Journal of Integer Sequences, Vol. 15 (2012),
Article 12.1.1

Counting Dyck Paths According to the Maximum Distance Between Peaks and Valleys

Toufik Mansour
Department of Mathematics
University of Haifa
31905 Haifa
Israel
toufik@math.haifa.ac.il

Mark Shattuck
Department of Mathematics
University of Tennessee
Knoxville, TN 37996
USA
shattuck@math.utk.edu

Abstract

A *Dyck path* of length $2n$ is a lattice path from $(0,0)$ to $(2n,0)$ consisting of up-steps $u=(1,1)$ and down-steps $d=(1,-1)$ which never passes below the x-axis. Let $\mathcal{D}_n$ denote the set of Dyck paths of length $2n$. A *peak* is an occurrence of ud (an upstep immediately followed by a downstep) within a Dyck path, while a *valley* is an occurrence of du. Here, we compute explicit formulas for the generating functions which count the members of $\mathcal{D}_n$ according to the maximum number of steps between any two peaks, any two valleys, or a peak and a valley. In addition, we provide closed expressions for the total value of the corresponding statistics taken over all of the members of $\mathcal{D}_n$. Equivalent statistics on the set of 231-avoiding permutations of length n are also described.

1 Introduction

A *Dyck path* of length $2n$ is a lattice path from $(0,0)$ to $(2n,0)$ consisting of up-steps $u=(1,1)$ and down-steps $d=(1,-1)$ which never passes below the x-axis. That is, a Dyck path is a word of length $2n$ over the alphabet $\{u,d\}$ in which there are equal numbers of occurrences of the letters u and d, with at least as many occurrences of the letter u in any initial segment of the word as the letter d. See Figure 1 below. Let $\mathcal{D}_n$ denote the set of Dyck paths of length $2n$. Dyck paths are well-known combinatorial objects that have been widely studied in the literature. Stanley [10] presents many structures equivalent to Dyck paths of length $2n$, all of which are counted by the *Catalan sequence* $c_n=\frac{1}{n+1}\binom{2n}{n}$ (see also A000108 in [9]). Various statistics have been studied on the set of Dyck paths, such as *area* [1, 4, 6, 14], *pyramid weight* [2], and the number of *udu*'s [11]. See also [3, 5, 7] for other related statistics.

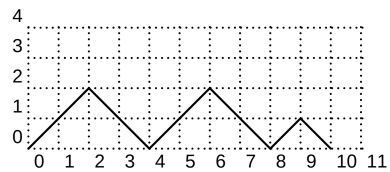

Figure 1: The Dyck path *uudduuddud*.

Let w be any word over the alphabet $\{u,d\}$. We say that the Dyck path $P\in\mathcal{D}_n$ *contains* the word w if it can be written as $P=P'wP''$. In this context, w is said to be the *left-most occurrence* (resp., *right-most occurrence*) if P' (resp., P'') does not contain any occurrences of w. For instance, if P=u**ud**uu**ud**ddu**ud**dd $\in\mathcal{D}_7$, then P contains three occurrences of ud indicated in bold. The left-most and right-most occurrences of ud correspond, respectively, to the second and third and to the eleventh and twelfth letters of P.

Let w,w' be any two words over the alphabet $\{u,d\}$. We say that the Dyck path $P\in\mathcal{D}_n$ contains the (ordered) pair (w,w') if it can be decomposed as $P=P'wP''w'P'''$. Moreover, we say that P contains the pair (w,w') with distance r if, in the corresponding decomposition $P'wP''w'P'''$, the occurrence of w is left-most, the occurrence of w' is right-most, and the number of letters in P'' is exactly r. In this context, we define $d_P(w,w')=r$ whenever such a decomposition $P'wP''w'P'''$ exists and let $d_P(w,w')=0$ otherwise. For instance, if $P=uuduuuddduuddd\in\mathcal{D}_7$, then $d_P(ud,du)=5$ (the length of P'', which is $uuudd$ in this case). We denote the generating function for the number of Dyck paths of length $2n$ according to the statistic $d_P(w,w')$ by $F_{w,w'}(x,q)$, that is,

$$F_{w,w'}(x,q)=\sum_{n\geq 0}x^n\sum_{P\in\mathcal{D}_n}q^{d_P(w,w')}.$$

In this note, we compute $F_{w,w'}(x,q)$ in the case when $w,w'\in\{ud,du\}$. This yields the generating function for the statistics on $\mathcal{D}_n$ recording the maximum distance between any two peaks (ud's), any two valleys (du's), or a peak and a valley. We also provide closed expressions for the total value of these statistics taken over all of the members of $\mathcal{D}_n$. Equivalent statistics may be described on the set of 231-avoiding permutations of length n. In addition, a Catalan number identity, which seems to be new, results from our analysis.

2 The pair (ud,ud)

Recall that the generating function for the Catalan sequence $\{c_n\}_{n\geq 0}$ is given by

$$\sum_{n\geq 0} c_n x^n = \frac{1-\sqrt{1-4x}}{2x},$$

which we will denote simply by $c(x)$.

In this section, we will find an explicit formula for the generating function $F_{ud,ud}(x,q)$. Note that any nonempty Dyck path $P \in \mathcal{D}_n$ may be decomposed as either $P = uP^{(1)}d$ or

$$\underbrace{uu\cdots u}_{s\text{ times}} dP^{(1)}dP^{(2)}\cdots dP^{(s-1)}dP^{(s)}uQ^{(r-1)}\cdots uQ^{(2)}uQ^{(1)}u\underbrace{dd\cdots d}_{r\text{ times}},$$

where $P^{(j)}$ and $Q^{(i)}$ are themselves possibly empty Dyck paths (suitably translated) and $r,s>0$. See Figure 2 below. Rewriting this in terms of generating functions, we have

$$F_{ud,ud}(x,q) = 1 + xF_{ud,ud}(x,q) + c(xq^2)\sum_{s\geq 1} x^s q^{s-1} c^{s-1}(xq^2) \sum_{r\geq 1} x^r q^{r-1} c^{r-1}(xq^2),$$

where the $c(xq^2)$ factors account for the number of steps occurring in intermediate Dyck paths $P^{(i)}$ and $Q^{(j)}$ in the above decomposition.

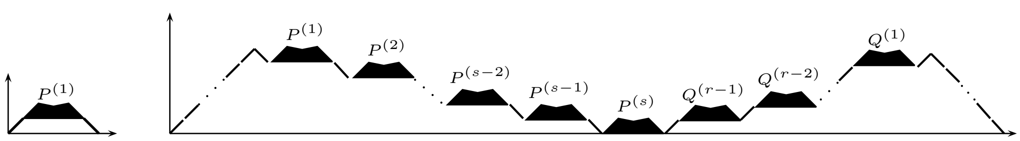

Figure 2: Decomposition of a Dyck path.

Hence, we can state the following result.

Proposition 1. *The generating function $F_{ud,ud}(x,q)$ is given by*

$$F_{ud,ud}(x,q) = \frac{1}{1-x}\left(1 + \frac{x^2 c(xq^2)}{(1-xqc(xq^2))^2}\right).$$

One can show that this expression for $F_{ud,ud}(x,q)$ reduces to $c(x)$ when $q=1$ using the fact that $yc^2(y) = c(y)-1$. Next let us find an explicit formula for the number of Dyck paths P of length $2n$ such that $d_P(ud,ud) = m$. To do so, we expand the generating function $F_{ud,ud}(x,q)$ as follows:

$$\begin{aligned}
F_{ud,ud}(x,q) &= \sum_{n\geq 0} x^n + \frac{x}{1-x}\sum_{j\geq 1} jx^j q^{j-1} c^j(xq^2) \\
&= \sum_{n\geq 0} x^n + \frac{x}{1-x}\sum_{j\geq 1}\sum_{i\geq 0} \frac{j^2(2i+j-1)!}{i!(i+j)!} x^{i+j} q^{2i+j-1} \\
&= \sum_{n\geq 0} x^n + \sum_{j\geq 0}\sum_{i\geq 0}\sum_{k\geq 0} \frac{(j+1)^2(2i+j)!}{i!(i+j+1)!} x^{i+j+k+2} q^{2i+j},
\end{aligned}$$

where in the second equality, we have used the identity

$$c(y)^j = \sum_{i\geq 0} \frac{j(2i+j-1)!}{i!(i+j)!} y^i, \qquad j \geq 1,$$

which occurs as Equation 2.5.16 in [13].

Collecting the coefficient of $x^n q^m$ in this expansion of $F_{ud,ud}(x,q)$ yields the following result.

Theorem 2. *The number of Dyck paths $P \in \mathcal{D}_n$ with $d_P(ud,ud) = m$, $0 \leq m \leq 2n-4$, is given by*

$$\delta_{m,0} + \sum_{i=\max\{m+2-n,0\}}^{\lfloor m/2 \rfloor} \frac{(m+1-2i)^2}{m+1} \binom{m+1}{i}.$$

When $n \geq 3$ and $m = 2n-4$ in the above theorem, we see that the number of Dyck paths P in $\mathcal{D}_n$ with $d_P(ud,ud) = 2n-4$ is given by $\frac{1}{n-1}\binom{2n-4}{n-2}$, which equals the number of Dyck paths of length $2n-4$. This agrees with the fact that any Dyck path $P \in \mathcal{D}_n$ with $d_P(ud,ud) = 2n-4$ must be of the form $P = udP'ud$, where P' is any member of $\mathcal{D}_{n-2}$.

As a consequence of the above result, we obtain the following Catalan number identity, which seems to be new.

Corollary 3. *For all $n \geq 1$,*

$$\frac{1}{n+1}\binom{2n}{n} = 1 + \sum_{m=0}^{2n-4} \frac{\sum_{i=\max\{m+2-n,0\}}^{\lfloor m/2 \rfloor} (m+1-2i)^2 \binom{m+1}{i}}{m+1}.$$

Differentiating the formula for $F_{ud,ud}(x,q)$ in Proposition 1 above with respect to q, and substituting $q = 1$, implies that the generating function for $\sum_{P\in\mathcal{D}_n} d_P(ud,ud)$ is given by $\frac{x^2+2x-1}{(1-x)x^2} - \frac{2x^2+3x-1}{x^2\sqrt{1-4x}}$. Using the fact $\frac{1}{\sqrt{1-4x}} = \sum_{n\geq 0}\binom{2n}{n}x^n$ yields the following result.

Corollary 4. *For all $n \geq 1$,*

$$\sum_{P\in\mathcal{D}_n} d_P(ud,ud) = \frac{2(n^2-2n-2)}{(n+2)(n+1)}\binom{2n}{n} + 2.$$

We conclude this section by considering an equivalent statistic on permutations. If $\pi = \pi_1\pi_2\cdots\pi_n \in \mathcal{S}_n$, then π is said to be *stack-sortable* if it can be put in the natural order $1, 2, \ldots, n$ with the aid of a "stack" in which each element, starting with the first, is moved to a stack and then to the output. The stack-sortable permutations are precisely those which have no occurrences of 231; i.e., there exist no indices $i < j < k$ with $\pi_k < \pi_i < \pi_j$. The set of such permutations is denoted by $\mathcal{S}_n(231)$ and has cardinality c_n (see, e.g., [8]). As pointed out by West [12], one can encode a sorting by writing an up-step whenever an element is put on the stack and writing a down-step whenever it is taken off. For instance, the permutations 123 and 132 in $\mathcal{S}_3(231)$ would be sorted by the sequence of moves encoded by *ududud* and *uduudd*, respectively. It may be shown that this encoding yields a $1-1$ correspondence between $\mathcal{D}_n$ and $\mathcal{S}_n(231)$, which we will denote by α.

The statistic $d_P(ud,ud)$ on $\mathcal{D}_n$ then translates into a statistic on $\mathcal{S}_n(231)$ as follows.

Definition 5. Given $\pi = \pi_1\pi_2\dots\pi_n \in \mathcal{S}_n(231)$, suppose $\pi_\ell = 1$ for some index ℓ and let $\pi_n = M$. Define the statistic r on $\mathcal{S}_n(231)$ by

$$r(\pi) = \begin{cases} n - \ell + M - 3, & if \quad M > 1; \\ 0, & if \quad M = 1. \end{cases}$$

For example, if

$$\lambda = uududduudd \in \mathcal{D}_5,$$

then $\alpha(\lambda) = 31254 \in \mathcal{S}_5(231)$, with $d_\lambda(ud, ud) = 4 = r(\alpha(\lambda))$. Indeed, one can show the following result.

Proposition 6. *If $n \geq 1$, then*

$$d_\lambda(ud, ud) = r(\alpha(\lambda))$$

for all $\lambda \in \mathcal{D}_n$.

3 The pair (du, du)

Let us write an equation for the generating function $F_{du,du}(x, q)$. Note that any nonempty Dyck path $P \in \mathcal{D}_n$ can be decomposed as either

- $P = uP'd$,
- $P = \underbrace{uu\cdots u}_{k}\underbrace{dd\cdots d}_{k}P'\underbrace{uu\cdots u}_{k'}\underbrace{dd\cdots d}_{k'}$,
- $uQ'dP'\underbrace{uu\cdots u}_{k}\underbrace{dd\cdots d}_{k}$,
- $\underbrace{uu\cdots u}_{k}\underbrace{dd\cdots d}_{k}P'uQ'd$,
- $uQ'dP'uQ''d$,

where k and k' are positive, P' is any Dyck path, and Q', Q'' are any Dyck paths having at least one valley.

Expressing these cases in terms of generating functions, we obtain

$$\begin{aligned} F_{du,du}(x, q) &= 1 + xF_{du,du}(x, q) + \frac{x^2}{q^2(1-x)^2}(c(xq^2) - 1 + q^2) \\ &+ \frac{2x}{1-x}G(x, q)c(xq^2) + q^2G^2(x, q)c(xq^2), \end{aligned}$$

where $G(x, q)$ is the generating function for the number of Dyck paths uPd of length $2n$, in which P itself is a Dyck path having at least one valley, according to the number of steps between the left-most valley and the last step of the path. That is,

$$G(x, q) = \sum_{n\geq 0} x^n \sum q^{d_{Qud}(du,du)},$$

where the internal sum is over all Dyck paths Q of length $2n$ of the form $Q = uPd$ in which P has at least one valley.

In order to find an explicit formula for the generating function $G(x,q)$, we decompose each Dyck path of the form $uPdud$, where P is a Dyck path having at least one valley, as follows:

$$uPdud = u\underbrace{uu\cdots u}_{k}\underbrace{dd\cdots d}_{\ell}uP^{(1)}dP^{(2)}d\cdots P^{(k-\ell+2)}dud,$$

where $\ell = 1,2,\ldots,k$ and each $P^{(j)}$ is a Dyck path. Translating this in terms of generating functions yields

$$\begin{aligned}
G(x,q) &= \sum_{k\geq 1} x^{k+2}\sum_{\ell=1}^{k} q^{k-\ell+1}c^{k-\ell+2}(xq^2)\\
&= \sum_{k\geq 1} x^{k+2}\frac{qc^2(xq^2)-q^{k+1}c^{k+2}(xq^2)}{1-qc(xq^2)}\\
&= \frac{x^3qc^2(xq^2)}{(1-x)(1-qc(xq^2))} - \frac{x^3q^2c^3(xq^2)}{(1-qc(xq^2))(1-xqc(xq^2))}\\
&= \frac{x^3qc^2(xq^2)}{(1-x)(1-xqc(xq^2))}. \qquad (1)
\end{aligned}$$

Substituting this expression for $G(x,q)$ into the one above for $F_{du,du}(x,q)$ yields the following result.

Theorem 7. *The generating function $F_{du,du}(x,q)$ is given by*

$$F_{du,du}(x,q) = \frac{1}{1-x}\left(1+\frac{x^2(q^2-1)}{q^2(1-x)^2}+\left(\frac{x}{q(1-x)}+qG(x,q)\right)^2 c(xq^2)\right),$$

where $G(x,q) = \frac{x^3qc^2(xq^2)}{(1-x)(1-xqc(xq^2))}$.

Differentiating this formula for $F_{du,du}(x,q)$ with respect to q, and substituting $q=1$, yields the following formulas for the total value of $d_P(du,du)$ taken over all of the members of $\mathcal{D}_n$.

Corollary 8. *The generating function $\sum_{n\geq 0}\left(\sum_{P\in\mathcal{D}_n} d_P(du,du)\right)x^n$ is given by*

$$\frac{1}{x^2(1-x)^3}\left(10x^4-17x^3+7x^2+2x-1+\frac{6x^5-32x^4+31x^3-5x^2-4x+1}{\sqrt{1-4x}}\right).$$

Moreover,

$$\sum_{P\in\mathcal{D}_n} d_P(du,du) = \sum_{k=3}^{n}\frac{2(27k^4-151k^3+310k^2-212k-280)}{k(k-1)(k+1)(k+2)}\binom{2k-6}{k-4}\binom{n-k+2}{2}.$$

4 The pairs (ud,du) and (du,ud)

First observe that the $d_P(ud,du)$ and $d_P(du,ud)$ statistics are identically distributed on $\mathcal{D}_n$, upon writing members of D_n in reverse order and replacing u with d and d with u since this transforms the left-most peak (resp., right-most valley) into the right-most peak (resp., left-most valley). Thus, we only find the generating function for the former. To do so, note that any nonempty Dyck path $P \in \mathcal{D}_n$ can be decomposed as either

- $P = uP'd$,
- $P = \underbrace{uu\cdots u}_{k} dP^{(1)}dP^{(2)}\cdots dP^{(k)} \underbrace{uu\cdots u}_{\ell}\underbrace{dd\cdots d}_{\ell}$,
- $P = \underbrace{uu\cdots u}_{k} dP^{(1)}dP^{(2)}\cdots dP^{(k)}uQd$,

where k and ℓ are positive, $P^{(j)}$, $j = 1,2,\dots,k$, is any Dyck path, and Q is any Dyck path having at least one valley.

Expressing these cases in terms of generating functions (note that we need to distinguish when $k = 1$ and $k > 1$ in the second decomposition above), we obtain

$$\begin{aligned}F_{ud,du}(x,q) &= 1 + xF_{ud,du}(x,q) + \frac{x^2}{1-x}\left(1+\frac{c(xq^2)-1}{q}\right)\\ &\quad + \frac{x^3c^2(xq^2)}{(1-x)(1-xqc(xq^2))} + \frac{xqc(xq^2)}{1-xqc(xq^2)}G(x,q),\end{aligned}$$

where $G(x,q)$ is the generating function for the number of Dyck paths uPd of length $2n$, in which P itself is a Dyck path having at least one valley, according to the number of steps between the left-most valley and the last step of the path (or, equivalently, according to the number of steps between the right-most valley and the first step of the path). Plugging the expression for $G(x,q)$ given in (1) into the above equation, and using $yc^2(y) = c(y) - 1$, we obtain the following result.

Theorem 9. *The generating function $F_{ud,du}(x,q)$ is given by*

$$\frac{1}{1-x}\left(1+\frac{x^2(q-1)}{q(1-x)}+\frac{x^2c(xq^2)(1-xq)}{q(1-x)(1-xqc(xq^2))^2}\right).$$

Note that this expression for $F_{ud,du}(x,q)$ reduces to $c(x)$ when $q = 1$. Differentiating the expression with respect to q, and substituting $q = 1$, yields the following formulas for the total value of $d_P(ud,du)$ taken over all of the members of $\mathcal{D}_n$.

Corollary 10. *The generating function $\sum_{n\ge0}\left(\sum_{P\in\mathcal{D}_n} d_P(ud,du)\right)x^n$ is given by*

$$\frac{2x^4-2x^3+x^2+5x-2}{2x^2(1-x)^2}+\frac{4x^3-2x^2-7x+2}{2x^2(1-x)\sqrt{1-4x}}.$$

Moreover,

$$\sum_{P\in\mathcal{D}_n} d_P(ud,du) = \sum_{k=3}^{n}\frac{3(3k^4-15k^3+23k^2-11k-2)(2k-2)!(n+1-k)}{(2k-3)(k-1)!(k+2)!}.$$

One may also express the $d_P(ud, du)$ statistic on $\mathcal{D}_n$ in terms of a statistic on $\mathcal{S}_n(231)$ as in the second section above.

Definition 11. Given $\pi = \pi_1\pi_2\cdots\pi_n \in \mathcal{S}_n(231)$, suppose $\pi_\ell = 1$ for some index ℓ and let $\pi_n = M$. Let m denote the index (if it exists) of the right-most entry between π_ℓ and π_n in π and strictly between 1 and M in value. Define the statistic s on $\mathcal{S}_n(231)$ by

$$s(\pi) = \begin{cases} m - \ell, & \textit{if } m \textit{ exists;} \\ 0, & \textit{otherwise.} \end{cases}$$

For example, if $\pi = 216354 \in \mathcal{S}_6(231)$, then $s(\pi) = 4 - 2 = 2$.

Definition 12. Given $\pi = \pi_1\pi_2\cdots\pi_n \in \mathcal{S}_n(231)$, suppose $\pi_\ell = 1$ for some index ℓ and let $\pi_n = M$. Define the statistic t on $\mathcal{S}_n(231)$ by

$$t(\pi) = \begin{cases} s(\pi) + M - 3, & if \quad M > 2; \\ 0, & if \quad M = 1, 2. \end{cases}$$

For example, if

$$\lambda = uudduuduuddd \in \mathcal{D}_6,$$

then $\alpha(\lambda) = 216354 \in \mathcal{S}_6(231)$, with $d_\lambda(ud, du) = 3 = t(\alpha(\lambda))$. Indeed, one can show the following result.

Proposition 13. *If* $n \geq 1$, *then*

$$d_\lambda(ud, du) = t(\alpha(\lambda))$$

for all $\lambda \in \mathcal{D}_n$.

References

[1] M. Delest and X. G. Viennot, Algebraic languages and polyominoe enumeration, *Theoret. Comput. Sci.* **34** (1984), 169–206.

[2] A. Denise and R. Simion, Two combinatorial statistics on Dyck paths, *Discrete Math.* **187:1-3** (1998), 71–96.

[3] E. Deutsch, Dyck path enumeration, *Discrete Math.* **204** (1999), 167–202.

[4] J. M. Fédou, Grammairs et q-énumération de polyominos, Ph.D. Thesis, Université de Bordeaux I, 1989.

[5] T. Mansour, Statistics on Dyck paths, *J. Integer Seq.* **9:1** (2006), Article 06.1.5.

[6] D. Merlini, R. Sprugnoli, and M. C. Verri, The area determined by underdiagonal lattice paths, In Proc. of CAAP'96, *Lecture Notes in Comput. Sci.* **1059** (1996), 59–71.

[7] D. Merlini, R. Sprugnoli, and M. C. Verri, Some statistics on Dyck paths, *J. Statist. Plann. Inference* **101** (2002), 211–227.

[8] R. Simion and F. W. Schmidt, Restricted permutations, *European J. Combin.* **6** (1985), 383–406.

[9] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, http://oeis.org.

[10] R. P. Stanley, *Enumerative Combinatorics, Vol. 1*, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1986; second printing, Cambridge University Press, Cambridge, 1996.

[11] Y. Sun, The statistic "number of *udu*'s" in Dyck paths, *Discrete Math.* **287:1-3** (2004), 177-186.

[12] J. West, Sorting twice through a stack, *Theoret. Comput. Sci.* **117** (1993), 303–313.

[13] H. Wilf, *generatingfunctionology*, Academic Press, New York, 1990.

[14] W. Woan, Area of Catalan paths, *Discrete Math.* **226** (2001), 439–444.

2010 *Mathematics Subject Classification*: Primary 05A15; Secondary 05A05.
Keywords: Dyck paths, generating functions, statistics, peaks, valleys.

(Concerned with sequence A000108.)

Received August 8 2011; revised version received November 6 2011. Published in *Journal of Integer Sequences*, December 15 2011.

Return to Journal of Integer Sequences home page.