Discrete Mathematics 12 (1975) 359–372.

THE DISTRIBUTION OF DEGREES IN A LARGE RANDOM TREE *

Robert W. ROBINSON *
University of Michigan, Ann Arbor, Mich. 48104, USA

and

Allen J. SCHWENK **
Mathematical Institute, Oxford University, Oxford, England

Received 14 August 1974

For labeled trees, Rényi showed that the probability that an arbitrary point of a random tree has degree k approaches $1/e(k-1)!$. For unlabeled trees, the answer is different because the number of ways to label a given tree depends on the order of its automorphism group. Using arguments involving combinatorial enumeration and asymptotics, we evaluate the corresponding probabilities for large unlabeled trees.

1. Introduction

Rényi [7] proved that as the number n of points increases, the probability that a point of a labeled tree of order n has degree k approaches $1/e(k-1)!$ (see also Moon [4, p. 73]). In the next section, we introduce the necessary notation for the various counting series needed to enumerate unlabeled trees with the number of points of degree k as a second parameter. In Section 3, functional relations amongst these series are derived using the methods of Otter [5] and Harary and Prins [3].

Section 4 presents recurrence relations derived from the functional relations of Section 3, and a table of numerical values for trees of order 36 or less. In Section 5 the functional relations are made to yield asymp-

* Research supported in part by Grant 73-2502 from the U.S. Air Force Office of Scientific Research.

* Present address: Department of Mathematics, The University of Newcastle, Newcastle, N.S.W. 2308, Australia.

** Present address: Department of Mathematics, Michigan State University, East Lansing, Mich. 48824, USA.

totic values following the analytic methods of Pólya [6] and building on asymptotic results of Otter [5]. We conclude with a brief account of related problems.

For completeness, we include a proof of Rényi's result for endpoints in labeled trees, namely, that the probability approaches $1/e$. It has long been known that the number of labeled trees of order n is n^{n-2}. (An account of the history of this result can be found in [4, p. 3].) Therefore the total number of points among these labeled trees is n^{n-1}.

On the other hand, the total number of endpoints may be calculated as follows. There are n ways to label the prospective endpoint, which when removed leaves $(n-1)^{n-3}$ possible labeled trees of order $n-1$. But there are now $n-1$ points to which the endpoint can be attached. The product of these three factors $n(n-1)^{n-2}$ is thus the total number of endpoints among the labeled trees of order n. Hence the probability of an endpoint is $((n-1)/n)^{n-2}$, which obviously approaches $1/e$ as n increases.

2. Counting series for trees in terms of both the total number of points and points of degree k

As in [1, 2, 3, 6], let t_n be the number of (unlabeled unrooted) trees with n points and let T_n be the corresponding number of rooted trees. The generating functions (counting series) for these are

$$(1)\qquad t(x)=\sum_{n=1}^{\infty} t_n\, x^n\ ,$$

$$(2)\qquad T(x)=\sum_{n=1}^{\infty} T_n\, x^n\ .$$

A *planted tree* is a tree rooted at an endpoint. It shall be convenient to take the rather unusual convention of not counting the plant, so that the generating function

$$(3)\qquad P(x)=\sum_{n=1}^{\infty} P_n\, x^n$$

for planted trees is identical to $T(x)$. Why then introduce $P(x)$ at all? We shall soon see the reason in the uses of the series $P(x, y)$ derived from $P(x)$.

We now refine these three counting series to include the number of points of degree k as a second parameter. We write $P^{(k)}(x, y)$ for the series in which the coefficient of $x^n y^m$ is the number of planted trees with n points other than the root, of which m have degree k; this convention effectively ignores the endpoint which is the root, even when $k = 1$. On the other hand, we write $T^{(k)}(x, y)$ and $t^{(k)}(x, y)$ for counting rooted trees and trees respectively, the coefficient of $x^n y^m$ here giving the number with n points of which m have degree k. For convenience, we shall usually suppress the superscript (k) when there is no possible confusion.

The effect of setting $y = 1$ in these series is to ignore the special status of points of degree k, leaving only the total number of points as the enumeration parameter, so that we have the equations

$$t(x, 1) = t(x) \ ,$$
(4) $$T(x, 1) = T(x) \ ,$$
$$P(x, 1) = P(x) = T(x) \ .$$

We require two more ordinary generating functions $D^{(k)}(x)$ and $d^{(k)}(x)$ whose coefficients for $n > 1$ are as follows. Let $D_n^{(k)}$ be the number of points of degree k (excluding the root in the case $k = 1$) occurring in all planted trees containing n points in addition to the root. Similarly, let $d_n^{(k)}$ be the total number of points of degree k in the trees of order n. The principal purpose of this paper is to determine the asymptotic value of the ratio $d_n^{(k)}/nt_n$ as n approaches infinity, that is, the probability that a point at random will have degree k.

From the definitions of the coefficients of the series $P(x, y)$ and $D(x)$, it follows that the partial derivative with respect to y, $P_y(x, y)$, when evaluated at $y = 1$ results precisely in $D(x)$:

(5) $$D(x) = P_y(x, 1) \ ,$$

and similarly,

(6) $$d(x) = t_y(x, 1) \ .$$

3. Functional relations for the counting series

Our object is to obtain functional equations first for $D(x)$ in terms of the known series $T(x)$, then for $d(x)$ in terms of $D(x)$ and $T(x)$. In the process we will derive equations for $P(x, y)$, $T(x, y)$ and $t(x, y)$.

We begin by presenting a functional relation satisfied by $P(x, y)$ which is analogous to Pólya's equation [6] for rooted trees (see also [1, p. 187]), viz.,

$$T(x) = x \exp\left[\sum_{i=1}^{\infty} T(x^i)/i\right] . \tag{7}$$

In the next equation, there appears an expression of the form $Z(S_n; f(x, y))$, which is the substitution of the counting series $f(x, y)$ into the cycle index $Z(S_n)$ of the symmetric group S_n. This involves replacing each variable a_i in $Z(S_n)$ by $f(x^i, y^i)$. According to the classical Pólya enumeration theorem of [6], the resulting counting series enumerates the number of functions from n indistinguishable elements into a collection of objects having $f(x, y)$ as its figure counting series (see also [2, pp. 45–48]). Now the method of proof of the equation

$$P(x, y) = x \exp\left[\sum_{i=1}^{\infty} P(x^i, y^i)/i\right] + (xy - x)\, Z(S_{k-1}; P(x, y)) \tag{8}$$

is the same as that of (7) with just two minor modifications. The first is an innocuous alteration, in that two variables are required, to carry along the number of points of degree k as well as the total number of points. The second change is the addition of the term $(xy - x)\, Z(S_{k-1}; P(x, y))$, which serves to count the point adjacent to the root as having degree k in the only case in which this occurs.

After differentiating each term of (8) with respect to y and simplifying by an appeal to (8), we substitute $y = 1$ and apply (4) and (5) to obtain

$$D(x) = T(x) \sum_{i=1}^{\infty} D(x^i) + xZ(S_{k-1}; T(x)) . \tag{9}$$

Now $T_y(x, 1)$ enumerates the points of degree k among all rooted trees with n points. Another way to count all these points of degree k is as follows. If the point of degree k is the root, then it is counted by the term $xZ(S_k; T(x))$. The remaining points of degree k may be counted by $D(x)$ provided we add a new point (which becomes the plant) adjacent to the old root. One final correction is needed, since if the old root had degree $k - 1$, then $D(x)$ counts it as having degree k. Thus, we must subtract $xZ(S_{k-1}; T(x))$ to obtain the equation

$$T_y(x, 1) = D(x) + xZ(S_k; T(x)) - xZ(S_{k-1}; T(x)) . \tag{10}$$

To write a formula for $t(x, y)$, we merely express Otter's equation for $t(x)$ (see [5] and [1, p. 189]),

$$t(x) = T(x) - \tfrac{1}{2}\, T^2(x) + \tfrac{1}{2}\, T(x^2) \tag{11}$$

in two variables:

$$t(x, y) = T(x, y) - \tfrac{1}{2}\, P^2(x, y) + \tfrac{1}{2}\, P(x^2, y^2)\ . \tag{12}$$

The last two terms of (12) involve $P(x, y)$ instead of $T(x, y)$ since a tree rooted at a line yields two planted trees when the root-line is cut in half and then each half-line is given a root at its dangling point. In the case $k = 1$, this process does not change the number of endpoints which are given cognizance in $P(x, y)$. We differentiate equation (12) with respect to y and then set $y = 1$ to get

$$d(x) = D(x) - D(x)\, T(x) + D(x^2) + xZ(S_k\,;\, T(x)) - xZ(S_{k-1}; T(x))\ , \tag{13}$$

using (6), (10), (4) and (5).

We now derive a rather different looking expression for $d(x)$, which will be useful later in asymptotic considerations. This is accomplished by solving (9) for $D(x)$ and substituting it into (13). In this way we obtain the formula

$$d(x) = T(x) \sum_{i=2}^{\infty} D(x^i) + xZ(S_k\,;\, T(x)) + D(x^2)\ . \tag{14}$$

4. Recurrence relations and numerical values

We begin by presenting recurrence relations for T_n, D_n, t_n and d_n. These are used to compile tables of numerical values of these numbers.

On taking the derivative of (7) and equating coefficients, we immediately obtain the recurrence relation

$$T_n = \frac{1}{n-1} \sum_{k=1}^{n-1} T_{n-k} \sum_{m \mid k} mT_m\ , \tag{15}$$

which holds for $n > 1$; whereas $T_1 = 1$. This result was noted by Otter [5]. Directly from (9) we see that $D_k = 1$, and for $n > k$,

$$D_n = \sum_{k=1}^{n-1} T_{n-k} \sum_{m \mid k} D_m + \text{coef. } x^{n-1} \text{ in } Z(S_{k-1}; T(x))\ . \tag{16}$$

For $k = 1$, this formula is particularly simple because $Z(S_0; T(x)) = 1$, and so for all $n > 1$, the coefficient of x^{n-1} in the last expression must

Table 1
The number of endpoints in planted and ordinary trees.

n	T_n	$D_n^{(1)}$	$D_n^{(1)}/nT_n$	t_n	$d_n^{(1)}$	$d_n^{(1)}/nt_n$
1	1	1	0.9999	1	0	0.0000
2	1	1	0.5000	1	2	0.9999
3	2	3	0.5000	1	2	0.6666
4	4	8	0.5000	2	5	0.6250
5	9	22	0.4888	3	9	0.5999
6	20	58	0.4833	6	21	0.5833
7	48	160	0.4761	11	43	0.5584
8	115	434	0.4717	23	101	0.5489
9	286	1204	0.4677	47	226	0.5342
10	719	3341	0.4646	106	556	0.5245
11	1842	9363	0.4620	235	1333	0.5156
12	4766	26308	0.4599	551	3365	0.5089
13	12486	74376	0.4582	1301	8500	0.5025
14	32973	210823	0.4567	3159	22007	0.4976
15	87811	599832	0.4553	7741	57258	0.4931
16	235381	1710803	0.4542	19320	151264	0.4893
17	634847	4891876	0.4532	48629	401761	0.4859
18	1721159	14015505	0.4523	123867	1077063	0.4830

Table 1 (continued)

n	T_n	$D_n^{(1)}$	$D_n^{(1)}/nT_n$	t_n	$d_n^{(1)}$	$d_n^{(1)}/nt_n$
19	4688676	40231632	0.4516	317955	2902599	0.4804
20	12826228	115669419	0.4509	823065	7871250	0.4781
21	35221824	333052242	0.4502	2144505	21440642	0.4760
22	97055168	960219974	0.4497	5623756	58672581	0.4742
23	268282848	2771707295	0.4491	14828074	161155616	0.4725
24	743724800	8009222197	0.4487	39299888	444240599	0.4709
25	2067174400	23166562539	0.4482	104636880	1228400524	0.4695
26	5759635456	67069287793	0.4478	279793408	3406668395	0.4682
27	16083730432	194332828851	0.4475	751065344	9472306719	0.4671
28	45007065088	563508783566	0.4471	2023442944	26402202159	0.4660
29	126186553344	1635169961530	0.4468	5469564928	73755057244	0.4649
30	354426814464	4747998793705	0.4465	14830870528	206463799289	0.4640
31	997171462144	13795078415447	0.4462	40330825728	579066749482	0.4631
32	2809934118912	40103826806692	0.4460	109972357120	1627010890978	0.4623
33	7929819299840	116648604165221	0.4457	300628836352	4579060860092	0.4615
34	22409528737792	339461739328210	0.4455	823779590144	12907454930530	0.4608
35	63411719241728	988340282199886	0.4453	2262365634560	36436809346364	0.4601
36	179655915077632	2878820009893684	0.4451	6226305875968	103000084831282	0.4595

be zero. For larger k, eq. (16) becomes increasingly more involved as the cycle index of S_{k-1} becomes more complicated.

On equating coefficients in (11) and simplifying slightly, we see that

$$t_n = T_n - \left[\binom{T_{n/2}}{2} + \sum_{k=1}^{[(n-1)/2]} T_k \, T_{n-k}\right] , \tag{17}$$

a formula essentially contained in [5].

Finally, eq. (13) similarly yields for $n > k$,

$$d_n = D_n + D_{n/2} - \sum_{k=1}^{n-1} T_k \, D_{n-k} + \text{coef. } x^{n-1} \text{ in } (Z(S_k; T(x)) - Z(S_{k-1}; T(x))) . \tag{18}$$

Of course, for $n \leq k$, the coefficient $d_n = 0$.

With the aid of a computer, these four recurrence relations provided the data in Table 1 for endpoints ($k = 1$) and $n \leq 36$ points. The table also includes numerical values for $D_n^{(1)}/nT_n$, the probability, among all planted trees with n points besides the root, that a point chosen from these is an endpoint. Similarly $d_n^{(1)}/nt_n$ is the corresponding probability for unrooted trees of order n. The convergence of these ratios, as shown in the table, is slow but suggestive. Accurate limiting values of these ratios, which turn out to be equal, will be derived in the next section with the aid of asymptotic formulas.

5. Asymptotic formulas and the limiting probability that a point has degree k

Asymptotic values of T_n and t_n were first developed by Otter [5]. He showed

$$T_n \sim \frac{b\rho^{1/2}}{2\pi^{1/2}} n^{-3/2} \rho^{-n} , \tag{19}$$

$$t_n \sim \frac{b^3\rho^{3/2}}{4\pi^{1/2}} n^{-5/2} \rho^{-n} , \tag{20}$$

where ρ is the radius of convergence of the power series $T(x)$ and b is a second constant. Otter calculated

$$\rho = 0.3383219 \;,$$

(21) $$\frac{b\rho^{1/2}}{2\pi^{1/2}} = 0.4399239 \;,$$

$$\frac{b^3\rho^{3/2}}{4\pi^{1/2}} = 0.5349485 \;.$$

Otter's derivation of (19) and (20) followed the outline of Pólya's proof in [6] of analogous results for trees in which every point has degree 1 or 4. Pólya's approach was to show that $0 < \rho < 1$, that $x = \rho$ is the only singularity of $T(x)$ on the circle of convergence $|x| = \rho$, and that ρ is in fact a branch point of order 2 for the continuation of $T(x)$. For the specific case of rooted trees Otter was able to show that $T(\rho) = 1$, so that near $x = \rho$, $T(x)$ has an expansion of the form

(22) $$T(x) = 1 - b(\rho - x)^{1/2} + c(\rho - x) + \ldots$$

which determines the constant b in (19)–(21). The details of the computation of b need not concern us; they may be found in the exposition of Otter's work by Harary and Palmer [2, Chapter 9]. However, once the value of $b > 0$ is known, the asymptotic expansion (19) follows at once from (22) by the useful lemma of Pólya presented below, along with the observation that $\Gamma(-\frac{1}{2}) = -2\pi^{1/2}$.

Lemma (Pólya [6, p. 240]). *Let the power series*

(23) $$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots$$

have the finite radius of convergence $\alpha > 0$, *with* $x = \alpha$ *the only singularity on its circle of convergence. Suppose also that* $f(x)$ *can be expanded near* $x = \alpha$ *in the form*

(24) $$f(x) = (1 - x/\alpha)^{-s} g(x) + (1 - x/\alpha)^{-t} h(x) \;,$$

where $g(x)$ *and* $h(x)$ *are analytic at* $x = \alpha$, $g(\alpha) \neq 0$, s *and* t *are real,* $s \neq 0, -1, -2, \ldots$, *and either* $t < s$ *or* $t = 0$. *Then*

(25) $$a_n \sim \frac{g(\alpha)}{\Gamma(s)}\, n^{s-1}\, \alpha^{-n} \;.$$

The same line of reasoning provides asymptotic values for D_n and d_n. From (9) and (14) it is easy to see that $D(x)$ and $d(x)$ have the same radius of convergence ρ as $T(x)$ and likewise have $x = \rho$ as the only singularity on the circle of convergence. From (22) it is clear that

$$(26)\qquad T^m(x) = 1 - mb(\rho - x)^{1/2} + \left(\binom{m}{2}b^2 + mc\right)(\rho - x) + \dots \ .$$

That is,

$$(27)\qquad (T^m)_n \sim mT_n \ .$$

Moreover, since $T(\rho) = 1$, we may conveniently express this asymptotic behaviour as

$$(28)\qquad (T^m)_n \sim T_n \frac{\mathrm{d}}{\mathrm{d}T}(T^m(x))\Big|_{x=\rho} \ .$$

Thus, since the factors of $T(\alpha^i)$ for $i > 1$ are clearly analytic at $x = \rho$, the asymptotic behaviour of the coefficients in the cycle index $Z(S_k\,;T(x))$ is given by

$$(29)\qquad Z(S_k\,;T(x)) \sim T(x)\frac{\partial}{\partial a_1} Z(S_k\,;T(\rho)) \ ,$$

where a_1 is the variable of $Z(S_k)$ which is replaced by $T(x)$. But this partial derivative is known to be $Z(S_{k-1})$ (see [2, eq. (8.5.3)]). Thus we see that near $x = \rho$,

$$(30)\qquad Z(S_k\,;T(x)) \sim T(x)\, Z(S_{k-1}\,;T(\rho)) \ .$$

Applying (25) and (30) to eq. (14), we conclude that

$$(31)\qquad d_n \sim T_n\left(\sum_{i=2}^{\infty} D(\rho^i) + \rho Z(S_{k-1}\,;T(\rho))\right) \ .$$

In view of (19) and (20), this yields

$$(32)\qquad \frac{d_n^{(k)}}{nt_n} \sim \frac{2}{b^2\rho}\left(\sum_{i=2}^{\infty} D^{(k)}(\rho^i) + \rho Z(S_{k-1}\,;T(\rho))\right) \ .$$

Regrouping the infinite sum provides a form more suitable for computation:

$$(33)\qquad \frac{d_n^{(k)}}{nt_n} \sim \frac{2}{b^2\rho}\left(\sum_{n=1}^{\infty} D_n^{(k)}\rho^{2n}/(1-\rho^n) + \rho Z(S_{k-1}\,;T(\rho))\right) \ .$$

Surprisingly, the limiting value of the probability of an endpoint given by (33) takes the interesting form of a weighted average:

$$(34)\qquad \frac{d_n^{(1)}}{nt_n} \sim \frac{\rho + \sum_{n=1}^{\infty} D_n^{(1)}\rho^{2n}/(1-\rho^n)}{1 + \sum_{n=1}^{\infty} nT_n\rho^{2n}/(1-\rho^n)} \ .$$

The evaluation of this expression by computer gives the limit

$$\frac{d_n^{(1)}}{nt_n} \sim 0.438156235664 \ldots \tag{35}$$

accurately to the indicated twelve places. As Table 1 demonstrates, the convergence of d_n/nt_n to this value is relatively slow.

To verify that $D_n/nT_n \sim d_n/nt_n$ we need an asymptotic expression for D_n. Eq. (9) can be written

$$D(x) = \frac{T(x)\,\Sigma_{i=2}^{\infty} D(x^i) + xZ(S_{k-1}; T(x))}{1 - T(x)} \tag{36}$$

in which the numerator takes the value

$$\sum_{i=2}^{\infty} D(\rho^i) + \rho Z(S_{k-1}; T(\rho))$$

at $x = \rho$. By (22) the denominator has the expansion

$$\frac{1}{1 - T(x)} = \frac{1}{b}(\rho - x)^{-1/2} + \ldots \tag{37}$$

near $x = \rho$, the remaining terms being of higher order in $(\rho - x)$. Thus

$$D_n \sim \frac{\Sigma_{i=2}^{\infty} D(\rho^i) + \rho Z(S_{k-1}; T(\rho))}{b\rho^{1/2}\pi^{1/2}}\, n^{-1/2}\rho^{-n} \tag{38}$$

by Pólya's lemma and the fact that $\Gamma(\frac{1}{2}) = \pi^{1/2}$. Combined, (38) and (19) yield

$$\frac{D_n^{(k)}}{nT_n} \sim \frac{2}{b^2\rho}\left(\sum_{i=2}^{\infty} D^{(k)}(\rho^i) + \rho Z(S_{k-1}; T(\rho))\right), \tag{39}$$

the same limit obtained in (32).

The recurrence relation (16) plus knowledge of $Z(S_{k-1}; T(x))$ permit the numerical evaluation of (33) by computer. This has been done for small k and is reported in the second column of Table 2. By (39), this is also the asymptotic probability that a point in a planted tree has degree k; it is similarly easy to find the same limit for rooted trees. For convenient comparison, the first column displays $1/e(k-1)!$, the asymptotic probability for degree k in labeled trees. The entries in the remaining columns will be described in the concluding section.

Table 2
Asymptotic probabilities for degree k in large random trees.

k	Point in a labeled tree	Point in an unlabeled tree	Similarity class in an unlabeled tree	Average size of a similarity class in an unlabeled tree
1	0.367 879	0.438 156	0.338 322	1.574 832
2	0.367 879	0.293 998	0.338 322	1.056 695
3	0.183 940	0.159 114	0.191 404	1.010 762
4	0.061 313	0.068 592	0.083 174	1.002 803
5	0.015 328	0.026 027	0.031 623	1.000 839
6	0.003 066	0.009 259	0.011 256	1.000 269
7	0.000 511	0.003 198	0.003 888	1.000 089
8	0.000 073	0.000 985 [a]	0.001 200 [a]	1.000 033
9	0.000 009	0.000 355 [a]	0.000 432 [a]	1.000 011
≥ 10	0.000 001	0.000 316 [a]	0.000 379 [a]	1.000 004

[a] These 6 entries are uncertain because estimates for $Z(S_{k-1}; T(\rho))$ had to be used for $k \geq 8$.

6. Related problems

The probability that a similarity class of points has degree k is closely related. Let $s^{(k)}(x)$ be the generating function whose nth coefficient $s_n^{(k)}$ gives the number of similarity classes whose points have degree k. Clearly, the total number of similarity classes of unrooted trees is simply T_n, for each selection of a similarity class for a given tree gives one possible rooting for that tree. Similarly, each rooted tree with root degree k corresponds to a similarity class with degree k, and so

$$(40) \qquad s^{(k)}(x) = xZ(S_k; T(x)) \ .$$

Thus, considering (30),

$$(41) \qquad s(x) \sim \rho\ T(x)\ Z(S_{k-1}; T(\rho)) \ .$$

Thus, the asymptotic probability is

$$(42) \qquad \frac{s_n}{T_n} \sim \rho Z(S_{k-1}; T(\varphi)) \ .$$

These limiting probabilities comprise the third column of Table 2. Notice that only for degree 1 is this probability smaller than the corresponding probability for a point in an unlabeled tree. The other probabilities for similarity classes are somewhat larger than those for points, but nevertheless can be shown to decrease at the same rate as k increases. This suggests that endpoints must tend to lie in larger similarity classes than points of higher degree. Indeed, the average size of a similarity class of degree k is just $d_n^{(k)}/s_n^{(k)}$ which can be evaluated by means of (31) and (42) from which we conclude that

$$(43)\qquad \frac{d_n^{(k)}}{s_n^{(k)}} \sim 1 + \frac{\Sigma_{i=2}^{\infty} D(\rho^i)}{\rho Z(S_{k-1}; T(\rho))} \;.$$

These numbers are displayed in the last column of Table 2. We observe that the size of endpoint equivalence classes is notably larger as suggested above. For comparison, we note that the average size of an arbitrary similarity class is simply T_n/nt_n which is approximated by (19) and (20) to obtain

$$(44)\qquad \frac{T_n}{nt_n} \sim \frac{b^2\rho}{2} = 1.216\ 004 \ldots \;.$$

In passing, let us mention that while producing these tables, we improved Otter's estimate of the radius of convergence to obtain

$$(45)\qquad \rho = 0.338\ 321\ 856\ 899 \ldots \;.$$

Variations of these methods can be applied to numerous problems. For example, if a large tree is chosen at random, how close to the average value of $(0.438156)n$ is its number of endpoints likely to be? This question is answered by using the second partial derivative $P_{yy}(x, 1)$ to compute the variance of the number of endpoints in each tree with n points. The asymptotics are much more complicated because the two largest asymptotic terms sum to precisely zero. The net result is that the variance of this distribution is asymptotic to $(0.19198)n$. Thus, as n increases, the distribution is becoming relatively more sharp because the standard deviation is only increasing at the rate of the square root of n. Rényi [7] found a similar result for the variance of the number of endpoints in a random labeled tree; in this case the constant factor of n is $(e - 2)e^{-2}$, which is about 0.09721. Moon [4, p. 73] presents a proof that for any $k > 1$, the variance of the distribution of points of degree k among labeled trees is asymptotic to $np(1-p) - np^2(k-2)^2$.

There is no doubt that the methods of the present paper can be easily applied to finding the distribution of degrees in various other kinds of trees, such as plane trees, achiral trees, steric trees, homeomorphically irreducible trees, and trees with a forbidden limb [8].

Acknowledgment

The authors are grateful to Professor Frank Harary who suggested the original problem and guided the revisions of early versions of this article.

References

[1] F. Harary, Graph Theory (Addison-Wesley, Reading, Mass., 1969).
[2] F. Harary and E.M. Palmer, Graphical Enumeration (Academic Press, New York, 1973).
[3] F. Harary and G. Prins, The number of homeomorphically irreducible trees, and other species, Acta Math. 101 (1959) 141–162.
[4] J.W. Moon, Counting Labelled Trees (Can. Math. Congress, Montreal, 1970).
[5] R. Otter, The number of trees, Ann. Math. 49 (1948) 583–599.
[6] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, Acta Math. 68 (1937) 145–254.
[7] A. Rényi, Some remarks on the theory of trees, Publ. Math. Inst. Hungar. Acad. Sci. 4 (1959) 73–85.
[8] A.J. Schwenk, Almost all trees are cospectral, in: F. Harary, ed., New Directions in the Theory of Graphs (Academic Press, New York, 1973) 275–308.