

A RECURRENCE RELATED TO TREES

DONALD E. KNUTH AND BORIS PITTEL

(Communicated by William J. Davis)

ABSTRACT. The asymptotic behavior of the solutions to an interesting class of recurrence relations, which arise in the study of trees and random graphs, is derived by making uniform estimates on the elements of a basis of the solution space. We also investigate a family of polynomials with integer coefficients, which may be called the "tree polynomials."

There are $n^{n-2}(n-1)!$ sequences of edges between vertices

$$(0.1) \quad u_1 - v_1, \dots, u_{n-1} - v_{n-1}, \quad 1 \leq u_k < v_k \leq n,$$

that define a free tree on $\{1, \dots, n\}$, because there are n^{n-2} free trees on n labeled vertices and every such tree has $n-1$ edges. If we consider each of these $n^{n-2}(n-1)!$ sequences to be equally likely, the probability that u_{n-1} and v_{n-1} belong respectively to components of sizes k and $n-k$ based on the first $n-2$ edges is

$$(0.2) \quad p_{nk} = \frac{1}{2(n-1)} \binom{n}{k} \left(\frac{k}{n}\right)^{k-1} \left(\frac{n-k}{n}\right)^{n-k-1}, \quad 0 < k < n.$$

Knuth and Schönhage [9, §§9–12] considered tree-construction algorithms whose analysis depended on the solution of the recurrence

$$(0.3) \quad x_n = c_n + \sum_{0 < k < n} p_{nk}(x_k + x_{n-k})$$

for various sequences $\langle c_n \rangle$. The purpose of the present note is to extend the results of [9] and to consider related sequences of functions whose exact and asymptotic values arise in a variety of algorithms.

Much of the analysis below, as in [9], depends on properties of the formal power series

$$(0.4) \quad T(z) = \sum_{n \geq 1} \frac{n^{n-1} z^n}{n!} = z + z^2 + \frac{3}{2} z^3 + \frac{8}{3} z^4 + \frac{125}{24} z^5 + \dots,$$

Received by the editors March 18, 1988.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 41A60, 12E10; Secondary 05A15, 68R99, 05C80.

This research was supported in part by the National Science Foundation under grant CCR-86-10181, and by Office of Naval Research contract N00014-87-K-0502.

which is the exponential generating function for labeled, rooted trees. §1 discusses the connection between $T(z)$ and functions that generate the sequences $\langle c_n \rangle$, $\langle x_n \rangle$ of (0.3). §2 derives fundamental properties of the *tree polynomials* $t_n(y)$, defined by

$$(0.5) \quad \frac{1}{(1 - T(z))^y} = \sum_{n \geq 0} t_n(y) \frac{z^n}{n!},$$

for fixed y as $n \rightarrow \infty$. Asymptotic properties of the tree polynomials and related quantities are derived in §3. These results are applied in §4 to derive asymptotic properties of $\langle x_n \rangle$ when the asymptotic behavior of $\langle c_n \rangle$ is given.

We prove in particular that if $c_n = n^\beta$ and $\beta > 1/2$ then

$$(0.6) \quad x_n = \frac{\Gamma(\beta - 1/2)}{\sqrt{2}\Gamma(\beta)} n^{\beta+1/2} + O(n^\beta) + O(n \log n).$$

On the other hand, if $c_n = O(n^\beta)$ and $\beta < 1/2$ then

$$(0.7) \quad x_n = n\kappa + O(n^{\beta+1/2}) + O(1),$$

where

$$(0.8) \quad \kappa = \sum_{m \geq 1} c_m m^{-2} \left(1 + \frac{m^m e^{-m}}{m!} (m - Q(m)) \right)$$

and $Q(m)$ is Ramanujan's function. Thus x_n is superlinear in n or linear in n dependent upon whether $\beta > 1/2$ or $\beta < 1/2$.

1. BASIC SOLUTIONS

If $\langle x_n^{(m)} \rangle$ is the solution of (0.3) in the special case $c_n = \delta_{mn}$, the general solution is

$$(1.1) \quad x_n = x_n^{(1)} c_1 + x_n^{(2)} c_2 + \dots + x_n^{(n)} c_n.$$

It is easy to verify that $x_n^{(1)} = n$ and that $x_n^{(1)} + x_n^{(2)} + \dots = 2n - 1$. Knuth and Schönhage [9] proved that

$$(1.2) \quad x_n^{(2)} = \frac{n}{4} \left(1 + \left(1 - \frac{2}{n} \right)^{n-2} \right)$$

and, for general m ,

$$(1.3) \quad x_n^{(m)} = \frac{n}{m^2} \left(m \binom{n}{m} \left(1 - \frac{m}{n} \right)^{n-m} \left(\frac{m}{n} \right)^m + 1 - \sum_{j=0}^{m-1} \binom{n-1}{j} \left(1 - \frac{m}{n} \right)^{n-1-j} \left(\frac{m}{n} \right)^j \right).$$

Notice that $x_n^{(m)} = 0$ when $m > n$, and $x_n^{(n)} = 1$.

A new family of special solutions to the recurrence also proves to be important, namely

$$(1.4) \quad x_n(y) = t_n(y)/n^{n-1}$$

where $t_n(y)$ is the tree polynomial of (0.5). We can find the corresponding leading terms $c_n(y)$ by plugging into (0.3):

$$(1.5) \quad \frac{t_n(y)}{n^{n-1}} = c_n(y) + \frac{1}{(n-1)n^{n-2}} \sum_{k=1}^{n-1} \binom{n}{k} k^{k-1} t_{n-k}(y).$$

(We have used the fact that $p_{nk} = p_{n(n-k)}$.) This sum is n^{n-1} less than the coefficient of $z^n/n!$ in $T(z)/(1-T(z))^y = 1/(1-T(z))^y - 1/(1-T(z))^{y-1}$, namely $t_n(y) - t_n(y-1)$. Therefore we have

$$(1.6) \quad c_n(y) = \frac{nt_n(y-1) - t_n(y) + n^n}{(n-1)n^{n-1}}, \quad \text{for } n \geq 2;$$

and $c_1(y) = t_1(y) = y$.

The value of $c_n(y)$ can also be expressed in another form, because of an interesting identity satisfied by the tree polynomials. The well-known equation

$$(1.7) \quad T(z) = ze^{T(z)},$$

due to Eisenstein [1], implies that

$$(1.8) \quad T'(z) = \frac{T(z)}{z(1-T(z))}.$$

Therefore

$$\sum_{n \geq 0} \frac{nt_n(y-1)z^n}{n!} = z \frac{d}{dz} \frac{1}{(1-T(z))^{y-1}} = \frac{(y-1)T(z)}{(1-T(z))^{y+1}}$$

and we have the difference equation

$$(1.9) \quad t_n(y+1) = t_n(y) + n \frac{t_n(y-1)}{y-1}.$$

It follows that

$$(1.10) \quad c_n(y) = \frac{(y-1)t_n(y+1) - yt_n(y) + n^n}{(n-1)n^{n-1}}.$$

2. TREE POLYNOMIALS

The first few cases of the polynomials $t_n(y)$ defined in (0.5) are

$$(2.1) \quad \begin{aligned} t_0(y) &= 1, \\ t_1(y) &= y, \\ t_2(y) &= y^2 + 3y, \\ t_3(y) &= y^3 + 9y^2 + 17y, \\ t_4(y) &= y^4 + 18y^3 + 95y^2 + 142y. \end{aligned}$$

Our derivation of (1.6) from (1.5) was, in essence, a proof of the difference equation

$$(2.2) \quad t_n(y) - t_n(y - 1) = \sum_{k=1}^n \binom{n}{k} k^{k-1} t_{n-k}(y).$$

Therefore, by (1.9), we have

$$(2.3) \quad t_n(y - 2) = \frac{y - 2}{n} \sum_{k=1}^n \binom{n}{k} k^{k-1} t_{n-k}(y),$$

a recurrence by which $t_n(y)$ can be obtained when $t_{n-1}(y), \dots, t_0(y)$ are known. Two analogous formulas can also be proved:

$$(2.4) \quad t_n(y + 1) = \sum_{k=0}^n \binom{n}{k} k^k t_{n-k}(y);$$

$$(2.5) \quad t'_n(y) = \sum_{k=1}^n \binom{n}{k} k^{k-1} Q(k) t_{n-k}(y).$$

Here $Q(n)$ is Ramanujan's function

$$(2.6) \quad \begin{aligned} Q(n) &= 1 + \frac{n-1}{n} + \frac{n-1}{n} \frac{n-2}{n} + \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} + \dots \\ &= \sqrt{\frac{\pi n}{2}} - \frac{1}{3} + \frac{1}{12} \sqrt{\frac{\pi}{2n}} - \frac{4}{135n} + O(n^{-3/2}), \end{aligned}$$

which arises in the analysis of many algorithms (see [5]). Equation (2.4) follows from the well-known identity (Riordan [12])

$$(2.7) \quad \frac{1}{1 - T(z)} = \sum_{n \geq 0} \frac{n^n z^n}{n!},$$

which gives the exponential generating function for n^n , the number of all mappings of the set $\{1, \dots, n\}$ into itself. Equation (2.5) follows from

$$(2.8) \quad \ln \frac{1}{1 - T(z)} = \sum_{n \geq 1} \frac{C(n) z^n}{n!}$$

where $C(n)$ is the total number of all one-cycle mappings [12], and from a formula

$$C(n) = n^{n-1} Q(n),$$

which is due to Rényi [11]. In general, we have the formal relation

$$(2.9) \quad \begin{aligned} &a_1 T(z) + a_2 T(z)^2 + a_3 T(z)^3 + \dots \\ &= \sum_{n \geq 1} \frac{n^{n-1} z^n}{n!} \left(a_1 + 2a_2 \frac{n-1}{n} + 3a_3 \frac{n-1}{n} \frac{n-2}{n} + \dots \right), \end{aligned}$$

by Lagrange's inversion formula; equations (2.7) and (2.8) are special cases of this identity.

Equation (2.4) is also a special case of the convolution formula

$$(2.10) \quad t_n(y_1 + y_2) = \sum_{k=0}^n \binom{n}{k} t_k(y_1) t_{n-k}(y_2).$$

If $F(z)$ is any formal power series such that $\ln F(z) = \sum_{n \geq 1} b_n z^n / n!$, the coefficient of $y^m z^n / n!$ in $F(z)^y = e^{y \ln F(z)}$ is a multinomial convolution

$$(2.11) \quad \begin{aligned} & \frac{1}{m!} \sum_{\substack{k_1 + \dots + k_m = n \\ k_1, \dots, k_m > 0}} \binom{n}{k_1, \dots, k_m} b_{k_1} \dots b_{k_m} \\ &= \sum_{\substack{r_1 + r_2 + \dots = m \\ r_1 + 2r_2 + \dots = n \\ r_1, r_2, \dots \geq 0}} \frac{n!}{1!^{r_1} r_1! 2!^{r_2} r_2! \dots} b_1^{r_1} b_2^{r_2} \dots \end{aligned}$$

(This is essentially Faà di Bruno’s formula [6, exercise 1.2.5–21].) In our case each $b_k = C(k)$ is an integer, so the coefficients of the tree polynomials $t_n(y)$ are integers.

Indeed, the coefficients of $t_n(y)$ have a natural combinatorial interpretation, as noted by Flajolet and Soria [4, Example 4]: The coefficient of y^k is the number of mappings of $\{1, 2, \dots, n\}$ into itself such that the corresponding functional digraph has exactly k components. (Equivalently, the mapping has exactly k different cycles; see [7, exercise 3.1–14].) The reason is simply that the corresponding exponential generating function, according to (2.8) and well-known techniques for enumerating labeled configurations, is $\exp(y \ln(1/(1 - T(z))))$, and this is just (0.5). Stepanov [13] showed that these coefficients are asymptotically normal with mean and variance $\frac{1}{2} \ln n + O(1)$; Flajolet and Soria [4] extended this to a general result on the number of components in random labeled structures.

We wish to study the asymptotic behavior of $t_n(y)$ for fixed y as $n \rightarrow \infty$, and for this we need another formula based on (2.9): The binomial expansion $(1 - T(z))^{-y} = \sum_{k \geq 0} \binom{y+k-1}{k} T(z)^k$ and equation (2.9) imply that

$$(2.12) \quad \begin{aligned} t_n(y) &= n^{n-1} \left(\frac{y}{0!} + \frac{y(y+1)}{1!} \frac{n-1}{n} + \frac{y(y+1)(y+2)}{2!} \frac{n-1}{n} \frac{n-2}{n} + \dots \right) \\ &= \frac{n^{n-1}}{\Gamma(y)} \sum_{k \geq 0} \binom{n-1}{k} \frac{\Gamma(y+k+1)}{n^k}. \end{aligned}$$

Integral formulas for $t_n(y)$ are also of interest:

$$(2.13) \quad \begin{aligned} t_n(y) &= n^n y \int_0^\infty \sum_k \binom{n-1}{k} \binom{y+k}{k} t^k e^{-nt} dt \\ &= n^n + n^{n+1} \int_0^\infty \sum_k \binom{n-1}{k} \binom{y+k-1}{k+1} t^k e^{-nt} dt. \end{aligned}$$

3. ASYMPTOTIC LEMMAS

The terms of (2.12) increase until k is approximately \sqrt{yn} , after which they begin to decrease rather rapidly. More explicitly, when y and $\varepsilon > 0$ are fixed and $0 < k \leq n^{1/2+\varepsilon}$, we have

$$(3.1) \quad \frac{\Gamma(y+k+1)}{k!} = k^y \left(1 + \frac{y^2+y}{2k} + O(k^{-2}) \right),$$

$$(3.2) \quad \frac{(n-1)!}{(n-1-k)!n^k} = e^{-k^2/2n} \left(1 - \frac{k}{2n} - \frac{k^3}{6n^2} + O(n^{-1+6\varepsilon}) \right),$$

by Stirling's approximation. Therefore the terms of (2.12) are exponentially small when $k \geq n^{1/2+\varepsilon}$, and we can replace them with other exponentially small terms to get

$$(3.3) \quad t_n(y) = \frac{n^{n-1}}{\Gamma(y)} \left(S_{2n}(y) + \frac{y^2+y}{2} S_{2n}(y-1) - \frac{1}{2n} S_{2n}(y+1) - \frac{1}{6n^2} S_{2n}(y+3) + R_n(y) \right),$$

where

$$(3.4) \quad S_n(y) = \sum_{k \geq 1} k^y e^{-k^2/n}$$

and where $R_n(y)$ is an error term to be determined below.

When y is a nonnegative integer, the sum $S_n(y)$ is essentially a theta function whose asymptotic behavior is well known. When $y = -1$, the analysis of bubble sort in Knuth [8, exercise 5.2.2-4] shows that $S_n(-1) = \frac{1}{2} \ln n + \frac{1}{2} \gamma + O(n^{-1})$. For our purposes we want to consider $S_n(y)$ when y is a real number larger than -1 . It turns out that $S_n(y) = \int_0^\infty x^y e^{-x^2/n} dx + O(1)$ in this case, and we can continue the asymptotic series explicitly:

Lemma 1. *If $y > -1$, the series $S_n(y)$ defined in (3.4) satisfies*

$$(3.5) \quad S_n(y) = \frac{1}{2} \Gamma\left(\frac{y+1}{2}\right) n^{(y+1)/2} + \sum_{j=0}^m \frac{(-1)^j \zeta(-y-2j)}{j! n^j} + O(n^{-m-1}),$$

for all $m \geq 0$, as $n \rightarrow \infty$.

Proof. Consider the related sum

$$(3.6) \quad \begin{aligned} \bar{S}_{n,N} &= \sum_{k=1}^{N\sqrt{n}-1} k^y \left(e^{-k^2/n} - \frac{1}{0!} + \frac{k^2}{1!n} - \dots + \frac{(-1)^{m+1} k^{2m}}{m! n^m} \right) \\ &= n^{y/2} \sum_{k=1}^{N\sqrt{n}-1} g_m\left(\frac{k}{\sqrt{n}}\right), \end{aligned}$$

where

$$(3.7) \quad g_m(x) = x^y \left(e^{-x^2} - \sum_{j=0}^m \frac{(-x^2)^j}{j!} \right).$$

Euler's summation formula tells us that

$$(3.8) \quad \bar{S}_{n,N} = n^{(y+1)/2} \int_0^N g_m(x) dx + \sum_{l=1}^p \frac{B_l}{l!} g_m^{(l-1)}(N) n^{(y+1-l)/2} + R_p,$$

where

$$(3.9) \quad R_p = \frac{(-1)^{p+1}}{p!} n^{(y+1-p)/2} \int_0^N B_p(\{x/\sqrt{n}\}) g_m^{(p)}(x) dx,$$

if p is chosen small enough that $\lim_{x \rightarrow 0} g_m^{(p-1)}(x) = 0$. The latter condition is satisfied if $p < y + 2m + 3$, since $g_m(x) = O(x^{y+2m+2})$ as $x \rightarrow 0$. If we choose p so that $y + 2m + 1 < p < y + 2m + 3$, we can in fact prove that the integral $\int_0^\infty |g_m^{(p)}(x)| dx$ is finite, because $g_m^{(p)}(x)$ is $O(x^{y+2m+2-p})$ as $x \rightarrow 0$ and its terms that do not involve e^{-x^2} are $O(x^{y+2m-p})$ as $x \rightarrow \infty$. Therefore $R_p = O(n^{(y+1-p)/2}) = O(n^{-m})$ for this choice of p , uniformly in N .

Euler's summation formula also implies the identity

$$(3.10) \quad \sum_{k=1}^{M-1} k^\alpha = \zeta(-\alpha) + \sum_{l=0}^p \binom{\alpha+1}{l} B_l \frac{M^{\alpha+1-l}}{\alpha+1} + O(M^{\alpha-p});$$

indeed, this is the traditional way to define the zeta function. Therefore if

$$(3.11) \quad h_m(x) = x^y \sum_{j=0}^m \frac{(-x^2)^j}{j!}$$

is the "smoothing function" included in $g_m(x)$, we have

$$(3.12) \quad n^{y/2} \sum_{k=1}^{N\sqrt{n}-1} h_m\left(\frac{k}{\sqrt{n}}\right) = n^{(y+1)/2} \int_0^N h_m(x) dx + \sum_{j=0}^m \frac{(-1)^j \zeta(-y-2j)}{j! n^j} + \sum_{l=1}^p \frac{B_l}{l!} h_m^{(l-1)}(N) n^{(y+1-l)/2} + O(n^{-m}).$$

Adding (3.8) to (3.12) and letting $N \rightarrow \infty$ completes the proof of (3.5), except that the error estimate is $O(n^{-m})$ instead of $O(n^{-m-1})$. To obtain the sharper estimate, we simply replace m by $m + 1$. \square

Lemma 2. *The tree polynomial $t_n(y)$ defined in (0.5) satisfies*

$$(3.13) \quad t_n(y) = \frac{\sqrt{2\pi} n^{n-1}}{2^{y/2}} \left(\frac{n^{(y+1)/2}}{\Gamma(y/2)} + \frac{\sqrt{2} y}{3} \frac{n^{y/2}}{\Gamma((y-1)/2)} + O(n^{(y-1)/2}) + O(1) \right),$$

for fixed y as $n \rightarrow \infty$.

Proof. We may assume that $y > -1$, for if $y = -1$ the stated result is consistent with the fact that $t_n(-1) = -n^{n-1}$ and if $y < -1$ we easily have $t_n(y) = O(n^{n-1})$.

The contribution to the error term $R_n(y)$ in (3.3) that arises from $O(n^{-1+6\epsilon})$ in (3.2) is $O(n^{(y-1)/2+6\epsilon})$, by Lemma 1; this can be improved to $O(n^{(y-1)/2})$ by carrying out the expansion in (3.2) a bit further. The contribution to $R_n(y)$ from $O(k^{-2})$ in (3.1) is $O(n^{(y-1)/2})$ if $y > 1$. If $y = 1$, the $O(k^{-2})$ of (3.1) is actually zero; and if $y < 1$ the contribution to $R_n(y)$ is $O(1)$ because $S_n(y-2) < \sum_{k \geq 1} k^{y-2}$.

To complete the estimation of $t_n(y)$ from (3.3), we need only the weak form $S_n(y) = \frac{1}{2}\Gamma((y+1)/2)n^{(y+1)/2} + O(1)$ of Lemma 1. Application of the duplication formula $\Gamma(y) = 2^{y-1}\Gamma(y/2)\Gamma((y+1)/2)/\sqrt{\pi}$ completes the proof. \square

Lemma 2 can be verified, for example, in the special cases

$$(3.14) \quad t_n(1) = n^n; \quad t_n(2) = n^n(Q(n) + 1).$$

Computer calculations show that the series in Lemma 2 continues as follows:

$$(3.15) \quad t_n(y) = n^{n-1} \left(f_n(y) + \frac{y}{3}f_n(y-1) + \frac{(y-1)(y+1)(2y-3)}{36(y-2)}f_n(y-2) \right. \\ \left. + \frac{y(y+2)(10y^2-35y+24)}{1620(y-3)}f_n(y-3) + O(n^{(y-3)/2}) + O(1) \right),$$

where $f_n(y) = \sqrt{2\pi} n^{(y+1)/2} / (2^{y/2}\Gamma(y/2))$. Stronger methods than those above will be needed to penetrate the $O(1)$ term.

Philippe Flajolet has pointed out, in a personal communication to the authors, that Lemma 2 can also be deduced directly from a singularity analysis as in [3]. Indeed, we have

$$(3.16) \quad T(z) = 1 - S(\sqrt{2-2ez}),$$

where the analytic function

$$(3.17) \quad S(z) = z - \frac{1}{3}z^2 + \frac{11}{72}z^3 - \frac{43}{540}z^4 + \frac{769}{17280}z^5 - \frac{221}{8505}z^6 + \dots$$

satisfies the functional relation

$$(3.18) \quad (1 - S(z))e^{S(z)} = 1 - \frac{z^2}{2}$$

and has a Maclaurin series that converges for $|z| < \sqrt{2}$. This leads to an asymptotic expansion of the coefficients of $(1 - T(z))^{-y} = S(\sqrt{2-2ez})^{-y}$. And in fact, the resulting expansion holds for all y ; we can actually *eliminate* the term $O(1)$ from (3.13) and (3.15). (We do not, however, require this strengthening of Lemma 2 in the calculations below.)

It is interesting to observe that (1.9) implies

$$(3.19) \quad t_n(y) = n \left(\frac{t_n(y-2)}{y-2} + \frac{t_n(y-3)}{y-3} + \dots + \frac{t_n(y-m)}{y-m} \right) + t_n(y-m+1);$$

the terms on the right decrease by factors of order \sqrt{n} until order n^{n-1} is reached.

Lemma 3. *If $y > 0$, we have*

$$(3.20) \quad \frac{t'_n(y)}{t_n(y)} = \frac{1}{2} \left(\ln \frac{n}{2} + \gamma - H_{y/2-1} \right) + O(n^{-1/2}),$$

where

$$(3.21) \quad H_y = \sum_{k \geq 1} \left(\frac{1}{k} - \frac{1}{k+y} \right).$$

Proof. The calculations in the proofs of Lemmas 1 and 2 also support differentiation by y . The right-hand side of (3.20) is $f'_n(y)/f_n(y)$, because of the well-known relation

$$(3.22) \quad H_y = \frac{\Gamma'(y+1)}{\Gamma(y+1)} + \gamma.$$

(See [6, exercise 1.2.7-23].) \square

The special case $y = 1$ of Lemma 3 is of particular interest. According to (0.5), (2.7), and (2.8), we have

$$(3.23) \quad \begin{aligned} t'_n(1) &= n! [z^n] \frac{1}{1-T(z)} \ln \frac{1}{1-T(z)} \\ &= \sum_{k=1}^n \binom{n}{k} k^{k-1} Q(k) (n-k)^{n-k}. \end{aligned}$$

By our combinatorial interpretation of $t_n(y)$, this is the total number of cycles among all mappings of $\{1, 2, \dots, n\}$ into itself. Lemma 3 and equation (3.14) imply that

$$(3.24) \quad t'_n(1) = \frac{1}{2} n^n (\ln n + \gamma + \ln 2 + O(n^{-1/2})),$$

a formula first obtained by Martin Kruskal [10]. The $O(n^{-1/2})$ can be refined further to $\frac{1}{6} \sqrt{2\pi/n} + O(n^{-1} \log n)$. In fact, the general formula (3.15) can be differentiated with respect to y ; the $O(1)$ remains $O(1)$.

The derivative of the formula in Lemma 1 with respect to y is sufficiently interesting to be stated explicitly. We have

$$(3.25) \quad \begin{aligned} \sum_{k=1}^{\infty} k^y \ln k e^{-k^2/n} &= \frac{1}{4} \Gamma \left(\frac{y+1}{2} \right) n^{(y+1)/2} (\ln n + \gamma + H_{(y-1)/2}) \\ &\quad + \sum_{j=0}^m \frac{(-1)^j \zeta'(-y-2j)}{j! n^j} + O(n^{-m-1}), \end{aligned}$$

for all $y > -1$.

We will also need asymptotic estimates of the basic functions $x_n^{(m)}$ of (1.3), valid for all m and n .

Lemma 4. *Let*

$$(3.26) \quad \alpha_m = \frac{1}{m^2} \left(1 + \frac{m^m e^{-m}}{m!} (m - Q(m)) \right),$$

where $Q(m)$ is Ramanujan's function (see (2.6)). The quantity $x_n^{(m)}$ defined in (1.3) satisfies

$$(3.27) \quad x_n^{(m)} = n\alpha_m + O(m^{-1/2}) + O((n-m)^{-1/2}),$$

uniformly for $1 \leq m \leq n-1$.

Proof. With Stirling's approximation, we can rewrite the main term of (1.3) as follows:

$$(3.28) \quad \binom{n}{m} \left(1 - \frac{m}{n}\right)^{n-m} \left(\frac{m}{n}\right)^m = \frac{n!}{n^n e^{-n}} \frac{m^m e^{-m}}{m!} \frac{(n-m)^{n-m} e^{-(n-m)}}{(n-m)!} \\ = \frac{m^m e^{-m}}{m!} \sqrt{\frac{n}{n-m}} \left(1 + O\left(\frac{1}{n-m}\right)\right).$$

Since

$$\sqrt{\frac{n}{n-m}} = 1 + \frac{m}{\sqrt{n-m}(\sqrt{n} + \sqrt{n-m})},$$

the right-hand side of (3.28) is

$$(3.29) \quad \frac{m^m e^{-m}}{m!} \left(1 + O\left(\frac{1}{n-m}\right) + O\left(\frac{m}{\sqrt{n(n-m)}}\right)\right).$$

The other terms of (1.3) are

$$\binom{n-1}{j} \left(1 - \frac{m}{n}\right)^{n-1-j} \left(\frac{m}{n}\right)^j \\ = \frac{m^j}{j!} \frac{n!}{n^n e^{-n}} \frac{(n-1-j)^{n-1-j}}{(n-1-j)!} \frac{e^{j+1-n}}{e^{j+1}} \left(\frac{n-m}{n-1-j}\right)^{n-1-j}.$$

If $m \leq n^{1/2}$ we have

$$\left(\frac{n-m}{n-1-j}\right)^{n-1-j} = e^{j+1-m} \left(1 + O\left(\frac{(m-1-j)^2}{n}\right)\right);$$

so the entire term is

$$\frac{m^j e^{-m}}{j!} \left(1 + O\left(\frac{(m-1-j)^2}{n}\right)\right),$$

uniformly in j and m . Summing over j gives

$$(3.30) \quad \sum_{j=0}^{m-1} \binom{n-1}{j} \left(1 - \frac{m}{n}\right)^{n-1-j} \left(\frac{m}{n}\right)^j = \frac{m^m e^{-m}}{m!} Q(m) \left(1 + O\left(\frac{m}{n}\right)\right),$$

because $\sum_{j=0}^{m-1} m^j/j! = m^m Q(m)/m!$ and

$$\begin{aligned} \sum_{j=0}^{m-1} \frac{m^j}{j!} (m-j-1)^2 &= \sum_{j=0}^{m-1} \frac{m^j}{j!} \left((m-1)^2 - (2m-3)j + j(j-1) \right) \\ &= \frac{m^m}{m!} \left((m+1)Q(m) - 2m \right). \end{aligned}$$

Hence, for $m \leq n^{1/2}$ we have (see (3.28), (3.29))

$$\begin{aligned} x_n^{(m)} &= \frac{n}{m^2} \left(m \frac{m^m e^{-m}}{m!} \left(1 + O\left(\frac{m}{n}\right) \right) + 1 - \frac{m^m e^{-m}}{m!} Q(m) \left(1 + O\left(\frac{m}{n}\right) \right) \right) \\ &= n\alpha_m + O(m^{-1/2}). \end{aligned}$$

We now consider the case $m > n^{1/2}$. The sum in (3.30) is the probability P_{nm} that the total number of successes in $n - 1$ Bernoulli trials does not exceed $m - 1$, where the probability of success in each trial equals $p = m/n$. By the Berry-Essén theorem (Feller [2]), we can write then

$$(3.31) \quad P_{nm} = \Phi \left(- \left(\frac{n-m}{m(n-1)} \right)^{1/2} \right) + O \left(\frac{\rho}{\sigma^3 \sqrt{n}} \right);$$

here $\Phi(\cdot)$ is the standard Gaussian distribution, and

$$\begin{aligned} \sigma^2 &= pq = \frac{m}{n} \left(1 - \frac{m}{n} \right), \\ \rho &= pq(p^2 + q^2) \leq pq. \end{aligned}$$

Since $\Phi(0) = 1/2$, the relation (3.31) leads to

$$(3.32) \quad P_{nm} = 1/2 + O((m(1 - m/n))^{-1/2}).$$

Applying the same theorem to the Poisson distribution, or using (2.6), we also obtain

$$(3.33) \quad \sum_{j=0}^{m-1} \frac{m^j e^{-m}}{j!} = 1/2 + O(m^{-1/2}).$$

According to (1.3), (3.29), (3.32), and (3.33), we can write then for $m > n^{1/2}$,

$$\begin{aligned} x_n^{(m)} &= \frac{n}{m^2} \left(m \frac{m^m e^{-m}}{m!} \left(1 + O \left(\frac{m}{n^{1/2}(n-m)^{1/2}} \right) \right) + 1 \right. \\ &\quad \left. - \sum_{j=0}^{m-1} \frac{m^j e^{-m}}{j!} + O \left(\frac{n^{1/2}}{m^{1/2}(n-m)^{1/2}} \right) \right) \\ &= n\alpha_m + O \left(\frac{n^{1/2}}{m^{1/2}(n-m)^{1/2}} \right). \end{aligned}$$

By considering two subcases $m < n/2$ and $n/2 \leq m < n$ separately, we immediately get (3.27).

The proof of Lemma 4 is now complete. \square

4. THE RECURRENCE WITH POLYNOMIAL GROWTH

We are now ready to return to the recurrence (0.3), and to derive the asymptotic behavior of x_n when c_n is a power of n .

Theorem 1. *Let $c_n = n^\beta$ for all $n > 0$, and let*

$$(4.1) \quad l(\beta) = \frac{\Gamma(\beta - 1/2)}{\sqrt{2}\Gamma(\beta)}.$$

Then the solution to (0.3) satisfies

$$(4.2) \quad x_n = \begin{cases} l(\beta) n^{\beta+1/2} + O(n^\beta), & \text{if } \beta > 1; \\ l(\beta) n^{3/2} + O(n \log n), & \text{if } \beta = 1; \\ l(\beta) n^{\beta+1/2} + O(n), & \text{if } \frac{1}{2} < \beta < 1; \\ (2\pi)^{-1/2} n \ln n + O(n), & \text{if } \beta = \frac{1}{2}; \\ O(n), & \text{if } \beta < \frac{1}{2}. \end{cases}$$

Proof. The special cases $\beta = 1$ and $\beta = \frac{1}{2}$ have already been considered by Knuth and Schönhage [9, (12.7) and (12.9)], who showed in fact that the recurrence with $c_n = n$ has a closed form solution

$$(4.3) \quad x_n = (n - 1)Q(n) + nK(n).$$

Here $K(n)$ is Kruskal's function from (3.24),

$$(4.4) \quad K(n) = \frac{t'_n(1)}{t_n(1)} = 1 + \frac{1}{2} \frac{n-1}{n} + \frac{1}{3} \frac{n-1}{n} \frac{n-2}{n} + \dots$$

Another closed form that works when $\beta = \frac{1}{2}$ is the pair of sequences

$$(4.5) \quad c_n = \frac{n}{n-1}(Q(n) - 1), \quad x_n = nK(n) - Q(n), \quad n \geq 2;$$

in this case we let $c_1 = x_1 = 0$.

For other values of β we use the sequences $c_n(y)$ and $x_n(y)$ of (1.5) and (1.10). Lemma 2 tells us that

$$(4.6) \quad c_n(y) = \frac{(y-1)t_n(y+1) - yt_n(y) + n^n}{(n-1)n^{n-1}} = \frac{\sqrt{4\pi} n^{y/2}}{2^{y/2}\Gamma((y-1)/2)} + O(n^{(y-1)/2}) + 1 + O(n^{-1});$$

$$(4.7) \quad x_n = \frac{t_n(y)}{n^{n-1}} = \frac{\sqrt{2\pi} n^{(y+1)/2}}{2^{y/2}\Gamma(y/2)} + O(n^{y/2}) + O(1).$$

Let $y = 2\beta$. If $\beta \leq 0$, so that $c_n = n^\beta = O(1)$, we clearly have $x_n = O(n)$, hence we may assume that $\beta > 0$. Therefore $y > 0$, and the terms $O(n^{-1})$ and $O(1)$ in (4.6) and (4.7) are irrelevant for our present purposes.

The special sequence $c_n = n^\beta$ is equal to $2^{y/2}\Gamma((y-1)/2)c_n(y)/\sqrt{4\pi} + c'_n$, where $c'_n = 1 + O(n^{\beta-1/2})$. Thus $x_n = 2^{y/2}\Gamma((y-1)/2)x_n(y)/\sqrt{4\pi} + x'_n = l(\beta)n^{\beta+1/2} + x'_n + O(n^\beta)$. If $0 < \beta < 1/2$, we have $x_n = O(n)$ because $x'_n = O(n)$. If $\beta > 1/2$, the estimates in (4.2) for $\beta > \frac{1}{2}$ now follow by induction on the integer $\lfloor 2\beta - 1 \rfloor$. \square

The asymptotic growth of x_n when $c_n = n^\beta \ln n$ can be obtained by differentiating the results of Theorem 1 with respect to β .

To complete our study, we would like to investigate the case $\beta < 1/2$ in greater detail, by showing that x_n/n often approaches a constant κ (depending on $\langle c_n \rangle$), and by estimating the difference $x_n - \kappa n$ for large n .

If $\langle c_n \rangle$ is any sequence such that $\sum_{m \geq 1} c_m/m^{3/2}$ converges absolutely, let us define

$$(4.8) \quad \kappa \langle c_n \rangle = \sum_{m \geq 1} c_m \alpha_m,$$

where α_m is the quantity defined in Lemma 4, equation (3.26).

Theorem 2. *If $\beta < \frac{1}{2}$ and $c_n = O(n^\beta)$, the solution x_n to (0.3) satisfies*

$$(4.9) \quad x_n = n\kappa \langle c_n \rangle + O(n^{\beta+1/2}).$$

Proof. We have $x_n = \sum_{m \geq 1} x_n^{(m)} c_m$, so we need to apply the estimates of Lemma 4 to the sum $\sum_{m \geq 1} \varepsilon_n^{(m)} c_m$ where

$$(4.10) \quad \varepsilon_n^{(m)} = x_n^{(m)} - n\alpha_m.$$

The sum splits into four parts, of which the first and second are sufficiently small because of (3.27):

$$\begin{aligned} \sum_{m=1}^{n/2} \varepsilon_n^{(m)} c_m &= \sum_{m=1}^{n/2} O(m^{\beta-1/2}) = O(n^{\beta+1/2}); \\ \sum_{m=n/2}^{n-1} \varepsilon_n^{(m)} c_m &= \sum_{m=n/2}^{n-1} O((n-m)^{-1/2} n^\beta) \\ &= \sum_{m=1}^{n/2} O(m^{1/2} n^\beta) = O(n^{\beta+1/2}); \\ \varepsilon_n^{(n)} c_n &= (1 - n\alpha_n) c_n = O(n^\beta) \\ \sum_{m>n} \varepsilon_n^{(m)} c_m &= \sum_{m>n} (-n)\alpha_m c_m = O(n^{\beta+1/2}). \quad \square \end{aligned}$$

The quantity $\kappa \langle c_n \rangle$ can, incidentally, be written as an integral instead of as a sum:

$$(4.11) \quad \kappa \langle c_n \rangle = c_1 + \int_0^{e^{-1}} \frac{c(w) dw}{w T(w)} = c_1 + \int_0^1 u^{-2} C(ue^{-u})(1-u) du,$$

where

$$(4.12) \quad C(z) = \sum_{n \geq 2} \frac{(n-1)n^{n-1}c_n}{n!} z^n.$$

Some sort of smoothness condition is necessary for the validity of (4.9); we cannot conclude that x_n/n approaches a limit if we know only that $\sum_{m \geq 1} c_m/m^{3/2}$ exists. For example, we might have

$$(4.13) \quad c_n = \begin{cases} 0, & \text{if } n \text{ is not a power of } 2; \\ 3^k/k^2, & \text{if } n = 2^k. \end{cases}$$

Then $x_n = n^{3/2}/(\lg n)^2 + O(n)$ when $n = 2^k$, but $x_n = O(n)$ when $n = 2^k - 1$.

It is, however, possible to prove that

$$(4.14) \quad \liminf_{n \rightarrow \infty} \frac{x_n}{n} = \kappa \langle c_n \rangle,$$

if $\sum_{m \geq 1} c_m/m^{3/2}$ exists and if all c_n are nonnegative. For in this case we can define $c'_n = c_n$ for $n \leq N$ and $c'_n = 0$ for $n > N$; then $x_n/n \geq x'_n/n \rightarrow \sum_{m=1}^N \alpha_m c_m$ for all N , and we can let $N \rightarrow \infty$ to obtain

$$\liminf_{n \rightarrow \infty} \frac{x_n}{n} \geq \kappa \langle c_n \rangle.$$

To prove that also

$$\liminf_{n \rightarrow \infty} \frac{x_n}{n} \leq \kappa \langle c_n \rangle,$$

we argue as follows. Set $c'_1 = 0$, and $c'_n = c_n$ for $n > 1$. Then $x_n = nc_1 + x'_n$. Introducing

$$G(z) = \sum_{n \geq 1} \frac{n^{n-1}x'_n}{n!} z^n$$

we have (see (4.12) and [9, equation (11.10)])

$$G(z) = \frac{T(z)}{1-T(z)} \int_0^z \frac{C(w)dw}{wT(w)}.$$

Since $1 - T(z) \sim 2^{1/2}(1 - ez)^{1/2}$ as $z \uparrow e^{-1}$, we have then

$$\lim_{z \uparrow e^{-1}} 2^{1/2}(1 - ez)^{1/2} G(z) = \int_0^{e^{-1}} \frac{C(w)dw}{wT(w)} = \kappa \langle c_n \rangle - c_1.$$

By the Tauberian theorem for power series (see [2], for instance), we can write subsequently

$$\sum_{j=1}^n \frac{j^{j-1}x'_j e^{-j}}{j!} = (1 + o(1))(2\pi)^{-1/2} 2n^{1/2}(\kappa \langle c_n \rangle - c_1).$$

According to Stirling's formula for factorials this leads to

$$\liminf_{n \rightarrow \infty} \frac{x'_n}{n} \leq \kappa \langle c_n \rangle - c_1.$$

Given a sequence c_n such that $\sum_{m \geq 1} |c_m|/m^{3/2}$ exists, we can define the residual sequence $r_n = x_n - n\kappa\langle c_n \rangle$. Residuals act linearly on $\langle c_n \rangle$, hence we can obtain further information about their asymptotic value by taking linear combinations of special sequences.

The residual of the constant sequence $\langle c_n \rangle$ with $c_n = 1$ is -1 , because $x_n = 2n - 1$. It follows from (4.6) that the residual of the sequence $\langle c_n(2\beta) \rangle$ is $l(\beta)n^{1/2}c_n(2\beta) + O(n^\beta) + O(1)$, when $\beta < \frac{1}{2}$. Therefore if $c_n = n^\beta$ we can put the estimates of Theorem 1 and Theorem 2 into a more precise form:

$$(4.15) \quad x_n = n\kappa\langle n^\beta \rangle + l(\beta)n^{\beta+1/2} + O(n^\beta) + O(1), \quad \text{if } \beta < \frac{1}{2}.$$

ACKNOWLEDGMENT

We wish to thank Philippe Flajolet for his penetrating remarks on a draft of this paper.

REFERENCES

1. Gotthold Eisenstein, *Entwicklung von $\alpha^{\alpha^{\alpha^{\dots}}}$* , J. Reine Angew. Math. **28** (1844), 49–52.
2. William Feller, *An introduction to probability theory and its applications*, Vol. 2, Wiley, New York, 1971.
3. Philippe Flajolet and Andrew Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math. (to appear).
4. Philippe Flajolet and Michèle Soria, *Gaussian limiting distributions for the number of components in combinatorial structures*, J. Combin. Theory Ser. A (to appear).
5. Donald E. Knuth, *An analysis of optimum caching*, J. Algorithms **6** (1985), 181–199.
6. —, *The art of computer programming*, Vol. 1: *Fundamental algorithms*, Addison-Wesley, Reading, Mass., 1973.
7. —, *The art of computer programming*, Vol. 2: *Seminumerical algorithms*, Addison-Wesley, Reading, Mass., 1981.
8. —, *The art of computer programming*, Vol. 3: *Sorting and searching*, Addison-Wesley, Reading, Mass., 1973.
9. Donald E. Knuth and Arnold Schönhage, *The expected linearity of a simple equivalence algorithm*, Theoret. Comput. Sci. **6** (1978), 281–315.
10. Martin Kruskal, *The expected number of components under a random mapping function*, Amer. Math. Monthly **61** (1954), 392–397.
11. Alfred Rényi, *Some remarks on the theory of trees*, Publ. Math. Inst. Hungar. Acad. Sci. **4** (1959), 73–85.
12. John Riordan, *Combinatorial identities*, Wiley, New York, 1968.
13. V. E. Stepanov, *Predel' nye raspredeleniia nekotorykh kharakteristik sluchainykh otobrazhenii*, Teor. Verofatnost. Primenen. **14** (1969), 639–653; English transl., *Limit distributions of certain characteristics of random mappings*, Theor. Probab. Appl. **14** (1969), 612–626.

DEPARTMENT OF COMPUTER SCIENCE, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305
(Current address of D. E. Knuth)

Current address (Boris Pittle): Department of Mathematics, Ohio State University, Columbus, Ohio 43210