Lagrange statistics on Dyck paths

Donatella Merlini, Renzo Sprugnoli, M. Cecilia Verri
Dipartimento di Sistemi e Informatica
via Lombroso 6/17, 50134 Firenze, Italia

Abstract

We study some statistics related to Dyck paths, whose explicit formulas are obtained by means of the Lagrange Inversion Theorem. There are five such statistics and one of them is well-known and owed to Narayana. The most interesting of the other 4 statistics is related to Euler's trinomial coefficients: we perform a study of that statistic proving a number of its properties.

1 Introduction

Dyck paths are very well-known combinatorial objects that have been widely studied from various points of view (see, e.g., [2, 3, 4, 6, 7]). Together with many other objects, they are counted by Catalan numbers. A surprisingly large number of 1-1-correspondences are known that relate these paths to other classes of objects, such as binary trees, parenthesized expressions, polygon triangulations and so on.

Other properties of Dyck paths, related to Catalan numbers, have also been studied. For example, the so-called Catalan triangle in Table 1 is defined by the fact that its generic element $c_{n,k}$ counts the number of partial Dyck paths arriving at the point $(n, n-k)$; in that case, Catalan numbers appear both as column 0 and as the row sums of the infinite triangle.

A common way of defining Dyck paths is through a context-free grammar: Dyck paths correpond to Dyck words, i.e., sequences of 0, 1 belonging to the language generated by the productions (written in Backus-Naur form)

$$D = \epsilon \mid 0D1D$$

n/k	0	1	2	3	4	5
0	1					
1	1	1				
2	2	2	1			
3	5	5	3	1		
4	14	14	9	4	1	
5	42	42	28	14	5	1

Table 1: The Catalan triangle

n/k	0	1	2	3	4	5
0	1					
1	0	1				
2	0	1	1			
3	0	1	3	1		
4	0	1	6	6	1	
5	0	1	10	20	10	1

Table 2: Narayana distribution

where ϵ denotes the empty word. In the usual representation as underdiagonal paths, 0 corresponds to an east step and 1 corresponds to a north step. If we wish to count Dyck paths according to their semilength n and to the number k of their valleys (i.e., two consecutive steps east and north), it is sufficient to count Dyck words according to their semi-length (as words) and according to the number of their subsequences 01. The resulting statistic, shown in Table 2, is called *Narayana distribution* and Sulanke [9, 10] describes a lot of properties of Dyck paths sharing the same distribution.

If $\nu_{n,k}$ is the generic element in the Narayana triangle, it is known that

$$\nu_{n,k} = \frac{1}{n}\binom{n}{k}\binom{n}{k-1} \tag{1}$$

This formula is easily proved by means of Lagrange Inversion Formula (LIF), (see, [2]), as we will show in the next section, and the same method can be used to find out formulas for other statistics related to Dyck paths. Because of that, we call them *Lagrange statistics*, and in this paper we particularly investigate one of them. This will be called *Trinomial statistics*, since it is related to the trinomial coefficients of Euler; we prove a number of properties, ranging from an explicit formula to generating functions and to unimodality.

The paper is organized in the following way: in Section 2 we classify and describe Lagrange statistics; Section 3 illustrates the trinomial statistics and prove some general properties; finally, in Section 4, we analyze more deeply the corresponding triangle, trying to characterize it by means of recurrences and generating functions.

2 Lagrange statistics

We are now going to use Narayana distribution as our introductory example to Lagrange statistics for Dyck paths. Let us suppose we wish to count Dyck paths of semi-length n (i.e., arriving at the point (n, n)), having exactly k valleys, or, equivalently, Dyck words of length $2n$ having exactly k occurences of the string 01. We can proceed by considering the ϵ-free context-free grammar generating Dyck words; according to standard methods (i.e., substituting ϵ in all the possible ways in the occurrences of the non-terminal symbol D) we obtain

$$D = 01 \mid 0D1 \mid 01D \mid 0D1D. \tag{2}$$

Since by this definition D does not contain the empty word, possible occurrences of 01 are found in D, in the word 01 and in all the words generated by $01D$. In fact, an occurence of 01 cannot be after the 0 in $0D1$ because D always begins with 0; for the same reason, no other occurrence of 01 can be found in $0D1D$, except the ones already present in D. We use Schützenberger methodology to pass from the grammar to the corresponding generating function. Let us use the indeterminate t for counting couples 0/1, and the indeterminate w for counting the occurrences of the string 01. The bivariate generating function $D(t, w)$ is given by the solution of

$$D(t, w) = tw + twD(t, w) + tD(t, w) + tD(t, w)^2. \tag{3}$$

An explicit form for $D(t, w)$ is easily obtained

$$D(t, w) = \frac{1 - t - tw - \sqrt{1 - 2t - 2tw + t^2 - 2t^2w + t^2w^2}}{2tw}.$$

This formula gives no hint on how the form of $[t^n w^k]D(t, w) = \nu_{n,k}$ should be. A classical application of the Lagrange Inversion Formula, however, gives the appropriate result. Formula (3) can be written as

$$D = t(w + wD + D + D^2) = t(1 + D)(w + D);$$

therefore, if we set $\phi(w) = (1 + D)(w + D)$, we have

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

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

This is the celebrated formula of Narayana for the present and many other statistics on Dyck paths. If we look at how this result has been obtained, we observe that

- the presence of a single couple 0/1 in the productions for D allows us to have a single t in the functional equation defining $D(t, w)$; t is collected from the various terms and this determines the application of the Lagrange Inversion Formula, which requires a relation $D = t\phi(w, D)$;
- the indeterminate w seems to play a secondary role, but its appearence in two of the terms in the functional equation determines the final form of the Narayana formula.

As a consequence, we may deduce that by varying the position of the indeterminate w in the functional equation for $D(t, w)$ other statistics are obtained with the same method of the Lagrange Inversion Formula. Let us give a name to the four productions in (2): (a) is the first one with 01 as a right hand member; (b) is the second one with $0D1$; (c) is the third one with $01D$ and finally, (d) is the fourth one with $0D1D$ as a right hand member. In this way, we can identify a distribution by the combination of letters corresponding to the terms in which the indeterminate w appears. For example, case (a, b) corresponds to the Narayana statistic, and we can investigate all the different possible situations. We observe that, from

n/k	0	1	2	3	4	5
0	1					
1	0	1				
2	0	2	0			
3	0	4	1	0		
4	0	8	6	0	0	
5	0	16	24	2	0	0

(i)

n/k	0	1	2	3	4	5
0	1					
1	1	0				
2	1	1	0			
3	2	2	1	0		
4	4	6	3	1	0	
5	9	16	12	4	1	0

(ii)

Table 3: Two Lagrange statistics

an algebraic point of view, the indeterminate w in position (b) is equivalent to w in position (c); therefore, only ten cases are possible:

$$(a), \quad (b) = (c), \quad (d), \quad (a, b) = (a, c), \quad (a, d), \quad (b, c), \quad (b, d) = (c, d),$$
$$(a, b, c), \quad (a, b, d) = (a, c, d), \quad (b, c, d).$$

Obviously, the two extreme cases: w nowhere and w in (a, b, c, d) are not interesting, and will be ignored. We also wish to point out that a symmetry exists and, for example, statistic (a) is symmetric to (b, c, d), where w appears in complementary positions. In this way, the number of possible statistics is reduced to five, and in this section we will find an explicit formula for each of them.

Let us begin with case (a). Suppose we change production (a) $D \to 01$ into $D \to XX$, where X is a new terminal symbol; we wish to count the new words according to their semilength and to the number of couples XX. The Schützenberger methodology gives the relation:

$$D = t(w + D + D + D^2) = t(w + D(2 + D)).$$

In this case $\phi(w, D) = (w + D(2 + D))$ and therefore the LIF gives:

$$[t^n w^k]D = \frac{1}{n}[D^{n-1} w^k](w + D(2 + D))^n = \frac{1}{n}\binom{n}{k}[D^{n-1}]D^{n-k}(2 + D)^{n-k} =$$
$$= \frac{1}{n}\binom{n}{k}[D^{k-1}](2 + D)^{n-k} = \frac{1}{n}\binom{n}{k}\binom{n-k}{k-1}2^{n-2k+1}.$$

In Table 3(i) we show the initial part of the infinite triangle corresponding to this statistic.

The second case is $(b) = (c)$. Suppose we wish to count Dyck words according to semilength and to the number of occurrences of the sequence 010. These sequences either occur inside a D or are generated by the production $D \to 01D$, since D always begins by 0. In all the other cases, new 010 sequences can never be generated. The Schützenberger methodology gives:

$$D = t(1 + wD + D + D^2) = t(wD + (1 + D + D^2)).$$

Here we have $\phi(w, D) = wD + (1 + D + D^2)$ and therefore by the LIF:

$$[t^n w^k]D = \frac{1}{n}[D^{n-1} w^k](wD + (1 + D + D^2))^n = \frac{1}{n}\binom{n}{k}[D^{n-1}]D^k(1 + D + D^2)^{n-k} =$$

$$= \frac{1}{n}\binom{n}{k}[D^{n-k-1}](1+D+D^2)^{n-k} = \frac{1}{n}\binom{n}{k}\overline{T}_{n-k}.$$

It is immediate to observe that $[D^{n-k-1}](1+D+D^2)^{n-k}$ is related to the trinomial coefficients of Euler, which are defined as $T_n = [t^n](1+t+t^2)^n$. In the next section we will investigate more deeply this connection; for the moment, let us denote by $\overline{T}_n$ the coefficient of t^{n-1} in $(1+t+t^2)^n$ and so obtain a formula for this statistics, which therefore will be called *trinomial distribution for Dyck paths*; it will be the main object of our studies in the next sections. The upper part of the corresponding infinite triangle is shown in Table 3(ii); its elements will be denoted by $\tau_{n,k}$.

The third case is (d). Suppose we wish to count Dyck words according to semilength and to the number of applications of the production $D \to 0D1D$ in the generation of the word (i.e., in its syntactic tree). We have:

$$D = t(1+D+D+wD^2) = t((1+2D)+wD^2).$$

Here $\phi(D,w) = (1+2D)+wD^2$ and therefore the LIF gives:

$$[t^n w^k]D = \frac{1}{n}[D^{n-1}w^k]((1+2D)+wD^2)^n = \frac{1}{n}\binom{n}{k}[D^{n-1}]D^{2k}(1+2D)^{n-k} =$$

$$= \frac{1}{n}\binom{n}{k}[D^{n-2k-1}](1+2D)^{n-k} = \frac{1}{n}\binom{n}{k}\binom{n-k}{n-2k-1}2^{n-2k-1} = \frac{1}{n}\binom{n}{k}\binom{n-k}{k+1}2^{n-2k-1}.$$

This distribution is strongly related to (a). In fact, if in the formula for (a) we perform the transformation $k \to k+1$ we have:

$$\frac{1}{n}\binom{n}{k+1}\binom{n-k-1}{k}2^{n-2k-1} = \frac{1}{n}\binom{n}{n-k-1}\binom{n-k-1}{k}2^{n-2k-1} =$$

$$= \frac{1}{n}\binom{n}{k}\binom{n-k}{n-2k-1}2^{n-2k-1}$$

which is distribution (d); therefore, this latter statistic is a simple translation of the former.

The fourth case is $(a,b) = (a,c)$ and, as we have already seen, it corresponds to Narayana statistic.

Finally, the fifth case is (a,d). Here we have:

$$D = t(w+D+D+wD^2) = t(w(1+D^2)+2D)$$

and the Lagrange Inversion Formula gives:

$$[t^n w^k]D = \frac{1}{n}[D^{n-1}w^k](w(1+D^2)+2D)^n = \frac{1}{n}\binom{n}{k}[D^{n-1}](1+D^2)^k 2^{n-k}D^{n-k} =$$

$$= \frac{1}{n}\binom{n}{k}2^{n-k}[D^{k-1}](1+D^2)^k = \frac{1}{n}\binom{n}{k}\binom{k}{(k-1)/2}2^{n-k}.$$

n/k	-5	-4	-3	-2	-1	0	1	2	3	4	5
0						1					
1					1	1	1				
2				1	2	3	2	1			
3			1	3	6	7	6	3	1		
4		1	4	10	16	19	16	10	4	1	
5	1	5	15	30	45	51	45	30	15	5	1

Table 4: The trinomial coefficients

As usual, a non-integer "denominator" in a binomial coefficient gives a zero result. This distribution is also related to (a); in fact, by substituting in the last formula $2k-1$ to k, we obtain:

$$\frac{1}{n}\binom{n}{2k-1}\binom{2k-1}{k-1}2^{n-2k+1} = \frac{1}{n}\binom{n}{k}\binom{n-k}{n-2k+1}2^{n-2k+1}$$

which is exactly the formula for distribution (a). The reader is invited to write down the upper part of the corresponding triangle. As is apparent from the formula, columns with odd indeces are all zero, while columns with even indeces coincide with the non-null columns of (a).

These are all the possible Lagrange statistics for Dyck paths (or words); they are reduced to three main cases: Narayana distribution, Trinomial distribution and (a) distribution. Obviously, they do not exhaust the possible statistics on Dyck paths, but surely they are relatively easy to study, since we have obtained explicit formulas for them. In the next section we will study the trinomial distribution and observe that a similar study could also be done for (a) distribution.

3 The trinomial statistics

Central trinomial coefficients $\{T_n\}_{n\in\mathbf{N}} = \{1, 1, 3, 7, 19, 51, \ldots\}$, i.e., coefficients $[t^n](1+t+t^2)^n$, $n \in \mathbf{N}$, were studied for the first time by Euler. In some way, they are related to binomial coefficients and, in fact, by taking $T_{n,k} = [t^{n-k}](1+t+t^2)^n$, we obtain an infinite triangle $T = \{T_{n,k}\}_{n,k\in\mathbf{N}}$ that is very similar to Pascal's triangle (see Table 4). Each of its elements not belonging to column 0 is obtained by summing up the three elements above it. Furthermore, by an obvious argument, column 0's elements are obtained by summing the previous element in the same column and the element in the previous row and column 1 multiplied by two. In the literature, the elements of this triangle are known as *trinomial coefficients* but, here on, we will use this name to denote central trinomial coefficients T_n.

Let us call $T = \{T_{n,k}\}_{n,k\in\mathbf{N}}$ this array; in the theory of Riordan Arrays (see [5, 8]) this means that T is actually a Riordan Array with A-sequence $1+t+t^2$ and Z-sequence $1+2t$; from these sequences the form of the Riordan Array is easily deduced:

$$T = \left(\frac{1}{\sqrt{1-2t-3t^2}}, \frac{1-t-\sqrt{1-2t-3t^2}}{2t^2}\right).$$

In other words, the generating function of column k is:

$$T_k(t) = \frac{1}{\sqrt{1-2t-3t^2}} \left(\frac{1-t-\sqrt{1-2t-3t^2}}{2t^2} \right)^k$$

and, in particular, the generating function for columns 0 and 1 are

$$T_0(t) = \frac{1}{\sqrt{1-2t-3t^2}}, \quad T_1(t) = \overline{T}(t) = \frac{1-t-\sqrt{1-2t-3t^2}}{2t\sqrt{1-2t-3t^2}}.$$

Before proceeding with this summary of trinomial coefficient properties, we wish to observe that the triangle T can also be obtained as a result in a lattice path problem. Let us consider underdiagonal lattice paths in $\mathbf{Z}^2$ composed by the coloured steps $(1, 0, black)$, $(1, 1, black)$, $(1, 2, black)$ and, for steps ending on the main diagonal $x-y=0$, by $(1, 1, black)$, $(1, 2, black)$, $(1, 2, red)$. Then $T_{n,k}$ represents the number of (coloured) paths starting at the origin and ending at $(n, n-k)$. In this interpretation, trinomial coefficients count the number of coloured paths ending on the main diagonal. Coming back to the original definition of the trinomial triangle, we immediately see that the elements in column 1 are defined as $[t^{n-1}](1+t+t^2)^n$, and therefore are just the coefficients appearing in our formula for the trinomial distribution of Dyck paths. This justifies our notation $\overline{T}(t)$ for $T_1(t)$. Since trinomial coefficients are better known than the coefficients $\overline{T}_n$ (which will be called *sub-trinomial coefficients*), let us begin our study by expressing $\overline{T}_n$ in terms of the T_n's:

Lemma 3.1 *The sub-trinomial coefficients $\overline{T}_n = [t^{n-1}](1+t+t^2)^n$ can be expressed in terms of the trinomial coefficients by the formula:*

$$\overline{T}_n = \frac{T_{n+1} - T_n}{2}.$$

Proof: As already observed and as an immediate consequence of the definitions, T_{n+1} is the sum of the previous element T_n plus twice the element of column 1 in the previous row, i.e., $\overline{T}_n$. This gives $T_{n+1} = T_n + 2\overline{T}_n$, which is just what we were looking for. ■

As a consequence of this lemma, the formula for trinomial distribution can be written in the form:

$$\tau_{n,k} = \frac{1}{n}\binom{n}{k}\frac{T_{n-k+1} - T_{n-k}}{2}.$$

We can now give a complete characterization of the trinomial statistics by finding both a bivariate generating function for it and some relations allowing us to build the whole triangle one element after the other.

Theorem 3.2 *The bivariate generating function $\tau(t, w)$ for the trinomial distribution is:*

$$\tau(t, w) = \frac{1-t-tw-\sqrt{1-2t-3t^2-2tw+2t^2w+t^2w^2}}{2t}$$

and therefore the generating function of column 0 is:

$$\tau(t) = \frac{1-t-\sqrt{1-2t-3t^2}}{2t}.$$

Proof: From Section 2 we know that the functional equation defining $\tau(t,w)$, there written D, is $\tau = t(1+w\tau+\tau+\tau^2)$; therefore, the function $\tau(t,w)$ is the solution of this second degree equation in τ having $\tau(0,0)=1$. By a simple application of the de l'Hôpital theorem, we see that the solution with a sign $-$ in front of the square root is the right one, and this proves the formula for $\tau(t,w)$. The generating function for the column 0 is now obtained[1] as $\tau(t,0)=\tau(t)$ and this proves the second formula. ∎

The bivariate generating function mathematically characterizes an infinite triangle like our trinomial statistics; however, it is difficult to try to extract from it useful information on the distribution. Some system of Computer Algebra is good for developing it into a McLauren series:

$$\tau(t,w) = 1 + t(1+w) + t^2(2+2w+w^2) + t^3(4+6w+3w^2+w^3) + O(t^4)$$

but a human would prefer some simpler characterization. For example, we can look for recurrence relations allowing us to build the elements $\tau_{n,k}$ one after the other, as the recurrence $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ does for binomial coefficients in the Pascal triangle.

We will now find out a recurrence for the elements in column 0; later on we'll be able to determine a relation between $\tau_{n+1,k+1}$ and $\tau_{n,k}$. Therefore, we can build the whole triangle by first expanding column 0 up to the index we are interested in, and then by extending the triangle to all the other columns. Let us denote $\{\tau_0, \tau_1, \tau_2, \ldots\}$ the coefficients of the column 0 in the triangle of the trinomial distribution; they will be called *co-trinomial coefficients*.

Theorem 3.3 *The co-trinomial coefficients $\{\tau_n\}_{n\in\mathbf{N}}$ are defined by the second order recurrence:*

$$\tau_n = \frac{1}{n+1}((2n-1)\tau_{n-1} + 3(n-2)\tau_{n-2})$$

and the initial conditions $\tau_0 = 1, \tau_1 = 1$.

Proof: The generating function $\tau(t)$ is algebraic and therefore its coefficients should satisfy a recurrence relation with polynomial coefficients. To find this relation we can follow this standard method. We try to find out a differential equation $p_1(t)\tau'(t) + p_2(t)\tau(t) = 1$ of which $\tau(t)$ is a solution. The two coefficients $p_1(t)$ and $p_2(t)$ should be rational functions in t. If we denote by Q the radical $\sqrt{1-2t-3t^2}$, we easily obtain:

$$\tau(t) = \frac{1-t-Q}{2t}, \quad \tau'(t) = \frac{(1-t)Q - (1-2t-3t^2)}{2t^2(1-2t-3t^2)}.$$

If we now substitute these expressions in the differential equation, we can separate the rational part (not depending on Q) from the coefficients of Q (the irrational part). We thus obtain a system of equations:

$$\begin{cases} (1-t)p_1(t) - t(1-2t-3t^3)p_2(t) = 0 \\ (1-2t-3t^2)p_1(t) - t(1-3t-t^2+3t^3)p_2(t) = -2t^2(1-2t-3t^2) \end{cases}$$

[1]If $F(t,w)$ is the generating function of a lower triangular array, we consistently use $F(t) = F(t,0)$ for the generating function of column 0.

and by solving it we have:

$$p_1(t) = \frac{1-2t-3t^2}{2}, \quad p_2(t) = \frac{1-t}{2t}.$$

The differential equation is therefore $(1-2t-3t^2)t\tau'(t)+(1-t)\tau(t) = 2t$ and we can extract the coefficients of $[t^n]$ from both sides:

$$[t^{n-1}]\tau'(t) - 2[t^{n-2}]\tau'(t) - 3[t^{n-3}]\tau'(t) + [t^n]\tau(t) - [t^{n-1}]\tau(t) = 2\delta_{n,1}$$

$$n\tau_n - 2(n-1)\tau_{n-1} - 3(n-2)\tau_{n-2} + \tau_n - \tau_{n-1} = 2\delta_{n,1}$$

The Kronecker δ on the right hand member has no meaning, because this relation should be valid for $n > 1$; therefore, we can substitute it by 0 and this immediately gives the desired recurrence. The initial conditions are easily determined by direct considerations. ■

We must now determine a relation connecting an element to the elements in the previous column. This is particularly easy for our trinomial statistics, for which we have:

Theorem 3.4 *The coefficients $\tau_{n,k}$ in the triangle of trinomial distribution are related to one another by the formula:*

$$\tau_{n+1,k+1} = \frac{n}{k+1}\tau_{n,k}$$

which allows us to build the whole triangle starting from column 0. *As a consequence, the generating functions for columns are related by:*

$$\tau_{k+1}(t) = \frac{t^2}{k+1}\tau_k'(t), \quad k \geq 0.$$

Proof: We immediately have from Section 1:

$$\tau_{n+1,k+1} = \frac{1}{n+1}\binom{n+1}{k+1}\overline{T}_{n+1-k-1} = \frac{1}{k+1}\binom{n}{k}\overline{T}_{n-k} = \frac{n}{k+1}\frac{1}{n}\binom{n}{k}\overline{T}_{n-k}$$

which is the relation desired. We can now look at this formula as at a relation between the coefficients of the two generating functions $\tau_k(t)$ and $\tau_{k+1}(t)$ for columns k and $k+1$. From the theory of formal power series, we know that if $f(t)$ is the generating function for a sequence $\{f_n\}_{n\in\mathbf{N}}$, then we have:

$$\mathcal{G}\{f_{n+1}\} = \frac{f(t)-f_0}{t} \quad \text{and} \quad \mathcal{G}\{nf_n\} = tf'(t)$$

where $\mathcal{G}$ denotes the "generating function operator". For $k > 0$, the coefficient of position 0 in every column is zero, and therefore the relation above translates into:

$$\frac{\tau_{k+1}(t)}{t} = \frac{1}{k+1}t\tau_k'(t)$$

which is the formula desired. Obviously, for $k = 0$ we have $\tau_1(t) = \overline{T}(t) = t^2\tau'(t)$. ■

Note the analogous formulas for Narayana and (a) distributions (denoted by $\tau_{n,k}$ and $\alpha_{n,k}$, respectively):

$$\tau_{n+1,k+1} = \frac{n(n+1)}{k(k+1)}\tau_{n,k}, \qquad \alpha_{n+1,k+1} = \frac{n(n-2k+1)}{2k(k+1)}\alpha_{n,k}.$$

We conclude our general considerations on the trinomial distribution by proving that it is unimodal. More specifically, we prove:

Theorem 3.5 *The trinomial distribution is unimodal and in row n it attains its maximum for $k \approx n/4$.*

Proof: Let us study the ratio between two consecutive coefficients of the trinomial statistics in row n. We obviously have:

$$\frac{\tau_{n,k+1}}{\tau_{n,k}} = \frac{\frac{1}{n}\binom{n}{k+1}\overline{T}_{n-k-1}}{\frac{1}{n}\binom{n}{k}\overline{T}_{n-k}} = \frac{k!(n-k)!}{(k+1)!(n-k-1)!}\frac{\overline{T}_{n-k-1}}{\overline{T}_{n-k}} = \frac{n-k}{k+1}\frac{\overline{T}_{n-k-1}}{\overline{T}_{n-k}}.$$

The generating function $\overline{T}(t)$ has two algebraic singularities at $t = 1$ and $t = 1/3$. This latter singularity is dominant and therefore the radius of convergence for $\overline{T}(t)$ is $1/3$. This means that the ratio of two consecutive coefficients of $\overline{T}(t)$ approaches $1/3$ as their indices tend to infinity. Therefore, for large values of n, we can substitute $1/3$ to the ratio $\frac{\overline{T}_{n-k-1}}{\overline{T}_{n-k}}$ and see when the corresponding ratio $\frac{\tau_{n,k+1}}{\tau_{n,k}}$ is (approximately) equal to 1 :

$$\frac{n-k}{k+1}\frac{1}{3} \approx 1 \quad \text{or} \quad n-k \approx 3k+3 \quad \text{or} \quad k \approx \frac{n-3}{4}.$$

Let us call $\overline{k}$ this special value of k. When $k < \overline{k}$ the ratio $\frac{\tau_{n,k+1}}{\tau_{n,k}}$ is greater then 1, and therefore the values in row n are increasing; when $k > \overline{k}$ the ratio is smaller then 1 and the values are decreasing. This shows that the distribution is unimodal and attains its maximum for $k = \overline{k} \approx n/4$, as desired. ∎

We conclude here our study of the basic properties of the trinomial distribution. In the next section we will prove some more specific properties and consequences of the considerations above.

4 Other properties of the trinomial distribution

In the previous sections we have taken into consideration three main sequences: trinomial coefficients $\{T_n\}_{n\in\mathbf{N}}$; sub-trinomial coefficients $\{\overline{T}_n\}_{n\in\mathbf{N}}$ and co-trinomial coefficients $\{\tau_n\}_{n\in\mathbf{N}}$. In order to complete our study, we now give some general information on these quantities, i.e., we show recurrence relations allowing us to compute these coefficients in an efficient way and give their asymptotic values, showing how they grow and providing an easy method to compute their approximations. These formulas are well known in the particular case of trinomial coefficients and we give them here only to show their connection and similarity with the other quantities.

Theorem 4.1 *Trinomial, sub-trinomial and co-trinomial coefficients satisfy the following recurrence relations with initial conditions:*

$$nT_n = (2n-1)T_{n-1} + 3(n-1)T_{n-2}, \quad T_0 = 1, T_1 = 1,$$

$$(n^2-1)\overline{T}_n = n(2n-1)\overline{T}_{n-1} + 3n(n-1)\overline{T}_{n-2}, \quad \overline{T}_0 = 0, \overline{T}_1 = 1,$$

$$(n+1)\tau_{n+1} = (2n-1)\tau_{n-1} + 3(n-2)\tau_{n-2}, \quad \tau_0 = 1, \tau_1 = 1.$$

Proof: The recurrence relation for co-trinomial coefficents was found in Theorem 3.3 of the previous section and is repeated here for the sake of completeness. The recurrence relation for trinomial coefficients is well-known (see, e.g., [1]) and is easily obtained with the method shown in the proof of Theorem 3.3. By Theorem 3.4, for $k = 0$ we have $\tau_{n+1} = n\tau_{n,0}$; however, by definition $\tau_{n,0} = \tau_n$ and as already observed $\tau_{n+1,1} = \overline{T}_n$. Therefore we have $\tau_n = \overline{T}_n/n$ and by substituting this formula in the recurrence relation for τ_n we immediately find the recurrence relation for sub-trinomial coefficients. ∎

Concerning the asymptotic values, we use the standard method of singularities in generating functions. In all three cases, the generating functions $T(t), \overline{T}(t)$ and $\tau(t)$ have a dominating singularity at $t = 1/3$:

Theorem 4.2 *Trinomial, sub-trinomial and co-trinomial coefficients have the following asymptotic expressions:*

$$T_n = \frac{\sqrt{3}}{2}\binom{2n}{n}\left(\frac{3}{4}\right)^n\left(1 - \frac{1}{8(2n-1)} + \frac{9}{128(2n-1)(2n-3)} + O(\frac{1}{n^3})\right),$$

$$\overline{T}_n = \frac{\sqrt{3}}{6}\binom{2n+2}{n+1}\left(\frac{3}{4}\right)^{n+1}\left(1 - \frac{5}{8(2n+1)} + \frac{33}{128(2n+1)(2n-1)} + O(\frac{1}{n^3})\right),$$

$$\tau_n = \frac{\sqrt{3}}{2n-1}\binom{2n}{n}\left(\frac{3}{4}\right)^n\left(1 - \frac{21}{8(2n-3)} + \frac{1665}{128(2n-3)(2n-5)} + O(\frac{1}{n^3})\right).$$

Proof: The three generating functions are to be developed around $t = 1/3$; this is easily done by hand and more easily done by some system of Computer Algebra:

$$T(t) = \frac{1}{\sqrt{1-2t-3t^2}} = \frac{\sqrt{3}}{2}\frac{1}{\sqrt{1-3t}} + \frac{\sqrt{3}}{16}\sqrt{1-3t} + \frac{3\sqrt{3}}{256}(1-3t)^{3/2} + \cdots$$

$$\overline{T}(t) = \frac{1-t-\sqrt{1-2t-3t^2}}{2t\sqrt{1-2t-3t^2}} = \frac{\sqrt{3}}{6}\frac{1}{\sqrt{1-3t}} - \frac{1}{2} + \frac{5\sqrt{3}}{48}\sqrt{1-3t} + \frac{11\sqrt{3}}{768}(1-3t)^{3/2} + \cdots$$

$$\tau(t) = 1 - \sqrt{3}\sqrt{1-3t} + \frac{3}{2}(1-3t) - \frac{7\sqrt{3}}{8}(1-3t)^{3/2} + \frac{3}{2}(1-3t)^2 - + \frac{111\sqrt{3}}{128}(1-3t)^{5/2} + \cdots$$

We can now extract the coefficient of t^n by remembering the classical identities connecting binomial coefficients with half-integer numerator and central binomial coefficients:

$$\binom{-1/2}{n} = \frac{(-1)^n}{4^n}\binom{2n}{n}, \quad \binom{3/2}{n} = \frac{(-1)^n 3}{4^n(2n-1)(2n-3)}\binom{2n}{n},$$

$$\binom{1/2}{n} = \frac{(-1)^{n-1}}{4^n(2n-1)}\binom{2n}{n}, \qquad \binom{5/2}{n} = \frac{(-1)^{n-1}15}{4^n(2n-1)(2n-3)(2n-5)}\binom{2n}{n}.$$

From the coefficients appearing in the corrections inside the parentheses, we deduce that the approximation for trinomial coefficients is good, the approximation for sub-trinomial coefficients is not so good and the one for co-trinomial coefficients is relatively poor. ■

The next result we wish to find is the average number of subsequences 010 in the Dyck words of length n. We begin with the following:

Lemma 4.3 *The total number of subsequences* 010 *in all the Dyck words of length n is given by:*

$$P_n = \frac{1}{2}\binom{2n}{n} - \binom{2n-1}{n-1} = \frac{n-1}{2(2n-1)}\binom{2n}{n}, \qquad n > 0$$

Proof: Let us consider the triangle of trinomial distribution; the number P_n is the weighted sum of row n or, in other words, if $\tau_n(w)$ is the generating function of row n, $P_n = [\tau_n'(w)|w = 1]$, i.e., the evaluation at $w = 1$ of the derivatives for $\tau_n(w)$. In a global way, we can find the generating function $P(t) = \mathcal{G}\{P_n\}_{n\in\mathbf{N}}$ by differentiating the bivariate generating function $\tau(t,w)$ with respect to w and then by substituting $w = 1$. We find:

$$\frac{\partial\tau(t,w)}{\partial w} = \frac{1}{2t}\left(\frac{t-t^2-t^2w}{\sqrt{1-2tw-2t+t^2w^2+2t^2w-3t^2}} - t\right),$$

and therefore:

$$P(t) = \left[\frac{\partial\tau(t,w)}{\partial w}\bigg| w = 1\right] = \frac{1-2t-\sqrt{1-4t}}{2\sqrt{1-4t}}.$$

From this generating function we find:

$$P_n = [t^n]\left(\frac{1}{\sqrt{1-4t}} - \frac{t}{\sqrt{1-4t}} - \frac{1}{2}\right) = \frac{1}{2}\binom{2n}{n} - \binom{2n-2}{n-1} - \frac{1}{2}\delta_{n,0}.$$

valid for every n. The last passage is obvious. ■

The announced average value is now easily found:

Theorem 4.4 *The average number of subsequences* 010 *in the Dyck paths of semilength n is given by:*

$$ave_n = \frac{n^2-1}{2(2n-1)} = \frac{n}{4} + \frac{1}{8} + O(\frac{1}{n}).$$

Proof: The total number of Dyck words of length n is the n-th Catalan number $C_n = \binom{2n}{n}/(n+1)$ and therefore ave_n is obtained by dividing the P_n found in the previous lemma by C_n. ■

We conclude our paper by solving the following problem. In the previous section we have seen that the trinomial distribution attains its maximun at $k \approx n/4$ in row n; we wish to find the asymptotic value of this maximum as $n \to \infty$. This is not difficult if we use the asymptotic value for $\overline{T}_n$ obtained in Theorem 4.1 and use Stirling approximation to evaluate binomial coefficients in the formula for $\tau_{n,k}$.

Theorem 4.5 *The asymptotic value for the maximum value in row n is given by:*

$$\tau_{n,(n-3)/4} \sim \frac{2\sqrt{\frac{2}{3}}4^n}{\pi n^2}.$$

Proof: The formula for $\tau_{n,(n-3)/4}$ is:

$$\tau_{n,(n-3)/4} = \frac{1}{n}\binom{n}{(n-3)/4}\overline{T}_{n-(n-3)/4} = \frac{1}{n}\binom{n}{(n-3)/4}\overline{T}_{3(n+1)/4}.$$

Theorem 4.1 now gives:

$$\overline{T}_{3(n+1)/4} \sim \frac{\sqrt{3}}{6}\binom{3(n+1)/2+2}{3(n+1)/4+1}\left(\frac{3}{4}\right)^{3(n+1)/4+1} \sim$$

$$\frac{\sqrt{3}}{6}\frac{4^{3(n+1)/4+1}}{\sqrt{\pi(1+3(n+1)/4)}}\left(\frac{3}{4}\right)^{3(n+1)/4+1} \sim \frac{3^{3(n+1)/4}}{\sqrt{\pi(n+7/3)}}.$$

The most difficult part is the binomial coefficient:

$$\binom{n}{(n-3)/4} \sim \frac{\sqrt{2\pi n}n^n/e^n}{\sqrt{2\pi(n-3)/4}((n-3)/4e)^{(n-3)/4}\sqrt{2\pi 3(n+1)/4}(3(n+1)/4e)^{3(n+1)/4}} =$$

$$= \frac{\sqrt{n}n^n}{\sqrt{6\pi(n-3)(n+1)/16}((n-3)/4)^{(n-3)/4}(3(n+1)/4)^{3(n+1)/4}} =$$

$$= \sqrt{\frac{8n}{3\pi(n^2-n-3)}}\frac{n^n 4^n}{(n-3)^{(n-3)/4}(n+1)^{3(n+1)/4}3^{3(n+1)/4}}.$$

We now take the logarithm of the parts with exponents:

$$\exp\left(n\ln n - \frac{n-3}{4}\ln(n-3) - \frac{3}{4}(n+1)\ln(n+1)\right) =$$

$$= \exp\left(n\ln n - \frac{n-3}{4}\ln\left(n\left(1-\frac{3}{n}\right)\right) - \frac{3}{4}(n+1)\ln\left(n\left(1+\frac{1}{n}\right)\right)\right) =$$

$$= \exp\left(n\ln n - \frac{n-3}{4}\ln n + \frac{n-3}{4}\frac{3}{n} - \frac{n-3}{4}\frac{9}{2n^2} + \cdots - \frac{3}{4}(n+1)\ln n + \right.$$

$$\left. -\frac{3}{4}(n+1)\frac{1}{n} + \frac{3}{4}\frac{n+1}{2n^2} + \cdots\right)$$

and obtain:

$$\binom{n}{(n-3)/4} \sim \sqrt{\frac{8}{3\pi(n-1)}}\frac{4^n e^{-15/4n}}{3^{3(n+1)/4}} \sim \sqrt{\frac{8}{3\pi n}}\frac{4^n}{3^{3(n+1)/4}}\left(1-\frac{15}{4n}+\cdots\right).$$

Finally, we can join these partial results:

$$\tau_{n,(n-3)/4} \sim \frac{3^{3(n+1)/4}}{\sqrt{\pi(n+7/3)}}\sqrt{\frac{8}{3\pi(n-1)}}\frac{4^n}{3^{3(n+1)/4}} \sim \frac{2\sqrt{\frac{2}{3}}4^n}{\pi n^2}$$

which is our final result. ∎

References

[1] E. Barcucci, R. Pinzani, and R. Sprugnoli. The Motzkin family. *Pure Mathematics and Applications*, 2:249–279, 1991.

[2] I. P. Goulden and D. M. Jackson. *Combinatorial Enumeration.* John Wiley & S., 1983.

[3] J. Labelle. On pairs of non-crossing generalized Dyck paths. *J. Statist. Plann. Inference*, 34:209–217, 1993.

[4] J. Labelle and Y. Yeh. Generalized Dyck paths. *Discrete mathematics*, 82:1–6, 1990.

[5] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri. On some alternative characterizations of Riordan arrays. *Canadian Journal of Mathematics*, 49(2):301–320, 1997.

[6] D. Merlini, R. Sprugnoli, and M. C. Verri. Algebraic and combinatorial properties of simple, coloured walks. In *Proceedings of CAAP'94*, volume 787 of *Lecture Notes in Computer Science*, pages 218–233, 1994.

[7] D. Merlini, R. Sprugnoli, and M. C. Verri. The area determined by underdiagonal lattice paths. In *Proceedings of CAAP'96*, volume 787 of *Lecture Notes in Computer Science*, pages 59–71, 1996.

[8] R. Sprugnoli. Riordan arrays and combinatorial sums. *Discrete Mathematics*, 132:267–290, 1994.

[9] R. A. Sulanke. Catalan path statistics having the Narayana distribution. *To appear in Discrete Mathematics.*

[10] R. A. Sulanke. Constraint sensitive Catalan path statistics having the Narayana distribution. Unpublished.