

THE LOCATION OF THE FIRST MAXIMUM IN THE FIRST SOJOURN OF A DYCK PATH

HELMUT PRODINGER

ABSTRACT. For Dyck paths (nonnegative symmetric) random walks, the location of the first maximum within the first sojourn is studied. Generating functions and explicit resp. asymptotic expressions for the average are derived. Related parameters are also discussed.

1. INTRODUCTION

In the recent paper [8], the distribution of the time at which a Brownian motion is maximal before its first-passage time was studied.

We undertake a related study, by investigating the *discrete* cousins, namely *Dyck paths* (nonnegative lattice paths, returning to the x -axis). These celebrated combinatorial objects, enumerated by *Catalan numbers* $\frac{1}{N+1} \binom{2N}{N}$, when their length is $2N$, are known by many names, including also Dyck words, nonnegative symmetric random walks, etc. They are also important in Computer Science, since they code planar trees (and binary trees), see e.g. [1].

We find an explicit form of the generating function $F(z, u)$, where the coefficient of $z^n u^k$ is the number of Dyck paths of length n such that the first maximum within the first sojourn occurs at time (=step) k . From it, we compute the *expectation* explicitly and asymptotically. It is roughly $1/4$ of the length of the walk.

We also compute the average height of the first sojourn, explicitly and asymptotically. (The average height of the whole path has been known for a long time.)

For reason of comparison, we also investigate the location of the first maximum of the whole path, which is roughly at $1/2$ of the length of the walk.

In a final section, we show how to extend these results from the first sojourn to the r -th sojourn. The parameters here are based on assuming all Dyck paths with at least r sojourns to be equally likely. Most of the old computations can be used and only have to be adapted slightly.

2. GENERATING FUNCTIONS

We study the parameter *position of the first maximum during the first sojourn of a Dyck path of length n* , which we denote by R . Since this parameter makes no sense for the empty path, we assume that $n \geq 1$.

$$+ C(z)\sqrt{1-4z^2} \sum_{h \geq 1} \frac{2hz^{2h+2}(\lambda_1^{h-1}(z) + \lambda_2^{h-1}(z))}{(\lambda_1^h(z) - \lambda_2^h(z))^2(\lambda_1^{h+1}(z) - \lambda_2^{h+1}(z))}.$$

With the substitution $z = v/(1+v^2)$, this expression becomes nicer:

$$\begin{aligned} G(z) &= -4v \sum_{h \geq 1} \frac{v^{2h+1}}{(1-v^{2h})(1-v^{2h+2})} + (1-v^2)^2 \sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h})(1-v^{2h+2})} \\ &\quad + (1-v^2) \sum_{h \geq 1} \frac{2hv^{2h+2}(1+v^{2h-2})}{(1-v^{2h})^2(1-v^{2h+2})} \\ &= -\frac{4v^4}{(1-v^2)^2} - \frac{(1+v^2)^2}{1-v^2} \sum_{h \geq 1} \frac{v^{2h}}{1-v^{2h}} + 2(1+v^2) \sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h})^2}. \end{aligned}$$

The last step was done by partial decomposition and a few rearrangements.

Now note that

$$\sum_{h \geq 1} \frac{v^{2h}}{1-v^{2h}} = \sum_{k \geq 1} d(k)v^{2k}$$

and

$$\sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h})^2} = \sum_{k \geq 1} k d(k)v^{2k}.$$

To read off coefficients, we observe that

$$[z^n]G(z) = [v^n](1-v^2)(1+v^2)^{n-1}G(z(v));$$

this can be justified either by the *Lagrange inversion formula* or *formal residue calculus*.

Now, since we have contributions only for even $n = 2N$, we can continue:

$$\begin{aligned} [z^n]G(z) &= [v^n](1-v^2)(1+v^2)^{n-1}G(z(v)) \\ &= [v^N](1+v)^{2N-1} \left(-\frac{4v^2}{1-v} - (1+v)^2 \sum_{h \geq 1} \frac{v^h}{1-v^h} + 2(1-v^2) \sum_{h \geq 1} \frac{hv^h}{(1-v^h)^2} \right) \\ &= -4 \sum_{k=0}^{N-2} \binom{2N-1}{k} - \sum_{k \geq 1} d(k) \binom{2N+1}{N-k} + 2 \sum_{k \geq 1} k d(k) \left[\binom{2N}{N-k} - \binom{2N}{N-1-k} \right]. \end{aligned}$$

The final formula follows from simplifying the first sum. \square

3. ASYMPTOTICS

In the paper [1] we find the procedure how to compute asymptotics of such sums. (See also [3].)

We use the following approximation

$$\begin{aligned} &\frac{-\binom{2N+1}{N-k} + 2k \binom{2N}{N-k} - 2k \binom{2N}{N-1-k}}{\binom{2N}{N}} \\ &\sim \exp(-x^2) \left(-2 + 4x^2 + \frac{1}{\sqrt{N}}(4x - 4x^3) + \frac{1}{N}(-9x^2 + 1 + \frac{19}{3}x^4 - \frac{2}{3}x^6) \right), \end{aligned}$$

where $x = k/\sqrt{n}$, and it holds for $|k| \leq n^{\frac{1}{2}+\epsilon}$, and for other values of k , the expression is negligible (exponentially small). Maple could produce this expansion, and more terms would be available. In [7], a procedure is described how to obtain such an approximation efficiently.

A key ingredient to the asymptotics is now the sum

$$S := \sum_{k \geq 1} d(k) e^{-k^2/n} \left(-2 + 4 \frac{k^2}{n} + \left(4 \frac{k}{N} - 4 \frac{k^3}{N^2} \right) + \frac{1}{N} (\dots) \right).$$

This can be attacked with the *Mellin transform*. Writing again $x = 1/\sqrt{N}$, it becomes a *harmonic sum* in the notation of [2];

$$S^* = \zeta^2(s) \left((s-1) \Gamma\left(\frac{s}{2}\right) + \text{further terms} \right).$$

(Note that $\sum_{k \geq 1} d(k) k^{-s} = \zeta^2(s)$.)

The asymptotics of S can be recovered by computing *residues* of $S^* x^{-s}$ (at $s = 1$ and $s = 0$):

$$S \sim \frac{\sqrt{\pi}}{x} - \frac{1}{2} + \text{further terms}.$$

The final asymptotic formula for the average R_N of the parameter R follows from this by multiplying by $(N+1)$, and taking care of the first two explicit terms, for which Stirling's formula applies. Most of this can be done by Maple. The final result is

Theorem 2. *The average value R_N of the parameter R when considering all Dyck paths of lengths $2N$ to be equally likely, is given by*

$$R_N \sim \frac{N}{2} - \frac{\sqrt{\pi N}}{6} - \frac{1}{12} \log N + \frac{10}{9} - \frac{\gamma}{4},$$

with more terms being available if necessary.

So, roughly, we expect the parameter to be about 1/4 of the length of the Dyck path.

Higher moments could also be computed, by higher derivatives of $F(z, u)$, but we did not do it, because the expressions become complicated.

4. THE HEIGHT OF THE FIRST SOJOURN

The height of Dyck paths has been studied for some 35 years: De Bruijn, Knuth and Rice [1] showed that the average height is asymptotic to $\sqrt{\pi N} - \frac{3}{2}$ (it is one less than the height of the corresponding planar tree).

However, it does not seem that the height of the first sojourn has been studied. Thus, we briefly address this here. Let $A_h(z)$ be the generating function of Dyck words of height $\leq h$, and $B_h(z)$ be the generating function of Dyck words where the first sojourn has height $\leq h$. Then, by the fundamental decomposition of Dyck words w.r.t. the first return to the origin,

$$B_h(z) = z^2 A_{h-1} C(z).$$

Now, it is known (and easy to derive via the system (1)) that

$$A_h(z) = \frac{\lambda_1^{h+1}(z) - \lambda_2^{h+1}(z)}{\lambda_1^{h+2}(z) - \lambda_2^{h+2}(z)}.$$

Consequently,

$$B_h(z) = v^2 \frac{1 - v^{2h}}{1 - v^{2h+2}},$$

again with the substitution $z = v/(1 + v^2)$. The generating function of Dyck words where the height of the first sojourn is $> h$, is then

$$C(z) - 1 - B_h(z) = v^2(1 - v^2) \frac{v^{2h}}{1 - v^{2h+2}}.$$

The generating function of the averages (apart from normalization by b_N) is obtained by summing this up over $h \geq 0$, viz.

$$(1 - v^2) \sum_{h \geq 1} \frac{v^{2h}}{1 - v^{2h}} = (1 - v^2) \sum_{k \geq 1} d(k) v^{2k}.$$

Now we must look at the coefficient of z^{2N} :

$$\begin{aligned} [z^{2N}](1 - v^2) \sum_{h \geq 1} \frac{v^{2h}}{1 - v^{2h}} &= [v^{2N}](1 - v^2)(1 + v^2)^{2N-1}(1 - v^2) \sum_{k \geq 1} d(k) v^{2k} \\ &= [v^N](1 - v)^2(1 + v)^{2N-1} \sum_{k \geq 1} d(k) v^k \\ &= \sum_{k \geq 1} d(k) \left[\binom{2N-1}{N-k} - 2 \binom{2N-1}{N-1-k} + \binom{2N-1}{N-2-k} \right]. \end{aligned}$$

Theorem 3. *The average height of the first sojourn of a Dyck path of length $2N$ is given by*

$$\frac{1}{b_N} \sum_{k \geq 1} d(k) \left[\binom{2N-1}{N-k} - 2 \binom{2N-1}{N-1-k} + \binom{2N-1}{N-2-k} \right].$$

A similar computation as before evaluates this asymptotically:

Theorem 4. *The average height of the first sojourn of a Dyck path of length $2N$ is asymptotic to*

$$\frac{1}{2} \sqrt{\pi N} + \frac{1}{4} \log N + \frac{3}{4}(\gamma - 1).$$

It is interesting to note that in the average height of the Dyck path, no such logarithmic term occurs.

5. THE LOCATION OF THE FIRST MAXIMUM

For reasons of comparison, it is also interesting to study the location of the first maximum of the whole path (not only the first sojourn). The computations are somewhat similar.

$$\begin{aligned} F(z, u) &= \sum_{h \geq 1} C_{h-1}(zu) zu C_h(z) \\ &= \sqrt{1-4z^2} \sqrt{1-4z^2 u^2} \sum_{h \geq 1} \frac{z^{2h} u^h}{(\lambda_1^{h+1}(zu) - \lambda_2^{h+1}(zu))(\lambda_1^{h+2}(zu) - \lambda_2^{h+2}(zu))}. \end{aligned}$$

We compute again the average, by differentiating this function:

$$\begin{aligned} G(z) &= -4z^2 \sum_{h \geq 1} \frac{z^{2h}}{(\lambda_1^{h+1}(z) - \lambda_2^{h+1}(z))(\lambda_1^{h+2}(z) - \lambda_2^{h+2}(z))} \\ &\quad + (1-4z^2) \sum_{h \geq 1} \frac{hz^{2h}}{(\lambda_1^{h+1}(z) - \lambda_2^{h+1}(z))(\lambda_1^{h+2}(z) - \lambda_2^{h+2}(z))} \\ &\quad + 2z^2 \sqrt{1-4z^2} \sum_{h \geq 1} \frac{(h+1)z^{2h}(\lambda_1^h(z) + \lambda_2^h(z))}{(\lambda_1^{h+1}(z) - \lambda_2^{h+1}(z))^2(\lambda_1^{h+2}(z) - \lambda_2^{h+2}(z))} \\ &= -4(1+v^2) \sum_{h \geq 2} \frac{v^{2h}}{(1-v^{2h})(1-v^{2h+2})} \\ &\quad + (1-v^2)^2(1+v^2) \sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h+2})(1-v^{2h+4})} \\ &\quad + 2(1-v^4) \sum_{h \geq 2} \frac{hv^{2h}(1+v^{2h-2})}{(1-v^{2h})^2(1-v^{2h+2})} \\ &= -\frac{(1+v^2)^3}{(1-v^2)^2} - \frac{(1+v^2)^3}{v^2(1-v^2)} \sum_{h \geq 1} \frac{v^{2h}}{1-v^{2h}} + \frac{2(1+v^2)^2}{v^2} \sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h})^2}. \end{aligned}$$

And we read off coefficients again:

$$\begin{aligned} [z^n]G(z) &= [v^n](1-v^2)(1+v^2)^{n-1}G(z(v)) \\ &= -[v^N] \frac{(1+v)^{2N+2}}{1-v} - [v^{N+1}](1+v)^{2N+2} \sum_{h \geq 1} \frac{v^h}{1-v^h} \\ &\quad + 2[v^{N+1}](1-v)(1+v)^{2N+1} \sum_{h \geq 1} \frac{hv^h}{(1-v^h)^2} \\ &= -\sum_{k=0}^N \binom{2N+2}{k} - [v^{N+1}](1+v)^{2N+2} \sum_{k \geq 1} d(k)v^k \end{aligned}$$

$$\begin{aligned}
& + 2[v^{N+1}](1-v)(1+v)^{2N+1} \sum_{k \geq 1} kd(k)v^k \\
& = -2^{2N+1} + \binom{2N+1}{N} - \sum_{k \geq 1} d(k) \binom{2N+2}{N+1-k} \\
& \quad + 2 \sum_{k \geq 1} kd(k) \left[\binom{2N+1}{N+1-k} - \binom{2N+1}{N-k} \right].
\end{aligned}$$

Theorem 5. *The average location of the first maximum, assuming that all $b_N = \frac{1}{N+1} \binom{2N}{N}$ Dyck paths of length $2N$ are equally likely, is given by*

$$\frac{1}{b_N} \left(-2^{2N+1} + \binom{2N+1}{N} - \sum_{k \geq 1} d(k) \binom{2N+2}{N+1-k} + 2 \sum_{k \geq 1} kd(k) \left[\binom{2N+1}{N+1-k} - \binom{2N+1}{N-k} \right] \right).$$

The asymptotic evaluation is exactly as before, with the result:

Theorem 6. *The average location of the first maximum, assuming that all $b_N = \frac{1}{N+1} \binom{2N}{N}$ Dyck paths of length $2N$ are equally likely, is asymptotic to*

$$N - \frac{1}{3} \sqrt{\pi N} + \frac{1}{2}.$$

6. THE GENERAL CASE

We don't have to stop here. Indeed, we can consider the height (location and value) within the r -th sojourn. The expected number of sojourns (=degree of the root of the associated planar tree) is just 2. Kemp has analyzed the height of trees with root degree exactly r , see [4, 6].

We consider Dyck paths of length n which have $\geq r$ sojourns, where r is a parameter. The generating function of them is $(zC(z)z)^r C(z) = z^{2r} C^{r+1}(z)$. A standard application of the Lagrange inversion formula enumerates the paths:

$$\frac{r+1}{N+1} \binom{2N-r}{N}.$$

Asymptotically, a proportion of $\frac{r+1}{2^r}$ of all paths has $\geq r$ sojourns.

Now, for the position of the first maximum within the r -th sojourn, we have to consider

$$F_r(z, u) = z^{2(r-1)} u^{2(r-1)} C^{r-1}(zu) F(z, u).$$

For the average, we compute

$$\begin{aligned}
G_r(z) & = 2(r-1)z^{2(r-1)}C^{r-1}(z)F(z, 1) \\
& \quad + z^{2(r-1)}(r-1)C^{r-2}(z)zC'(z)F(z, 1) \\
& \quad + z^{2(r-1)}C^{r-1}(z)G(z) \\
& = (r-1) \left(2 + \frac{zC'(z)}{C(z)} \right) z^{2r} C^{r+1}(z) + z^{2(r-1)} C^{r-1}(z) G(z)
\end{aligned}$$

$$\begin{aligned}
&= \frac{2(r-1)v^{2r}}{(1-v^2)(1+v^2)^{r+1}} \\
&+ \frac{v^{2(r-1)}}{(1+v^2)^{r-1}} \left[-\frac{4v^4}{(1-v^2)^2} - \frac{(1+v^2)^2}{1-v^2} \sum_{h \geq 1} \frac{v^{2h}}{1-v^{2h}} + 2(1+v^2) \sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h})^2} \right] \\
&= \frac{2(r-1)v^{2r}}{(1-v^2)(1+v^2)^{r+1}} - \frac{4v^{2(r+1)}}{(1-v^2)^2(1+v^2)^{r-1}} \\
&- \frac{v^{2(r-1)}}{(1-v^2)(1+v^2)^{r-3}} \sum_{h \geq 1} \frac{v^{2h}}{1-v^{2h}} + \frac{2v^{2(r-1)}}{(1+v^2)^{r-2}} \sum_{h \geq 1} \frac{hv^{2h}}{(1-v^{2h})^2}.
\end{aligned}$$

And thus

$$\begin{aligned}
[z^{2N}]G(z) &= 2(r-1) \binom{2N-r-2}{N-r} - 4 \sum_{k=1}^{N-r-1} \binom{2N-r}{k} \\
&- \sum_{k \geq 1} d(k) \binom{2N-r+2}{N-r+1-k} \\
&+ 2 \sum_{k \geq 1} k d(k) \left[\binom{2N-r+1}{N-r+1-k} - \binom{2N-r+1}{N-r-k} \right].
\end{aligned}$$

Note that

$$\sum_{k=1}^{N-r-1} \binom{2N-r}{k} = \frac{1}{2} \left(2^{2N-r} - \sum_{k=N-r}^N \binom{2N-r}{k} \right).$$

And we have

$$\begin{aligned}
&\frac{-\binom{2N-r+2}{N-r+1-k} + 2k \binom{2N-r+1}{N-r+1-k} - 2k \binom{2N-r+1}{N-r-k}}{\binom{2N-r}{N}} \sim e^{-x^2} \left(4(-1+2x^2) \right. \\
&\left. + \frac{8r(x-x^3)}{\sqrt{N}} + \frac{2-6(r^2-r+3)x^2 + (4r^2-4r+\frac{38}{3})x^4 - \frac{4}{3}x^6}{N} \right)
\end{aligned}$$

Eventually, the same procedure as before produces the asymptotic formula for the position of the first maximum within the r -th sojourn:

$$\frac{N}{2} - \frac{\sqrt{\pi N}}{3(r+1)}.$$

And finally we will consider the height of the r -th sojourn. We must multiply the old $B_h(z)$ function by $(z^2 C(z))^{r-1}$, or by $\frac{v^{2(r-1)}}{(1+v^2)^{2(r-1)}}$.

The generating function of paths where the maximum in the r -th sojourn is $> h$, is given by

$$\frac{v^{2r}(1-v^2)}{(1+v^2)^{r-1}} \frac{v^{2h}}{1-v^{2h+2}}.$$

And for the expectation, we must consider

$$\frac{v^{2(r-1)}(1-v^2)}{(1+v^2)^{r-1}} \sum_{h \geq 1} \frac{v^{2h}}{1-v^{2h}}.$$

Apart from normalization, we look at

$$\sum_{k \geq 1} d(k) \left[\binom{2N-r}{N+1-r-k} - 2 \binom{2N-1}{N-r-k} + \binom{2N-1}{N-1-r-k} \right].$$

We approximate again:

$$\frac{\binom{2N-r}{N+1-r-k} - 2 \binom{2N-1}{N-r-k} + \binom{2N-1}{N-1-r-k}}{\binom{2N-r}{N}} \sim \frac{1}{N} e^{-k^2/N} \left(-2 + 4 \frac{k^2}{N} + \text{further terms} \right).$$

Eventually, the average evaluates asymptotically to

$$\frac{\sqrt{\pi N}}{r+1} + \frac{r}{2(r+1)} \log N - \frac{2r+1}{2(r+1)} + \frac{3r\gamma}{2(r+1)}.$$

7. CONCLUSION

We leave limiting distributions, higher moments and asymmetric random walks to further research.

REFERENCES

- [1] N. G. De Bruijn, D. E. Knuth, and S. O. Rice. The average height of planted plane trees. In R. C. Read, editor, *Graph Theory and Computing*, pages 15–22. Academic Press, 1972.
- [2] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.
- [3] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Addison-Wesley. available at <http://algo.inria.fr/flajolet/Publications/>.
- [4] R. Kemp. The average height of r -tuply rooted planted plane trees. *Computing*, 25:209–232, 1980.
- [5] W. Panny and H. Prodinger. The expected height of paths for several notions of height. *Studia Scientiarum Mathematicarum Hungarica*, 20:119–132, 1985.
- [6] H. Prodinger. A note on a result of R. Kemp on r -tuply rooted planted plane trees. *Computing*, 28:363–366, 1982.
- [7] H. Prodinger. Some analytic techniques for the investigation of the asymptotic behaviour of tree parameters. 47:180–199, 1992.
- [8] J. Randon-Furling and S.N. Majumdar. Distribution of the time at which a Brownian motion is maximal before its first-passage time. *arXiv:0708.2101v1*, page 16 pages, 2007.

HELMUT PRODINGER, MATHEMATICS DEPARTMENT, STELLENBOSCH UNIVERSITY, 7602 STELLENBOSCH, SOUTH AFRICA.

E-mail address: hproding@sun.ac.za