

# Down-step statistics in generalized Dyck paths

Andrei Asinowski<sup>1</sup>   Benjamin Hackl<sup>1,2</sup>   Sarah J. Selkirk<sup>1\*</sup>

<sup>1</sup> *Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Austria*

<sup>2</sup> *Department of Mathematics, Uppsala University, Sweden*

received 8<sup>th</sup> Feb. 2021, revised 14<sup>th</sup> Oct. 2021, 19<sup>th</sup> Apr. 2022, accepted 21<sup>st</sup> Apr. 2022.

The number of down-steps between pairs of up-steps in  $k_t$ -Dyck paths, a generalization of Dyck paths consisting of steps  $\{(1, k), (1, -1)\}$  such that the path stays (weakly) above the line  $y = -t$ , is studied. Results are proved bijectively and by means of generating functions, and lead to several interesting identities as well as links to other combinatorial structures. In particular, there is a connection between  $k_t$ -Dyck paths and perforation patterns for punctured convolutional codes (binary matrices) used in coding theory. Surprisingly, upon restriction to usual Dyck paths this yields a new combinatorial interpretation of Catalan numbers.

**Keywords:** Lattice path, bijection, generating function, Fuss–Catalan number, binary matrix, convolutional code

## 1 Introduction

This article presents a comprehensive investigation of the number of down-steps between pairs of up-steps in a family of generalized Dyck paths known as  $k_t$ -Dyck paths. In particular, we obtain an explicit formula for this parameter, and explore its asymptotic growth. Throughout, we further establish properties of  $k_t$ -Dyck paths and their use in finding combinatorial interpretations of identities. To begin, we define these paths and associated objects, and provide background and motivation for studying this parameter.

**Definition 1** ( $k$ -Dyck path). Let  $k$  be a positive integer. A  $k$ -Dyck path is a lattice path that consists of up-steps  $(1, k)$  and down-steps  $(1, -1)$ , starts at  $(0, 0)$ , stays weakly above the line  $y = 0$  and ends on the line  $y = 0$ .

Notice that if a  $k$ -Dyck path has  $n$  up-steps, then it has  $kn$  down-steps, and thus has length  $(k + 1)n$ . The family of  $k$ -Dyck paths in this form were introduced by Hilton and Pedersen [HP91] and have since been the topic of several studies involving statistics in these paths and their relations to other mathematical objects (see, for example, [CM16, HHP18, JVK16, XZ19]). However, note that equivalent lattice path families, such as the family of paths consisting of horizontal and vertical unit steps staying weakly below the line  $y = kx$ , have been under consideration for much longer — see, for example, [Moh79, Section 1.4]

\*A. Asinowski, B. Hackl, and S. J. Selkirk are supported by the Austrian Science Fund (FWF): P 28466-N35, *Analytic Combinatorics: Digits, Automata and Trees*, S. J. Selkirk is supported by Austrian Science Fund (FWF): DOC 78, and A. Asinowski is supported by Austrian Science Fund (FWF): P 32731, *Generic Rectangulations: Enumerative and Structural Aspects*.

where this particular family is discussed, including references to older literature. More information can also be found in the recent survey article [Kra15] by Krattenthaler.

In particular,  $k$ -Dyck paths are a special case of Łukasiewicz paths. A *Łukasiewicz path* of length  $n$  is a lattice path consisting of steps  $(1, i)$  where  $i \in \{-1, 0, 1, 2, \dots\}$ , that starts at  $(0, 0)$ , stays weakly above the line  $y = 0$ , and terminates at  $(n, 0)$ . There is a well-known natural bijective mapping between Łukasiewicz paths and plane trees [Lot97, Chapter 11] (see also [BF02, Example 3]). Under this mapping,  $k$ -Dyck paths correspond to  $(k + 1)$ -ary trees, and thus  $k$ -Dyck paths of length  $(k + 1)n$  are enumerated by *Fuss–Catalan numbers* (see [FS09, Example I.14]) which are given by

$$C_n^{(k)} = \frac{1}{kn + 1} \binom{(k + 1)n}{n}. \quad (1)$$

In [GPW10], Gu, Prodinger, and Wagner introduced a class of  $k$ -coloured plane trees called  *$k$ -plane trees*, which are plane trees where each vertex is labeled  $i$  with  $i \in \{1, 2, \dots, k\}$  and the sum of labels along any edge is at most  $k + 1$  (with the root labeled  $k$ ). They construct a bijection between this class and  $k$ -Dyck paths as well as  $(k + 1)$ -ary trees. Furthermore, they note that removing the condition that the root is labeled  $k$  gives a bijection with generalized  $k$ -Dyck paths which can go below the  $x$ -axis until some fixed negative height. There they mention that only few enumeration problems lead to the numbers corresponding to the total number of generalized  $k$ -Dyck paths,

$$\frac{k}{n + 1} \binom{(k + 1)n}{n}.$$

We provide a new combinatorial interpretation for these numbers in (4).

Recently, this generalization of  $k$ -Dyck paths was further developed and enumerated by Selkirk in [Sel19], where such paths were given the name  $k_t$ -Dyck paths. It is this family of paths that will be the main object of interest in our article.

**Definition 2** ( $k_t$ -Dyck paths). Let  $k$  and  $t$  be integers with  $k > 0$  and  $t \geq 0$ . A  $k_t$ -Dyck path is a lattice path that consists of up-steps  $(1, k)$  and down-steps  $(1, -1)$ , starts at  $(0, 0)$ , stays weakly above the line  $y = -t$ , and ends on the line  $y = 0$ . The combinatorial class of  $k_t$ -Dyck paths will be denoted by  $S_t^{(k)}$ .

As in the case of  $k$ -Dyck paths, a  $k_t$ -Dyck path with  $n$  up-steps has  $kn$  down-steps, and thus its length is  $(k + 1)n$ .

For  $0 \leq t \leq k$ ,  $k_t$ -Dyck paths are enumerated by *generalized Fuss–Catalan numbers* (see for example A001764, A006013, A002293–A002296, A069271, A006632, A118969–A118971 in the OEIS [OEI22]), as shown in the next proposition.

**Proposition 3** ([Sel19, Proposition 2.2.2]). For  $0 \leq t \leq k$ , the generating function for  $k_t$ -Dyck paths (with the variable  $x$  counting the number of up-steps) is given by

$$S_t^{(k)}(x) = \sum_{n \geq 0} C_{n,t}^{(k)} x^n,$$

where the generalized Fuss–Catalan numbers  $C_{n,t}^{(k)}$  are given by

$$C_{n,t}^{(k)} = \frac{t + 1}{(k + 1)n + t + 1} \binom{(k + 1)n + t + 1}{n}. \quad (2)$$

For ease of notation we omit the superscript and write  $\mathcal{S}_t$ ,  $S_t(x)$ , or  $C_{n,t}$  if  $k$  is fixed and clear from the context. Note that for  $t = 0$  we obtain the “classical” Fuss–Catalan numbers as in (1).

This can be proved in a number of ways, as discussed in [Sel19], the simplest of those discussed being an application of the Cycle Lemma (see [DM47]; or [BF02, Section 4.1], [DZ90], [Lot97] for lattice path applications). The result can also be found in literature: see, for example, [Moh79, Section 1.4, Theorem 3; in particular (1.11)] for just the enumeration, or [Kra15, Theorem 10.4.5] where both the enumeration as well as the generating function representation is proved. We choose to include the proof using the Cycle Lemma here nonetheless, because it also sets the stage for some results later in this article.

**Proof of Proposition 3:** Trivially,  $k_t$ -Dyck paths of length  $(k+1)n$  are in bijection with reverse  $k_t$ -Dyck paths of length  $(k+1)n$  (with step set  $\{(1, 1), (1, -k)\}$ ). Using the Cycle Lemma, we will show that reverse  $k_t$ -Dyck paths of length  $(k+1)n$  are enumerated by the generalized Fuss–Catalan numbers given in (2). As outlined in [DZ90], a sequence  $p_1 \cdots p_n$  of boxes and circles is  $k$ -dominating if for every  $j$ ,  $1 \leq j \leq n$  the number of boxes in  $p_1 \cdots p_j$  is more than  $k$  times the number of circles.

The Cycle Lemma states that: “For any sequence  $p_1 \cdots p_{m+n}$  of  $m$  boxes and  $n$  circles,  $m \geq kn$ , there exist exactly  $m - kn$  (out of  $m + n$ ) cyclic permutations  $p_j \cdots p_{m+n} p_1 \cdots p_{j-1}$ ,  $1 \leq j \leq m + n$ , that are  $k$ -dominating.”

We claim that the number of unique  $k$ -dominating sequences of  $n$  circles and  $kn + t + 1$  boxes is equal to the number of reverse  $k_t$ -Dyck paths, where a box represents a  $(1, 1)$  step, and a circle represents a  $(1, -k)$  step in order. A  $k$ -dominating sequence of boxes and circles can be drawn as a path starting at  $(0, -(t+1))$  and ending at  $((k+1)n + t + 1, 0)$ , and its subpath from  $(t+1, 0)$  to  $((k+1)n + t + 1, 0)$  is a valid reverse  $k_t$ -Dyck path. We can take this subpath because every  $k$ -dominating sequence must begin with at least  $t+1$  boxes (or  $(1, 1)$  steps). Since the sequence is  $k$ -dominating, we have that  $k$  times the number of  $(1, 1)$  steps is always strictly greater than the number of  $(1, -k)$  steps, i. e.  $k \cdot \#(1, 1) + 1 \geq \#(1, -k)$ , after removing the initial  $t+1$  steps of type  $(1, 1)$  the path will possibly have a  $(1, -k)$  step at height  $k-t$  and we have the condition that the path stays weakly above the line  $y = -t$ . Thus unique  $k$ -dominating sequences of  $n$  circles and  $kn + t + 1$  boxes and reverse  $k_t$ -Dyck paths are in bijection.

The number of sequences of  $n$  circles and  $kn + t + 1$  boxes that can be constructed is

$$\binom{(k+1)n + t + 1}{n}.$$

By applying the Cycle Lemma we see that for each sequence there are exactly  $(kn + t + 1) - kn = t + 1$  starting points for cyclic permutations of the sequence which are  $k$ -dominating. However, these starting points might not give unique sequences due to periodicity of the sequence. On the other hand, all non-unique  $k$ -dominating cyclic permutations are periodic with a period that divides the length of the sequence. Therefore proportionally there are  $(t+1)/((k+1)n + t + 1)$  (unique)  $k$ -dominating cyclic permutations per sequence. Hence there are

$$\frac{t+1}{(k+1)n + t + 1} \binom{(k+1)n + t + 1}{n}$$

unique  $k$ -dominating sequences, which concludes the proof.  $\square$

*Remark 1.* In a recent preprint by Prodinger [Pro19] an approach for investigating  $k_t$ -Dyck paths with  $t > k$  using generating functions was discussed — note that the enumeration of these paths is known,

and can be obtained from, e.g., [Kra15, Theorem 10.4.7]. Combinatorially, removing the restriction of  $0 \leq t \leq k$  results in the paths behaving differently with respect to certain decompositions. As we will see later, the approaches presented in our paper cannot be applied directly to the case  $t > k$ : this is a subject of our ongoing research. Therefore throughout this work we assume that  $0 \leq t \leq k$ , unless stated otherwise.

In this paper, we study the number of down-steps between pairs of up-steps in  $k_t$ -Dyck paths.

**Definition 4** (Number of down-steps). Let  $n \geq 1$  and  $0 \leq r \leq n$  be two non-negative integers. Let  $s_{n,t,r}^{(k)}$  be the total number of all down-steps between the  $r$ -th and the  $(r+1)$ -th up-steps in all  $k_t$ -Dyck paths of length  $(k+1)n$ . For  $r = 0$  this quantity enumerates all down-steps before the first up-step, and for  $r = n$  all down-steps after the last up-step.

Whenever  $k$  is clear from the given context, we write  $s_{n,t,r}$  instead of  $s_{n,t,r}^{(k)}$ , to improve readability. Figure 1 illustrates all  $3_1$ -Dyck paths with  $n = 2$  up-steps. By counting the down-steps of the paths in the figure we can see that  $s_{2,1,0}^{(3)} = 4$ ,  $s_{2,1,1}^{(3)} = 16$ , and  $s_{2,1,2}^{(3)} = 34$ .

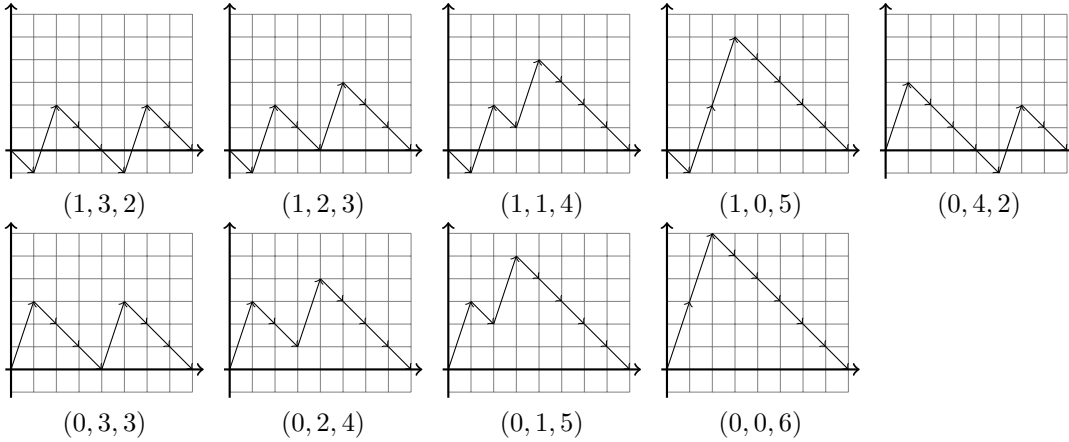


Fig. 1: All  $3_1$ -Dyck paths with two up-steps. The triple below each path indicates its contribution to  $s_{2,1,0}^{(3)}$ ,  $s_{2,1,1}^{(3)}$ ,  $s_{2,1,2}^{(3)}$  — that is, the number of down-steps before the first up-step, between the first and second up-step, and after the second up-step, respectively.

The motivation for studying down-steps in  $k_t$ -Dyck paths comes from trying to better understand and enumerate generalized Kreweras walks. Walks in the quarter plane have been and remain an active area of research, with many unresolved questions. Kreweras walks [Kre65] are walks in the quarter plane,  $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ , that consist of steps  $\{(1, 1), (-1, 0), (0, -1)\}$ , and start and end at  $(0, 0)$ .

We consider a generalized Kreweras model with step set  $\{(a, b), (-1, 0), (0, -1)\}$ ,  $a, b \in \mathbb{N}$ . These  $(a, b)$ -Kreweras walks can be decomposed into pairs consisting of an  $a$ -Dyck path and a  $b$ -Dyck path using an observation in [BBMM21, Proposition 21]. Given an arbitrary  $(a, b)$ -Kreweras walk, we sequentially consider the non-zero  $x$ -coordinates of steps in the walk, which will be either  $a$  or  $-1$ , and thus form a sequence which represents an  $a$ -Dyck path. Similarly, non-zero  $y$ -coordinates are either  $b$  or  $-1$  and thus form a  $b$ -Dyck path.

However, because not all the conditions of [BBMM21, Proposition 21] are met, this decomposition is not unique — two different walks can result in the same pair of  $a$ -Dyck and  $b$ -Dyck path. For example, consider the decomposition in Figure 2.

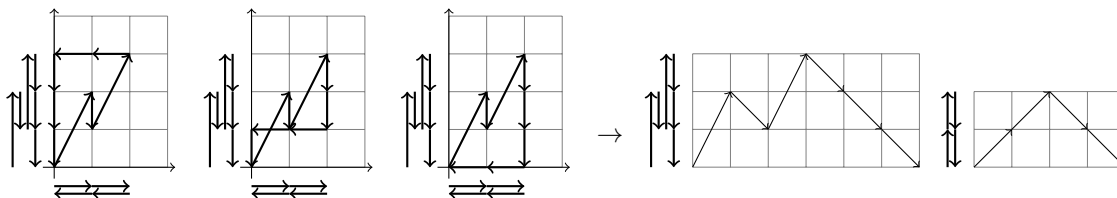


Fig. 2: Three different  $(1, 2)$ -Kreweras walks whose  $x$ - and  $y$ -projections yield the same pair of paths. In total there are ten  $(1, 2)$ -Kreweras walks yielding this pair of paths.

This is because there is no way of encoding the relative positions of  $(-1, 0)$  and  $(0, -1)$  steps between pairs of  $(a, b)$  steps, but in the  $a$ -Dyck and  $b$ -Dyck path the corresponding steps are down-steps. Since  $(a, b)$  steps result in an up-step in both generalized Dyck paths, the number of down-steps between pairs of up-steps in each path allows us to enumerate the number of combinations of  $(-1, 0)$  and  $(0, -1)$  steps that give the same  $a$ -Dyck and  $b$ -Dyck path, thus providing a basis for the enumeration of generalized Kreweras walks. We hope that a better understanding of the down-step statistics in  $k_t$ -Dyck paths will assist in the enumeration of this family of generalized Kreweras walks in the quarter plane.

## 2 Structure of the paper and summary of results

There are two different approaches used for investigating the number of down-steps. Firstly, in Section 3 we bijectively prove a recursion which gives an explicit sum formula for the total number of down-steps between pairs of up-steps. The second approach in Section 5 makes use of the symbolic decomposition of  $k_t$ -Dyck paths and uses bivariate generating functions to count the number of down-steps between pairs of up-steps. Combining results from both approaches leads to several interesting identities and observations. See Table 1 for a concise summary including references to related sequences in the OEIS [OEI22].

Section 4 outlines connections of both  $k_t$ -Dyck paths and their down-step statistics to another combinatorial object used in coding theory — a family of perforation patterns of punctured convolutional codes — which were originally studied by Bégin [Bég92]. These are binary matrices with a prescribed number of 1's and 0's, and we show that they are bijective to  $k_t$ -Dyck paths for  $0 \leq t < k$  (Theorem 15). As a consequence of this new link between binary matrices and lattice paths, a new interpretation of Catalan numbers is obtained; see Corollary 16 and Figure 8.

The generating function for the number of down-steps between the  $r$ -th and  $(r + 1)$ -th up-steps in  $k_t$ -Dyck paths can be found in Theorem 25. The asymptotic behavior related to the results in the table below can be found in a concise format in Section 6. The results are derived from the explicit formulas determined in Sections 3 and 5.

## 3 Down-step statistics: A bijective approach

The main result of this section is the recurrence relation for  $s_{n,t,r}$  given in (5) together with relevant initial values given in (3) and (4). This recurrence fully explains the combinatorial structure behind the quantities

Tab. 1: Summary of explicit formulas for  $s_{n,t,r}^{(k)}$ , the total number of down-steps between the  $r$ -th and  $(r+1)$ -th up-steps in  $k_t$ -Dyck paths of length  $(k+1)n$ , where  $k, n \geq 1$ ,  $0 \leq t \leq k$  and  $r$  are integers. Restrictions on  $r$  are listed in the table.

|  |                   |   |
|--|-------------------|---|
| Theorem 5 (1) and Theorem 21   | $r = 0$           | OEIS A001764, A002293, A002294,<br>A334785–A334787          |
| $s_{n,t,0}^{(k)} = \sum_{j=0}^{t-1} \frac{j+1}{(k+1)n+j+1} \binom{(k+1)n+j+1}{n} = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n}$  |                   |   |
| Corollary 12 and Corollary 27  | $0 \leq r < n$    | OEIS A007226, A007228, A124724,<br>A334640–A334651          |
| $s_{n,t,r}^{(k)} = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n} + \sum_{j=1}^r \frac{t+1}{(k+1)j+t+1} \frac{k}{n-j+1} \binom{(k+1)j+t+1}{j} \binom{(k+1)(n-j)}{n-j}$  |                   |   |
| Corollary 14   | $1 \leq r \leq n$ | OEIS A334976–A334980  |
| $s_{n,t,n-r}^{(k)} = \frac{t+1}{(k+1)(n+1)+t+1} \binom{(k+1)(n+1)+t+1}{n+1} - \frac{(k+1)(t+1)}{(k+1)n+t+1} \binom{(k+1)n+t+1}{n} - \sum_{j=2}^r \frac{1}{j-1} \binom{(j-1)(k+1)}{j} \frac{t+1}{(k+1)(n-j+1)+t+1} \binom{(k+1)(n-j+1)+t+1}{n-j+1}$ |                   |   |
| Proposition 13 and Proposition 18  | $r = n$           | OEIS A030983, A334609–A334612,<br>A334680, A334682, A334719 |
| $s_{n,t,n}^{(k)} = \frac{t+1}{(k+1)(n+1)+t+1} \binom{(k+1)(n+1)+t+1}{n+1} - \frac{(t+1)^2}{(k+1)n+t+1} \binom{(k+1)n+t+1}{n}$  |                   |   |

given by  $s_{n,t,r}$ .

**Theorem 5.** Let  $C_{n,j}$  be the generalized Fuss–Catalan numbers from (2),  $k \geq 1$  and  $0 \leq t \leq k$  be fixed integers, and  $n \geq 1$  be a positive integer. Then the following statements hold:

1. The total number of down-steps before the first up-step in all  $k_t$ -Dyck paths of length  $(k+1)n$  is given by

$$s_{n,t,0} = \sum_{j=0}^{t-1} C_{n,j}. \quad (3)$$

2. The total number of down-steps between the first and second up-steps in all  $k_0$ -Dyck paths is given by

$$s_{n,0,1} = \sum_{j=0}^{k-1} C_{n-1,j} = \frac{k}{n} \binom{(k+1)(n-1)}{n-1}. \quad (4)$$

3. For all positive integers  $r$  with  $1 \leq r \leq n$  and  $(t,r) \neq (0,1)$ , the following recurrence relation<sup>(i)</sup> holds:

$$s_{n,t,r} = s_{n,t,r-1} + C_{r,t}(s_{n-r+1,0,1} - t\llbracket r = n \rrbracket). \quad (5)$$

We prove all of the statements of this theorem by means of combinatorial arguments such as bijections and double counting. Additionally, we verify the recurrence relation for small values of  $n$ ,  $k$ , and  $t$  in an associated SageMath [The20] worksheet<sup>(ii)</sup>.

Let  $r$  be an integer where  $0 \leq r \leq n$ . We introduce  $r$ -marked  $k_t$ -Dyck paths as  $k_t$ -Dyck paths where exactly one of the down-steps between the  $r$ -th and the  $(r+1)$ -th up-steps is marked. By construction, the number of  $r$ -marked  $k_t$ -Dyck paths of length  $(k+1)n$  is equal to  $s_{n,t,r}$ .

Also, notice that for  $t = 0$  we have  $s_{n,0,0} = 0$ , and (5) does not produce  $s_{n,0,1}$ . Therefore for  $t = 0$  initial values  $s_{n,0,1}$  are explicitly given in (4).

**Proof of Theorem 5 (1):** For  $t = 0$  we have  $s_{n,0,0} = 0$  since  $k_0$ -Dyck paths cannot start with a down-step. Accordingly, the sum in the right-hand side of (3) is empty.

For  $t > 0$ , we prove (3) via a cyclic shift argument. Consider a 0-marked  $k_t$ -Dyck path where the  $j$ -th down-step at the beginning of the path is marked. We cut out the first  $j$  down-steps from the beginning of the path and attach them to the end of the path, removing the marking, and shift the path by  $j$  units upwards (to make it start and end at the  $x$ -axis). In this way the  $k_t$ -Dyck path is transformed to an unmarked  $k_{t-j}$ -Dyck path. Conversely, given a  $k_{t-j}$ -Dyck path, the corresponding 0-marked  $k_t$ -Dyck path can be constructed by marking the last step (which is always a down-step since  $j > 0$  and  $t - j < t \leq k$ ), removing the final  $j$  down-steps, adding them back to the beginning of the path, and relocating the newly constructed path such that it starts at the origin.

This proves that for  $t > 0$ , 0-marked  $k_t$ -Dyck paths are in bijection with the tuples in  $\cup_{1 \leq j \leq t} \{j\} \times \mathcal{S}_{t-j}$ , where  $j$  is required to make the union disjoint and to remember the number of steps moved to the end. Since  $\mathcal{S}_j$  is enumerated by the generalized Fuss–Catalan number  $C_{n,j}$  (see Proposition 3), this proves (3).  $\square$

<sup>(i)</sup> We use the Iverson bracket popularized in [GKP94, Chapter 2] where  $\llbracket A \rrbracket = 1$  if  $A$  is true, and  $\llbracket A \rrbracket = 0$  if  $A$  is false.

<sup>(ii)</sup> The interactive worksheets related to this article can be found at <https://gitlab.aau.at/behack1/kt-dyck-downstep-code>.

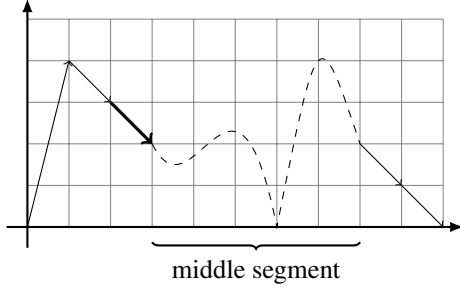


Fig. 3: Decomposition of a marked  $k_0$ -Dyck path: The middle segment corresponds to an elevated  $k_0$ -Dyck path.

Within the following proof, we show that the numbers  $s_{n,0,1}$  also enumerate an associated class of lattice walks which occurs in the proof of the recurrence in (5) (this also explains why these special values occur prominently in the recurrence relation). This class is introduced in the following definition.

**Definition 6** (Elevated  $k_0$ -Dyck path). An *elevated  $k_0$ -Dyck path* is a lattice path consisting of *up-steps*  $(1, k)$  and *down-steps*  $(1, -1)$ , that starts and ends at the line  $y = j$  for any integer  $0 \leq j < k$ , and stays weakly above the  $x$ -axis. The number  $j$  is called the *elevation* of the path.

The following observation relating  $k_t$ -Dyck paths to elevated  $k_0$ -Dyck paths is straightforward; nevertheless we choose to state it formally because we will reference it several times later on.

**Observation 7.** For each  $0 \leq j < k$ , the vertical shift by  $j$  units is a bijection between  $k_j$ -Dyck paths and elevated  $k_0$ -Dyck paths of elevation  $j$ . Therefore, the number of elevated  $k_0$ -Dyck paths of elevation  $j$  is  $C_{n,j}$  (see Proposition 3).

**Proof of Theorem 5 (2):** The path decomposition illustrated in Figure 3 shows that 1-marked  $k_0$ -Dyck paths with  $n$  up-steps are in bijection with elevated  $k_0$ -Dyck paths<sup>(iii)</sup> with  $n - 1$  up-steps: the middle segment starting at the end of the marked down-step and reaching until the final return to the same height is an elevated  $k_0$ -Dyck path with  $n - 1$  up-steps.

This proves that there are equally many down-steps after the first up-step in all  $k_0$ -Dyck paths with  $n$  up-steps as elevated  $k_0$ -Dyck paths with  $n - 1$  up-steps. By Observation 7 and summing over all possible elevations  $0 \leq j < k$ , this number is  $\sum_{j=0}^{k-1} C_{n,j}$ . Moreover, it was proved by Gu, Prodinger, and Wagner in [GPW10, Section 3] that the number of elevated  $k_0$ -Dyck paths with  $n - 1$  up-steps and elevation  $j$  with  $0 \leq j < k$  is

$$\frac{k}{n} \binom{(k+1)(n-1)}{n-1}, \quad (6)$$

which is equal to the expression on the right-hand side of (4). This completes the proof.  $\square$

The introduction of elevated  $k_0$ -Dyck paths together with the bijection constructed in the previous proof allows us to extend the statement in Theorem 5 (2).

<sup>(iii)</sup> Observe that as an elevated path with elevation  $j$  starts at  $(0, j)$ , the elevation is inherently stored within the path itself. This is in contrast to the paths from  $\mathcal{S}_j$  discussed in part (1) where we formally needed to construct a tuple storing the number of shifted steps  $j$  to make the union over several path classes  $\mathcal{S}_{t-j}$  disjoint.



**Corollary 8.** *Let  $k \geq 1$  be a fixed integer, and let  $n \geq 1$  be a positive integer. Then  $s_{n,0,1} = \frac{k}{n} \binom{(k+1)(n-1)}{n-1}$  enumerates 1-marked  $k_0$ -Dyck paths with  $n$  up-steps. Furthermore, these paths are in bijection with elevated  $k_0$ -Dyck paths with  $n - 1$  up-steps.*

*Remark 2.* Notice that our reasoning contains a bijective proof of the nice summation formula

$$\sum_{j=0}^{k-1} C_{n,j} = \sum_{j=0}^{k-1} \frac{j+1}{(k+1)n+j+1} \binom{(k+1)n+j+1}{n} = \frac{k}{n+1} \binom{(k+1)n}{n}. \quad (7)$$

As can be expected from its simple form, this identity is well-known and can also be proved using algebraic techniques. For some fixed values of  $k$  and  $n$ , these numbers can be found in OEIS entries [A007226](#), [A007228](#), [A124724](#).

Moreover, in the formula  $s_{n,t,0} = \sum_{j=0}^{t-1} C_{n,j}$  we have a partial sum of  $s_{n+1,0,1} = \sum_{j=0}^{k-1} C_{n,j}$ . In the special case  $t = k$  this yields  $s_{n,k,0} = s_{n+1,0,1}$ . In Proposition 11 we return to and generalize this equality, and provide a bijective proof.

Now that we have proved the initial values of  $s_{n,t,r}$ , we proceed with the proof of the recurrence.

**Proof of Theorem 5 (3):** As before, we represent the left-hand side of (5),  $s_{n,t,r}$ , by the number of  $r$ -marked  $k_t$ -Dyck paths of length  $(k+1)n$  (consisting of  $n$  up-steps). In Claims 9 and 10 we prove that the right-hand side of (5) is also equal to  $s_{n,t,r}$ , with each summand counting paths whose marked step ends in a specified height interval.

**Claim 9.** *The term  $s_{n,t,r-1}$  enumerates  $r$ -marked  $k_t$ -Dyck paths where the marked step ends on height  $h$  where  $h \geq k - t$ .*

**Proof:** We construct a bijection that maps  $(r-1)$ -marked  $k_t$ -Dyck paths to  $r$ -marked  $k_t$ -Dyck paths where the marked down-step ends at a height of at least  $k - t$ . We do this by shifting the  $(r-1)$ -th up-step so that it becomes the  $r$ -th one. Effectively, this moves a peak<sup>(iv)</sup> in the path to another position, thus we will refer to the procedure as the “peak shift” bijection.

Consider any  $(r-1)$ -marked  $k_t$ -Dyck path. Let  $h$  be the height on which the last down-step between the  $(r-1)$ -th and the  $r$ -th up-steps ends, then traverse the path from the end of this last downstep towards the beginning. The traversal ends if

- we reach a height of  $h$  or below again (as in Figure 4a), or
- we reach the beginning of the path without doing so (as in Figure 4b).

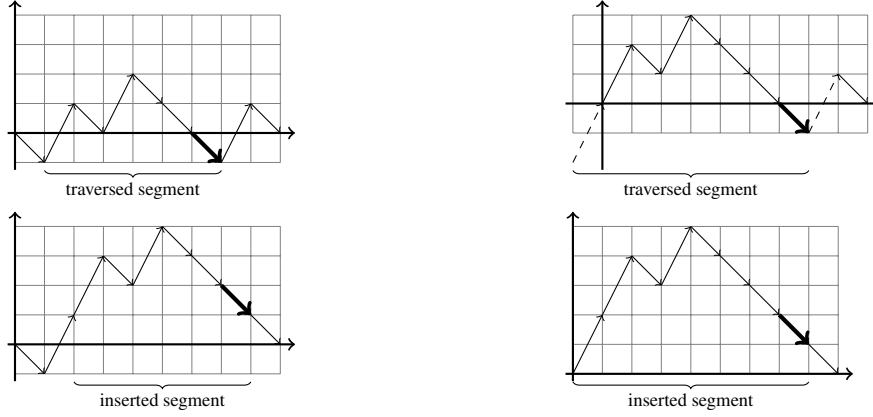
In the second case, prepend the  $r$ -th up-step to the beginning of the traversed segment. This ensures that the traversed segment starts at a lower or equal height compared to its end.

The path is then transformed by removing the traversed segment and inserting it at the rightmost position located above height  $k - t$  after the former  $r$ -th up-step and before the  $(r+1)$ -th up-step (should it exist). This results in an  $r$ -marked path, with the marked step guaranteed to end above or at height  $k - t$ .

This transformation can always be reversed: consider an  $r$ -marked  $k_t$ -Dyck path whose marked down-step ends on a height of at least  $k - t$ . By construction, in the maximal consecutive sequence of down-steps which contains the marked down-step, all steps ending at a height of at least  $k - t$  must have been

---

<sup>(iv)</sup> A peak in a  $k_t$ -Dyck path is an up-step followed by a down-step.



(a) The ending height of the down-step sequence containing the marked down-step is reached earlier in the path.

(b) The ending height of the down-step sequence containing the marked down-step is not reached earlier in the path; the traversed segment extends until the beginning of the path. The dashed arrow represents the “borrowed” up-step.

Fig. 4: Two examples of the “peak shift” bijection mapping  $(r-1)$ -marked paths to  $r$ -marked paths whose marked step ends on height  $k-t$  or above. The parameters in the illustrations correspond to  $k=2$ ,  $t=1$ , and  $n=r=3$ .

included in the relocated traversed segment; let  $h$  be this height. Start at the down-step ending at height  $h$  and traverse the path towards the beginning until you reach height  $h$  (or below) again.

If the traversed segment spans until the beginning of the path, or if there are no further up-steps before the traversed segment, then this instance corresponds to the second case — otherwise it corresponds to the first case. Then, the transformation can be reversed by relocating the traversed segment (and for the second case additionally adjusting the initial up-step).

This establishes that the transformation is indeed a bijection, and thus proves the claim.  $\square$

**Claim 10.** *The term  $C_{r,t}(s_{n-r+1,0,1} - t\llbracket r = n \rrbracket)$  enumerates  $r$ -marked  $k_t$ -Dyck paths where the marked step ends on height  $h$  where  $-t \leq h < k-t$ .*

**Proof:** We use a similar “middle segment”-decomposition as in the proof of Theorem 5 (2).

Let us first consider the case of  $r < n$ . If the marked step ends above or on the  $x$ -axis, we decompose the path into a “middle segment” starting after the marked down-step (which means it has a height between 0 and  $k-t-1$ ) and ending with the final return to the same height it started at. If the marked step ends below the  $x$ -axis, the middle segment extends until the end of the path. To have the segment end on the same height it starts at in both cases, we remove the down-steps from the  $x$ -axis until the start of the middle segment and attach them again at the end of the middle segment. This middle segment is an elevated  $k_0$ -Dyck path with  $n-r > 0$  up-steps which has been shifted downwards by  $t$  units, which is visualized in Figure 5. By Corollary 8, the number of such paths is  $s_{n-r+1,0,1}$ . Now consider the path obtained after removing the middle segment: it is precisely a  $k_t$ -Dyck path with  $r$  up-steps, and there are  $C_{r,t}$  such paths. Note that we can ignore the marking because its location is already encoded by the elevation of the middle segment.

For the case of  $r = n$  we can use the same decomposition — however, in this case the “middle segment” is empty. There are  $s_{1,0,1} = k$  such paths (the empty path with different elevations) in total, but as the final sequence of down-steps in the original path ends at height  $y = 0$ , the shifted elevated paths with elevation  $0, 1, \dots, t - 1$  cannot occur — which is why these  $t$  paths have to be excluded. This explains the occurrence of the Iverson bracket in (5).

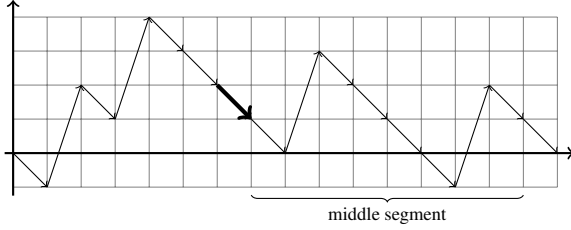


Fig. 5: Illustration of the “middle segment” (an elevated  $3_0$ -Dyck path of elevation 2) in a 2-marked  $3_1$ -Dyck path. Removing the segment yields a  $3_1$ -Dyck path with 2 up-steps. The marked step can be identified uniquely from the elevation of the middle segment, which allows us to forget the marking.

By following the steps above in reverse order, the original path can be reconstructed and so the decomposition is a bijection between an  $r$ -marked  $k_t$ -Dyck path and both a shorter  $k_t$ -Dyck path and an elevated  $k_0$ -Dyck path. This proves the claim.  $\square$

Now, with the statements of Claims 9 and 10 we see that the right-hand side of (5) enumerates all  $r$ -marked  $k_t$ -Dyck paths — as does the corresponding left-hand side. This proves the recurrence and completes the proof of Theorem 5.  $\square$

*Remark 3.* As mentioned in Remark 1, Theorem 5 is not valid for  $t > k$ . This is because for  $t > k$  the combinatorial structure of these paths changes: it is no longer guaranteed that the path is on or above the  $x$ -axis after an up-step. For the same reason, the formula from Proposition 3 is invalid for  $t > k$ . In this case,  $k_t$ -Dyck paths are not enumerated by generalized Fuss–Catalan numbers, but rather by weighted sums thereof (see [Pro19] for details).

There is more about the case of  $t = k$  that is worth mentioning from a combinatorial point of view.

*Remark 4.* Observe that for  $t = k$  and  $r = n$ , the recurrence (5) degenerates to  $s_{n,k,n} = s_{n,k,n-1}$ . Combinatorially, this happens because the summand treated by Claim 10 corresponds to the number of down-steps before the last return to the axis that are at the same time located below the axis — which is not possible. This formula also has a bijective interpretation: observe that if one exchanges the down-steps before and after the last up-step in a  $k_k$ -Dyck paths, another valid  $k_k$ -Dyck path is obtained.

The cases  $t = 0$  and  $t = k$  are inherently linked, as the next proposition shows.

**Proposition 11.** *Let  $n \geq 1$ . Then the relation  $s_{n+1,0,r+1} = s_{n,k,r}$  holds for all integers  $r$  with  $0 \leq r < n$ , and for  $r = n$  we have  $s_{n+1,0,n+1} = s_{n,k,n} + kC_{n,k}$ .*

**Proof:** Let  $0 \leq r \leq n$ . Notice that if we take a  $k_k$ -Dyck path of length  $(k+1)n$ , add a new up-step to the beginning and  $k$  down-steps to the end, the result is a  $k_0$ -Dyck path of length  $(k+1)(n+1)$ . Removing

the initial up-step and the final  $k$  down-steps from a  $k_0$ -Dyck path reverses this construction and yields a  $k_k$ -Dyck path, thus forming a bijection.

This bijection transforms the  $r$ -th up-step in a  $k_k$ -Dyck path to the  $(r + 1)$ -th up-step in the resulting  $k_0$ -Dyck path, which directly proves  $s_{n+1,0,r+1} = s_{n,k,r}$  for  $0 \leq r < n$ . For  $r = n$ , the additional summand  $kC_{n,k}$  reflects the fact that  $k$  additional down-steps are appended to the end of every  $k_k$ -Dyck path.  $\square$

With the recurrence from Theorem 5 we can find an explicit summation formula for  $s_{n,t,r}$ .

**Corollary 12.** *For positive integers  $k$  and  $n$ , and integers  $r$  and  $t$  with  $0 \leq r < n$  and  $0 \leq t \leq k$ , the number of down-steps between the  $r$ -th and  $(r + 1)$ -th up-steps in  $k_t$ -Dyck paths of length  $(k + 1)n$  can be expressed explicitly as*

$$s_{n,t,r} = \sum_{j=0}^{t-1} C_{n,j} + \sum_{j=1}^r C_{j,t} s_{n-j+1,0,1}. \quad (8)$$

The above expression can be found in terms of binomial coefficients in Table 1. These sequences for small cases of  $k$  and  $t$  have been added to the OEIS, see entries A007226, A007228, A124724, A334640–A334651.

Observe that for easier reading, we restricted ourselves to the case of  $r < n$ . A similar explicit formula can also be given for  $r = n$ . However, we can find a better formula for this boundary case via a different combinatorial argument.

**Proposition 13.** *The number of down-steps following the last up-step in  $k_t$ -Dyck paths satisfies*

$$s_{n,t,n} = C_{n+1,t} - (t + 1)C_{n,t} \quad (9)$$

for  $n \geq 1$ , with initial value  $s_{0,t,0} = 1$ .

**Proof:** Consider the rearranged form

$$C_{n+1,t} = (t + 1)C_{n,t} + s_{n,t,n}. \quad (10)$$

The left-hand side of (10) enumerates  $k_t$ -Dyck paths with  $n + 1$  up-steps. We claim that the right-hand side does the same, where the summands correspond to different ways of expanding a  $k_t$ -Dyck path with  $n$  up-steps by one additional up-step.

In particular, the quantity  $s_{n,t,n}$  enumerates  $n$ -marked  $k_t$ -Dyck paths with  $n$  up-steps. Such a path can be extended by replacing the marked down-step by an up-step, and adding  $k + 1$  down-steps after it.

The longer paths that cannot be constructed in this way are those where the  $(n + 1)$ -th up-step starts on height  $h$  with  $-t \leq h \leq 0$ . However, we are able to construct them by starting with an arbitrary  $k_t$ -Dyck path with  $n$  up-steps (enumerated by  $C_{n,t}$ ), appending between 0 and  $t$  down-steps to the end of the path (which gives  $t + 1$  possibilities) followed by the  $(n + 1)$ -th up-step and sufficiently many down-steps to make it a  $k_t$ -Dyck path again.

This proves (10), and thus (9).  $\square$

Finally, we combine Proposition 13 with Theorem 5 (2, 3) to obtain another formula for  $s_{n,t,n-r}$ .

**Corollary 14.** For  $1 \leq r \leq n$  we have

$$s_{n,t,n-r} = C_{n+1,t} - (k+1)C_{n,t} - \sum_{j=2}^r \frac{1}{j-1} \binom{(j-1)(k+1)}{j} C_{n-j+1,t}. \quad (11)$$

**Proof:** For  $r = 1$ , we can combine the expressions in (4), (5), and (9) to obtain

$$s_{n,t,n-1} = s_{n,t,n} - (s_{1,0,1} - t)C_{n,t} = C_{n+1,t} - (t+1)C_{n,t} - (k-t)C_{n,t} = C_{n+1,t} - (k+1)C_{n,t}.$$

Similarly, when  $r > 1$  we can use induction to obtain

$$\begin{aligned} s_{n,t,n-r} &= s_{n,t,n-(r-1)} - s_{r,0,1}C_{n-(r-1),t} = C_{n+1,t} - (k+1)C_{n,t} - \\ &\quad - \sum_{j=2}^{r-1} \frac{1}{j-1} \binom{(j-1)(k+1)}{j} C_{n-j+1,t} - \frac{1}{r-1} \binom{(r-1)(k+1)}{r} C_{n-(r-1),t}, \end{aligned}$$

where we use the identity  $\frac{1}{b+1} \binom{a}{b} = \frac{1}{a-b} \binom{a}{b+1}$  for positive integers  $a > b$ . This yields (11).  $\square$

Asymptotic investigations into the quantities discussed in this section will be conducted in Section 6.

## 4 A connection to coding theory: Binary matrices and punctured codes

In this section we point out a surprising link between the down-step parameter  $s_{n,t,r}$ , and punctured convolutional codes which are an important concept in coding theory.

### 4.1 Perforation patterns of punctured convolutional codes

Error correcting codes are a tool for adding redundancy to (binary) data before transmitting it via some channel that may introduce transmission errors. There are two large families of codes: block codes, which map a fixed number of input bits to their respective encoding; and convolutional codes, which process an input stream by producing an output for overlapping input blocks (these “windows” move one bit at a time). See [JZ99] for more basic details on error correcting codes.

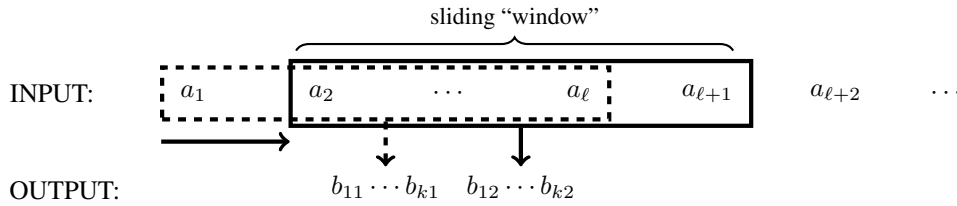


Fig. 6: Diagram of the input-output process of convolutional codes: input bits  $a_1 a_2 \dots a_l$  produce output bits  $b_{11} b_{21} \dots b_{k1}$ .

Comparing these two basic code design concepts under the assumption that both codes produce a similar number of redundancy bits per block, convolutional codes will be much longer and contain more

redundancy. This is because the input blocks overlap, and thus each bit is used multiple times when creating the output.

This is why for convolutional codes, the idea of simply deleting some of the output bits (in a systematic way) is very practical (see [JZ99, Chapter 4.1]) as many error correction and performance properties are preserved, and yet the code is made more efficient. Let  $k, n, r \in \mathbb{N}$ . We refer to convolutional codes that generate  $k$  output bits per input window as  $(k, 1)$  convolutional codes. By collecting  $n$  successive input blocks, encoding them to obtain  $kn$  output bits and deciding for all  $1 \leq i \leq k$  and  $1 \leq j \leq n$  whether the  $i$ -th bit resulting from the  $j$ -th input window should be kept or deleted, a *punctured  $(r, n)$  convolutional code* is created. Here,  $r$  refers to the number of output bits that are selected to remain, and  $n$  indicates the number of input windows.

The systematic way of choosing the bits that are kept and form a punctured  $(r, n)$  convolutional code can be represented by a binary  $k \times n$  matrix

$$P = (p_{ij})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq n}} \quad \text{where} \quad p_{ij} = \begin{cases} 1 & \text{if } i\text{-th output bit of } j\text{-th block is kept,} \\ 0 & \text{otherwise,} \end{cases}$$

that has exactly  $r$  entries that are 1. This matrix is called the *perforation pattern* of the code.

As an example, consider the perforation pattern

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

When applied to a  $(3, 1)$  convolutional code, a punctured  $(6, 4)$  code is created which assembles its output by deleting the second bit in the first output block, the first and third bit from the second block, the first two bits from the third block, and the second bit from the fourth block. Considering output similar to that in Figure 6, the output  $b_{11}b_{21}b_{31}b_{12}b_{22}b_{32}b_{13}b_{23}b_{33}b_{14}b_{24}b_{34}$  would be ‘punctured’ to form the new output  $b_{11}b_{31}b_{22}b_{33}b_{14}b_{34}$ .

In [BH89] two perforation patterns are considered to be equivalent if they can be obtained from each other by a cyclic permutation of columns, inducing an equivalence relation on the set of perforation patterns (and therefore also on the corresponding binary matrices). This is because two different punctured codes that are constructed from a given convolutional code by applying equivalent perforation patterns have exactly the same distance properties, and thus the same error correction capabilities.

The OEIS [OEI22] contains several sequences concerned with the enumeration of non-equivalent perforation patterns of punctured convolutional codes, namely A007223–A007229. According to these OEIS entries, the corresponding sequences were investigated by B egin and presented at a conference in 1992 [B eg92], but not published otherwise.

Note once again, the objects of interest — non-equivalent punctured convolutional codes — can be described by equivalence classes of binary matrices. With the coding theoretic background in mind, we take a closer look at these matrices to find the connection between them and our down-step parameter.

## 4.2 Equivalence classes of binary matrices

In this section we construct a bijection between elevated  $k_0$ -Dyck paths (Definition 6) with  $n$  up-steps and  $\{0, 1\}$ -matrices of dimension  $(k + 1) \times n$  with precisely  $n + 1$  many 1’s, considered up to cyclic

permutation of columns. This will imply that  $\frac{1}{n} \binom{(k+1)n}{n+1}$  enumerates<sup>(v)</sup> these paths. Therefore we re-establish the formula (6) due to [GPW10], while contributing a simple combinatorial interpretation of the factor  $\frac{1}{n}$  in the first formula. For the codes, it follows that for each fixed  $k$ , the sequence  $(s_{n,0,1}^{(k)})_{n \geq 1}$  (Corollary 8) enumerates perforation patterns for punctured convolutional  $(n+1, n)$  codes: see A007226 for  $k=2$ , A007228 for  $k=3$ , A124724 for  $k=4$ ; these sequences are also columns of the rectangular array A241262. As a by-product, we obtain **a new result in the spirit of the Cycle Lemma**, and **a new structure enumerated by Catalan numbers** (see details below).

We introduce the following notions.

- Let  $\mathcal{E}_n^k$  be the family of elevated  $k_0$ -Dyck paths with  $n$  up-steps and  $kn$  down-steps.
- Let  $\mathcal{M}_n^k$  be the family of  $\{0, 1\}$ -matrices of order  $(k+1) \times n$  with  $n+1$  many 1's and  $kn-1$  many 0's.
- We say that two matrices in  $\mathcal{M}_n^k$  are *CPC-equivalent* if they can be obtained from each other by a *cyclic permutation of columns* (it is easy to see that this is indeed an equivalence relation). For example, the matrices  $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ ,  $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$  are CPC-equivalent in  $\mathcal{M}_3^1$ .
- Let  $\overline{\mathcal{M}}_n^k$  be the quotient set of  $\mathcal{M}_n^k$  with respect to the CPC equivalence relation.

**Theorem 15.** *For each  $k \geq 1$ ,  $n \geq 1$  we have*

$$|\mathcal{E}_n^k| = \frac{1}{n} \binom{(k+1)n}{n+1}. \quad (12)$$

**Proof:** The proof consists of two parts: First, we show that  $|\overline{\mathcal{M}}_n^k| = \frac{1}{n} \binom{(k+1)n}{n+1}$ ; and then we construct a bijection between  $\overline{\mathcal{M}}_n^k$  and  $\mathcal{E}_n^k$ .

To begin with, we clearly have  $|\mathcal{M}_n^k| = \binom{(k+1)n}{n+1}$ . Since in a matrix  $A \in \mathcal{M}_n^k$  we have  $n+1$  many 1's distributed over  $n$  columns, it is not possible that a non-trivial cyclic permutation of the columns of  $A$  will yield  $A$  itself. (Indeed, if  $d$  is the smallest positive number such that cyclic shifting of  $A$  by  $d$  columns yields  $A$ , then  $A$  splits into  $c = n/d$  identical blocks, and we have  $c \mid n$ ; but considering the number of 1's in each block we also have  $c \mid n+1$ , which necessarily yields  $c = 1$  and  $d = n$ .) Therefore each equivalence class consists of precisely  $n$  matrices, and we have  $|\overline{\mathcal{M}}_n^k| = \frac{1}{n} \binom{(k+1)n}{n+1}$ .

Next, we consider two mappings,  $f$  and  $g$ , from  $\mathcal{M}_n^k$  to the set of lattice paths with up-steps  $(1, k)$  and down-steps  $(1, -1)$ : a specialization of the latter one will yield the desired bijection between  $\overline{\mathcal{M}}_n^k$  and  $\mathcal{E}_n^k$ . For  $A \in \mathcal{M}_n^k$ , we define  $f(A)$  and  $g(A)$  as follows:

- Read the entries of  $A$  column by column (read each column from top to bottom, and the columns from left to right). Starting at  $(0, k)$ , draw an up-step for each 1 entry, a down-step for each 0 entry. This yields a path from  $(0, k)$  to  $((k+1)n, 2k+1)$ : we denote this path by  $f(A)$ .
- In  $f(A)$ , replace the first up-step by a down-step. The part of the path to the left of this step does not change, and the part to the right of this step is accordingly translated  $k+1$  units downwards. At this point we obtain a path from  $(0, k)$  to  $((k+1)n, k)$ .

---

<sup>(v)</sup> Observe that this is compatible with Corollary 8 as  $\frac{1}{n} \binom{(k+1)n}{n+1} = \frac{k}{n+1} \binom{(k+1)n}{n}$ .

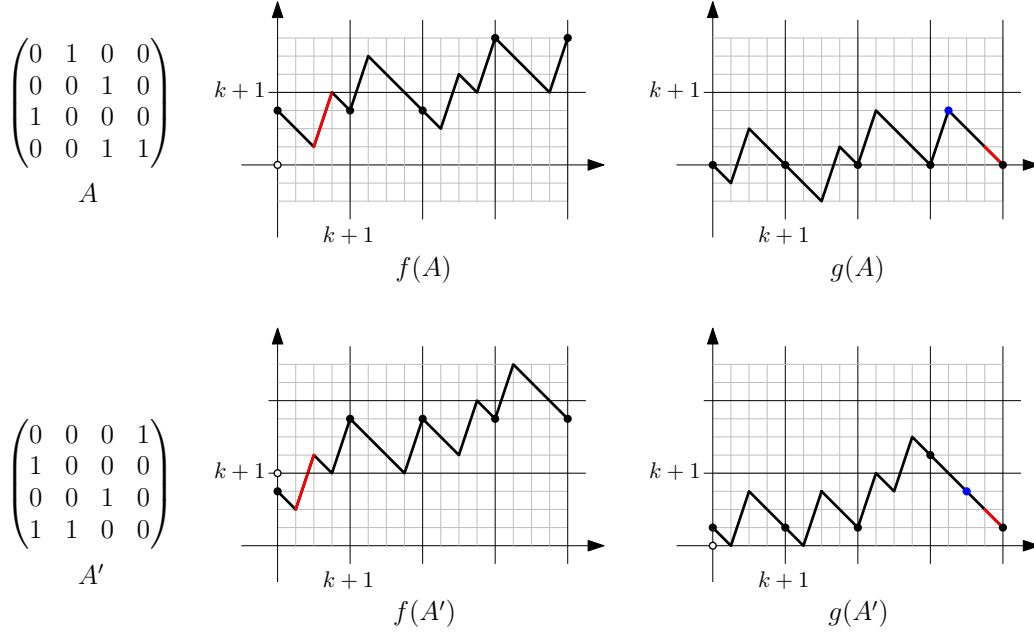


Fig. 7: An illustration to the proof of Theorem 15. The mappings  $f$  and  $g$  from matrices to paths are defined in the proof. The matrices  $A$  and  $A'$  are CPC-equivalent.  $A'$  is valid, hence  $g(A')$  is an elevated  $k_0$ -Dyck path.

- Shift this path to the left so that the right endpoint of the modified down-step will lie on the  $y$ -axis. Finally, remove the part of the path to the left of the  $y$ -axis, and attach it to the right of the remaining path. This yields a path from  $(0, j)$  to  $((k+1)n, j)$  for some  $j \leq k-1$ : we denote this path by  $g(A)$ .

See Figure 7 for two examples.

For  $\alpha = 1, 2, \dots, n+1$ , we call a node in  $f(A)$  (or in  $g(A)$ ) with  $x$ -coordinate  $(\alpha-1)(k+1)$  the  $\alpha$ -th principal node: these nodes are marked in the figure. Notice that the mapping  $f$  preserves the cyclic structure of the matrix, in the following sense. If  $A'$  is obtained from  $A$  by the cyclic permutation of columns such that the  $\alpha$ -th column of  $A$  is the first column of  $A'$ , then one obtains  $f(A')$  from  $f(A)$  by cutting  $f(A)$  at the  $\alpha$ -th principal node, exchanging the left and the right parts and joining them, and positioning the path so that its starting point is again  $(0, k)$ . This is also illustrated in Figure 7.

It follows directly from the definitions of  $f$  and  $g$  that these mappings are injective. By definition,  $g(A)$  is a path from  $(0, j)$  to  $((k+1)n, j)$  for some  $j \leq k-1$ . However it is not always the case that  $g(A)$  lies (weakly) above the  $x$ -axis (in fact,  $j$  can even be negative). We say that  $B \in \mathcal{M}_n^k$  is valid if  $g(B)$  is an elevated  $k_0$ -Dyck path ( $g(B) \in \mathcal{E}_n^k$ ). We will prove that each CPC-equivalence class in  $\mathcal{M}_n^k$  has a unique valid representative. In other words: for each  $A \in \mathcal{M}_n^k$  there is a unique valid  $B$  obtained from  $A$  by a (possibly trivial) cyclic permutation of columns.

First we notice that  $B \in \mathcal{M}_n^k$  is valid if and only if  $f(B)$  is weakly above the line  $y = k+1$  after its



first up-step. Moreover, each principal node of  $f(B)$  has  $y$ -coordinate of the form  $\beta(k+1) - 1$ , and this implies one more equivalent condition:  $B \in \mathcal{M}_n^k$  is valid if and only if all the principal nodes of  $f(B)$  (except for the starting point of the path) have  $y$ -coordinates of the form  $\beta(k+1) - 1$  with  $\beta \geq 2$ .

Let  $A \in \mathcal{M}_n^k$ . Let  $\beta_0$  be the *smallest* number such that some principal node (including the starting point) of  $f(A)$  has  $y$ -coordinate  $\beta_0(k+1) - 1$ . Next, let

$$\alpha_0 = \max \{ \alpha : ((\alpha - 1)(k + 1), \beta_0(k + 1) - 1) \text{ is a principal node} \},$$

and then consider  $A'$  to be the matrix that is CPC-equivalent to  $A$  where the  $\alpha_0$ -th column of  $A$  is the first column of  $A'$ . As a consequence, all the principal nodes in  $f(A')$  except for the starting point have  $y$ -coordinates of the form  $\beta(k+1) - 1$  with  $\beta \geq 2$ . Therefore  $A'$  is a valid matrix.

It is also easy to show that in each CPC-equivalence class there is at most one valid matrix. Indeed, assume that  $A$  is valid. Then all the principal nodes of  $f(A)$  (except for the starting point) have  $y$ -coordinates of the form  $\beta(k+1) - 1$  with  $\beta \geq 2$ . If some non-trivial shift of  $A$  is also valid, then, starting at some  $\alpha < n + 1$ , all the principal nodes have  $y$ -coordinates of the form  $\beta(k+1) - 1$  with  $\beta \geq 3$ . However, this contradicts the fact that  $f(A)$  ends at  $((k+1)n, 2k+1)$ .

Finally, we can reverse the process by taking  $P \in \mathcal{E}_n^k$ , cutting it at the point with  $y$ -coordinate  $k$  which lies at the last string of down-steps, modify/reconnect pieces accordingly, and read out the matrix. Such a cutting point (it is shown by blue colour in Figure 7) always exists, and it is not the end-point of the path: this follows from the fact that  $P$  terminates at height  $j$ , where  $0 \leq j \leq k - 1$ , and that the last string of down-steps in  $P$  starts at height at least  $k$ . Thus we have a bijection between the set of valid matrices in  $\mathcal{M}_n^k$  and  $\mathcal{E}_n^k$ . This shows that each CPC-equivalence class contains exactly one valid matrix, and completes the proof.  $\square$

*Remark 5.* (a) The claim that each CPC-equivalence class contains precisely one valid matrix is a *Cycle Lemma*-type result. The classical Cycle Lemma [DM47] determines bounds on the number  $N$  of those cyclic permutations of a given  $\{0, 1\}$ -sequence that satisfy the property: In each prefix, we have  $\#1 - \alpha \cdot \#0 > 0$  (or  $\#1 - \alpha \cdot \#0 \geq 0$ ) for a fixed number  $\alpha$ , where  $\#0$  and  $\#1$  denote the number of zeros and ones in the prefix, respectively. It is possible to interpret such results in terms of lattice paths, where the inequality translates to some version of non-negativity of a path. In our case the condition is that the sequence (extracted from  $A$  column-by-column) is such that in each prefix *that contains at least one 1*, we have  $k + (k \cdot \#1 - \#0) \geq k + 1$  (see Figure 7), or, equivalently,  $k \cdot \#1 - \#0 > 0$ . So, if we have a sequence of  $n + 1$  many 1's and  $kn - 1$  many 0's, and partition it into  $n$  consecutive blocks of  $k + 1$  entries, then there is precisely one cyclic permutation *of the blocks* that satisfies the property above. To summarize, the difference between our result and the classical Cycle Lemma is that in our case (1) the property holds *with delay* (meaning that the prefix has to contain at least one 1), and (2) we cyclically permute *blocks*.

- (b) Recall that for the number of elevated  $k_0$ -Dyck paths we also have the formula  $\sum_{j=0}^{k-1} C_{n,j}$ , where  $C_{n,j}$  enumerates elevated  $k_0$ -Dyck paths of elevation  $j$  ( $0 \leq j \leq k - 1$ ), or equivalently,  $k_j$ -Dyck paths (see Corollary 8). In our bijection between elevated  $k_0$ -Dyck paths and valid binary matrices, the elevation of a path corresponds to  $k$  minus the position of the first 1 in the matrix (converting it into a sequence as above). This yields a formula for valid matrices with fixed position of the first 1.
- (c) For  $k = 1$ , the CPC-equivalence classes are in bijection with paths from  $\mathcal{E}_n^1$ , which are just classical Dyck paths of length  $2n$ , and, thus, enumerated by Catalan numbers (see A000108 in the OEIS). This

means that we have found yet another combinatorial family enumerated by Catalan numbers. To the best of our knowledge, this interpretation has not been mentioned before in literature.

**Corollary 16 (New interpretation of Catalan numbers).** *The family of equivalence classes of binary  $2 \times n$ -matrices with  $n + 1$  many 1's, where two matrices are equivalent if they can be obtained from each other by a cyclic permutation of columns, is enumerated by the  $n$ -th Catalan number  $C_n$ .*

The bijection between Dyck paths with  $n$  up-steps and valid matrices in  $\mathcal{M}_n^1$  can even be described in a simplified form. Given a Dyck path of length  $2n$ , the associated valid  $2 \times n$  matrix is constructed as follows: first, put a 1 in the first row and first column. The remaining entries are filled column-by-column, from top to bottom by reading the Dyck path from left to right and putting a 1 for an up-step, and a 0 for a down-step. The last step of the path is ignored. See Figure 8 for an illustration.

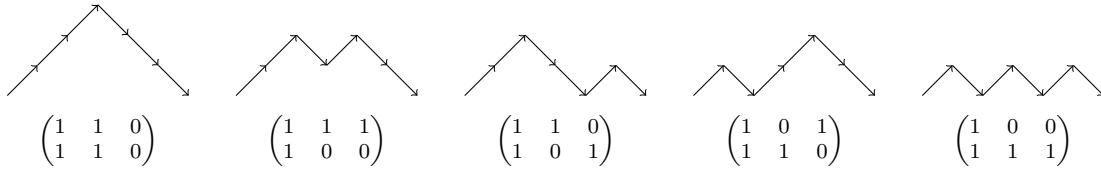


Fig. 8: The  $C_3 = 5$  Dyck-paths with 3 up-steps, each with their associated valid matrix from  $\mathcal{M}_3^1$ .

The above investigations were for the case where we have precisely  $n + 1$  many 1's in the binary matrices. However, we can also handle the case where we have an arbitrary number of 1's in the matrix by making use of the Pólya theory framework (see [Pól37] for Pólya's original work and [FS09, Remark I.60] for a useful statement in terms of generating functions and combinatorial classes). This has also been observed in [Bég92].

**Proposition 17.** *Let  $k, n \in \mathbb{N}$  be fixed positive integers and let  $z$  be a symbolic variable. Then with  $\varphi$  as Euler's totient function, the generating polynomial*

$$M_{k,n}(z) = \frac{1}{n} \sum_{d|n} \varphi(d) (1 + z^d)^{(k+1)n/d} \quad (13)$$

*enumerates all equivalence classes of  $(k + 1) \times n$  binary matrices under the CPC-equivalence where the power of  $z$  corresponds to the number of 1's.*

**Proof:** Let  $x_1, \dots, x_n$  be symbolic variables. It is a well-known fact [FS09, Remark I.60] that the cycle index of the group of cyclic permutations of length  $n$  is given by

$$Z(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{d|n} \varphi(d) x_d^{n/d}. \quad (14)$$

The Pólya–Redfield theorem assumes the following setting: We consider a family of combinatorial objects (binary matrices of dimension  $(k + 1) \times n$ ) that are constructed from  $n$  other “building blocks” (the columns of the binary matrices). We wish to enumerate the composed objects, given that two of them

are considered to be the same if their building blocks are a cyclic shift of each other. Then, the theorem asserts that the generating function enumerating the composed objects is given by

$$Z(B(z), B(z^2), \dots, B(z^n)),$$

where  $B(z)$  is the generating function enumerating the “building block” objects.

Translating this to our setting we find that the  $(k + 1) \times 1$  binary columns are enumerated by  $B(z) = (1 + z)^{k+1}$  (as for every entry we can either choose 1, which contributes  $z$ , or we can choose 0, which contributes  $z^0 = 1$ ). Thus, applying the Pólya–Redfield theorem to this situation yields (13).  $\square$

From this enumerating polynomial, formulas for the number of equivalence classes with an arbitrary number of 1’s can be derived by extracting the coefficient of the corresponding monomial.

## 5 Down-step statistics: A generating function approach

In this section we will consider the symbolic decomposition of  $k_t$ -Dyck paths, making use of the symbolic method (see [FS09, Chapter I.1]). From this, we then calculate down-step statistics using bivariate generating functions with the variables  $x$  and  $u$  to count the number of up-steps and the number of down-steps, respectively.

Recall that the definition of  $\mathcal{S}_t$ , the class of  $k_t$ -Dyck paths, is given in Definition 2. Let  $k \geq 2$  and  $0 \leq t \leq k$ . The symbolic decomposition of  $k_t$ -Dyck paths is given by

$$\mathcal{S}_t = \varepsilon + \sum_{i=0}^t (\{\setminus\}^i \times \{/ \} \times \mathcal{S}_{k-1-i} \times \{\setminus\}^{k-i} \times \mathcal{S}_t), \tag{15}$$

where  $\setminus$  represents a  $(1, -1)$  step,  $/$  represents a  $(1, k)$  step,  $\varepsilon$  represents the empty path, and by convention we set  $\mathcal{S}_{-1} = \varepsilon$ . This is obtained via a first return decomposition (with the return being to  $y = 0$ ) of the path, and illustrated in Figure 9. Here we have  $i$  down-steps, where  $0 \leq i \leq t$ , followed by an up-step, bringing the height to  $k - i$ . In order to not introduce any returns in this part of the path, we have an arbitrary  $k_{k-i-1}$ -Dyck path, again ending at a height of  $k - i$ . Finally, we introduce the first return to  $y = 0$  of the path with  $k - i$  down-steps, and the remainder of the path is an arbitrary  $k_t$ -Dyck path.

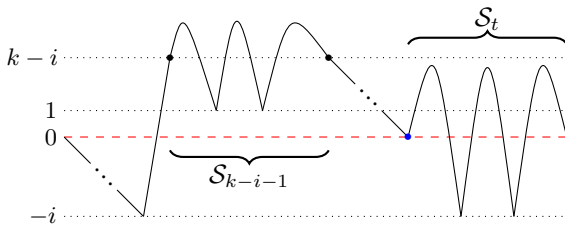


Fig. 9: The symbolic decomposition of a  $k_t$ -Dyck path for fixed  $i$ .

*Remark 6.* Note that this decomposition does not work for  $t > k$ , as for  $k < i \leq t$  one up-step  $(1, k)$  would not be enough to return to or above the  $x$ -axis, and thus a decomposition based on the first return to the  $x$ -axis would be much more involved, as has been outlined in [Pro19].

This translates into the following system of functional equations for  $0 \leq t \leq k$ , where  $S_t(x)$  is the generating function for  $k_t$ -Dyck paths with  $x$  counting the number of up-steps,

$$S_t(x) = 1 + xS_t(x) \sum_{i=0}^t S_{k-i-1}(x) \quad \text{and} \quad S_{-1}(x) = 1.$$

With the substitution as used in [Ges16, Section 3.3] to deal with the Fuss–Catalan family,

$$x = z(1 - z)^k \tag{16}$$

we can show (see [Sel19, Chapter 6]) that the generating function for  $k_t$ -Dyck paths simplifies to the expression  $S_t(x) = (1 - z)^{-t-1}$ . Throughout this section this substitution will be used to simplify expressions and thus extract coefficients in an easier manner.

### 5.1 The number of down-steps at the end of $k_t$ -Dyck paths

Before calculating down-step statistics generally, we first consider a special case which will be used in all other calculations of the total number of down-steps: the total number of down-steps at the end of all  $k_t$ -Dyck paths of a given length.

**Proposition 18.** *The total number of down-steps at the end of all  $k_t$ -Dyck paths of length  $(k + 1)n$  is equal to*

$$s_{n,t,n} = \frac{t+1}{n+1} \binom{(k+1)(n+1)+t}{n} - \frac{(t+1)^2}{n} \binom{(k+1)n+t}{n-1},$$

where  $s_{n,t,r}$  is used as in Definition 4.

Note that these numbers are, as expected, the same as those of (9) in Proposition 13, however, using this generating function approach allows us to later obtain the variance — a statistic which is more difficult to obtain via the bijective approach. Interestingly, the only pre-existing combinatorial interpretation of these numbers, related to non-crossing trees, occurred for  $k = 2$  and  $t = 1$  as entry A030983 in the OEIS. We have added entries for additional parameter combinations as A334609–A334612, A334680, A334682, A334719.

Before proving this, we will prove a lemma that will be used to extract coefficients from generating functions under the substitution  $x = z(1 - z)^k$ . This will be done to prevent repetition of very similar calculations.

**Lemma 19.** *For integers  $n, k, t, a, b, c,$  and  $d,$  we have*

$$[x^n] \frac{1}{z^a(1-z)^{bk+ct+d}} = \frac{k(b-a) + ct + d}{n+a} \binom{(k+1)n + kb + ct + d + a - 1}{n+a-1}.$$

**Proof:** We will extract coefficients by means of Cauchy’s integral formula (see [FS09, Theorem IV.4]). Let  $\gamma$  be a small contour that winds around the origin once, and let  $\tilde{\gamma}$  be the image of  $\gamma$  under the substitution  $x := z(1 - z)^k$  from (16). Then

$$[x^n] \frac{1}{z^a(1-z)^{bk+ct+d}} = \frac{1}{2\pi i} \oint_{\tilde{\gamma}} \frac{1}{z^a(1-z)^{bk+ct+d}} \frac{1}{x^{n+1}} dx$$

$$\begin{aligned}
&= \frac{1}{2\pi i} \oint_{\tilde{\gamma}} \frac{1}{z^a(1-z)^{bk+ct+d}} (1 - (k+1)z)(1-z)^{k-1} dz \\
&= \frac{1}{2\pi i} \oint_{\tilde{\gamma}} \frac{1}{z^{n+a+1}(1-z)^{k(n+b)+ct+d+1}} dz \\
&\quad - \frac{k+1}{2\pi i} \oint_{\tilde{\gamma}} \frac{1}{z^{n+a}(1-z)^{k(n+b)+ct+d+1}} dz \\
&= [z^{n+a}] \frac{1}{(1-z)^{k(n+b)+ct+d+1}} - (k+1)[z^{n+a-1}] \frac{1}{(1-z)^{k(n+b)+ct+d+1}} \\
&= \frac{k(b-a) + ct + d}{n+a} \binom{(k+1)n + kb + ct + d + a - 1}{n+a-1}.
\end{aligned}$$

□

**Proof of Proposition 18:** We use  $u$  to count the number of down-steps after the last up-step, and  $x$  to count the number of up-steps in the paths. With Figure 9 as reference: if the final  $k_t$ -Dyck path (denoted by  $S_t$  in Figure 9) is empty, then the contribution to the sequence of down-steps at the end of the path comes from the  $k-i$  down-steps as well as the final down-steps from the  $k_{k-i-1}$ -Dyck path. Otherwise, just the final down-steps of the  $k_t$ -Dyck path contribute towards the total number of down-steps. For  $0 \leq t \leq k$  this leads to the system  $S_{-1}(x, u) = 1$  and

$$S_t(x, u) = 1 + x \sum_{i=0}^t S_{k-i-1}(x, u) u^{k-i} + x \sum_{i=0}^t S_{k-i-1}(x, 1) (S_t(x, u) - 1), \quad (17)$$

where  $S_{k-i-1}(x, 1) = (1-z)^{-k+i}$ . We use the notation  $\partial_u$  to represent the partial derivative with respect to  $u$ . Differentiating with respect to  $u$ , setting  $u = 1$ , and substituting  $x = z(1-z)^k$  gives

$$\begin{aligned}
\partial_u S_t(x, u) \Big|_{u=1} &= z(1-z)^k \sum_{i=0}^t \partial_u S_{k-i-1}(x, u) \Big|_{u=1} - (k-t)(1-z)^{t+1} + k + 1 \\
&\quad - \frac{1}{z} + \frac{(1-z)^{t+1}}{z} + \partial_u S_t(x, u) \Big|_{u=1} (1 - (1-z)^{t+1}).
\end{aligned}$$

Collecting the  $\partial_u S_t(x, u) \Big|_{u=1}$  terms on the left-hand side, we obtain<sup>(vi)</sup>

$$\partial_u S_t(x, u) \Big|_{u=1} = \frac{z}{(1-z)^{t-k+1}} \sum_{i=0}^t \partial_u S_{k-i-1}(x, u) \Big|_{u=1} - k + t + \frac{k+1}{(1-z)^{t+1}} - \frac{1 - (1-z)^{t+1}}{z(1-z)^{t+1}}. \quad (18)$$

It can be proved by induction on  $t$  and substitution into the above equation that the generating function for the total number of down-steps at the end of a  $k_t$ -Dyck path is given by

$$\partial_u S_t(x, u) \Big|_{u=1} = -\frac{t+1}{(1-z)^{t+1}} + \frac{1}{z(1-z)^{k+t+1}} - \frac{1}{z(1-z)^k}.$$

---

<sup>(vi)</sup> For specific values these computations are carried out in the interactive SageMath worksheet at <https://gitlab.aau.at/behack1/kt-dyck-downstep-code>.

Uniqueness of this solution can be shown using bootstrapping. Since everything on the right-hand side in (18) except for the sum of partial derivatives can be expanded in terms of  $z$ , using the initial estimate  $\partial_u S_t(x, u)|_{u=1} = O(z)$  and expanding leads to the new estimate

$$\partial_u S_t(x, u)|_{u=1} = (k+1)(t+1)z + \frac{(t+1)t}{2}z - (t+1)^2z + O(z^2),$$

and this can be continued indefinitely. Throughout this section there will be several expressions which can be verified to be a solution to a given functional equation, and with similar bootstrapping the uniqueness can be shown. Finally, applying Lemma 19 for each of the terms in the generating function, we obtain

$$[x^n] \frac{1}{(1-z)^{t+1}} = \frac{t+1}{n} \binom{(k+1)n+t}{n-1}; \quad [x^n] \frac{1}{z(1-z)^{k+t+1}} = \frac{t+1}{n+1} \binom{(k+1)(n+1)+t}{n},$$

from which the result follows. Note that we do not need to compute coefficients of  $1/z(1-z)^k$  nor the positive integer powers of this term, since these are equal to  $1/x$  and powers thereof and thus do not contribute to the coefficients.  $\square$

**Proposition 20.** *The coefficients of the second derivative of  $S_t(x, u)$  with respect to  $u$  at  $u = 1$  are given by*

$$\begin{aligned} [x^n] \partial_u^2 S_t(x, u)|_{u=1} &= \frac{2(t+1)}{n+2} \binom{(k+1)(n+2)+t}{n+1} + \frac{(t+1)^2(t+2)}{n} \binom{(k+1)n+t}{n-1} \\ &\quad - \frac{2(t+k+2)(t+1)}{n+1} \binom{(k+1)(n+1)+t}{n}. \end{aligned}$$

The proof of the above proposition can be found in Appendix A.

## 5.2 The number of down-steps between pairs of up-steps

Here we again use the parameter  $u$  to count the number of down-steps in between pairs of up-steps. Let  $S_{t,r}(x, u)$  be the bivariate generating function for the number of down-steps between the  $r$ -th and the  $(r+1)$ -th up-steps, or in the case where an  $(r+1)$ -th up-step does not exist, after the  $r$ -th up-step. Using this we can easily set up a system of functional equations using (15) to determine  $S_{t,r}(x, u)$ . The case where  $r = 0$  is given by  $S_{-1,0}(x, u) = 1$  and

$$S_{t,0}(x, u) = 1 + xS_t(x, 1) \sum_{i=0}^t u^i S_{k-i-1}(x, 1). \quad (19)$$

**Theorem 21.** *Let  $n \geq 1$ . The total number of down-steps before the first up-step in  $k_t$ -Dyck paths of length  $(k+1)n$  is given by*

$$s_{n,t,0} = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n}.$$

Again, by definition these numbers are equal to those from (3), and have been added to the OEIS as entries A001764, A002293, A002294, A334785–A334787. Together, the bijective and generating function approaches yield the following interesting summation identity related to (7),

$$\sum_{j=0}^{t-1} C_{n,j} = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n}. \quad (20)$$

This identity can be interpreted combinatorially by counting  $k_t$ -Dyck paths. On the left-hand side we have the total number of  $k_j$ -Dyck paths of length  $(k+1)n$  for  $0 \leq j \leq t-1$ . By Observation 7, these are in bijection with elevated  $k_0$ -Dyck paths of elevation  $j$  where  $0 \leq j \leq k-1$ , which we know from (6) to be enumerated by

$$\frac{k}{n+1} \binom{(k+1)n}{n},$$

minus the elevated  $k_0$ -Dyck paths of elevation  $j$  where  $t \leq j \leq k-1$ . It can be verified by symbolic summation that summing the number of elevated  $k_0$ -Dyck paths over  $t \leq j \leq k-1$  gives

$$\sum_{j=t}^{k-1} \frac{j+1}{(k+1)n+j+1} \binom{(k+1)n+j+1}{n} = \frac{k-t}{n+1} \binom{(k+1)n+t}{n}. \quad (21)$$

With this, we have that the summation on the left-hand side of (20) is equal to the right-hand side.

Alternatively, (20) can be proved by considering the following identity. This is a consequence of (21) or can be verified by algebraic manipulation.

$$C_{n,j} = \frac{k-j}{n+1} \binom{(k+1)n+j}{n} - \frac{k-(j+1)}{n+1} \binom{(k+1)n+(j+1)}{n}.$$

Summing over  $0 \leq j \leq t-1$  yields the following telescoping sum

$$\begin{aligned} \sum_{j=0}^{t-1} C_{n,j} &= \sum_{j=0}^{t-1} \frac{k-j}{n+1} \binom{(k+1)n+j}{n} - \sum_{j=0}^{t-1} \frac{k-(j+1)}{n+1} \binom{(k+1)n+(j+1)}{n} \\ &= \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n}. \end{aligned}$$

**Proof of Theorem 21:** The system in (19) can be further simplified with the substitution  $x = z(1-z)^k$  and the expression  $S_t(x) = (1-z)^{-t-1}$ :

$$S_{t,0}(x, u) = 1 + \frac{z(1-z)^k}{(1-z)^{k+t+1}} \sum_{i=0}^t u^i (1-z)^i = 1 + \frac{z}{(1-z)^{t+1}} \frac{1-u^{t+1}(1-z)^{t+1}}{1-u+uz}. \quad (22)$$

Differentiating this system and setting  $u = 1$  yields

$$\partial_u S_{t,0}(x, u) \Big|_{u=1} = \frac{z}{(1-z)^{t+1}} \left( \frac{-(t+1)(1-z)^{t+1}z + (1-z)(1-(1-z)^{t+1})}{z^2} \right)$$

$$= -(t+1) + \frac{1}{z(1-z)^t} - \frac{1-z}{z} = \frac{1}{z(1-z)^t} - \frac{1}{z} - t.$$

Applying Lemma 19 to extract coefficients yields the expression given in the theorem.  $\square$

**Proposition 22.** *Let  $n \geq 1$ . The coefficients of the second derivative (with respect to  $u$ ) of  $S_{t,0}(x, u)$  are*

$$\begin{aligned} [x^n] \partial_u^2 S_{t,0}(x, u) \Big|_{u=1} &= \frac{2(t-2k-1)}{n+2} \binom{(k+1)n+t}{n+1} + \frac{2(t-1)k}{n+1} \binom{(k+1)n}{n} \\ &+ \frac{4k}{n+2} \binom{(k+1)n+1}{n+1}. \end{aligned}$$

**Proof:** Differentiating the expression in (22) twice with respect to  $u$  and setting  $u = 1$  gives

$$\partial_u^2 S_{t,0}(x, u) \Big|_{u=1} = \frac{2}{z^2(1-z)^{t-1}} - \frac{t(t-1)z^2 + 2(t-1)z + 2}{z^2}$$

The coefficients of this function can then be extracted using Lemma 19 to obtain the result.  $\square$

For  $r = 1$ , with reference to Figure 9 and the symbolic decomposition (15) we must consider two cases:

- the sequence of down-steps before the second up-step lies entirely within the path from  $\mathcal{S}_{k-i-1}$ , and
- the path from  $\mathcal{S}_{k-i-1}$  is the empty path and so the sequence of down-steps before the second up-step includes the  $k-i$  down-steps after the path from  $\mathcal{S}_{k-i-1}$  as well as the down-steps before the first up-step in the path from  $\mathcal{S}_t$ .

This gives the system of functional equations for  $0 \leq t \leq k$ , setting  $S_{-1,1}(x, u) = 1$ ,

$$S_{t,1}(x, u) = 1 + x \sum_{i=0}^t (S_{k-i-1,0}(x, u) - 1) S_t(x, 1) + x \sum_{i=0}^t u^{k-i} S_{t,0}(x, u).$$

**Theorem 23.** *The total number of down-steps between the first and second up-steps in  $k_t$ -Dyck paths of length  $(k+1)n$  is equal to*

$$s_{n,t,1} = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n} + \frac{k(t+1)}{n} \binom{(k+1)(n-1)}{n-1}. \quad (23)$$

**Proof:** With the substitution  $x = z(1-z)^k$  and using the generating function for  $S_{t,0}(x, u)$  in terms of  $z$  and  $u$  as given in (22) we have that

$$\begin{aligned} S_{t,1}(x, u) &= 1 + z(1-z)^k \frac{1}{(1-z)^{t+1}} \left( \frac{1 - (1-z)^{t+1}}{(1-z)^k(1-u+uz)} - \frac{zu^k}{1-u+uz} \sum_{i=0}^t u^{-i} \right) \\ &+ z(1-z)^k \sum_{i=0}^t u^{k-i} \left( 1 + \frac{z}{(1-z)^{t+1}} \frac{1 - u^{t+1}(1-z)^{t+1}}{1-u+uz} \right) \end{aligned}$$



$$= 1 + \frac{z(1 - (1 - z)^{t+1})}{(1 - z)^{t+1}(1 - u + uz)} + \left( z(1 - z)^k u^k - \frac{z^2 u^{k+t+1} (1 - z)^k}{1 - u + uz} \right) \sum_{i=0}^t u^{-i}. \quad (24)$$

Differentiating this with respect to  $u$  results in

$$\begin{aligned} \partial_u S_{t,1}(x, u) &= \frac{z(1 - (1 - z)^{t+1})}{(1 - z)^t(1 - u + uz)^2} - \left( z(1 - z)^k u^k - \frac{z^2 u^{k+t+1} (1 - z)^k}{1 - u + uz} \right) \sum_{i=0}^t i u^{-i-1} \\ &\quad + \left( k z (1 - z)^k u^{k-1} - \frac{z^2 u^{k+t} (1 - z)^k ((k + t)(1 - u + uz) + 1)}{(1 - u + uz)^2} \right) \sum_{i=0}^t u^{-i}, \end{aligned}$$

and finally setting  $u = 1$  gives the expression

$$\partial_u S_{t,1}(x, u)|_{u=1} = \frac{1 - (1 - z)^{t+1}}{z(1 - z)^t} - (t + 1)(1 - z)^k (tz + 1).$$

Extracting the  $n$ -th coefficient (for  $n \geq 2$ ) gives the formula for the total number of down-steps between the first and second up-steps in  $k_t$ -Dyck paths of length  $(k + 1)n$  as given in (23).  $\square$

**Proposition 24.** *Let  $n \geq 2$ . The coefficients of the second derivative with respect to  $u$  of  $S_{t,1}(x, u)$  are*

$$\begin{aligned} [x^n] \partial_u^2 S_{t,1}(x, u)|_{u=1} &= \frac{2(t - 2k - 1)}{n + 2} \binom{(k + 1)n + t}{n + 1} + \frac{4(k + 1)}{n + 2} \binom{(k + 1)n - 1}{n + 1} \\ &\quad + \frac{k(t + 1)(2k + t - 2)}{n} \binom{(k + 1)(n - 1)}{n - 1} + \frac{4(t + 1)k}{n + 1} \binom{(k + 1)n - k}{n}. \end{aligned}$$

**Proof:** Differentiating the expression in (24) twice with respect to  $u$  and setting  $u = 1$  yields

$$\partial_u^2 S_{t,1}(x, u)|_{u=1} = \frac{2}{z^2(1 - z)^{t-1}} - \frac{2(1 - z)^2}{z^2} - \frac{(1 - z)^k (t + 1)(t(2k - 1)z^2 + (2k + t - 2)z + 2)}{z}.$$

Extracting coefficients of each term via Lemma 19 gives the result in the proposition.  $\square$

When  $r \geq 2$  the system of functional equations as well as the calculations become trickier — here there are now three possible configurations with reference to the symbolic decomposition in (15): the sequence of down-steps between the  $r$ -th and  $(r + 1)$ -th up-steps

1. lies entirely within the path from  $S_{k-i-1}$ ,
2. lies entirely within the path from  $S_t$ , and
3. consists of the final down-steps of the path from  $S_{k-i-1}$ , the  $k - i$  intermediate down-steps, and the initial down-steps of the path from  $S_t$  (illustrated in Figure 10).

This leads to the general functional equation for  $r \geq 2$ ,

$$S_{t,r}(x, u) = 1 + x S_t(x, 1) \sum_{i=0}^t \sum_{n \geq r} x^n ([x^n] S_{k-i-1, r-1}(x, u)) \quad (25)$$



$$+ x^r S_t(x, 1) \sum_{i=0}^t \frac{(k-i)^2}{(k+1)(r-1)+k-i} \binom{(k+1)(r-1)+k-i}{r-1}.$$

**Theorem 25.** For  $r \geq 2$ , the generating function for  $S_{t,r}(x, u)$  after differentiating with respect to  $u$  and setting  $u = 1$  is

$$\begin{aligned} \partial_u S_{t,r}(x, u)|_{u=1} &= \frac{1-(1-z)^{t+1}}{z(1-z)^t} - \sum_{j=1}^r z^{j-1}(1-z)^{kj} \frac{t+1}{(k+1)j+t+1} \binom{(k+1)j+t+1}{j} \\ &+ \sum_{j=1}^{r-1} z^j(1-z)^{jk} \frac{t+1}{(k+1)j+t+1} \binom{(k+1)j+t+1}{j} \\ &- z^r(1-z)^{rk} \frac{t(t+1)}{(k+1)r+t+1} \binom{(k+1)r+t+1}{r}. \end{aligned}$$

The key to the proof of the above theorem is the following identity, which follows from the symbolic decomposition of  $k_t$ -Dyck paths given in (15). The remainder of the proof of Theorem 25 consists of algebraic manipulations and is detailed in Appendix B.

**Proposition 26.** For  $n \geq 0$ , and  $0 \leq t \leq k$  we have that

$$\begin{aligned} \sum_{j=0}^{n-1} \frac{t+1}{(k+1)(n-1-j)+t+1} \binom{(k+1)(n-1-j)+t+1}{n-1-j} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j} \\ = \frac{t+1}{(k+1)n+t+1} \binom{(k+1)n+t+1}{n}. \end{aligned}$$

This result can be seen as a generalized Chu–Vandermonde-type identity, and is similar (though not equivalent) to the family of identities given in [Rot93] and mentioned in [Gou99], known as Rothe–Hagen identities. In particular, it can be proved algebraically using these identities, but we still provide a bijective argument here, as we are not aware of this combinatorial interpretation in the literature.

**Proof of Proposition 26:** We enumerate  $k_t$ -Dyck paths in two ways to obtain this identity. We have from Proposition 3 that the number of  $k_t$ -Dyck paths of length  $(k+1)n$  is

$$\frac{t+1}{(k+1)n+t+1} \binom{(k+1)n+t+1}{n}.$$

However, the symbolic decomposition of  $k_t$ -Dyck paths given in Figure 9 shows that we can decompose a  $k_t$ -Dyck path of length  $(k+1)n$  into a  $k_{k-i-1}$ -Dyck path of length  $(k+1)j$  (where  $0 \leq i \leq t$  and  $0 \leq j \leq n-1$ ) and a  $k_t$ -Dyck path of length  $(k+1)(n-1-j)$ . Using identity (21) to simplify the double sum results in

$$\begin{aligned} \sum_{j=0}^{n-1} \left[ \sum_{i=0}^t \frac{k-i}{(k+1)j+k-i} \binom{(k+1)j+k-i}{j} \right] \\ \times \frac{t+1}{(k+1)(n-1-j)+t+1} \binom{(k+1)(n-1-j)+t+1}{n-1-j} \end{aligned}$$

$$= \sum_{j=0}^{n-1} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j} \frac{t+1}{(k+1)(n-1-j)+t+1} \binom{(k+1)(n-1-j)+t+1}{n-1-j}.$$

Since these are both valid ways to enumerate  $k_t$ -Dyck paths of length  $(k+1)n$ , the identity holds.  $\square$

**Corollary 27.** *The total number of down-steps between the  $r$ -th and the  $(r+1)$ -th up-steps in  $k_t$ -Dyck paths of length  $(k+1)n$  for  $r < n$  is*

$$s_{n,t,r} = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n} \\ + \sum_{j=1}^r \frac{t+1}{(k+1)j+t+1} \frac{k}{n-j+1} \binom{(k+1)j+t+1}{j} \binom{(k+1)(n-j)}{n-j}$$

By using (20), this result corresponds to that of Corollary 12. In fact, we can even rewrite this result in the form

$$s_{n,t,r} = s_{n,t,0} + \sum_{j=1}^r \frac{t+1}{(k+1)j+t+1} \frac{k}{n-j+1} \binom{(k+1)j+t+1}{j} \binom{(k+1)(n-j)}{n-j}.$$

**Proof of Corollary 27:** We begin by calculating  $[x^n]z^{j-1}(1-z)^{kj}$  for  $j \geq 1$ , using Lemma 19,

$$[x^n]z^{j-1}(1-z)^{kj} = -\frac{k}{n-j+1} \binom{(k+1)(n-j)}{n-j}.$$

Additionally, since  $\frac{1-(1-z)^{t+1}}{z(1-z)^t} = \frac{1}{z(1-z)^t} - \frac{1}{z} + 1$ , in a similar way we have

$$[x^n] \frac{1}{z(1-z)^t} = \binom{(k+1)n+t}{n} \frac{t-k}{n+1},$$

as well as

$$[x^n] \frac{1}{z} = -\frac{k}{n+1} \binom{(k+1)n}{n}.$$

Combining these with the result in Theorem 25, we have that for  $r < n$

$$[x^n] \partial_u S_{t,r}(x, u) = [x^n] \frac{1-(1-z)^{t+1}}{z(1-z)^t} \\ - \sum_{j=1}^r \frac{t+1}{(k+1)j+t+1} \binom{(k+1)j+t+1}{j} [x^n] z^{j-1}(1-z)^{kj} \\ = \frac{k}{n+1} \binom{(k+1)n}{n} - \frac{k-t}{n+1} \binom{(k+1)n+t}{n} \\ + \sum_{j=1}^r \frac{t+1}{(k+1)j+t+1} \binom{(k+1)j+t+1}{j} \frac{k}{n-j+1} \binom{(k+1)(n-j)}{n-j}.$$

□

In this section we have shown that the generating function approach is just as suitable as the bijective approach for the investigation of the down-step parameter, with the added benefit of providing a strategy for computing the variance for all fixed values of  $r$ . In fact, combining these two approaches reveals interesting summation identities like (20).

## 6 Asymptotics for down-step statistics

In this section we calculate asymptotics for the average number of down-steps as well as the variance in the number of down-steps in  $k_t$ -Dyck paths of length  $(k+1)n$ . These multivariable computations can become quite involved, and thus were performed<sup>(vii)</sup> using the asymptotic expansions module [HHK15] in SageMath [The20].

Formally, we conduct this analysis by considering the random variable  $X_{n,t,r}$  which models the number of down-steps between the  $r$ -th and the  $(r+1)$ -th up-steps in a  $k_t$ -Dyck path of length  $(k+1)n$  chosen uniformly at random.

### 6.1 Expected Value

From the explicit formulas for  $s_{n,t,r}$  provided in Corollary 12 and Corollary 27 (for the case where  $0 \leq r < n$ ) as well as Proposition 13 and Proposition 18 (for  $r = n$ ) the expected number of down-steps can be obtained immediately by the relation

$$\mathbb{E}X_{n,t,r} = \frac{s_{n,t,r}}{C_{n,t}}.$$

From there, we can determine asymptotic expansions (for  $n \rightarrow \infty$  and with  $k, t$  fixed) of down-steps between the  $r$ -th and  $(r+1)$ -th up-steps.

An asymptotic expansion with error term  $O(1/n)$  can be obtained directly from the expressions for  $s_{n,t,r}$  and  $C_{n,t}$ . Expansions of arbitrary precision can be obtained by rewriting the products as factorials and applying Stirling's formula, cf. [DLMF20, 5.11.3]; the first few terms are contained in the SageMath worksheet.

**Proposition 28.** *For  $k_t$ -Dyck paths of length  $(k+1)n$ , the following results hold:*

1. *The expected number of down-steps before the first up-step is*

$$\begin{aligned} \mathbb{E}X_{n,t,0} &= \frac{k(kn+t+1) \cdots (kn+1)}{(n+1)(t+1)((k+1)n+t) \cdots ((k+1)n+1)} - \frac{(k-t)(kn+t+1)}{(n+1)(t+1)} \\ &= \frac{k}{t+1} \left( \frac{k^{t+1}}{(k+1)^t} - k+t \right) + O\left(\frac{1}{n}\right). \end{aligned}$$

2. *For fixed  $1 \leq r < n$ , the expected number of down-steps between the  $r$ -th and  $(r+1)$ -th up-steps is*

$$\mathbb{E}X_{n,t,r} = \frac{k(kn+t+1) \cdots (kn+1)}{(n+1)(t+1)((k+1)n+t) \cdots ((k+1)n+1)} - \frac{(k-t)(kn+t+1)}{(n+1)(t+1)}$$

---

<sup>(vii)</sup> A worksheet containing these computations can be found at <https://gitlab.aau.at/behack1/kt-dyck-downstep-code>.

$$\begin{aligned}
& + \sum_{j=1}^r \frac{t+1}{(k+1)j+t+1} \binom{(k+1)j+t+1}{j} \\
& \quad \times \frac{k n \cdots (n-j+2) (kn+t+1) \cdots (k(n-j)+1)}{(t+1) ((k+1)n+t) \cdots ((k+1)(n-j)+1)} \\
& = \frac{k^{t+2}}{(k+1)^t} \sum_{j=1}^r \frac{1}{(k+1)j+t+1} \binom{(k+1)j+t+1}{j} \left( \frac{k^k}{(k+1)^{k+1}} \right)^j \\
& \quad + \frac{k}{t+1} \left( \frac{k^{t+1}}{(k+1)^t} - k+t \right) + O\left(\frac{1}{n}\right).
\end{aligned}$$

Analogously, we obtain the following result for  $\mathbb{E}X_{n,t,n}$  as well as  $\mathbb{E}X_{n,t,n-r}$ .

**Proposition 29.** For  $k_t$ -Dyck paths of length  $(k+1)n$  we find:

1. The expected number of down-steps after the last up-step is

$$\begin{aligned}
\mathbb{E}X_{n,t,n} & = \frac{((k+1)(n+1)+t) \cdots ((k+1)n+t+1)}{(n+1) (k(n+1)+t+1) \cdots (kn+t+2)} - (t+1) \\
& = \frac{(k+1)^{k+1}}{k^k} - (t+1) + O\left(\frac{1}{n}\right).
\end{aligned}$$

2. For fixed  $1 \leq r \leq n$ , the expected number of down-steps between the  $(n-r)$ -th and the  $(n-r+1)$ -th up-steps is

$$\begin{aligned}
\mathbb{E}X_{n,t,n-r} & = \frac{((k+1)(n+1)+t) \cdots ((k+1)n+t+1)}{(n+1) (k(n+1)+t+1) \cdots (kn+t+2)} - (k+1) \\
& \quad - \sum_{j=2}^r \binom{(j-1)(k+1)}{j} \frac{n \cdots (n-j+2) (kn+t+1) \cdots (k(n-j+1)+t+2)}{(j-1) ((k+1)n+t) \cdots ((k+1)(n-j+1)+t+1)} \\
& = \frac{(k+1)^{k+1}}{k^k} - k - 1 - \sum_{j=2}^r \frac{1}{j-1} \binom{(j-1)(k+1)}{j} \frac{k^{k(j-1)}}{(k+1)^{(k+1)(j-1)}} + O\left(\frac{1}{n}\right).
\end{aligned}$$

*Remark 7.* Note that extracting the asymptotic behaviour from the explicit formulas is, of course, not the only viable strategy: alternatively, we could extract the growth directly from the generating functions by investigating the substitution  $x = z(1-z)^k$  from (16), where the variable  $x$  symbolically corresponds to the number of up-steps in  $k_t$ -Dyck paths.

To be more precise, instead of using Lemma 19 for translating between the “generating function world” and the “coefficient world”, we could also observe that  $\varphi(z) := z(1-z)^k$  has an analytic inverse as long as  $\varphi'(z) \neq 0$ . And indeed:  $z_0 = 1/(k+1)$  is the zero of smallest modulus of the derivative  $\varphi'(z)$ . Singular inversion [FS09, Chapter VI.7] (which basically expands  $\varphi(z)$  as a power series in  $z - z_0$  to find that the inverse function has a square root singularity at  $x_0 = \varphi(z_0) = \frac{k^k}{(k+1)^{k+1}}$ ) then allows to obtain the asymptotic growth of the coefficients of the inverse function of  $\varphi(z)$  (or compositions thereof, which is what obtained in the formulas throughout Section 5) by means of singularity analysis. These calculations can also be carried out with SageMath’s module on asymptotic expansions [HHK15].

It is worthwhile to take a closer look at the asymptotic expansions given in Propositions 28 and 29. In particular, if we divide these expansions by  $k$  and let  $k \rightarrow \infty$ , a very particular behaviour (illustrated in Figures 11, 12a, 12b) arises. Note that dividing by  $k$  is fairly intuitive as  $k_t$ -Dyck paths need to have  $k$  down-steps for every up-step, normalization thus yields a parameter for down-steps that is on the same scale as the number of up-steps.

For the *beginning of paths* (Proposition 28), the limiting behaviour depends on  $t$ . In particular, for  $t = 0$  we obtain

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{E}X_{n,0,r}}{k} = \sum_{j=1}^r \frac{j^{j-1}}{j!} \frac{1}{e^j}, \quad (29)$$

and for  $t = k$ ,

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{E}X_{n,k,r}}{k} = \sum_{j=1}^{r+1} \frac{j^{j-1}}{j!} \frac{1}{e^j}. \quad (30)$$

Notice that (30) is the same sum as in (29) with a shifted upper limit of  $r + 1$  instead of  $r$ . This means that for fixed  $r$  and  $k \rightarrow \infty$ , the quantity  $\mathbb{E}X_{n,k,r}/k$  has the same asymptotic behaviour as  $\mathbb{E}X_{n,0,r+1}/k$ . This is also indicated by Figure 12c, compare the values represented by the red dots (corresponding to  $\mathbb{E}X_{n,k,r}/k$ ) to the values represented by the blue dots one unit to the right (corresponding to  $\mathbb{E}X_{n,0,r+1}/k$ ).

In contrast, analogous limiting expressions for the *end of paths* (Proposition 29) do not depend on  $t$ , with the exception of down-steps after the last up-step. Namely, for  $k \rightarrow \infty$  we have

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}X_{n,t,n}}{k} = e - \frac{2t + 2 - e}{2k} + O\left(\frac{1}{k^2}\right).$$

Furthermore, for fixed  $r \geq 1$  we have

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\mathbb{E}X_{n,t,n-r}}{k} = e - \sum_{j=1}^r \frac{(j-1)^{j-1}}{j!} \frac{1}{e^{j-1}}. \quad (31)$$

Together with the observation that the asymptotic main term in  $n$  given in Proposition 29 (2) does not depend on  $t$ , these considerations explain why the rightmost points in Figure 12d (i.e., those corresponding to  $r = n$ ) are evenly distributed, and the points for  $r < n$  nearly coincide for  $0 \leq t \leq k$ .

The expressions in (29) and (31) are related to  $T(x)$ , the *Cayley tree function*<sup>(viii)</sup> which satisfies the equation  $T(x) = x \exp(T(x))$  (see [FS09, VI.16]), as follows. We have  $T(x) = \sum_{j \geq 1} \frac{j^{j-1}}{j!} x^j$  and  $\frac{1}{T(x)} = \frac{1}{x} - \sum_{j \geq 0} \frac{j^j}{(j+1)!} x^j$  (see A000169 in the OEIS [OEI22]). The right-hand side expressions in (29) and (31) are partial sums of these power series, upon substitution of  $x = 1/e$ . Since we have  $T(1/e) = 1$ , and since the sequence  $(s_{n,t,r})_{0 \leq r \leq n}$  is increasing (strongly, except for  $t = k$ ,  $r = n - 1$ ), for each  $0 < \beta < 1$  we have

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{s_{n,t, \lfloor \beta n \rfloor}}{k C_{n,t}} = 1.$$

This observation is illustrated in Figure 11.

---

<sup>(viii)</sup> The Cayley tree function  $T(x)$  is also strongly related to the *Lambert W function*  $W(x)$  via  $-W(-x) = T(x)$ .

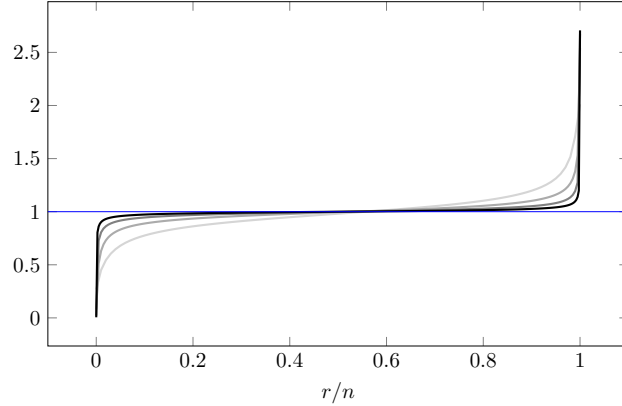


Fig. 11: Distribution of  $X_{n,t,r}/k$  for  $t = 7$  and several combinations of values for  $n$  and  $k$ :  $(n, k) \in \{(10, 100), (40, 400), (160, 1600), (640, 6400)\}$ . Darker lines correspond to larger values of  $n$  and  $k$ .

## 6.2 Variance

Fueled by the generating function approach presented in Section 5 we are able to obtain asymptotic expansions for the variance  $\mathbb{V}X_{n,t,r}$  as well. To do so, we use the known asymptotic behaviour for the expected value  $\mathbb{E}X_{n,t,r}$  as discussed by Propositions 28 and 29 together with the explicit formulas for the second factorial moment  $\mathbb{E}(X_{n,t,r}(X_{n,t,r} - 1))$  that are obtained by normalizing the coefficients extracted from the second partial derivatives of the corresponding bivariate generating functions (Propositions 20, 22, and 24). Combining these quantities by means of the well-known formula

$$\mathbb{V}X_{n,t,r} = \mathbb{E}(X_{n,t,r}(X_{n,t,r} - 1)) + \mathbb{E}X_{n,t,r} - (\mathbb{E}X_{n,t,r})^2$$

yields the variance of the number of down-steps between the  $r$ -th and  $(r + 1)$ -th down-steps in  $k_t$ -Dyck paths of length  $(k + 1)n$ .

While the calculations are not complicated per se, the involved asymptotic expansions are rather large which makes carrying out these calculations by hand a tedious task. Thus, we used SageMath [The20] and included our calculations in one of our associated worksheets<sup>(ix)</sup>. The following theorem summarizes our findings.

**Theorem 30.** *Let  $k$  and  $t$  be fixed integers with  $k \geq 1$  and  $0 \leq t \leq k$ , and consider  $n \rightarrow \infty$ . Then the variance of the random variables  $X_{n,t,0}$ ,  $X_{n,t,1}$ , and  $X_{n,t,n}$  modeling the number of down-steps before the first up-step, between the first and second up-steps, and after the last up-step respectively, admit the asymptotic expansions*

$$\mathbb{V}X_{n,t,0} = \frac{-\alpha_{t,k}^2 k^4 + 2(t+2)\alpha_{t,k} k^3 + ((2t+3)(t+1)\alpha_{t,k} + 3t^2 + 4t)k^2 + t(t+1)k}{(t+1)^2} + O\left(\frac{1}{n}\right), \quad (32)$$

<sup>(ix)</sup> Our computations are publicly available at <https://gitlab.aau.at/behackl/kt-dyck-downstep-code>.



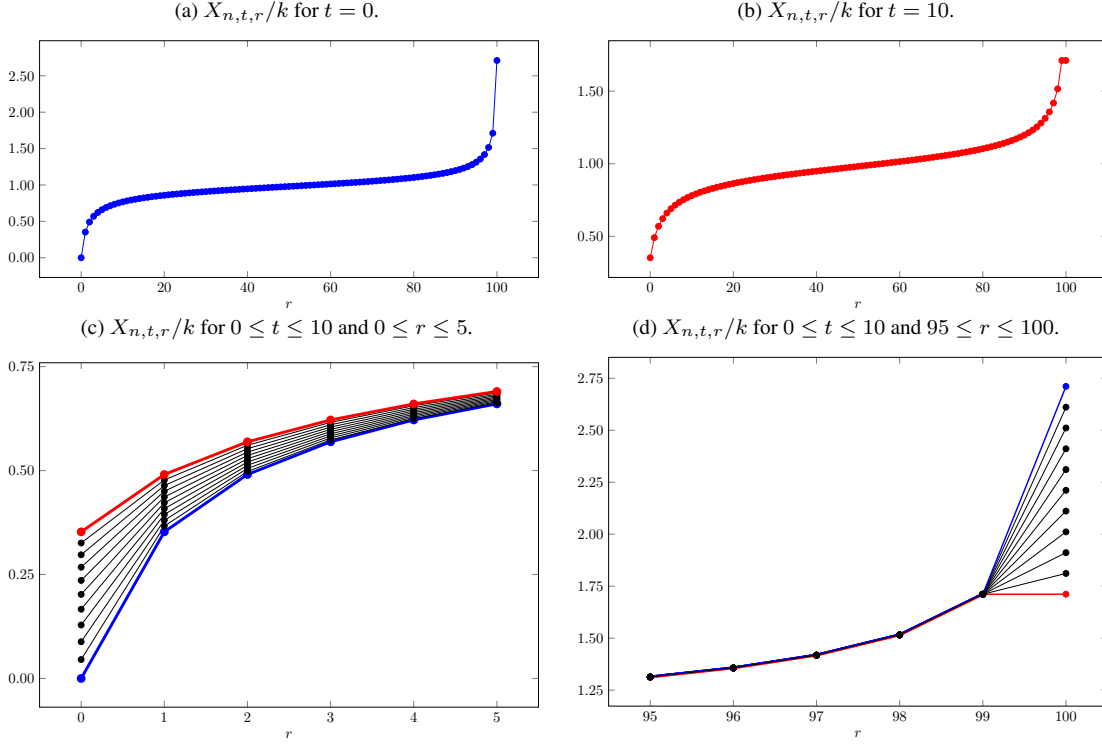


Fig. 12: Distribution of  $X_{n,t,r}/k$  for  $n = 100$ ,  $k = 10$ , and varying values of  $t$  and  $r$ .

where  $\alpha_{t,k} = \left(\frac{k}{k+1}\right)^t - 1$ ,

$$\begin{aligned} \mathbb{V}X_{n,t,1} &= \frac{(6k+t+3)k^{k+t+2}}{(k+1)^{k+t+1}} + \frac{(4k+1)k^{t+2}}{(k+1)^t(t+1)} - \frac{(4k^2-2kt+3k-t)k}{(t+1)} \\ &\quad - \left( \frac{(t+1)k^{k+t+2} + k^{t+2}(k+1)^{k+1} - k(k-t)(k+1)^{k+t+1}}{(k+1)^{k+t+1}(t+1)} \right)^2 + O\left(\frac{1}{n}\right) \end{aligned} \quad (33)$$

and finally,

$$\mathbb{V}X_{n,t,n} = \frac{(k+1)^{2k+2} - (2k+1)(k+1)^{k+1}k^k}{k^{2k}} + O\left(\frac{1}{n}\right). \quad (34)$$

*Remark 8.* Observe that the main contribution in the asymptotic expansions of both the expected value and the variance is of constant order and independent of  $n$ . This hints towards the parameter admitting a discrete limiting distribution; details are subject to further investigations.

## Conclusion

In this article we began investigations into statistics concerning down-steps in  $k_t$ -Dyck paths under the assumption  $0 \leq t \leq k$ . This led to new combinatorial identities, as well as new bijective proofs for known

combinatorial identities. Furthermore, within the study of applications of this statistic in coding theory, we found a new *Cycle Lemma*-type result which also led us to a novel interpretation of Catalan numbers.

There are several open questions related to the down-step statistic that we intend to address in further work. This includes the more involved case of  $t > k$ , a closer investigation to extended Kreweras walks as mentioned in the introduction, as well as further study of the asymptotic behaviour.

## Acknowledgements

The authors would like to thank Clemens Heuberger for comments on the full version of this paper, as well as anonymous referees for their thorough and valuable reviews.

## References

- [BBMM21] Alin Bostan, Mireille Bousquet-Mélou, and Stephen Melczer. Counting walks with large steps in an orthant. *J. Eur. Math. Soc. (JEMS)*, 23(7):2221–2297, 2021.
- [Bég92] Guy Bégin. On the enumeration of perforation patterns for punctured convolutional codes. Poster presentation at the 4th international conference on Formal Power Series and Algebraic Combinatorics, 1992.
- [BF02] Cyril Banderier and Philippe Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoret. Comput. Sci.*, 281(1–2):37–80, 2002.
- [BH89] Guy Bégin and David Haccoun. High-rate punctured convolutional codes: structure properties and construction technique. *IEEE Trans. Comm.*, 37(12):1381–1385, 1989.
- [CM16] Naomi T. Cameron and Jillian E. McLeod. Returns and hills on generalized Dyck paths. *J. Integer Seq.*, 19(2):3, 2016.
- [DLMF20] NIST Digital Library of Mathematical Functions. <http://dlmf.nist.gov/>, Release 1.0.26 of 2020-03-15, 2020. Frank W. J. Olver, Adri B. Olde Daalhuis, Daniel W. Lozier, Barry I. Schneider, Ronald F. Boisvert, Charles W. Clark, Bruce R. Miller, Bonita V. Saunders, Howard S. Cohl, and Marjorie A. McClain, eds.
- [DM47] Aryeh Dvoretzky and Theodore Motzkin. A problem of arrangements. *Duke Math. J.*, 14:305–313, 1947.
- [DZ90] Nachum Dershowitz and Shmuel Zaks. The cycle lemma and some applications. *European J. Combin.*, 11(1):35–40, 1990.
- [FS09] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.
- [Ges16] Ira M. Gessel. Lagrange inversion. *J. Combin. Theory Ser. A*, 144:212–249, 2016.
- [GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics. A foundation for computer science*. Addison-Wesley, second edition, 1994.

- [Gou99] Henry W. Gould. The Girard–Waring power sum formulas for symmetric functions and Fibonacci sequences. *Fibonacci Quart.*, 37(2):135–140, 1999.
- [GPW10] Nancy S. S. Gu, Helmut Prodinger, and Stephan Wagner. Bijections for a class of labeled plane trees. *European J. Combin.*, 31(3):720–732, 2010.
- [HHK15] Benjamin Hackl, Clemens Heuberger, and Daniel Krenn. Asymptotic expansions in SageMath. <http://trac.sagemath.org/17601>, 2015. Module in SageMath 6.10.
- [HHP18] Benjamin Hackl, Clemens Heuberger, and Helmut Prodinger. Ascents in non-negative lattice paths. arXiv:1801.02996 [math.CO], 2018.
- [HP91] Peter Hilton and Jean Pedersen. Catalan numbers, their generalization, and their uses. *Math. Intelligencer*, 13(2):64–75, 1991.
- [JVK16] Matthieu Josuat-Vergès and Jang Soo Kim. Generalized Dyck tilings. *European J. Combin.*, 51:458–474, 2016.
- [JZ99] Rolf Johannesson and Kamil S. Zigangirov. *Fundamentals of convolutional coding*. IEEE Series on Digital & Mobile Communication. IEEE Press, New York, 1999.
- [Kra15] Christian Krattenthaler. Lattice path enumeration. In *Handbook of enumerative combinatorics*, Discrete Math. Appl. (Boca Raton), pages 589–678. CRC Press, Boca Raton, FL, 2015.
- [Kre65] Germain Kreweras. Sur une classe de problèmes de dénombrement liés au treillis des partitions des entiers. *Cahiers Bur. Univ. Rech. Opér.*, 6:9–107, 1965.
- [Lot97] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.
- [Moh79] Sri Gopal Mohanty. *Lattice path counting and applications*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London-Toronto, Ont., 1979. Probability and Mathematical Statistics.
- [OEI22] The On-Line Encyclopedia of Integer Sequences. <http://oeis.org>, 2022.
- [Pó137] George Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68:145–254, 1937.
- [Pro19] Helmut Prodinger. On  $k$ -Dyck paths with a negative boundary. arXiv:1912.06930 [math.CO], 2019.
- [Rot93] Heinrich August Rothe. Formulæ de serierum reversione demonstratio universalis signis localibus combinatorio-analyticorum vicariis exhibita. Dissertatio academica, Universität Leipzig, 1793.
- [Sel19] Sarah J. Selkirk. On a generalisation of  $k$ -Dyck paths. Master’s thesis, Stellenbosch University, 2019.

[The20] The SageMath Developers. *SageMath Mathematics Software (Version 9.0)*, 2020. <http://www.sagemath.org>.

[XZ19] Guoce Xin and Yingrui Zhang. On the sweep map for  $\vec{k}$ -Dyck paths. *Electron. J. Combin.*, 26(3):Paper 3.63, 24, 2019.

## A Proof of Proposition 20

**Proof:** The second derivative of the functional equation given in (17) is

$$\begin{aligned} \partial_u^2 S_t(x, u) &= x \sum_{i=0}^t \partial_u^2 S_{k-i-1}(x, u) u^{k-i} + 2x \sum_{i=0}^t \partial_u S_{k-i-1}(x, u) (k-i) u^{k-i-1} \\ &\quad + x \sum_{i=0}^t S_{k-i-1}(x, 1) \partial_u^2 S_t(x, u) + x \sum_{i=0}^t S_{k-i-1}(x, u) (k-i)(k-i-1) u^{k-i-2}. \end{aligned}$$

Simplifying this expression, setting  $u = 1$ , and applying relevant substitutions yields

$$\begin{aligned} \partial_u^2 S_t(x, u) \Big|_{u=1} (1-z)^{t+1} &= z(1-z)^k \sum_{i=0}^t \partial_u^2 S_{k-i-1}(x, u) \Big|_{u=1} - z(1-z)^k \sum_{i=0}^t \frac{(k-i)^2}{(1-z)^{k-i}} \\ &\quad + (2-z(1-z)^k) \sum_{i=0}^t \frac{k-i}{(1-z)^{k-i}} - 2 \sum_{i=0}^t (k-i). \end{aligned} \quad (35)$$

We show that the second derivative at  $u = 1$  is equal to

$$\partial_u^2 S_t(x, u) \Big|_{u=1} = \frac{2(1-(1-z)^{t+1})}{z^2(1-z)^{2k+t+1}} - \frac{2(t+k+2)}{z(1-z)^{k+t+1}} + \frac{2(k+1)}{z(1-z)^k} + \frac{(t+1)(t+2)}{(1-z)^{t+1}} \quad (36)$$

by substituting it into the occurrences of  $\partial_u^2 S_{k-i-1}(x, u) \Big|_{u=1}$  on the right-hand side of (35), and simplifying this to obtain

$$\begin{aligned} \partial_u^2 S_t(x, u) \Big|_{u=1} (1-z)^{t+1} &= z(1-z)^k \sum_{i=0}^t \frac{2(1-(1-z)^{k-i})}{z^2(1-z)^{3k-i}} - z(1-z)^k \sum_{i=0}^t \frac{2(2k-i+1)}{z(1-z)^{2k-i}} \\ &\quad + z(1-z)^k \sum_{i=0}^t \frac{2(k+1)}{z(1-z)^k} + z(1-z)^k \sum_{i=0}^t \frac{(k-i)(k-i+1)}{(1-z)^{k-i}} \\ &\quad - z(1-z)^k \sum_{i=0}^t \frac{(k-i)^2}{(1-z)^{k-i}} + (2-z(1-z)^k) \sum_{i=0}^t \frac{(k-i)}{(1-z)^{k-i}} \\ &\quad - 2 \sum_{i=0}^t (k-i) \\ &= \frac{2(1-(1-z)^{t+1})}{z^2(1-z)^{2k}} - \frac{2(t+k+2)}{z(1-z)^k} + 2(k+1) \frac{(1-z)^{t+1}}{z(1-z)^k} \end{aligned}$$

$$+ (t+2)(t+1),$$

thus proving (36). We then extract coefficients from (36) using Lemma 19 to obtain the result, where

$$[x^n] \frac{1}{z^2(1-z)^{2k+t+1}} = \frac{t+1}{n+2} \binom{(k+1)(n+2)+t}{n+1}, \quad [x^n] \frac{1}{(1-z)^{t+1}} = \frac{t+1}{n} \binom{(k+1)n+t}{n-1},$$

and

$$[x^n] \frac{1}{z(1-z)^{k+t+1}} = \frac{t+1}{n+1} \binom{(k+1)(n+1)+t}{n}.$$

□

## B Proof of Theorem 25

**Proof:** To make the proof simpler, we compute each term of (28) individually, starting with the first.

$$\begin{aligned} xS_t(x, 1) \sum_{i=0}^t \partial_u S_{k-i-1, r-1}(x, u) \Big|_{u=1} &= z(1-z)^k S_t(x, 1) \sum_{i=0}^t \frac{1 - (1-z)^{k-i}}{z(1-z)^{k-i-1}} \\ &- z(1-z)^k S_t(x, 1) \sum_{i=0}^t \sum_{j=1}^{r-1} z^{j-1} (1-z)^{jk} \frac{k-i}{(k+1)j+k-i} \binom{(k+1)j+k-i}{j} \\ &+ z(1-z)^k S_t(x, 1) \sum_{i=0}^t \sum_{j=1}^{r-2} z^j (1-z)^{jk} \frac{k-i}{(k+1)j+k-i} \binom{(k+1)j+k-i}{j} \\ &- z(1-z)^k S_t(x, 1) \sum_{i=0}^t z^{r-1} (1-z)^{(r-1)k} \frac{(k-i-1)(k-i)}{(k+1)(r-1)+k-i} \binom{(k+1)(r-1)+k-i}{r-1}. \end{aligned}$$

Simplifying the sum that is a geometric series (and including its excess terms into the  $j = 0$  cases of the second and third terms) as well as evaluating sums in the second and third sums by using (21), we obtain

$$\begin{aligned} xS_t(x, 1) \sum_{i=0}^t \partial_u S_{k-i-1, r-1}(x, u) \Big|_{u=1} &= \frac{1 - (1-z)^{t+1}}{z(1-z)^t} \\ &- \left( \frac{1}{z} S_t(x, 1) - S_t(x, 1) \right) \sum_{j=0}^{r-1} z^{j+1} (1-z)^{(j+1)k} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j} \\ &- z^r (1-z)^{rk} S_t(x, 1) \sum_{i=0}^t \frac{(k-i)^2}{(k+1)(r-1)+k-i} \binom{(k+1)(r-1)+k-i}{r-1}. \end{aligned}$$

Set  $B_{j,k,t} := \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j}$ , the second term from (28) simplifies to

$$x \sum_{j=0}^{r-1} x^j \partial_u S_{t, r-1-j}(x, u) \Big|_{u=1} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j}$$

$$\begin{aligned}
&= \sum_{j=0}^{r-1} z^{j+1} (1-z)^{k(j+1)} \frac{1-(1-z)^{t+1}}{z(1-z)^t} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j} \\
&\quad - \sum_{j=0}^{r-1} \sum_{\ell=1}^{r-1-j} z^{\ell+j} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\
&\quad + \sum_{j=0}^{r-1} \sum_{\ell=1}^{r-2-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\
&\quad - \sum_{j=0}^{r-1} z^r (1-z)^{rk} \frac{t(t+1)}{(k+1)(r-1-j)+t+1} \binom{(k+1)(r-1-j)+t+1}{r-1-j} B_{j,k,t}.
\end{aligned}$$

This is further simplified by applying Proposition 26 and rewriting  $(1-(1-z)^t)/(z(1-z)^t)$  in terms of  $S_t(x, 1)$ ,

$$\begin{aligned}
&x \sum_{j=0}^{r-1} x^j \partial_u S_{t,r-1-j}(x, u) \Big|_{u=1} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j} \\
&= \left( \frac{1}{z} S_t(x, 1) - S_t(x, 1) - \frac{1}{z} + 1 \right) \sum_{j=0}^{r-1} z^{j+1} (1-z)^{k(j+1)} B_{j,k,t} \\
&\quad - \frac{1}{z} \sum_{j=0}^{r-1} \sum_{\ell=1}^{r-1-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\
&\quad + \sum_{j=0}^{r-1} \sum_{\ell=1}^{r-2-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\
&\quad - z^r (1-z)^{rk} \frac{t(t+1)}{(k+1)r+t+1} \binom{(k+1)r+t+1}{r},
\end{aligned}$$

and finally, including the terms for  $\ell = 0$  into each of the sums,

$$\begin{aligned}
&x \sum_{j=0}^{r-1} x^j \partial_u S_{t,r-1-j}(x, u) \Big|_{u=1} \frac{t+1}{j+1} \binom{(k+1)j+k-t-1}{j} \\
&= \left( \frac{1}{z} S_t(x, 1) - S_t(x, 1) \right) \sum_{j=0}^{r-1} z^{j+1} (1-z)^{k(j+1)} B_{j,k,t} \\
&\quad - \frac{1}{z} \sum_{j=0}^{r-1} \sum_{\ell=0}^{r-1-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\
&\quad + \sum_{j=0}^{r-1} \sum_{\ell=0}^{r-2-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t}
\end{aligned}$$

$$- z^r (1-z)^{rk} \frac{t(t+1)}{(k+1)r+t+1} \binom{(k+1)r+t+1}{r}.$$

Combining these terms and the final term of (28) which has not been simplified gives us

$$\begin{aligned} \partial_u S_{t,r}(x,u)|_{u=1} &= \frac{1-(1-z)^{t+1}}{z(1-z)^t} \\ &- \left( \frac{1}{z} S_t(x,1) - S_t(x,1) \right) \sum_{j=0}^{r-1} z^{j+1} (1-z)^{(j+1)k} B_{j,k,t} \\ &- z^r (1-z)^{rk} S_t(x,1) \sum_{i=0}^t \frac{(k-i)^2}{(k+1)(r-1)+k-i} \binom{(k+1)(r-1)+k-i}{r-1} \\ &+ \left( \frac{1}{z} S_t(x,1) - S_t(x,1) \right) \sum_{j=0}^{r-1} z^{j+1} (1-z)^{k(j+1)} B_{j,k,t} \\ &- \frac{1}{z} \sum_{j=0}^{r-1} \sum_{\ell=0}^{r-1-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\ &+ \sum_{j=0}^{r-1} \sum_{\ell=0}^{r-2-j} z^{\ell+j+1} (1-z)^{k(\ell+j+1)} \frac{t+1}{(k+1)\ell+t+1} \binom{(k+1)\ell+t+1}{\ell} B_{j,k,t} \\ &- z^r (1-z)^{rk} \frac{t(t+1)}{(k+1)r+t+1} \binom{(k+1)r+t+1}{r} \\ &+ z^r (1-z)^{rk} S_t(x,1) \sum_{i=0}^t \frac{(k-i)^2}{(k+1)(r-1)+k-i} \binom{(k+1)(r-1)+k-i}{r-1}. \end{aligned}$$

From this, there are some clear cancellations of factors and we re-index the double sums, which leads to

$$\begin{aligned} \partial_u S_{t,r}(x,u)|_{u=1} &= \frac{1-(1-z)^{t+1}}{z(1-z)^t} \\ &- \frac{1}{z} \sum_{m=1}^r z^m (1-z)^{km} \sum_{j=0}^{m-1} \frac{t+1}{(k+1)(m-1-j)+t+1} \binom{(k+1)(m-1-j)+t+1}{m-1-j} B_{j,k,t} \\ &+ \sum_{m=1}^{r-1} z^m (1-z)^{km} \sum_{j=0}^{m-1} \frac{t+1}{(k+1)(m-1-j)+t+1} \binom{(k+1)(m-1-j)+t+1}{m-1-j} B_{j,k,t} \\ &- z^r (1-z)^{rk} \frac{t(t+1)}{(k+1)r+t+1} \binom{(k+1)r+t+1}{r}, \end{aligned}$$

and finally, once again using Proposition 26 we have obtained an expression which is in the form of Theorem 25, concluding the proof

$$\partial_u S_{t,r}(x,u)|_{u=1} = \frac{1-(1-z)^{t+1}}{z(1-z)^t} - z^r (1-z)^{kr} \frac{t(t+1)}{(k+1)r+t+1} \binom{(k+1)r+t+1}{r}$$

$$\begin{aligned}
& - \sum_{m=1}^r z^{m-1} (1-z)^{km} \frac{t+1}{(k+1)m+t+1} \binom{(k+1)m+t+1}{m} \\
& + \sum_{m=1}^{r-1} z^m (1-z)^{km} \frac{t+1}{(k+1)m+t+1} \binom{(k+1)m+t+1}{m}.
\end{aligned}$$

□