RETURNS AND HILLS ON t-ARY TREES

HELMUT PRODINGER

ABSTRACT. A recent analysis of returns and hills of generalized Dyck paths is carried over to the language of t-ary trees, from which, by explicit bivariate generating functions, all the relevant results follow quickly and smoothly. A conjecture about the (discrete) limiting distribution of hills is settled in the affirmative.

1. INTRODUCTION

In the recent paper [2], *generalized Dyck paths* where investigated: they have an up-step $\mathsf{u} = (1, 1)$, a down-step $\mathsf{d} = (1, -t + 1)$, where $t \geq 2$, start at the origin, end on the x-axis, and never go below the x-axis. A general reference for such lattice paths is [1].

Two parameters were investigated (with the help of Riordan arrays): the number of returns to the x-axis (the origin itself does not count), and the number of (contiguous) subpaths of the form $\mathsf{u}^{t-1}\mathsf{d}$, that sit on the x-axis.

In the present note, I would like to emphasize that the language of trees, in particular t-ary trees, is favorable here, because it allows to write the relevant generating functions with ease, without any mentioning of Riordan arrays, and also leads to settling a conjecture mentioned in [2].

The family of t-ary trees is recursively described: such a tree is either an external node (depicted as a square), or a root (an internal node, depicted as a circle), followed by subtrees (in this order) $T_1, \ldots, T_t$. For this and many other concepts, we refer to the universal book [3]. The generating function $T(z) = \sum_{n \geq 0} a_n z^n$, where a_n is the number of trees of size n (n internal nodes) is, following the recursive definition, given by $T(z) = 1 + zT^t(z)$. Extracting coefficients is efficiently done by setting $z = u/(1+u)^t$, thus $T = 1 + u$, and contour integration; the method is closely related to the Lagrange inversion formula. Here is an example:

$$\begin{aligned}
[z^n]T^k(z) &= \frac{1}{2\pi i} \oint \frac{dz}{z^{n+1}} T^k(z) \\
&= \frac{1}{2\pi i} \oint \frac{du(1+u-tu)(1+u)^{t(n+1)}}{(1+u)^{t+1}u^{n+1}} (1+u)^k \\
&= [u^n](1+u-tu)(1+u)^{tn+k-1}
\end{aligned}$$

Date: July 12, 2016.

1991 *Mathematics Subject Classification.* Primary: 05A15, 05A16; Secondary: 60C05.

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

This produces in particular the numbers $a_n = \frac{1}{n}\binom{tn}{n-1}$.

There is a natural bijection between the family of generalized Dyck paths and the family of t-ary trees. It is based on the decomposition of paths according to the *first* return to the x-axis. The first part of the Dyck paths is (recursively) responsible for the first $t-1$ subtrees, and the rest of the Dyck path for the remaining t-th subtree. It is then apparent that the number of down-steps is the same as the number of internal nodes of the associated tree. Here is the situation depicted for $t = 3$.

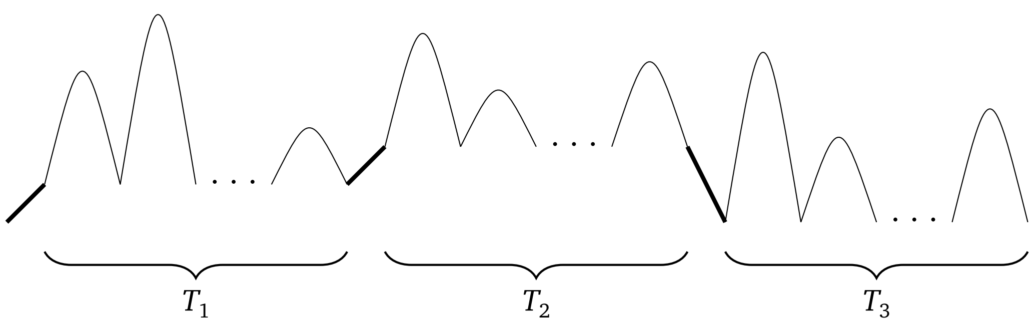

FIGURE 1. The decomposition of generalized Dyck paths leading (recursively) to a ternary tree with subtrees T_1, T_2, T_3.

Now a little reflection convinces us that the number of returns is the same as the number of (internal) nodes on the rightmost path from the root to the rightmost leaf. And, further: the number of hills is the number of nodes on this rightmost path with the property that its first $t-1$ subtrees are empty (are the empty subtree, consisting only of an external node).

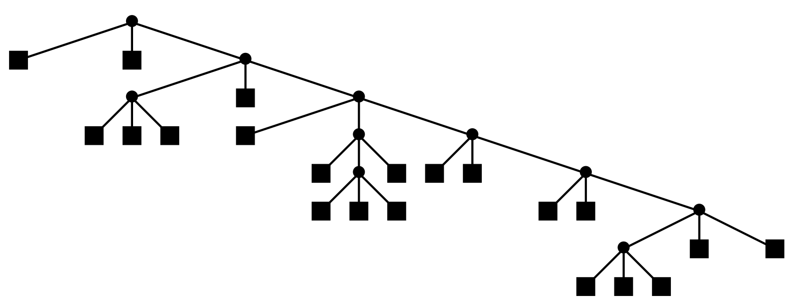

FIGURE 2. A ternary tree with 10 (internal) nodes. It has 6 returns and 3 hills.

In what follows, we will analyze these parameters in terms of t-ary trees.

In [2], the *negative binomial distribution* is defined via

$$\mathbb{P}\{Y=k\}=\binom{k-1}{r-1}p^r(1-p)^{k-r}.$$

This is somewhat in contrast with [3] and *Wikipedia*, as it is a shifted version, and the roles of p and $1-p$ are interchanged from the more common definitions. Nevertheless, we will stick to this definition here, for the reason of comparisons. The numbers r and p are called the parameters of the distribution.

2. The number of returns on t-ary trees

Let $F(z,v)$ be the generating function with respect to the size and the number of returns, i. e., the coefficient of $z^n v^k$ is the number trees with n internal nodes and k returns. Then we find the equation

$$F(z,v)=1+zT^{t-1}(z)vF(z,v).$$

Since $zT^{t-1}(z)=\frac{T(z)-1}{T(z)}$, this leads to the explicit form

$$F(z,v)=\frac{1}{1-v\frac{T(z)-1}{T(z)}}.$$

Therefore

$$[v^k]F(z,v)=\Big(\frac{T(z)-1}{T(z)}\Big)^k=\Big(\frac{u}{1+u}\Big)^k.$$

Furthermore

$$\begin{aligned}
[z^n][v^k]F(z,v)&=[z^n]\Big(\frac{u}{1+u}\Big)^k\\
&=\frac{1}{2\pi i}\oint\frac{dz}{z^{n+1}}\Big(\frac{u}{1+u}\Big)^k\\
&=\frac{1}{2\pi i}\oint\frac{du(1+u-tu)(1+u)^{t(n+1)}}{u^{n+1}(1+u)^{t+1}}\Big(\frac{u}{1+u}\Big)^k\\
&=[u^{n-k}](1+u-tu)(1+u)^{tn-1-k}\\
&=\binom{tn-1-k}{n-k}-(t-1)\binom{tn-1-k}{n-1-k}\\
&=\frac{k}{n}\binom{tn-1-k}{n-k}.
\end{aligned}$$

Division by a_n gives the probability that a random tree of size n has k returns:

$$p_k(n)=k\frac{\binom{tn-1-k}{n-k}}{\binom{tn}{n-1}}\to\frac{k(t-1)^2}{t^{k+1}},\quad \text{fixed } k,\quad n\to\infty.$$

In order to compute the d-th (factorial) moment, we evaluate

$$\frac{\partial^d}{\partial v^d}F(z,v)\Big|_{v=1} = d!T(z)(T(z)-1)^d = d!(1+u)u^d.$$

Furthermore,

$$\begin{aligned}[z^n]\frac{\partial^d}{\partial v^d}F(z,v)\Big|_{v=1} &= [u^{n-d}]d!(1+u-tu)(1+u)^{tn}\\ &= d!\binom{tn}{n-d} - d!(t-1)\binom{tn}{n-1-d} = \frac{(td+1)d!}{n-d}\binom{tn}{n-1-d}.\end{aligned}$$

For the expected value, we consider $d=1$ and divide by a_n, with the result

$$\frac{(t+1)n}{n(t-1)+2} \sim \frac{t+1}{t-1}.$$

The second factorial moment is obtained via $d=2$, with the result

$$\frac{2(2t+1)n(n-1)}{(tn-n+3)(tn-n+2)} \sim \frac{2(2t+1)}{(t-1)^2}.$$

This leads to the variance:

$$2\frac{n(t-1)(n-1)(tn+1)}{(tn-n+3)(tn-n+2)^2} \sim \frac{2t}{(t-1)^2}.$$

This section reproved and extended the results of [2] on the number of returns. Note that the quantity $\frac{k(t-1)^2}{t^{k+1}}$ is $\mathbb{P}\{Y=k+1\}$, where Y is a random variable, distributed according to the negative binomial distribution for $r=2$ and $p=\frac{t-1}{t}$.

3. The number of hills on t-ary trees

Let $G(z,v)$ be the generating function with respect to the size (variable z) and the number of hills (variable v). Then we find the recursion

$$G(z,v) = 1 + zT^{t-1}(z)G(z,v) + z(v-1)G(z,v).$$

Since $zT^{t-1}(z) = 1-1/T(z)$, we find the explicit solution

$$G(z,v) = \frac{T(z)}{1-(v-1)zT(z)} = \sum_{k\ge0}(v-1)^k z^k T^{k+1}(z).$$

By d-fold differentation, followed by setting $v=1$, we get the generating function of the d-th factorial moments (apart from normalization):

$$d!z^dT^{d+1}(z).$$

Furthermore,

$$[z^n]d!z^dT^{d+1}(z) = \frac{d!}{2\pi i}\oint\frac{dz}{z^{n+1-d}}T^{d+1}(z)$$

$$
\begin{aligned}
&= \frac{d!}{2\pi i}\oint \frac{du(1+u-tu)(1+u)^{t(n-d)+d}}{u^{n+1-d}}\\
&= d![u^{n-d}](1+u-tu)(1+u)^{t(n-d)+d}\\
&= d!\binom{tn-(t-1)d}{n-d} - d!(t-1)\binom{tn-(t-1)d}{n-1-d}\\
&= \frac{(d+1)!}{n-d}\binom{tn-(t-1)d}{n-1-d}.
\end{aligned}
$$

For $d=1$, this leads to the expected value:

$$
\frac{n}{\binom{tn}{n-1}}\frac{2}{n-1}\binom{tn-t+1}{n-2} = \frac{2(tn-t+1)!(tn-n+1)!}{t(tn-1)!(tn-n-t+3)!} \to \frac{2(t-1)^{t-2}}{t^{t-1}}.
$$

The variance evaluates to

$$
\frac{n}{\binom{tn}{n-1}}\frac{6}{n-2}\binom{tn-2t+2}{n-3} + \frac{2(tn-t+1)!(tn-n+1)!}{t(tn-1)!(tn-n-t+3)!} - \left[\frac{2(tn-t+1)!(tn-n+1)!}{t(tn-1)!(tn-n-t+3)!}\right]^2,
$$

which we do not attempt to simplify any further.

Writing

$$
G(z,v) = \sum_{n,k} g_{n,k} z^n v^k,
$$

it is possible to derive an explicit form for the coefficients $g_{n,k}$, but they are not as nice as the corresponding quantities in the previous section:

$$
\begin{aligned}
G(z,v) &= \sum_{k\ge 0}(v-1)^k z^k T^{k+1}(z)\\
&= \sum_{n\ge 0} z^n \sum_{k\ge 0}(v-1)^k \frac{k+1}{n-k}\binom{tn-(t-1)k}{n-1-k}\\
&= \sum_{n\ge 0} z^n \sum_{k\ge 0}\sum_{0\le j\le k}\binom{k}{j} v^j (-1)^{k-j}\frac{k+1}{n-k}\binom{tn-(t-1)k}{n-1-k}.
\end{aligned}
$$

This leads to

$$
g_{n,j} = \sum_{j\le k\le n}\binom{k}{j}(-1)^{k-j}\frac{k+1}{n-k}\binom{tn-(t-1)k}{n-1-k}.
$$

The limiting distribution of $g_{n,j}/a_n$, for j fixed, must thus be determined in a different way.

We need a crash course of asymptotic tree enumeration here; all this can be found in [3], but compare also [4], in particular the notion of *simply generated families of trees*. The procedure that we describe here is closely related to the discussion in [3], Section IX-3, where very similar parameters were analyzed.

We start from $u = z\phi(u)$, with $\phi(u) = (1+u)^t$. The quantity τ is determined via the equation $\phi(\tau) = \tau\phi'(\tau)$. In our case this leads to $\tau = \frac{1}{t-1}$. Then there is the quantity

$\rho = \frac{\tau}{\phi(\tau)}$, which here evaluates to

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

Then one knows by general principles that the function $u(z)$ has a square-root singularity around $z = \rho$, with the local expansion

$$u \sim \tau - \sqrt{\frac{2\tau}{\rho\phi''(\tau)}}\sqrt{1-z/\rho}.$$

This is here

$$T(z) = 1 + u \sim \frac{t}{t-1} - \sqrt{\frac{2t}{(t-1)^3}}\sqrt{1-z/\rho}.$$

This expansion will now be used inside of $G(z, v)$, with the result (Maple):

$$G(z,v) \sim a - \frac{\sqrt{2}t^{2t-3/2}}{(t-1)^{3/2}\Big(t^{t-1}+(t-1)^{t-2}-(t-1)^{t-2}v\Big)^2}\sqrt{1-z/\rho},$$

with a being an unimportant constant. Note that

$$\frac{\sqrt{2}t^{2t-3/2}}{(t-1)^{3/2}\Big(t^{t-1}+(t-1)^{t-2}-(t-1)^{t-2}v\Big)^2}\Bigg|_{v=1} = \sqrt{\frac{2t}{(t-1)^3}}.$$

Thus, the limiting distribution is given by the probability generating function

$$\frac{t^{2t-2}}{\Big(t^{t-1}+(t-1)^{t-2}-(t-1)^{t-2}v\Big)^2} = \frac{\frac{t^{2t-2}}{(t^{t-1}+(t-1)^{t-2})^2}}{\Big(1-\frac{(t-1)^{t-2}}{t^{t-1}+(t-1)^{t-2}}v\Big)^2}.$$

The coefficient of v^k in it given by

$$(k+1)\frac{t^{2t-2}(t-1)^{(t-2)k}}{(t^{t-1}+(t-1)^{t-2})^{k+2}}.$$

which is $\mathbb{P}\{Y = k+2\}$, for a random variable Y, which follows the negative binomial distribution with parameters $r = 2$ and $p = \frac{t^{t-1}}{t^{t-1}+(t-1)^{t-2}}$, as conjectured in [2].

Acknowledgement. Thanks are due to Stephan Wagner for stimulating feedback.

References

[1] C. Banderier and P. Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science*, 281:37–80, 2002.

[2] N. T. Cameron and J. E. McLeod. Returns and hills on generalized Dyck paths. *Journal of Integer Sequences*, 19:Article 16.6.1 (28 pages), 2016.

[3] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009.

[4] A. Meir and J. W. Moon. On the altitudes of nodes in random trees. *Canad. Math. J.*, 30:997–1015, 1978.

H. Prodinger, Department of Mathematical Sciences, Mathematics Division, Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa

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