MOMENTS OF LEVEL NUMBERS OF LEAVES IN BINARY TREES

ALOIS PANHOLZER AND HELMUT PRODINGER

ABSTRACT. For the random variable "height of leaf j in a binary tree of size n" we derive closed formulæ for all moments. The more general case of t–ary trees is also considered.

1. Introduction

We reexamine the heights of leaves (enumerated from left to right) in binary trees and, more generally, t–ary trees.

The expectations go back to Kirschenhofer [5], and the variance to Gutjahr [3]. For higher moments, Gutjahr had to resort to analytic techniques. However, we are able to produce an explicit formula for the s–th moment, for general s. The result is "closed form," as it comprised only a number of summands that depends on s, not on n or j, where n denotes the size of the tree and j the number of the leaf in consideration.

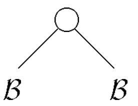
From this explicit result, asymptotic formulæ, as given by Gutjahr, are of course corollaries for which we don't need more than Stirling's approximation for the factorials.

In a follow–up paper to [5], Kirschenhofer considered in [7] also t–ary trees (binary trees are the instance $t = 2$). He treated the expectation with a complicated analytic technique ("diagonalization method") because he believed that explicit enumeration is not feasible. However, we disproof this claim by giving an exact formula for the s-th (factorial) moment. This time, it is not "closed form," but we can get asymptotics for the expectations by rather elementary techniques (Stirling's formula and approximation of a sum by an integral).

2. The height of leaves in binary trees

We denote by $\mathcal{B}$ the family of binary trees, sometimes called extended binary trees, which is defined by the formal identity

$$\mathcal{B} \quad = \quad \square \quad + $$



(1)

where $\bigcirc$ is the symbol for an internal node and $\square$ is the symbol for a leaf or external node.

The family of objects from $\mathcal{B}$ with exactly n internal nodes is denoted by $\mathcal{B}_n$. Further we write B_n for the number of elements from $\mathcal{B}_n$. Finally we denote with $B(z) = \sum_{n\geq 0} B_n z^n$ the ordinary generating function of the B_n's.

Let us state the following classic results that we will need in the sequel:

$$B(z) = 1 + zB^2(z) = \frac{1-\sqrt{1-4z}}{2z}; \qquad B_n = \frac{1}{n+1}\binom{2n}{n}.$$

Assume that the leaves are labelled with the numbers $0, 1, \ldots, n$ from left to right and that the distance (depth, height) of a leaf is given by the number of internal nodes on the path from the root to the leaf. This defines random variables $X_{n,j}$ assuming that all binary trees of size n are equally likely.

[1]This research was supported by the Austrian Research Society (FWF) under the project number P12599-MAT

Date: March 17, 1998.

Kirschenhofer [5, 6] has obtained the following explicit formula for the expectations (see [8] for more results);

$$\mathbb{E}(X_{n,j}) = \frac{2(2j+1)(2n-2j+1)}{n+2} \frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}} - 1 \qquad \text{for } \ 0 \le j \le n\,.$$

Gutjahr obtained the variance $\mathbb{V}(X_{n,j})$ in [3]; unfortunately, this paper was overlooked during the preparation of our paper [8].

Studying the higher moments, Gutjahr writes:

"There is little hope to find a closed expression for the third *moment."*

This statement is too pessimistic, since we will find explicit formulæ for *all* (factorial) moments in the sequel. Since the usual moments can be obtained from them by linear combinations with Stirling numbers as coefficients, we have also explicit formulæ for all usual moments.

Let $f_{n,j,l}$ denote the number of binary trees of size n, where the leaf labelled with j has height l. Then we get for the trivariate generating function $F(z,u,v)$ of the $f_{n,j,l}$, defined by $\sum_{n,j,l} f_{n,j,l} z^n u^j v^l$ the functional equation

$$F(z,u,v) = 1 + zvF(z,u,v)B(z) + zuvF(z,u,v)B(zu)\,. \tag{2}$$

This leads immediately to the solution

$$F(z,u,v) = \frac{1}{1 - zvB(z) - zuvB(zu)}\,. \tag{3}$$

This appears in [8] and in [3]. The s–th factorial moment $\mathbb{E}^{\underline{s}}(X_{n,j})$ is given by

$$\mathbb{E}^{\underline{s}}(X_{n,j}) = \frac{1}{B_n}[z^n u^j] \frac{\partial^s}{\partial v^s} F(z,u,v)\bigg|_{v=1}.$$

Now, a routine calculation (that is easy to verify by induction) shows that

$$\frac{\partial^s}{\partial v^s} F(z,u,v)\bigg|_{v=1} = 2s! \frac{\big(2-(X+Y)\big)^s}{(X+Y)^{s+1}} = 2s! \sum_{l=0}^{s} \binom{s}{l} (-1)^{s-l} 2^l \frac{1}{(X+Y)^{l+1}},$$

with

$$X = \sqrt{1-4z} \qquad \text{and} \qquad Y = \sqrt{1-4zu},$$

so that our task is reduced to find explicit formulæ for

$$[z^n u^j] \frac{1}{(X+Y)^q}.$$

First we compute the coefficients of $1/(1+\sqrt{1-4x}\,)^q$; for that, we use the substitution $x = t/(1+t)^2$ and the Lagrange inversion formula:

$$\begin{aligned}
[x^k] \frac{1}{(1+\sqrt{1-4x}\,)^q} &= \frac{1}{k}[t^{k-1}](1+t)^{2k} \frac{d}{dt} \frac{1}{\left(1+\sqrt{1-\frac{4t}{(1+t)^2}}\,\right)^q} \\
&= \frac{1}{k}[t^{k-1}](1+t)^{2k} \frac{d}{dt} \left(\frac{1+t}{2}\right)^q \\
&= \frac{q}{2^q} \frac{1}{k+q} \binom{2k+q-1}{k}.
\end{aligned}$$

Hence

$$\frac{1}{(X+Y)^q} = \frac{1}{(1-4z)^{q/2}\left(1+\sqrt{1-\frac{4z(u-1)}{1-4z}}\right)^q}$$
$$= \sum_{k\ge 0} \frac{q}{2^q}\frac{1}{k+q}\binom{2k+q-1}{k}\frac{z^k(u-1)^k}{(1-4z)^{k+\frac{q}{2}}}.$$

Thus

$$[z^n u^j]\frac{1}{(X+Y)^q} = \sum_{j\le k\le n}(-1)^{k-j}\binom{k}{j}\frac{q}{2^q}\frac{1}{k+q}\binom{2k+q-1}{k}4^{n-k}\binom{n-1+\frac{q}{2}}{n-k}.$$

To evaluate the remaining sum we use Zeilberger's algorithm [2, 9]. (Alternative proofs, e. g. by the use of hypergeometric series, would also be possible.) Let

$$F(n,k) = (-1)^{k-j}\binom{k}{j}\frac{q}{2^q}\frac{1}{k+q}\binom{2k+q-1}{k}4^{n-k}\binom{n-1+\frac{q}{2}}{n-k},$$

then we find

$$(n-j+1)(n+q+1)F(n+1,k) - (2n+q)(2n+q-2j+1)F(n,j) = G(n,k+1) - G(n,k)$$

with

$$R(n,k) = -\frac{2(2n+q)(k+q)(k-j)}{n+1-k} \qquad \text{and} \qquad G(n,k) = F(n,k)R(n,k).$$

Now set

$$f(n) = \sum_{k\in\mathbb{Z}} F(n,k),$$

then we obtain by summing (since the right side telescopes)

$$(n-j)(n+q)f(n) = (2n+q-2)(2n+q-2j-1)f(n-1),$$

with the initial value

$$f(j) = \frac{q}{2^q}\frac{1}{j+q}\binom{2j+q-1}{j},$$

which gives by iteration

$$f(n) = \frac{q}{2^q}\frac{1}{j+q}\binom{2j+q-1}{j}\prod_{k=j+1}^{n}\frac{(2k+q-2)(2k+q-2j-1)}{(k-j)(k+q)}$$
$$= \frac{q}{2^q}\frac{1}{j+q}\binom{2j+q-1}{j}\frac{(j+q)!}{(n-j)!(n+q)!}\prod_{k=j+1}^{n}(2k+q-2)(2k+q-2j-1)$$
$$= q\,2^{2n-2j-q}\frac{(2j+q-1)!}{j!(n-j)!(n+q)!}\frac{\Gamma\big(n+\frac{q}{2}\big)\Gamma\big(n-j+\frac{q+1}{2}\big)}{\Gamma\big(j+\frac{q}{2}\big)\Gamma\big(\frac{q+1}{2}\big)}.$$

Of course, these Γ–functions can always be written by factorials (or binomial coefficients), by distinguishing the cases q even and odd. Summarizing,

$$[z^n u^j]\frac{1}{(X+Y)^q} = q\,2^{2n-2j-q}\frac{(2j+q-1)!}{j!(n-j)!(n+q)!}\frac{\Gamma\big(n+\frac{q}{2}\big)\Gamma\big(n-j+\frac{q+1}{2}\big)}{\Gamma\big(j+\frac{q}{2}\big)\Gamma\big(\frac{q+1}{2}\big)}$$

or citing the cases q even and odd explicitly

$$[z^n u^j]\frac{1}{(X+Y)^{2p}} = \frac{2p(p-1)!(j+2p)!(2n-2j+2p-1)!(2j+2p-1)!(n+p-1)!}{2^{2p}(2p-1)!(j+p-1)!(n-j)!(n-j+p-1)!j!(j+2p)!(n+2p)!},$$
$$[z^n u^j]\frac{1}{(X+Y)^{2p+1}} = \frac{(2p+1)(j+p)!(2n+2p-1)!(n-j+p)!}{2^{2p}p!j!(n+p-1)!(n-j)!(n+2p+1)!}.$$

Hence we have the following theorem.

Theorem 1. *The s–th factorial moment $\mathbb{E}(X_{n,j}^{\underline{s}})$ is given by the following explicit formula.*

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \frac{s!}{B_n}\sum_{l=0}^{s}\binom{s}{l}(-1)^{s-l}4^{n-j}\frac{(l+1)\,(2j+l)!}{j!(n-j)!(n+l+1)!}\frac{\Gamma\big(n+\frac{l+1}{2}\big)\Gamma\big(n+1-j+\frac{l}{2}\big)}{\Gamma\big(j+\frac{l+1}{2}\big)\Gamma\big(1+\frac{l}{2}\big)}. \tag{4}$$

□

With Stirling's asymptotic formula for the factorials, we get for a fixed ratio $\rho = j/n$

$$\frac{(l+1)\,(2j+l)!}{j!(n-j)!(n+l+1)!}\frac{\Gamma\big(n+\frac{l+1}{2}\big)\Gamma\big(n+1-j+\frac{l}{2}\big)}{\Gamma\big(j+\frac{l+1}{2}\big)\Gamma\big(1+\frac{l}{2}\big)} = 4^{\rho n}n^{\frac{l-3}{2}}\frac{(1-\rho)^{\frac{l}{2}}\rho^{\frac{l}{2}}2^l(l+1)}{\sqrt{\pi}\,\Gamma\big(1+\frac{l}{2}\big)}\left(1+\mathcal{O}\Big(\frac{1}{n}\Big)\right).$$

From this we see that the summand with $l = s$ gives the main term in the asymptotic expansion of the factorial moments. Using the reflection law of the Gamma function we get

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \frac{(s+1)!(1-\rho)^{\frac{s}{2}}\rho^{\frac{s}{2}}2^s}{\Gamma\big(1+\frac{s}{2}\big)}n^{\frac{s}{2}} + \mathcal{O}\left(n^{\frac{s-1}{2}}\right) = \frac{2}{\sqrt{\pi}}4^s\rho^{\frac{s}{2}}(1-\rho)^{\frac{s}{2}}\Gamma\left(\frac{s+3}{2}\right)n^{\frac{s}{2}} + \mathcal{O}\left(n^{\frac{s-1}{2}}\right).$$

Thus we have the following corollary.

Corollary 2. *The s–th factorial moment $\mathbb{E}(X_{n,j}^{\underline{s}})$ (and thus also the ordinary s–th moment $\mathbb{E}(X_{n,j}^{s})$) is asymptotically for $n \to \infty$ given by*

$$\mathbb{E}(X_{n,j}^{\underline{s}}) = \frac{2}{\sqrt{\pi}}4^s\rho^{\frac{s}{2}}(1-\rho)^{\frac{s}{2}}\Gamma\left(\frac{s+3}{2}\right)n^{\frac{s}{2}} + \mathcal{O}\left(n^{\frac{s-1}{2}}\right)$$

Together with the theorem of Fréchet and Shohat (second central limit theorem) [1] and the fact, that the Maxwell distribution is uniquely determined by its moments, we get the following result, which was first proved in [4].

Corollary 3. *The limiting distribution of the normalized height $\frac{X_{n,j}}{\sqrt{n}}$ is for a fixed ratio $\rho = \frac{j}{n}$ with $0 < \rho < 1$ a Maxwell distribution with parameter $\sigma = \sqrt{8\rho(1-\rho)}$. That means it has for $x > 0$ the density function*

$$f_\rho(x) = \frac{x^2}{16\sqrt{\pi}\big(\rho(1-\rho)\big)^{\frac{3}{2}}}e^{-\frac{x^2}{16\rho(1-\rho)}}. \tag{5}$$

For the reader's convenience we mention that the Maxwell distribution with parameter σ is defined as the distributions $Y = \sqrt{X_1^2+X_2^2+X_3^2}$, where the X_i are independently normally distributed random variables $\mathcal{N}(0,\sigma^2)$ with mean 0 and variance σ^2. It has the following density function $f(x)$ and moments $M_s = \mathbb{E}(Y^s)$

$$f(x) = \frac{\sqrt{2}x^2}{\sqrt{\pi}\sigma^3}e^{-\frac{x^2}{2\sigma^2}} \quad \text{for } x > 0\,; \quad M_s = \frac{2}{\sqrt{\pi}}2^{\frac{s}{2}}\sigma^s\Gamma\left(\frac{s+3}{2}\right).$$

3. t–ARY TREES

Let us consider t–ary trees for general $t \ge 2$. In this instance, we can also report some progress, although the results are not so nice. The random variable $X_{n,j}$ is again defined as the height of the leaf j in a t–ary tree of size n, where the leaves are labelled with the numbers $0, 1, \dots, (t-1)n$ from left to right and all trees are equally likely.

Recall that the generating function of t–ary trees satisfies $T(z) = 1 + zT^t(z)$, and that

$$T_n := [z^n]T(z) = \frac{1}{(t-1)n+1}\binom{tn}{n} \quad \text{and} \quad T_n^{(p)} := [z^n]T^p(z) = \frac{p}{(t-1)n+p}\binom{tn+p-1}{n},$$

which is classical and easy to see by means of Lagrange's inversion formula: With $T(z) = 1 + U(z)$ we get

$$[z^n]T^p(z) = [z^n]\big(1+U(z)\big)^p = \frac{p}{n}[U^{n-1}](1+U)^{tn+p-1} = \frac{p}{(t-1)n+p}\binom{tn+p-1}{n}.$$

The trivariate generating function

$$F(z,u,v) = \sum_{n,j,l}\Big[\textsf{Number of trees of size } n \textsf{ with leaf } j \textsf{ on level } l\Big] z^n u^j v^l$$

is now given by

$$F(z,u,v) = \frac{1}{1 - zv\sum_{k=0}^{t-1} u^k\, T^{t-1-k}(z)\, T^k(zu^{t-1})}.$$

Let us define the following abbreviations: $X = T(z)$ and $Y = T(zu^{t-1})$ (in the previous section X and Y had a slightly different meaning). Now note that

$$z\sum_{k=0}^{t-1} u^k X^{t-1-k} Y^k = z\frac{X^t - u^tY^t}{X-uY} = \frac{X - uY - (1-u)}{X-uY} = 1 - \frac{1-u}{X-uY}$$

so that we have

$$F(z,u,v) = \frac{1}{1 - v\big(1 - \frac{1-u}{X-uY}\big)}. \tag{6}$$

3.1. **The Expectation.** At first we want to find an explicit expression for the expectations $\mathbb{E}(X_{n,j})$. Here Kirschenhofer writes [7]:

"For general $t \ge 2$, the method, that led to an explicit expression for $t = 2$, is not feasible."

Again, this statement is too pessimistic; in the following we will find an explicit formula for the expectations, from which it is also possible to get an asymptotic result without the laborious "diagonalization method," by only using very elementary techniques.

Introducing

$$G_1(z,u) = \frac{\partial}{\partial v}F(z,u,v)\bigg|_{v=1}$$

we get by differentiating (6) w.r.t. v and evaluating at $v = 1$ the following result, which appeared in [5].

$$G_1(z,u) = \left(\frac{X-uY}{1-u}\right)^2 - \frac{X-uY}{1-u}.$$

Extracting the coefficients for $0 \le j \le (t-1)n$ leads immediately to

$$[z^nu^j]\frac{X-uY}{1-u} = T_n \quad \text{and} \quad [z^nu^j]\left(\frac{X-uY}{1-u}\right)^2 = [z^nu^j]\frac{X^2}{(1-u)^2} - 2[z^nu^j]\frac{XuY}{(1-u)^2}.$$

For $0 \le j \le (t-1)n$ we get then

$$[z^n u^j]\frac{X^2}{(1-u)^2} = \frac{2(j+1)}{(t-1)n+2}\binom{tn+1}{n}$$

and

$$[z^n u^j]\frac{XuY}{(1-u)^2} = \sum_{0\le k\le \lfloor\frac{j-1}{t-1}\rfloor} \frac{1}{(t-1)k+1}\binom{tk}{k}\frac{1}{(t-1)(n-k)+1}\binom{t(n-k)}{n-k}\bigl(j-(t-1)k\bigr).$$

Combining these results leads to

$$\begin{aligned}[z^n u^j]\left(\frac{X-uY}{1-u}\right)^2 &= \frac{2(j+1)}{(t-1)n+2}\binom{tn+1}{n} \\ &\quad - 2\sum_{0\le k\le \lfloor\frac{j-1}{t-1}\rfloor} \frac{j-(t-1)k}{\bigl((t-1)k+1\bigr)\bigl((t-1)(n-k)+1\bigr)}\binom{tk}{k}\binom{t(n-k)}{n-k}.\end{aligned} \tag{7}$$

To get asymptotic equivalents it is helpful to manipulate this expression in the following way. Partial fraction expansion leads to

$$\begin{aligned}[z^n u^j]\left(\frac{X-uY}{1-u}\right)^2 &= \frac{2(j+1)}{(t-1)n+2}\binom{tn+1}{n} - \frac{2(j+1)}{(t-1)n+2}\sum_{0\le k\le \lfloor\frac{j-1}{t-1}\rfloor} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k} \\ &\quad + \frac{2\bigl((t-1)n-j+1\bigr)}{(t-1)n+2}\sum_{0\le k\le \lfloor\frac{j-1}{t-1}\rfloor} \frac{1}{(t-1)(n-k)+1}\binom{tk}{k}\binom{t(n-k)}{n-k}.\end{aligned}$$

In the following we use the identity

$$\sum_{k=0}^{n} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k} = \binom{tn+1}{n}.$$

This is easy to prove, when considering the equation

$$(t-1)zT'(z) + T(z) = \sum_{n\ge 0}\binom{tn}{n} z^n.$$

The sum can then be simplified by extracting the coefficients of a convolution of two functions:

$$\begin{aligned}\sum_{k=0}^{n} &\frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k} = [z^n]\,T(z)\,\bigl((t-1)zT'(z)+T(z)\bigr) \\ &= [z^n]\,T^2(z) + \frac{t-1}{2}[z^{n-1}]\,\bigl(T^2(z)\bigr)' = \left(1+\frac{t-1}{2}n\right)[z^n]\,T^2(z) = \binom{tn+1}{n}.\end{aligned}$$

When the first sum is running over the whole range from 0 to n it cancels the extra summand, which is asymptotically of higher order. Adding the remaining summands leads to the following expression

$$\begin{aligned}[z^n u^j]\left(\frac{X-uY}{1-u}\right)^2 &= \frac{2(j+1)}{(t-1)n+2}\sum_{\lfloor\frac{j-1}{t-1}\rfloor+1\le k\le n} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k} \\ &\quad + \frac{2\bigl((t-1)n-j+1\bigr)}{(t-1)n+2}\sum_{0\le k\le \lfloor\frac{j-1}{t-1}\rfloor} \frac{1}{(t-1)(n-k)+1}\binom{tk}{k}\binom{t(n-k)}{n-k}.\end{aligned}$$

This leads to the following exact formula for the expectation $\mathbb{E}(X_{n,j}) = \frac{1}{T_n}[z^n u^j]\,G_1(z,u)$.

Theorem 4. *The expectation $\mathbb{E}(X_{n,j})$ of the height of leaf j in a t–ary tree with n internal nodes is given by*

$$\mathbb{E}(X_{n,j}) = \frac{2(j+1)\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{\lfloor\frac{j-1}{t-1}\rfloor+1\le k\le n} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k}$$
$$+ \frac{2\big((t-1)n-j+1\big)\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{0\le k\le\lfloor\frac{j-1}{t-1}\rfloor} \frac{1}{(t-1)(n-k)+1}\binom{tk}{k}\binom{t(n-k)}{n-k} - 1\,. \tag{8}$$

□

In the following we will obtain an asymptotic equivalent of this expression. First we want to evaluate the sum

$$\frac{2(j+1)\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{\lfloor\frac{j-1}{t-1}\rfloor+1\le k\le n} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k}$$

asymptotically for a fixed ratio $\rho = \frac{j}{(t-1)n}$. With Stirling's asymptotic formula for the binomial coefficients

$$\binom{tn}{n} = \frac{\sqrt{t}\left(\frac{t^t}{(t-1)^{t-1}}\right)^n}{\sqrt{t-1}\sqrt{2}\sqrt{\pi}} n^{-\frac{1}{2}} \left(1+\mathcal{O}\Big(\frac{1}{n}\Big)\right)$$

we obtain the following expansion for the summands from $k=\rho n$ to $n-1$;

$$\frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k}$$
$$= \frac{t}{2\pi(t-1)}\left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1}\frac{1}{\sqrt{k}\sqrt{n-k}}\left(1+\mathcal{O}\Big(\frac{1}{k}\Big)\right)\left(1+\mathcal{O}\Big(\frac{1}{n-k}\Big)\right)$$
$$= \frac{t}{2\pi(t-1)}\left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1}\frac{1}{\sqrt{k}\sqrt{n-k}} + \frac{t}{2\pi(t-1)}\left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1}\frac{c_t}{\sqrt{k}(n-k)^{\frac{3}{2}}}.$$

In the last expression, the constant c_t is independent from k and n.

Therefore we get

$$\frac{2(j+1)\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{\lfloor\frac{j-1}{t-1}\rfloor+1\le k\le n} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k}$$
$$= \sum_{k=\rho n}^{n-1} \frac{t}{2\pi(t-1)}\left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1}\frac{1}{\sqrt{k}\sqrt{n-k}}$$
$$+ \sum_{k=\rho n}^{n-1} \frac{t}{2\pi(t-1)}\left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1}\frac{c_t}{\sqrt{k}(n-k)^{\frac{3}{2}}} + \mathcal{O}(1).$$

Interpreting the first sum as a Riemann sum, we obtain the following expansion;

$$\begin{aligned}
&\frac{2(j+1)\big((t-1)n+1\big)}{((t-1)n+2)\binom{tn}{n}} \sum_{k=\rho n}^{n-1} \frac{t}{2\pi(t-1)} \left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1} \frac{1}{\sqrt{k}\sqrt{n-k}} \\
&= \frac{\sqrt{2}\sqrt{t}\rho\sqrt{n}}{\sqrt{\pi}\sqrt{t-1}} \sum_{\frac{k}{n}=\rho}^{1-\frac{1}{n}} \frac{1}{n\left(\frac{k}{n}\right)^{\frac{3}{2}}\sqrt{1-\frac{k}{n}}} \left(1+\mathcal{O}\Big(\frac{1}{n}\Big)\right) = \frac{\sqrt{2}\sqrt{t}\rho\sqrt{n}}{\sqrt{\pi}\sqrt{t-1}} \int_{x=\rho}^{1} \frac{dx}{x^{\frac{3}{2}}\sqrt{1-x}} + \mathcal{O}(1) \\
&= \frac{\sqrt{2}\sqrt{t}\rho\sqrt{n}}{\sqrt{\pi}\sqrt{t-1}} \left(-\frac{2\sqrt{1-x}}{\sqrt{x}}\right)\Bigg|_{x=\rho}^{1} + \mathcal{O}(1) = \frac{2\sqrt{2}\sqrt{t}\sqrt{\rho}\sqrt{1-\rho}}{\sqrt{\pi}\sqrt{t-1}}\sqrt{n} + \mathcal{O}(1).
\end{aligned}$$

The remainder term leads to the following order estimation;

$$\begin{aligned}
&\frac{2(j+1)\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{k=\rho n}^{n-1} \frac{t}{2\pi(t-1)} \left(\frac{t^t}{(t-1)^{t-1}}\right)^n \frac{1}{(t-1)k+1} \frac{c_t}{\sqrt{k}(n-k)^{\frac{3}{2}}} \\
&= cn^{\frac{3}{2}} \sum_{k=\rho n}^{n-1} \frac{1}{k^{\frac{3}{2}}(n-k)^{\frac{3}{2}}} = c\sum_{k=\rho n}^{n-1} \left(\frac{1}{k}+\frac{1}{n-k}\right)^{\frac{3}{2}} \le 2c\sum_{k=1}^{\frac{n}{2}} \left(\frac{1}{k}+\frac{1}{n-k}\right)^{\frac{3}{2}} \le 2c\sum_{k=1}^{\frac{n}{2}} \left(\frac{1}{k}+\frac{2}{n}\right)^{\frac{3}{2}} \\
&\le 2c\sum_{k=1}^{\frac{n}{2}} \left(\frac{2}{k}\right)^{\frac{3}{2}} = \mathcal{O}(1).
\end{aligned}$$

So we have the asymptotic expansion

$$\frac{2\,(j+1)\,\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{\lfloor\frac{j-1}{t-1}\rfloor+1\le k\le n} \frac{1}{(t-1)k+1}\binom{tk}{k}\binom{t(n-k)}{n-k} = \frac{2\sqrt{2}\sqrt{t}\sqrt{\rho}\sqrt{1-\rho}}{\sqrt{\pi}\sqrt{t-1}}\sqrt{n} + \mathcal{O}(1).$$

An analogous procedure leads to

$$\begin{aligned}
&\frac{2\,\big((t-1)n-j+1\big)\,\big((t-1)n+1\big)}{\big((t-1)n+2\big)\binom{tn}{n}} \sum_{0\le k\le\lfloor\frac{j-1}{t-1}\rfloor} \frac{1}{(t-1)(n-k)+1}\binom{tk}{k}\binom{t(n-k)}{n-k} \\
&= \frac{2\sqrt{2}\sqrt{t}\sqrt{\rho}\sqrt{1-\rho}}{\sqrt{\pi}\sqrt{t-1}}\sqrt{n} + \mathcal{O}(1).
\end{aligned}$$

So we get finally the asymptotic result for the expectations of the leaf heights in t–ary trees, as it was stated in [7].

Theorem 5. *The expectation $\mathbb{E}(X_{n,j})$ of the height of the leaf j in a t–ary tree with n internal nodes is asymptotically for a fixed ratio $\rho = \frac{j}{(t-1)n}$ with $0<\rho<1$ for $n\to\infty$ given as*

$$\mathbb{E}(X_{n,j}) = \frac{4\sqrt{2}\sqrt{t}\sqrt{\rho}\sqrt{1-\rho}}{\sqrt{\pi}\sqrt{t-1}}\sqrt{n} + \mathcal{O}(1). \tag{9}$$

3.2. **Higher Moments.** We are also able to find explicit expressions for the higher factorial moments, although they are not so nice.

Differentiating $F(z,u,v)$ s-times w.r.t. v and evaluating at $v=1$ leads to

$$\frac{\partial^s}{\partial v^s}F(z,u,v)\bigg|_{v=1} = \frac{s!\big(1-\frac{1-u}{X-uY}\big)^s}{\big(\frac{1-u}{X-uY}\big)^{s+1}} = \sum_{p=1}^{s+1}(-1)^{s+1-p}s!\binom{s}{p-1}\left(\frac{X-uY}{1-u}\right)^p.$$

The coefficients from $\left(\frac{X-uY}{1-u}\right)^p$ are obtained as follows.

$$\begin{aligned}[z^n u^j]\left(\frac{X-uY}{1-u}\right)^p &= \sum_{l=0}^{p}[z^n u^j](-1)^l\binom{p}{l}\frac{(uY)^l X^{p-l}}{(1-u)^p}\\ &= T_n^{(p)}\binom{j+p-1}{p-1}+\sum_{l=1}^{p-1}(-1)^l\binom{p}{l}\sum_{0\le k\le\lfloor\frac{j-l}{t-1}\rfloor}T_k^{(l)}T_{n-k}^{(p-l)}\binom{j-l-(t-1)k+p-1}{p-1}.\end{aligned}$$

Therefore we get

$$\begin{aligned}\frac{\partial^s}{\partial v^s}F(z,u,v)\Big|_{v=1} &= \sum_{p=1}^{s+1}(-1)^{s+1-p}s!\binom{s}{p-1}\Bigg[T_n^{(p)}\binom{j+p-1}{p-1}\\ &\quad+\sum_{l=1}^{p-1}(-1)^l\binom{p}{l}\sum_{0\le k\le\lfloor\frac{j-l}{t-1}\rfloor}T_k^{(l)}T_{n-k}^{(p-l)}\binom{j-l-(t-1)k+p-1}{p-1}\Bigg].\end{aligned}$$

Hence the higher moments are given by the following theorem.

Theorem 6. *The s–th factorial moment $\mathbb{E}(X_{n,j}^{\underline{s}})$ of the height of the leaf j in a t–ary tree with n internal nodes is given by*

$$\begin{aligned}\mathbb{E}(X_{n,j}^{\underline{s}}) &= \frac{1}{T_n}\sum_{p=1}^{s+1}(-1)^{s+1-p}s!\binom{s}{p-1}\Bigg[T_n^{(p)}\binom{j+p-1}{p-1}\\ &\quad+\sum_{l=1}^{p-1}(-1)^l\binom{p}{l}\sum_{0\le k\le\lfloor\frac{j-l}{t-1}\rfloor}T_k^{(l)}T_{n-k}^{(p-l)}\binom{j-l-(t-1)k+p-1}{p-1}\Bigg].\end{aligned}\tag{10}$$

One could work out the asymptotics from this formula, at least for the variance, but the computational effort would be considerable, because of the alternating sign in the inner sum. Since according formulæ are already stated in [3], we refrain from doing that.

References

[1] M. Fisz. *Probability Theory and Mathematical Statistics.* John Wiley, 1963.

[2] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics (Second Edition).* Addison Wesley, 1994.

[3] W. Gutjahr. The variance of level numbers in certain families of trees. *Random Structures and Algorithms*, 3:361–374, 1992.

[4] W. Gutjahr and G. Pflug. The asymptotic distribution of leaf heights in binary trees. *Graphs and Combinatorics*, 8:243–251, 1992.

[5] P. Kirschenhofer. On the height of leaves in binary trees. *Journal of Combinatorics, Information and System Sciences*, 8:44 –60, 1983.

[6] P. Kirschenhofer. Some new results on the average height of binary trees. *Ars Combinatoria*, 16 A:255–260, 1983.

[7] P. Kirschenhofer. Asymptotische Untersuchungen zur durchschnittlichen Gestalt gewisser Graphenklassen. In E. Hlawka, editor, *Zahlentheoretische Analysis*, volume 1114 of *Lecture Notes in Mathematics*, pages 40–54, 1985.

[8] A. Panholzer and H. Prodinger. Descendants and ascendants in binary trees. *Discrete Mathematics and Theoretical Computer Science*, 1:247–266, 1997.

[9] M. Petkovšek, H. Wilf, and D. Zeilberger. $A = B$. A.K. Peters, Ltd., 1996.

Institut für Algebra und Diskrete Mathematik, Technical University of Vienna, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria.

E-mail address: e9125354@fbma.tuwien.ac.at (Alois Panholzer)

E-mail address: Helmut.Prodinger@tuwien.ac.at

WWW-address: http://info.tuwien.ac.at/theoinf/proding.htm