily check that $v\mu\lambda$ is a border of w , if $\nu \geq 2$, or $v\mu\lambda = w$, if $\nu = 1$. Since $|v\mu\lambda| > |\lambda\mu|$, by Proposition 2 we deduce that $v\mu\lambda = \lambda\mu\lambda$, which implies that $v = \lambda$.

ii) If $r(v)$ starts with $r(v')$, then it can be easily shown that $vr(v')$ is a border of w . Clearly, since by Proposition 2 $|v'| \leq |\mu\lambda|$, we obtain that

$$|r(v')| = |(\lambda\mu)^\nu\lambda| - |v'| \geq (\nu + 1)|\lambda| + \nu|\mu| - |\mu\lambda| = |v_{\nu-1}|.$$

Then, $|vr(v')| > |v_{\nu-1}|$, which is a contradiction. □

3 Counting Dyck prefixes

In this section, we consider the string τ being a Dyck prefix, and we evaluate the associated generating function F .

Proposition 7. *The generating function F which counts the occurrences of a Dyck prefix τ , satisfies the equation*

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_u}F^{|\tau|_u - |\tau|_d} \left(F + (F - 1 - xF^2) \sum_{v \in \mathcal{V}} x^{-|v|_u} F^{|v|_d - |v|_u} \right),$$

where \mathcal{V} is the set of all borders of τ .

Proof. Firstly, we write $\tau = wp$, where p is a Dyck prefix and $w = u$, if τ does not return to the x -axis, or w is a prime Dyck path, otherwise.

Using the first return decomposition $\alpha = u\beta d\gamma$, we obtain that α has an occurrence of τ which does not lie entirely inside β or γ , iff $w = u$ and p is a prefix of β (resp. $w = u\beta d$ and p is a prefix of γ), when τ does not (resp. does) return to the x -axis. Thus, it follows easily that

$$F = 1 + xF^2 + (y - 1)x^{|w|_u}F^{|w|_u - |w|_d}A_p. \tag{3.1}$$

For the evaluation of A_p , we consider the following cases:

- i)* The string τ is non-periodic.

A Dyck path α with prefix p can be decomposed as $\alpha = p\beta$, where

$$\beta = \beta_0 d \beta_1 d \cdots d \beta_{\xi-1} d \beta_{\xi}, \quad \xi = |p|_u - |p|_d, \quad \beta_0, \beta_1, \dots, \beta_{\xi} \in \mathcal{D}.$$

Clearly, since τ is non-periodic, every occurrence of τ in α must lie entirely in β and furthermore, since τ is a Dyck prefix, it must lie entirely in a single β_i , for some $i \in [\xi]$. Thus,

$$A_p = x^{|p|_u} F^{|p|_u - |p|_d + 1}.$$

Substituting in relation (3.1), we obtain that

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_u} F^{|\tau|_u - |\tau|_d + 1} \tag{3.2}$$

and since in this case $\mathcal{V} = \emptyset$, we deduce the required result.

ii) The string τ is periodic.

Let $\tau = \lambda(\mu\lambda)^\nu$, $\nu \in \mathbb{N}^*$, be the canonical form of the string τ .

It follows easily that $|w| \leq |\lambda\mu|$, so that $v_{\nu-1}$ is a suffix of p .

If α is a Dyck path with prefix p , then, since $v_{\nu-1}$ is the greatest border of τ , every occurrence of τ starting from some point of p in α , must start from a point of $v_{\nu-1}$; (see Fig. 3).

It follows that

$$A_p = x^{|p|_u - |v_{\nu-1}|_u} F^{|p|_u - |v_{\nu-1}|_u - (|p|_d - |v_{\nu-1}|_d)} A_{v_{\nu-1}},$$

or equivalently

$$A_p = x^{-|w|_u} F^{|w|_d - |w|_u} G A_{v_{\nu-1}}, \tag{3.3}$$

where $G = x^{|\lambda\mu|_u} F^{|\lambda\mu|_u - |\lambda\mu|_d}$.

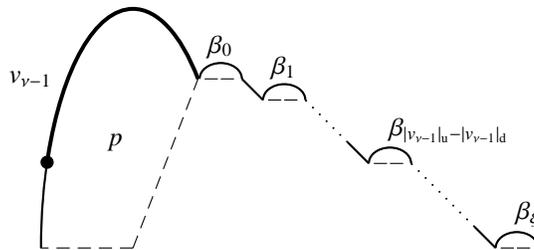


Figure 3: A Dyck path α with prefix p .

Let E_k be the generating function of the set \mathcal{E}_k of all Dyck paths starting with $\lambda(\mu\lambda)^k$ but not with $\lambda(\mu\lambda)^{k+1}$, where $k \in \mathbb{N}^*$ and let E be the generating function of the set \mathcal{E} of all Dyck paths starting with $\mu_2\lambda$ but not with $\mu_2\lambda\mu\lambda$, where $\mu = \mu_1\mu_2$ is the leftmost lowest point decomposition of μ .

Every Dyck path $\beta \in \mathcal{E}_k$, $k \in \mathbb{N}^*$, can be uniquely decomposed as follows:

$$\beta = \lambda(\mu\lambda)^{k-1} \mu_1 \beta_0 d \beta_1 \cdots d \beta_{\xi},$$

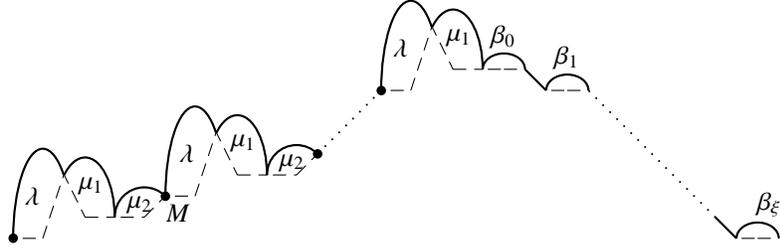


Figure 4: A Dyck path $\beta \in \mathcal{E}_k$, where $\beta_0 \in \mathcal{E}$.

where $\xi = k(|(\mu\lambda)^{k-1}\mu_1|_u - |(\mu\lambda)^{k-1}\mu_1|_d)$, $\beta_i \in \mathcal{D}$, $i \in [\xi]$ and $\beta_0 \in \mathcal{E}$; (see Fig. 4).

Every occurrence of τ in β not lying entirely in some β_i must start from a point of $\lambda(\mu\lambda)^{k-1}$. Any such point M should be an initial point of some λ in the expression $\lambda(\mu\lambda)^{k-1}$, (i.e., one of the bold vertices in Fig. 4) since otherwise the path v starting from M and ending at the first on the right terminal point of some λ of $\lambda(\mu\lambda)^{k-1}$ would be a border of $\lambda\mu\lambda$, while $\mu\lambda$ would be a prefix of $r(v)$, which contradicts Proposition 6.

Moreover, since β_0 does not start with $\mu_2\lambda\mu\lambda$, we deduce that, for $k \geq \nu$, among these points M , an occurrence of τ can only start from the $k - \nu + 1$ leftmost ones, while if $k < \nu$, no occurrence of τ starts before β_0 .

It follows that

$$E_k = x^{k|\lambda\mu|_u - |\mu_2|_u} F^{k(|\lambda\mu|_u - |\lambda\mu|_d) - (|\mu_2|_u - |\mu_2|_d)} y^{(k-\nu+1)^+} E,$$

or equivalently

$$E_k = G^k x^{-|\mu_2|_u} F^{-(|\mu_2|_u - |\mu_2|_d)} y^{(k-\nu+1)^+} E, \quad k \in \mathbb{N}^*. \quad (3.4)$$

It follows that

$$A_{v_{\nu-1}} = \sum_{k=\nu-1}^{\infty} E_k = x^{-|\mu_2|_u} F^{-(|\mu_2|_u - |\mu_2|_d)} \sum_{k=\nu-1}^{\infty} G^k y^{k-\nu+1} E,$$

which gives that

$$A_{v_{\nu-1}} = \frac{x^{-|\mu_2|_u} F^{-(|\mu_2|_u - |\mu_2|_d)} G^{\nu-1} E}{1 - yG} \quad (3.5)$$

and for $\nu \geq 2$

$$A_{v_1} = \sum_{k=1}^{\nu-2} E_k + A_{v_{\nu-1}} = x^{-|\mu_2|_u} F^{-(|\mu_2|_u - |\mu_2|_d)} \left(\sum_{k=1}^{\nu-2} G^k + \frac{G^{\nu-1}}{1 - yG} \right) E,$$

which gives that

$$A_{v_1} = x^{-|\mu_2|_u} F^{-(|\mu_2|_u - |\mu_2|_d)} \frac{G(1 - yG) + (y - 1)G^\nu}{(1 - G)(1 - yG)} E. \quad (3.6)$$

From relations (3.1), (3.3) and (3.5), we obtain that

$$E = (F - 1 - xF^2) x^{|\mu_2|_u} F^{|\mu_2|_u - |\mu_2|_d} G^{-\nu} \frac{1 - yG}{y - 1}. \quad (3.7)$$

In the following, we give another formula for the generating function E .
 Every Dyck path $\beta \in \mathcal{E}$ can be uniquely decomposed as follows:

$$\beta = \mu_2\lambda\gamma,$$

where $\gamma = \gamma_0 d \gamma_1 \cdots d \gamma_t$, $t = |\mu_2\lambda|_u - |\mu_2\lambda|_d$, $\gamma_i \in \mathcal{D}$, $i = 0, 1, \dots, t$ and γ does not start with $\mu\lambda$.

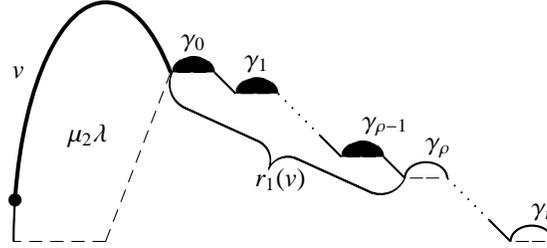


Figure 5: A Dyck path $\beta \in \mathcal{E}$ containing an occurrence of τ which starts at some point of the initial $\mu_2\lambda$.

Every occurrence of τ in β not lying entirely in some γ_i , must start with some $v \in \mathcal{V}' = \mathcal{V} \setminus \{v_i : i = 0, 1, \dots, \nu - 1\}$ (which is a suffix of $\mu_2\lambda$) and it occurs iff $r(v)$ is a prefix of γ i.e., if $r_1(v) = \gamma_0 d \gamma_1 \cdots d \gamma_{\rho-1} d$ and $r_2(v)$ is a prefix of γ_ρ , where $\rho = |r_1(v)|_d - |r_1(v)|_u$. Here, $r(v) = r_1(v)r_2(v)$ is the leftmost lowest point decomposition of $r(v)$; (see Fig. 5).

Furthermore, since by Proposition 6, γ can start with $r(v)$ for at most one $v \in \mathcal{V}'$, it follows that

$$E = x^{|\mu_2\lambda|_u} \left(F^{|\mu_2\lambda|_u - |\mu_2\lambda|_d + 1} - x^{|\mu_1|_u} F^{|\mu_2\lambda|_u - |\mu_2\lambda|_d - (|\mu_1|_d - |\mu_1|_u)} A_{\mu_2\lambda} + (y - 1) \sum_{v \in \mathcal{V}'} x^{|r_1(v)|_u} F^{|\mu_2\lambda|_u - |\mu_2\lambda|_d - (|r_1(v)|_d - |r_1(v)|_u)} A_{r_2(v)} \right). \quad (3.8)$$

For $\nu \geq 2$, we have that

$$A_{\mu_2\lambda} = E + A_{\mu_2 v_1} = E + x^{|\mu_2|_u} F^{|\mu_2|_u - |\mu_2|_d} A_{v_1},$$

and, using relation (3.6), we deduce that

$$A_{\mu_2\lambda} = \frac{1 - yG + (y - 1)G^\nu}{(1 - G)(1 - yG)} E. \quad (3.9)$$

We note that, for $\nu = 1$, relation (3.9) follows automatically from relation (3.5).

Furthermore, using similar ideas as before, we obtain that

$$\begin{aligned} A_{r_2(v)} &= x^{|r_2(v)|_u - |v_{\nu-1}|_u} F^{|r_2(v)|_u - |r_2(v)|_d - (|v_{\nu-1}|_u - |v_{\nu-1}|_d)} A_{v_{\nu-1}} \\ &= x^{|\lambda\mu|_u - |r_1(v)|_u - |v|_u - |\mu_2|_u} F^{|\lambda\mu|_u - |r_1(v)|_u - |v|_u - |\mu_2|_u - (|\lambda\mu|_d - |r_1(v)|_d - |v|_d - |\mu_2|_d)} \frac{G^{\nu-1}}{1 - yG} E, \end{aligned} \quad (3.10)$$

for every $v \in \mathcal{V}'$.

From relations (3.8), (3.9) and (3.10), we deduce that

$$E = x^{|\mu_2\lambda|_u} F^{|\mu_2\lambda|_u - |\mu_2\lambda|_d + 1} - \frac{1 - yG + (y-1)G^\nu}{(1-G)(1-yG)} GE + (y-1)x^{|\lambda|_u} F^{|\lambda|_u - |\lambda|_d} \frac{G^\nu E}{1-yG} \sum_{v \in \mathcal{V}'} x^{-|v|_u} F^{|v|_d - |v|_u}, \quad (3.11)$$

If we set $T = \sum_{v \in \mathcal{V}} x^{-|v|_u} F^{|v|_d - |v|_u}$, then we have that

$$\sum_{v \in \mathcal{V}'} x^{-|v|_u} F^{|v|_d - |v|_u} = T - x^{|\mu|_u} F^{|\mu|_u - |\mu|_d} \frac{G^{-\nu} - 1}{1-G}.$$

Then, by substituting in relation (3.11), we obtain after some simple manipulations that

$$\frac{E}{1-yG} = x^{|\mu_2\lambda|_u} F^{|\mu_2\lambda|_u - |\mu_2\lambda|_d + 1} + (y-1)x^{|\lambda|_u} F^{|\lambda|_u - |\lambda|_d} \frac{TG^\nu E}{1-yG}$$

Finally, by substituting the above expression for E in relation (3.7), we easily obtain the required result. \square

We note that the above result has been proved in [25], for non-periodic τ .

Applications

1. If $\tau = p^\xi$, where p is a non-periodic Dyck prefix, and $\xi \in \mathbb{N}^*$, $\xi \geq 2$, then $\mathcal{V} = \{p^i : i \in [\xi - 1]\}$ and

$$\sum_{v \in \mathcal{V}} x^{-|v|_u} F^{|v|_d - |v|_u} = \frac{G^{1-\xi} - 1}{1-G},$$

where $G = x^{|\rho|_u} F^{|\rho|_u - |\rho|_d}$. It follows from Proposition 7, that the associated generating function satisfies the equation

$$F = 1 + xF^2 + (y-1)G \left(F + (GF - 1 - xF^2) \frac{1 - G^{\xi-1}}{1-G} \right). \quad (3.12)$$

Examples

- i) If $\tau = u^\xi$, then $G = xF$, so that from equation (3.12) we deduce that the associated generating function satisfies the equation

$$F = 1 + xF^2 + (y-1)xF \left(F - \frac{1 - (xF)^{\xi-1}}{1 - xF} \right).$$

ii) If $\tau = (u\sigma d)^\xi$, where $\sigma \in \mathcal{D}$ with $|\sigma|_u = r$, then since the path $u\sigma d$ is non-periodic and $G = x^{r+1}$, substituting in (3.12), we deduce that the associated generating function satisfies the equation

$$F = 1 + xF^2 + (y - 1)x^{r+1} \left(F + (x^{r+1}F - 1 - xF^2) \frac{1 - x^{(r+1)(\xi-1)}}{1 - x^{r+1}} \right).$$

2. If $\tau = pu^\xi$, where p is a non-periodic Dyck prefix, and $\xi \in \mathbb{N}^*$, then $\mathcal{V} = \{u^i : i \in [m]\}$, where $m = \min\{\xi, k\}$ and k is the length of the first ascent of p . It is easy to check that

$$\sum_{v \in \mathcal{V}} x^{-|v|_u} F^{|v|_d - |v|_u} = \frac{(xF)^{-m} - 1}{1 - xF},$$

so that, from Proposition 7, it follows that the associated generating function satisfies the equation

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_u - m} F^{|\tau|_d - |\tau|_u - m} \left(F - \frac{1 - (xF)^m}{1 - xF} \right).$$

Example

If $p = u^k d^\nu$, where $k, \nu \in \mathbb{N}^*$, with $\nu \leq k$, from the previous formula, we obtain that the generating function which counts the occurrences of the string $u^k d^\nu u^\xi$ satisfies the equation

$$F = 1 + xF^2 + (y - 1)x^M F^{M-\nu} \left(F - \frac{1 - (xF)^m}{1 - xF} \right), \quad (3.13)$$

where $M = \max\{k, \xi\}$ and $m = \min\{k, \xi\}$.

We note that this result has been proved firstly in [13], for $\nu = \xi = 1$ and it was extended in [20], for $\nu = 1$.

If $k, \xi \geq \nu$, then we can exchange the roles of k, ξ . It follows that the statistics $N_{u^k d^\nu u^\xi}$ and $N_{u^\xi d^\nu u^k}$ are equidistributed. To illustrate this result bijectively, we will construct an involution φ of \mathcal{D} such that

$$|\varphi(\alpha)|_u = |\alpha|_u \quad \text{and} \quad N_{u^k d^\nu u^\xi}(\varphi(\alpha)) = N_{u^\xi d^\nu u^k}(\alpha), \quad \text{for every } \alpha \in \mathcal{D}.$$

Indeed, firstly we define the involution ψ of the set \mathcal{B} of all paths

$$\beta = u^{\xi_1} d^\nu u^{\xi_2} \dots d^\nu u^{\xi_{k-1}} d^\nu u^{\xi_k},$$

where $k \geq 2$ and $\xi_i \geq \nu$, $i \in [k]$, by

$$\psi(\beta) = u^{\xi_k} d^\nu u^{\xi_{k-1}} \dots d^\nu u^{\xi_2} d^\nu u^{\xi_1}.$$

It is clear that every Dyck path α containing $u^\nu d^\nu u^\nu$ can be uniquely decomposed as $\alpha = \gamma_0 \beta_1 \gamma_1 \beta_2 \gamma_2 \dots \beta_\ell \gamma_\ell$, where β_i is a maximal subpath of α in \mathcal{B} and γ_i avoids the string $u^\nu d^\nu u^\nu$, $i \in [\ell]$. It follows that the required involution is given by

$$\varphi(\alpha) = \gamma_0 \psi(\beta_1) \gamma_1 \psi(\beta_2) \gamma_2 \dots \psi(\beta_\ell) \gamma_\ell.$$

Remark

For every Dyck suffix τ , applying Proposition 7 for the mirror string $\bar{\tau}$, we obtain that the generating function F which counts the occurrences of τ satisfies the equation

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_d}F^{|\tau|_d - |\tau|_u} \left(F + (F - 1 - xF^2) \sum_{v \in \mathcal{V}} x^{-|v|_d} F^{|v|_u - |v|_d} \right), \quad (3.14)$$

where \mathcal{V} is the set of all borders of τ .

This result can be generalized for ballot paths. For this, we evaluate the associated generating function $G = G(x, y, z)$, where x, y, z count the number of rises, the number of occurrences of τ and the height $h(\alpha)$ of a ballot path α respectively.

Indeed, every ballot path α with height $h(\alpha) = h$ is uniquely decomposed as

$$\alpha = \beta_0 u \beta_1 \cdots u \beta_h, \quad \beta_i \in \mathcal{D}, \quad 0 \leq i \leq h.$$

Since τ is a Dyck suffix, an occurrence of τ in α must be entirely contained in a single β_i , for some $0 \leq i \leq h$. It follows that

$$\begin{aligned} G &= \sum_{h=0}^{\infty} \sum_{\substack{\alpha \text{ ballot path} \\ h(\alpha)=h}} x^{|\alpha|_u} y^{|\alpha|_\tau} z^h = \sum_{h=0}^{\infty} \sum_{\substack{\beta_i \in \mathcal{D} \\ 0 \leq i \leq h}} x^{h + \sum_{i=0}^h |\beta_i|_u} y^{\sum_{i=0}^h |\beta_i|_\tau} z^h \\ &= \sum_{h=0}^{\infty} x^h z^h \prod_{i=0}^h \sum_{\beta_i \in \mathcal{D}} x^{|\beta_i|_u} y^{|\beta_i|_\tau} = \sum_{h=0}^{\infty} x^h z^h F^{h+1}(x, y). \end{aligned}$$

Thus,

$$G = \frac{F(x, y)}{1 - xzF(x, y)},$$

where F satisfies relation (3.14).

Sullivan [18], using a different approach, provided a recursive formula for the evaluation of the coefficients of G .

4 Counting strings with positive depth and height

Throughout this section, τ is a (δ, h) -string with $\delta, h > 0$, i.e., τ is neither a Dyck prefix nor a Dyck suffix. In this case, τ is uniquely decomposed as $\tau = sd p$, where s is a Dyck suffix of depth $\delta - 1$ and p is a Dyck prefix of height h . Using the first return decomposition, we deduce that

$$F = 1 + xF^2 + (y - 1)x B_s A_p. \quad (4.1)$$

For the evaluation of the generating functions B_s, A_p in terms of F , we will use the (Fibonacci-like) polynomials $p_i, q_i, i \geq -1$, (see [10, p. 327]) defined by

$$p_i(t) = p_{i-1}(t) - x p_{i-2}(t), \quad p_{-1}(t) = \frac{1}{x}, \quad p_0(t) = t, \quad (4.2)$$

(where x is considered as a parameter), and

$$q_i(x) = q_{i-1}(x) - xq_{i-2}(x), \quad q_{-1}(x) = 0, \quad q_0(x) = 1. \quad (4.3)$$

We note that

$$q_i(x) = \sqrt{x^i} U_i\left(\frac{1}{2\sqrt{x}}\right), \quad (4.4)$$

for $i \geq -1$, where $U_i(x)$ are the Chebyshev polynomials of the second kind (see sequence A053117 in [23]).

It is easy to check that these polynomials satisfy the following identities:

$$p_i(t) = xtp_{i-1}\left(\frac{t-1}{xt}\right), \quad (4.5)$$

$$(1 - xt)p_i(t) - (t - 1 - xt^2)q_i(x) = xp_{i-1}(t), \quad (4.6)$$

$$p_{i-1}(t)p_i(t) - p_i^2(t) - xp_{i-1}^2(t) = x^{i-1}(t - 1 - xt^2), \quad (4.7)$$

for every $i \in \mathbb{N}$.

We first give the following result, which will be used in the sequel.

Lemma 8. *For every (δ, h) -string τ , we have that*

$$B_i = p_i(F), \quad i \leq \min\{h + k, h + 2\} \quad (4.8)$$

and

$$\Gamma_{p,d^i} = q_i(x)A_p, \quad i \leq \min\{h + k, h + 2, |p|_u - |p|_d + t\}, \quad (4.9)$$

where p is a non-empty Dyck prefix and k (resp. t) is the number of all consecutive falls in the end of τ (resp. p).

Proof. Using the bijection of Fig. 6, under the inequality restrictions of relations (4.8) and (4.9) respectively, for $i \geq 2$, we have that

$$B_{i-1} - B_i = xB_{i-2} \quad \text{and} \quad \Gamma_{p,d^{i-1}} - \Gamma_{p,d^i} = x\Gamma_{p,d^{i-2}},$$

since, for $i \leq h + k$, the last peak of the Dyck path α does not belong to any occurrence of τ , while for $i \leq h + 2$, its deletion does not result to a new occurrence of τ in α' .

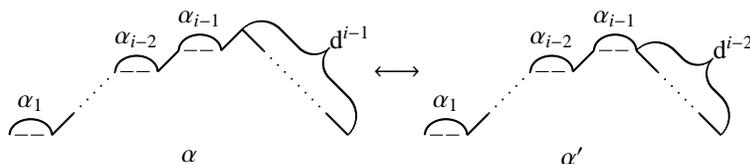


Figure 6: The Dyck path α ending with exactly $i - 1$ falls is mapped to α' ending with (at least) $i - 2$ falls.

Furthermore, since $B_0 = F$, $B_1 = F - 1$ and $\Gamma_{p,d^0} = \Gamma_{p,d} = A_p$, the result follows immediately from relations (4.2) and (4.3). \square

We note that if we apply the previous Lemma for the mirror string $\bar{\tau}$, it follows that

$$A_j = p_j(F), \quad j \leq \min\{\delta + k', \delta + 2\}, \quad (4.10)$$

and

$$\Gamma_{u^j, s} = q_j(x)B_s, \quad j \leq \min\{\delta + k', \delta + 2, |s|_d - |s|_u + t'\}, \quad (4.11)$$

where s is a non-empty Dyck suffix and k' (resp. t') is the number of all consecutive rises in the beginning of τ (resp. s).

In particular, we have that

$$\Gamma_{j, i} = \begin{cases} q_i(x)A_j, & i \leq \min\{h + k, h + 2, j\} \\ q_j(x)B_i, & j \leq \min\{\delta + k', \delta + 2, i\}. \end{cases} \quad (4.12)$$

In the following result we establish the equation of the generating function F , for a non-periodic string.

Proposition 9. *The generating function F which counts the occurrences of a non-periodic (δ, h) -string τ , satisfies the equation*

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_u - h + 1} \frac{p_m^{|\delta| + 1}(F)}{p_{m-1}^{|\delta| - 1}(F)},$$

where $m = \min\{h, \delta\}$.

Proof. Firstly, we write $\tau = sdp$, where $s = \beta_0 d \beta_1 \cdots d \beta_{\delta-1}$, $p = \gamma_h u \cdots \gamma_1 u \gamma_0$ and $\beta_i, \gamma_j \in \mathcal{D}$, $0 \leq i \leq \delta - 1$, $0 \leq j \leq h$.

Let $b_i = \beta_0 d \beta_1 \cdots d \beta_i$, $0 \leq i \leq \delta - 1$ and $c_j = \gamma_j u \cdots \gamma_1 u \gamma_0$, $0 \leq j \leq h$. Since τ is non-periodic, using the first return decomposition, we can easily show that

$$A_{c_j} = x^{|c_j|_u - |c_{j-1}|_u} \left(A_{c_{j-1}} F + (y - 1) \Gamma_{c_{j-1}, s} A_p \right), \quad (4.13)$$

for every $0 \leq j \leq h$, where $c_{-1} = \varepsilon$.

For every $j \leq \delta$, using the fact that τ is non-periodic, we can easily check that the bijection of Fig. 7 preserves the number of occurrences of τ , so that

$$A_{c_j} = x^{|c_j|_u - j} A_j, \quad j \leq m. \quad (4.14)$$

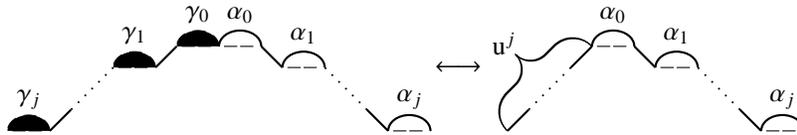


Figure 7: The bijection between Dyck paths starting with c_j and those starting with u^j .

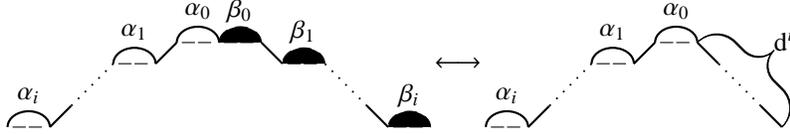


Figure 8: The bijection between Dyck paths ending with b_i and those ending with d^i .

Similarly, using the bijection of Fig. 8, we deduce that

$$B_{b_i} = x^{|b_i|_u} B_i, \quad i \leq m, \quad \text{and} \quad \Gamma_{c_j, b_i} = x^{|b_i|_u} \Gamma_{c_j, d^i}, \quad i \leq j. \quad (4.15)$$

Without loss of generality, we may assume that $\delta \leq h$, since, otherwise we replace τ by its mirror string $\bar{\tau}$. For $\delta \leq j \leq h$, using relations (4.9), (4.13) and (4.15), we deduce that

$$\frac{A_{c_j} x^{-|c_j|_u}}{A_{c_{j-1}} x^{-|c_{j-1}|_u}} = F + (y-1)x^{|s|_u} q_{\delta-1}(x) A_p.$$

It follows that

$$A_p = x^{|p|_u - |c_{\delta-1}|_u} A_{c_{\delta-1}} \left(\frac{A_{c_\delta} x^{-|c_\delta|_u}}{A_{c_{\delta-1}} x^{-|c_{\delta-1}|_u}} \right)^{h-\delta+1} = x^{|p|_u - h} \frac{p_\delta^{h-\delta+1}(F)}{p_{\delta-1}^{h-\delta}(F)}.$$

The last equality follows from relations (4.10) and (4.14).

Furthermore, since $\delta \leq h$, from relations (4.8) and (4.15), we obtain that

$$B_s = x^{|s|_u} B_{\delta-1} = x^{|s|_u} p_{\delta-1}(F).$$

Therefore, after substituting the above expressions for A_p and B_s in relation (4.1), we obtain the required result. \square

Example

The string $\tau = d^\nu u d^\nu u^2 \cdots d^\nu u^{2\nu}$, $\nu \in \mathbb{N}^*$, is non-periodic with $\delta = \frac{\nu(\nu+1)}{2}$ and $h = \frac{\nu(\nu+3)}{2}$. It follows from Proposition 9 that the associated generating function satisfies the equation

$$F = 1 + xF^2 + (y-1)x^{\frac{3\nu^2 - \nu + 2}{2}} \frac{p_{\frac{\nu^2 + \nu}{2}}^{\nu+1}(F)}{p_{\frac{\nu^2 + \nu - 2}{2}}^{\nu-1}(F)}.$$

The case of a periodic string seems very complex. We will examine some particular cases where the polynomials p_i are also used. Before that, in the next result we give an expression of the generating function A_p , where p is a Dyck prefix, in terms of the generating functions A_i , $i \in \mathbb{N}$.

Lemma 10. *Let τ be a (δ, h) -string starting with a fall and let p be a Dyck prefix such that $|p| < |\tau|$. Then, the generating function A_p with respect to the string τ is given by*

$$A_p = x^{|p|_d} A_{|p|_u - |p|_d} + \sum_{w \in \mathcal{W}_p} x^{|l_p(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}),$$

where \mathcal{W}_p is the set of all non-empty suffixes of p which are prefixes of τ , $l_p(w)$ is the complementary to w prefix of p (i.e., $p = l_p(w)w$) and $j_w = |l_p(w)|_u - |l_p(w)|_d$.

Proof. We will use induction with respect to $M_p = \max\{|w| : w \in \mathcal{W}_p\}$.

If $M_p = 0$, then $\mathcal{W}_p = \emptyset$ and the result follows immediately, since for every Dyck path with prefix p we can replace that prefix with $u^{|p|_u - |p|_d}$ without affecting the number of occurrences of τ in the path.

For $M_p > 0$, let q be the greatest element of \mathcal{W}_p , i.e., $|q| = M_p$. We first assume that p ends with a fall and we write $q = q'd$, $p_1 = u^\xi q'$ and $p_2 = u^\xi q'u$, where $\xi = |l_p(q)|_u - |l_p(q)|_d > 0$. Clearly, p_1, p_2 are Dyck prefixes such that $M_{p_1} = |q'| < |q| = M_p$ and $M_{p_2} < |q'u| = |q| = M_p$,

Let \mathcal{W} (resp. \mathcal{W}') be the set of all elements $w \in \mathcal{W}_{p_1}$ such that wd (resp. wu) is a prefix of τ . Clearly, the sets \mathcal{W} and \mathcal{W}' form a partition of \mathcal{W}_{p_1} such that

$$\mathcal{W}_p = \{wd : w \in \mathcal{W}\} \cup \{d\}, \quad \mathcal{W}_{p_2} = \{wu : w \in \mathcal{W}'\}$$

and

$$l_{p_1}(w) = \begin{cases} l_{p_1 d}(wd), & w \in \mathcal{W} \\ l_{p_2}(wu), & w \in \mathcal{W}'. \end{cases}$$

Using the induction hypothesis, we have that

$$\begin{aligned} A_p &= x^{|l_p(q)|_d} A_{p_1 d} = x^{|l_p(q)|_d} (A_{p_1} - A_{p_2}) \\ &= x^{|l_p(q)|_d} \left(x^{|p_1|_d} A_{|p_1|_u - |p_1|_d} + \sum_{w \in \mathcal{W}_{p_1}} x^{|l_{p_1}(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}) \right. \\ &\quad \left. - x^{|p_2|_d} A_{|p_2|_u - |p_2|_d} - \sum_{w \in \mathcal{W}_{p_2}} x^{|l_{p_2}(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}) \right) \\ &= x^{|l_p(q)|_d} \left(x^{|q|_d - 1} (A_{|p|_u - |p|_d + 1} - A_{|p|_u - |p|_d + 2}) + \sum_{w \in \mathcal{W}} x^{|l_{p_1}(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}) \right) \\ &= x^{|p|_d} A_{|p|_u - |p|_d} + x^{|p|_d - 1} (A_{|p|_u - |p|_d + 1} - xA_{|p|_u - |p|_d} - A_{|p|_u - |p|_d + 2}) \\ &\quad + \sum_{w \in \mathcal{W}} x^{|l_p(wd)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}) \\ &= x^{|p|_d} A_{|p|_u - |p|_d} + \sum_{w \in \mathcal{W}_p} x^{|l_p(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}). \end{aligned}$$

The proof of the result when p ends with a rise is similar, except when the height of p is equal to 1, since, in this case, the path obtained by replacing the last rise of $u^\xi q$ with a fall is a Dyck suffix, and the induction step cannot be applied.

This particular case, where $p = \alpha u$ and α is a Dyck path, is treated below separately. Clearly, in this case, every $w \in \mathcal{W}_\alpha$ is a Dyck suffix of depth $|w|_d - |w|_u \leq \delta$. Furthermore, if the depth of w is less than δ , we have that

$$j_w = |l_\alpha(w)|_u - |l_\alpha(w)|_d = |w|_d - |w|_u \leq \delta - 1,$$

so that, from relation (4.10), we deduce that

$$A_{j_w} - xA_{j_w-1} - A_{j_w+1} = 0.$$

In fact, the above equality holds for every $w \in \mathcal{W}_\alpha$, such that wd is a prefix of τ , which yields that

$$\sum_{w \in \mathcal{W}_\alpha} x^{|l_\alpha(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}) = \sum_{w \in \mathcal{W}_p} x^{|l_p(w)|_d} (A_{j_w} - xA_{j_w-1} - A_{j_w+1}). \quad (4.16)$$

Let q be the largest element of \mathcal{W}_α . If $|q|_d - |q|_u \leq \delta - 1$, then every Dyck path with prefix p has no occurrence of τ starting from a point of p , so that

$$A_p = x^{|\alpha|_u} (F - 1) = x^{|p|_d} A_{|p|_u - |p|_d}.$$

Since in this case the sums of relation (4.16) are equal to 0, we obtain the required result.

Finally, if $|q|_d - |q|_u = \delta$, then $qu \in \mathcal{W}_p$, so that $M_\alpha = |q| < |qu| = M_p$. Then, since $A_p = A_\alpha - x^{|\alpha|_d}$, using the induction hypothesis and relation (4.16), we obtain again the required result. \square

In the next Proposition we restrict ourselves to the string $\tau = d^\delta p$, where p is a Dyck prefix.

Proposition 11. *Let F be the generating function which counts the occurrences of the string $d^\delta p$, where p is a Dyck prefix of height h . Then, we have that*

i) if $\delta \leq \min\{h + k, h + 3\}$, then

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_u - h + 1} \left(\frac{p_\delta(F)}{p_{\delta-1}(F)} \right)^{h-\delta} \left(p_\delta(F)p_{\delta-1}(F) + (F - 1 - xF^2)x^{\delta-1} \sum_{v \in \mathcal{V}} x^{-|v|_d} \left(\frac{p_\delta(F)}{p_{\delta-1}(F)} \right)^{|v|_d - |v|_u} \right), \quad (4.17)$$

ii) if $h + k + 1 \leq \delta$ and $\mathcal{V} = \{d^i : i \in [k]\}$, then

$$F = 1 + xF^2 + (y - 1)x^{|\tau|_d - \delta + 1} \left(\frac{p_h(F)}{p_{h-1}(F)} \right)^{\delta-h} \left(p_h(F)p_{h-1}(F) + (F - 1 - xF^2)x^{h-1} \sum_{v \in \mathcal{V}} x^{-|v|_u} \left(\frac{p_h(F)}{p_{h-1}(F)} \right)^{|v|_u - |v|_d} \right), \quad (4.18)$$

where \mathcal{V} is the set of all borders of τ and k is the number of all consecutive falls in the end of τ .

Proof. i) Let $\delta \leq \min\{h + k, h + 3\}$. Then, using the first return decomposition of Dyck paths and relation (4.12), we obtain that

$$A_j = x(A_{j-1}F + (y - 1)\Gamma_{j-1, \delta-1}A_p) = xA_{j-1}(F + (y - 1)q_{\delta-1}(x)A_p), \quad j \geq \delta. \quad (4.19)$$

Using relation (4.10), it follows that

$$A_j = p_{\delta-1}(F) \left(\frac{p_{\delta}(F)}{p_{\delta-1}(F)} \right)^{j-\delta+1}, \quad j \geq \delta - 1. \quad (4.20)$$

Then, from Lemma 10, it follows that

$$\begin{aligned} A_p &= x^{|\tau|_u-h} \left(A_h + \sum_{v \in \mathcal{V}} x^{-|v|_d} (A_{h+|v|_d-|v|_u} - xA_{h+|v|_d-|v|_u-1} - A_{h+|v|_d-|v|_u+1}) \right) \\ &= x^{|\tau|_u-h} p_{\delta-1}(F) \left(\frac{p_{\delta}(F)}{p_{\delta-1}(F)} \right)^{h-\delta} \\ &\quad \cdot \left(\frac{p_{\delta}(F)}{p_{\delta-1}(F)} + \sum_{v \in \mathcal{V}} x^{-|v|_d} \left(\frac{p_{\delta}(F)}{p_{\delta-1}(F)} \right)^{|v|_d-|v|_u} \left(\frac{p_{\delta}(F)}{p_{\delta-1}(F)} - x - \left(\frac{p_{\delta}(F)}{p_{\delta-1}(F)} \right)^2 \right) \right). \end{aligned}$$

Finally, since from relation (4.8) we have that $B_{\delta-1} = p_{\delta-1}(F)$, the required result follows by substituting the expressions for A_p and B_s in relation (4.1) and by using relation (4.7).

ii) Let $h + k + 1 \leq \delta$. Then, since we have also assumed that each border of τ is of the form d^i , $i \in [k]$, we deduce that the bijection of Fig. 9 preserves the number of occurrences of τ . It follows that

$$A_p = x^{|\tau|_u-h} A_h \quad (4.21)$$

and

$$\Gamma_{p, d^i} = \begin{cases} x^{|\tau|_u-h} \Gamma_{h,i} + x^{|\tau|_u}, & h + 1 \leq i \leq h + k \\ x^{|\tau|_u-h} \Gamma_{h,i}, & i \geq h + k + 1. \end{cases} \quad (4.22)$$

For the justification of relation (4.22), notice also that d^i is a suffix of α iff d^i is a suffix of α' , except when $\alpha = pd^h$ and $h + 1 \leq i \leq h + k$; (see Fig. 9).

For the evaluation of $B_{\delta-1}$, first notice that, using the first return decomposition of Dyck paths, we obtain that

$$B_i = x(B_{i-1} + FB_i + (y - 1)B_{\delta-1}\Gamma_{p, d^i}), \quad i \geq 1. \quad (4.23)$$

Then, for $i \geq h + k + 1$, using relations (4.1), (4.10), (4.12), (4.21) and (4.22), we deduce that

$$((1 - xF)p_h(F) - (F - 1 - xF^2)q_h(x)) B_i = xp_h(F)B_{i-1},$$

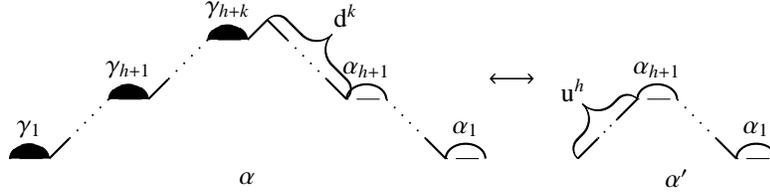


Figure 9: The Dyck path α is mapped to α' by substituting its prefix $p = \gamma_1 u \cdots \gamma_{h+k} u d^k$ with u^h .

or equivalently, using relation (4.6), that

$$\frac{B_i}{B_{i-1}} = \frac{p_h(F)}{p_{h-1}(F)}.$$

This shows that

$$B_i = B_{h+k} \left(\frac{p_h(F)}{p_{h-1}(F)} \right)^{i-h-k}, \quad i \geq h+k. \quad (4.24)$$

For the evaluation of B_i , for $h+1 \leq i \leq h+k$, we proceed as before, obtaining that

$$p_{h-1}(F)B_i - p_h(F)B_{i-1} = x^{h-1}(F-1-xF^2).$$

By solving the above linear recurrence equation with initial condition $B_{h+1} = p_{h+1}(F)$, and using relation (4.7), we obtain, after some simple manipulations, that

$$B_i = p_h(F) \left(1 - x \frac{p_{h-1}^2(F)}{p_h(F)} \cdot \frac{1 - \left(\frac{p_h(F)}{p_{h-1}(F)} \right)^{i-h}}{p_{h-1}(F) - p_h(F)} \right),$$

for every $h+1 \leq i \leq h+k$. Applying the previous equality for $i = h+k$ and relation (4.24) for $i = \delta-1$, we obtain that

$$B_{\delta-1} = \frac{p_h^{\delta-h-k}(F)}{p_{h-1}^{\delta-h-k-1}(F)} \left(1 - x \frac{p_{h-1}(F)}{p_h(F)} \cdot \frac{1 - \left(\frac{p_h(F)}{p_{h-1}(F)} \right)^k}{1 - \frac{p_h(F)}{p_{h-1}(F)}} \right)$$

which, after some simple manipulations and using relation (4.7), yields that

$$B_{\delta-1} = \frac{p_h^{\delta-h}(F)}{p_{h-1}^{\delta-h-1}(F)} \left(p_h(F)p_{h-1}(F) + (F-1-xF^2)x^{h-1} \sum_{v \in \mathcal{V}} x^{-|v|_u} \left(\frac{p_h(F)}{p_{h-1}(F)} \right)^{|v|_u - |v|_d} \right).$$

Finally, since from relations (4.10) and (4.21) we have that $A_p = x^{|\tau|_d - \delta} p_h(F)$, the required result follows, after substituting the expressions for A_p and B_s in relation (4.1). \square

Example

Let $\tau = d^\xi u^\nu d^k$, where $\xi, \nu \in \mathbb{N}^*$, $k \in \mathbb{N}$ and let F be the associated generating function.

If $\nu \leq k$, then, by the equidistribution of the statistics $N_{d^\xi u^\nu d^k}$ and $N_{u^k d^\nu u^\xi}$, it follows that F is given by relation (3.13).

If $\nu > k$, by applying Proposition 9, we obtain that

$$F = 1 + xF^2 + (y-1)x^{k-m+1} \frac{p_\xi^{\nu-M+1}(F)}{p_{\xi-1}^{\nu-M-1}(F)} \left(1 - x \frac{p_\xi(F)}{p_{\xi-1}(F)} \cdot \frac{1 - \left(x \frac{p_{\xi-1}(F)}{p_\xi(F)} \right)^m}{1 - x \frac{p_{\xi-1}(F)}{p_\xi(F)}} \right), \quad (4.25)$$

for $\xi \leq \min\{\nu, \nu - k + 3\}$, $m = \min\{\xi, k\}$ and $M = \max\{\xi, k\}$, and

$$F = 1 + xF^2 + (y-1)x^{k+1} \frac{p_{\nu-k}^{\xi-\nu+1}(F)}{p_{\nu-k}^{\xi-\nu-1}(F)} \left(1 - x \frac{p_{\nu-k-1}(F)}{p_{\nu-k}(F)} \cdot \frac{1 - \left(\frac{p_{\nu-k}(F)}{p_{\nu-k-1}(F)} \right)^k}{1 - \frac{p_{\nu-k}(F)}{p_{\nu-k-1}(F)}} \right), \quad (4.26)$$

for $\xi \geq \nu + 1$.

We do not know the equation of the generating function F when $\nu > k$ and $\nu - k + 4 \leq \xi \leq \nu$.

5 Occurrences at height at least j

The occurrence of strings at a specified height was introduced for certain strings in [12] and it has been studied extensively for arbitrary strings in [19]. It was shown that the generating function which counts the occurrences of a string τ at height j can be expressed via the Chebyshev polynomials of the second kind and the generating function which counts the low occurrences of τ ; (see Proposition 1 in [19]).

In this section, we study the occurrences of strings at height greater or equal to a given $j \in \mathbb{N}$. We say that the string τ *occurs at height at least j* in a Dyck path, if the minimum height of the points of τ in this occurrence is greater or equal to j . For example, the Dyck path of Fig. 1 has four occurrences of the string ud at height at least 1 (two at height 1 and two at height 3).

An occurrence of the string τ at height at least 1 is usually referred as a *high occurrence* of τ . It is known that the statistics “number of high $(ud)^r$ ’s” and “number of $(du)^r$ ’s” are equidistributed for every $r \in \mathbb{N}^*$ (for $r = 1$, see [6] and [8]).

We denote with $F_j = F_j(x, y)$ the generating function which counts the occurrences of the string τ at height at least j . Clearly, $F_0 = F$ (resp. F_1) is the generating function which counts all (resp. the high) occurrences of the string τ . Using the first return decomposition, we can easily deduce that

$$F_j = \frac{1}{1 - xF_{j-1}}, \quad j \in \mathbb{N}^*.$$

Furthermore, following the same procedure used in [19], and relation (4.4), we can express the generating function F_j in terms of F :

Proposition 12. For every string τ , the generating function F_j is given by

$$F_j = \frac{q_{j-1}(x)}{q_j(x)} + \frac{x^j}{q_j^2(x) \left(\frac{1}{F} - x \frac{q_{j-1}(x)}{q_j(x)} \right)}.$$

In the following result, an alternative way for the evaluation of F_j (without the use of F), when τ is a Dyck prefix, is presented.

Proposition 13. The generating function F_j for a Dyck prefix τ satisfies the equation

$$F_j = 1 + xF_j^2 + (y-1)x^{|\tau|_d-j+1} \left(\frac{p_j(F_j)}{p_{j-1}(F_j)} \right)^{|\tau|_u-|\tau|_d} \left(p_j(F_j)p_{j-1}(F_j) + (F_j - 1 - xF_j^2) x^{j-1} \sum_{v \in \mathcal{V}} x^{-|v|_d} \left(\frac{p_j(F_j)}{p_{j-1}(F_j)} \right)^{|v|_d-|v|_u} \right),$$

where \mathcal{V} is the set of all borders of τ .

For the proof of the above formula, we first show, using identity (4.5), that

$$\frac{p_j(F_j)}{p_{j-1}(F_j)} = xF \quad \text{and} \quad \frac{p_j(F_j)p_{j-1}(F_j)}{F_j - 1 - xF_j^2} x^{1-j} = \frac{F}{F - 1 - xF^2},$$

for every $j \in \mathbb{N}$, and then we substitute in the formula of Proposition 7.

Example

If $\tau = u^\xi d^\nu u^k$, where $\xi, \nu \in \mathbb{N}^*$, $k \in \mathbb{N}$ with $k < \nu \leq \xi$, then $\mathcal{V} = \{u^i : i \in [k]\}$, for $k > 0$, and $\mathcal{V} = \emptyset$, for $k = 0$. From Proposition 13, using identity (4.7) and after some simple manipulations, we deduce that

$$F_j = 1 + xF_j^2 + (y-1)x^{\nu-j+1} \frac{p_j^{\xi-\nu+1}(F_j)}{p_{j-1}^{\xi-\nu-1}(F_j)} \left(1 - x \frac{p_{j-1}(F_j)}{p_j(F_j)} \frac{1 - \left(\frac{p_j(F_j)}{p_{j-1}(F_j)} \right)^k}{1 - \frac{p_j(F_j)}{p_{j-1}(F_j)}} \right).$$

Furthermore, if $\nu < \xi$, applying the above relation for $j = \nu - k$ and using relation (4.26), we deduce that the statistics “number of $d^\xi u^\nu d^k$ ’s” and “number of $u^\xi d^\nu u^k$ ’s at height at least $\nu - k$ ” are equidistributed.

This remains true if $\nu = \xi$ and $k = 0$. In fact, using the mirror string $u^\nu d^\xi$, when $\nu > \xi$, we deduce that the statistics “number of $d^\xi u^\nu$ ’s” and “number of $u^\nu d^\xi$ ’s at height at least m ”, where $m = \min\{\xi, \nu\}$ are equidistributed.

We remark that the analogous equation for F_j , when τ is a Dyck suffix, follows by applying Proposition 13 for the mirror string $\bar{\tau}$.

6 Unification

All the equations of the generating function F obtained in this paper, for various strings, can be put under the same roof. For this, we consider the rational functions

$$R_i(t) = \frac{p_i(t)}{xp_{i-1}(t)}, \quad i \in \mathbb{N}$$

and the equations

$$R_i(F) = 1 + xR_i^2(F) + (y-1)x^{|\tau|_u}R_i^{|\tau|_u-|\tau|_d}(F) \left(R_i(F) + (R_i(F) - 1 - xR_i^2(F)) \sum_{v \in \mathcal{V}} x^{-|v|_u} R_i^{|v|_d-|v|_u}(F) \right), \quad (6.1)$$

$$R_i(F) = 1 + xR_i^2(F) + (y-1)x^{|\tau|_d}R_i^{|\tau|_d-|\tau|_u}(F) \left(R_i(F) + (R_i(F) - 1 - xR_i^2(F)) \sum_{v \in \mathcal{V}} x^{-|v|_d} R_i^{|v|_u-|v|_d}(F) \right). \quad (6.2)$$

If τ is a Dyck prefix (resp. Dyck suffix), then F satisfies equation (6.1) (resp. (6.2)), for $i = 0$.

If τ is a non-periodic (δ, h) -string, then F satisfies equation (6.1) (resp. (6.2)), if $h \geq \delta$ (resp. $h \leq \delta$), for $i = \min\{h, \delta\}$.

If $\tau = d^\delta p$, where p is a Dyck prefix of height h , then, under the corresponding inequality conditions of Proposition 11, F satisfies either of the equations (6.1), (6.2), for $i = \delta, h$ respectively.

Finally, notice that the generating function F_j for a Dyck prefix (resp. Dyck suffix) satisfies equation (6.1) (resp. (6.2)), for $i = j$.

References

- [1] Y. Le Borgne, Counting upper interactions in Dyck paths, *Sém. Lothar. Combin.* 54 (2005/06), Article B54f.
- [2] D. Callan, Two bijections for Dyck path parameters, Preprint, (2004), 4pp. <http://www.arxiv.org/abs/math.CO/0406381>.
- [3] D. Callan, Some bijections and identities for the Catalan and Fine numbers, *Sém. Lothar. Combin.* 5 (2006) Article B53e.
- [4] D. Callan, A bijection on Dyck paths and its cycle structure, *Electron. J. Combin.* 14 (2007), #R28.
- [5] A. Denise, R. Simion, Two combinatorial statistics on Dyck paths, *Discrete Math.* 137 (1995) 155–176.

- [6] E. Deutsch, A bijection on Dyck paths and its consequences, *Discrete Math.* 179 (1998), 253–256.
- [7] E. Deutsch, An involution on Dyck paths and its consequences, *Discrete Math.* 204 (1999), 163–166.
- [8] E. Deutsch, Dyck path enumeration, *Discrete Math.* 204 (1999), 167–202.
- [9] S.-P. Eu, S.-C. Liu and Y.-N. Yeh, Dyck paths with peaks avoiding or restricted to a given set, *Stud. Appl. Math.* 111 (2003), 453–465.
- [10] P. Flajolet and R. Sedgewick, 2009. *Analytic Combinatorics*. Cambridge University Press.
- [11] K. Manes, A. Sapounakis, I. Tasoulas, P. Tsikouras, String in Dyck paths, (abstract), *Lattice Paths Combinatorics and Applications*, Sienna, Italy, 2010, 231–233.
- [12] T. Mansour, Counting peaks at height k in a Dyck path, *J. Integer Seq.* 5 (2002), Article 02.1.1.
- [13] T. Mansour, Statistics on Dyck paths, *J. Integer Seq.* 9 (2006), Article 06.1.5.
- [14] D. Merlini, R. Sprungoli and M. Verri, Some statistics on Dyck paths, *J. Statist. Plann. and Infer.* 101 (2002), 211–227.
- [15] H. Niederhausen and S. Sullivan, Euler Coefficients and Restricted Dyck Paths, *Congr. Numer.* 188 (2007), 196–210.
- [16] H. Niederhausen and S. Sullivan, Pattern Avoiding Ballot Paths and Finite Operator Calculus, *J. Statist. Plann. and Infer.* 140 (2010), 2312–2320.
- [17] P. Peart and W.-J. Woan, Dyck paths with no peaks at height k , *J. Integer Seq.* 4 (2001), Article 01.1.3.
- [18] S. Sullivan, Counting strings in ballot paths, (abstract), *Lattice Paths Combinatorics and Applications*, Sienna, Italy, 2010, 247–252.
- [19] A. Sapounakis, I. Tasoulas and P. Tsikouras, Counting strings in Dyck paths, *Discrete Math.* 307 (2007), 2909–2924.
- [20] A. Sapounakis, I. Tasoulas and P. Tsikouras, Some strings in Dyck paths, *Australasian J. Combin.* 39 (2007), 49–72.
- [21] A. Sapounakis, I. Tasoulas and P. Tsikouras, Enumeration of strings in Dyck paths: A bijective approach, *Discrete Math.* 309 (2009), 3033–3039.
- [22] A. Sapounakis and P. Tsikouras, Counting peaks and valleys in k -colored Motzkin paths, *Electron. J. Combin.* 12 (2005), #R16.
- [23] N. J. A. Sloane, Online Encyclopedia of Integer Sequences, published electronically at <http://oeis.org>.
- [24] Y. Sun, The statistic “number of udu’s” in Dyck paths, *Discrete Math.* 287 (2004), 177–186.
- [25] I. Tasoulas, Strings in Dyck paths and ordered trees, PhD thesis (in Greek), University of Piraeus, Greece (2009).