Noname manuscript No.
(will be inserted by the editor)

Number of Dyck and ballot paths with a given number of “touchdowns” – a combinatorial derivation

Karl Schaettle · Nikola P. Petrov

Received: date / Accepted: date

Abstract Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references. *Keywords*: Please provide 4 to 6 keywords which can be used for indexing purposes.

Keywords keywords

Define a ballot path as a random walk from $(0,0)$ to (n,h) given by steps $(1,1)$ and $(1,-1)$ such that the second coordinate is always nonnegative. Moreover define a Dyck path to be a ballot path such that $h=0$. We desire to find a non-recursive formula for the number of ballot paths ending at (n,h) which contain d "touchdowns," or points $(x,0)$ such that $x \neq 0$. We define this function to be $N_{(n,h)}(d)$.

According to the well-known Ballot Theorem [1], for arbitrary n and $h>0$, the number of ballot paths such that all points of the path have positive second coordinate except the origin is $\frac{h}{n}\binom{n}{\frac{n+h}{2}}$. The main ingredient in the proof of this theorem is the *Reflection Principle*, stating that, if n_1, n_2, h_1, and h_2 are integers satisfying $0 \leq n_1 < n_2$, $h_1>0$, $h_2>0$, then the number of paths from (n_1,h_1) to (n_2,h_2) which touch or cross the t-axis is equal to the number of all paths from $(n_1,-h_1)$ to (n_2,h_2). Then the Ballot Theorem is obtained by This formula is obtained by applying the principle of reflection about the line

This research has been partially supported by NSF grant DMS 0807658.

K. Schaettle
Department of Chemical Engineering, University of Oklahoma, Norman, OK 73019, USA
E-mail: kschaettle@ou.edu

N. P. Petrov
Department of Mathematics, University of Oklahoma, Norman, OK 73019, USA; Tel.: +1-405-325-4316, Fax: +1-405-325-7484. E-mail: npetrov@ou.edu

$y = 0$ to find the number of paths which *do* intersect the x-axis nontrivially and subtracting them from the total number of random walks from $(0,0)$ to (n,h) given by steps $(1,1)$ and $(1,-1)$.

Similarly we reflect about $y = -1$ to obtain the total number of ballot paths ending at (n,h), which is

$$\frac{2+2h}{n+h+2}\binom{n}{\frac{n+h}{2}} = \frac{h+1}{n+1}\binom{n+1}{\frac{n+h}{2}+1} \tag{1}$$

We establish a general expression for $N_{(2n,0)}(d)$ by recursion, first finding $N_{(2n,0)}(1)$, which is of course the smallest value d may assume when $h = 0$ as the last point of the Dyck path is necessarily a touchdown. We note that the $N_{(2n,0)}(1)$ is equal to the number of Dyck paths of length $2n-2$; by expressing a Dyck path of length $2n$ and zero touchdowns as an initial upward step followed by a Dyck path of length $2n-2$ and concluded by a downward step, we see a clear bijection between the two quantities. Thus by using the latter part of equation (1) with the proper parameters, we have that

$$N_{(2n,0)}(1) = \frac{1}{2n-1}\binom{2n-1}{n} \tag{2}$$

A recursive relationship for $N_{(2n,0)}(d)$ can thus be obtained by the simple observation that a Dyck path of length $2n$ and d touchdowns may be described as the concatenation of a Dyck path of length $2i$ and 1 touchdown and a Dyck path of length $(2n-2i)$ and $(d-1)$ touchdowns. Summing over the number of such concatenations gives the recursion:

$$\begin{aligned} N_{(2n,0)}(d) &= \sum_{i=1}^{n-(d-1)} N_{(2i,0)}(1) N_{(2n-2i,0)}(t-1) \\ &= \sum_{i=1}^{n-(d-1)} \frac{1}{2i-1}\binom{2i-1}{i} N_{(2n-2i,0)}(d-1) \end{aligned} \tag{3}$$

Where the upper limit of summation is obtained by observing that $(d-1)$ touchdowns require a Dyck path of length at least $2(d-1)$. Equation (3) is helpful in verifying the general closed form of $N_{(2n,0)}(d)$, and gives us the following surprising result.

Proposition 1 *For $n > 0$, $N_{(2n,0)}(1) = N_{(2n,0)}(2) = \frac{1}{2n-1}\binom{2n-1}{n}$*

Proof

$$N_{(2n,0)}(2) = \sum_{i=1}^{n-1} \frac{1}{2i-1}\binom{2i-1}{i} N_{(2n-2i,0)}(1) \tag{4}$$

$$= \sum_{i=1}^{n-1} \frac{1}{2i-1}\binom{2i-1}{i} \frac{1}{2n-2i-1}\binom{2n-2i-1}{n-i} \tag{5}$$

It turns out that Equation (5) has a very simple closed-form expression if we sum from $i = 0$ to $i = n$, which [3]

$$\sum_{k=0}^{n} \frac{p+qk}{(a+ck)(b-ck)} \binom{a+ck}{k} \binom{b-ck}{n-k} = \frac{p(a+b-cn)+aqn}{a(a+b)(b-cn)} \binom{a+b}{n} \quad (6)$$

gives as $\frac{1}{1-n}\binom{2n-2}{n}$, and thus

$$\begin{aligned} N_{(2n,0)}(2) &= \frac{1}{1-n}\binom{2n-2}{n} - \frac{-2}{2n-1}\binom{2n-1}{n} \\ &= \frac{1}{1-n}\binom{2n-2}{n} + \frac{2}{n-1}\binom{2n-2}{n} \\ &= \frac{1}{n-1}\binom{2n-2}{n} \\ &= \frac{1}{2n-1}\frac{1}{n-1}\frac{(2n-1)!}{n!(n-2)!} \\ &= \frac{1}{2n-1}\binom{2n-1}{n} \end{aligned}$$

Which completes our proof.

Observe now that $N_{(2n,0)}(1) = \frac{1}{2n-1}\binom{2n-1}{n}$ and $N_{(2n,0)}(2) = \frac{1}{n-1}\binom{2n-2}{n} = \frac{2}{2n-2}\binom{2n-2}{n}$. This inspires us towards conjecture.

Theorem 1 *For $n > 0$ and $d \le n$, $N_{(2n,0)}(d) = \frac{d}{2n-d}\binom{2n-d}{n}$.*

The proof is purely algebraic once one applies the following lemma, which in turn comes fairly naturally from Equation (3) and Proposition 1.

Lemma 1 $N_{(2n,0)}(d) = N_{(2n,0)}(d-1) - N_{(2n-2,0)}(d-2)$

Proof

$$N_{(2n,0)}(d) = \sum_{i=2}^{n-(d-2)} N_{(2i,0)}(2) N_{(2n-2i,0)}(d-2)$$

which is obtained by considering a Dyck path of length $2n$ and d touchdowns to be a concatenation of a Dyck path of length $2i$, 2 touchdowns and a Dyck path of length $2n-2i$, $d-2$ touchdowns. By Proposition 1, we then have

$$\begin{aligned} N_{(2n,0)}(d) &= \sum_{i=2}^{n-(d-2)} N_{(2i,0)}(1) N_{(2n-2i,0)}(d-2) \\ &= \sum_{i=1}^{n-(d-2)} N_{(2i,0)}(1) N_{(2n-2i,0)}(d-2) - N_{(2,0)}(1) N_{(2n-2,0)}(d-2) \\ &= N_{(2n,0)}(d-1) - N_{(2n-2,0)}(d-2) \end{aligned} \quad (7)$$

noting that the final step is simply an application of Equation (3) and the observation that $N_{(2,0)}(1) = 1$, which completes the proof of the lemma.

To prove Theorem 1, we begin with the fact that it holds for $d = 1$ and $d = 2$, and proceed by strong induction on d. Suppose the Theorem to be true for all positive integers strictly less than some d. We wish to show, then, that $N_{(2n,0)}(d) = \frac{d}{2n-d}\binom{2n-d}{n}$. By Lemma 1 we have the recursion:

$$\begin{aligned}
N_{(2n,0)}(d) &= N_{(2n,0)}(d-1) - N_{(2n-2,0)}(d-2) \\
&= \binom{2n-d+1}{n}\frac{d-1}{2n-d+1} - \binom{2n-d}{n-1}\frac{d-2}{2n-d} \\
&= \frac{(2n-d)!(2n-d+1)}{n!(n-d)!(n-d+1)}\frac{d-1}{2n-d+1} - \frac{(2n-d)!}{n!\frac{1}{n}(n-d)!(n-d+1)}\frac{d-2}{2n-d} \\
&= \binom{2n-d}{n}\left(\frac{d-1}{n-d+1} - \frac{n(d-2)}{(2n-d)(n-d+1)}\right) \\
&= \binom{2n-d}{n}\frac{2nd-d^2-2n+d-nd+2n}{(n-d+1)(2n-d)} \\
&= \binom{2n-d}{n}\frac{d(n-d+1)}{(n-d+1)(2n-d)} \\
&= \binom{2n-d}{n}\frac{d}{2n-d}
\end{aligned}$$

Finally, we use the general formulas of $N_{(2n,0)}(d)$ and $N_{(n,h)}(0)$ to obtain a non-recursive formula for $N_{(n,h)}(d)$, where $h \neq 0$, $d \neq 0$. We observe that a ballot path of length n, final height h with d touchdowns can be described as the concatenation of a Dyck path of length $2i$ and d touchdowns with a ballot path of length $n - 2i$, ending height h and 0 touchdowns. In other words, for $h \neq 0$ and $d \neq 0$,

$$N_{(n,h)}(d) = \sum_{i=d}^{\frac{n-h}{2}} N_{(2i,0)}(d)N_{(n-2i,h)}(0) \tag{8}$$

$$= \sum_{i=d}^{\frac{n-h}{2}} \frac{d}{2i-d}\binom{2i-d}{i}\frac{h}{n-2i}\binom{n-2i}{\frac{n-h}{2}-i} \tag{9}$$

where the lower limit of summation comes from the fact that a Dyck path with d touchdowns must be of length at least $2d$, and similarly the upper limit ensures that the ballot path of ending height h with 0 touchdowns is at least h steps long.

Thus by using Equation ???, we can now write a closed-form expression for $Z_n(h)$, though it adopts a different form in the case $h = 0$. For $h \neq 0$, we have

$$
\begin{aligned}
Z_n(h) &= \sum_{j=0}^{\frac{n-h}{2}} N_{(n,h)}(j)\kappa^j \\
&= \frac{h}{n}\binom{n}{\frac{n+h}{2}} + \sum_{j=1}^{\frac{n-h}{2}} N_{(n,h)}(j)\kappa^j \\
&= \frac{h}{n}\binom{n}{\frac{n+h}{2}} + \sum_{j=1}^{\frac{n-h}{2}} \kappa^j \sum_{i=j}^{\frac{n-h}{2}} \frac{j}{2i-j}\binom{2i-j}{i}\frac{h}{n-2i}\binom{n-2i}{\frac{n-h}{2}-i}
\end{aligned}
$$

And of course we also have

$$
\begin{aligned}
Z_{2n}(0) &= \sum_{j=1}^{n} N_{(2n,0)}(j)\kappa^j \\
&= \sum_{j=1}^{n} \binom{2n-j}{n}\frac{j}{2n-j}\kappa^j
\end{aligned} \tag{10}
$$

Equation 12 seems a bit gruesome, but it is far easier to compute for large n, say $n > 1000$, than the equation given by Brak, Owczarek, and Rechnitzer [2]. It is also much easier to approximate, as it involves nothing more complicated than binomial coefficients, for which there are a plethora of approximation methods.

References

1. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, 3rd edn. (Wiley, New York, 1968), Sec. III.1
2. R. Brak, A. L. Owczarek, A. Rechnitzer, J. Math. Chem. 45, 49–57 (2009)
3. А. П. Прудников, Ю. А. Брычков, О. И. Маричев, Интегралы и ряды. Том 1: Элементарные функции. (Russian) [Integrals and series. Vol. 1: Elementary functions.] (Nauka, Moscow, 1981), Section 4.2.6, Equation (21), p. 624
4. O. Sofrygin, Counting the number of Dyck paths according to different types of parameters, http://people.brandeis.edu/ gessel/47a/sofrygin.pdf
5. E. Deutsch, A bijection on Dyck paths and its consequences, Discrete Mathematics 179, 253–256 (1998)
6. E. Deutsch, Dyck path enumeration, Discrete mathematics 204, 167–202 (1999) [1]
7. N. Dershowitz, S. Zaks, Enumerations of ordered trees, Discrete Mathematics 31, 9–28 (1980) [2]
8. D. Knuth, The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 3rd edn. (Addison-Wesley, Reading, MA, 1997), Sec. 1.2.6, Example 4 (pp. 62, 63)

[1] His equation (6.22) is our formula for the number of Dyck paths; he says that this result can be found in Theorem 4 and Corollary 4.1 of [7].

[2] Theorem 1 is Narayana's result; an alternative proof of it in terms of lattice paths is given in the Appendix.

9. R. L. Graham, D. E. Knuth, O. Patashnik, Concrete Mathematics (Addison-Wesley, Reading, MA, 1989), Eqn. (5.63)
10. T. V. Narayana, Lattice path combinatorics, with statistical applications (Univ of Toronto Press, 1979)