J. Riordan*

The Enumeration of Trees by Height and Diameter

Abstract: The enumeration of trees which was begun by Harary and Prins, is simplified and elaborated in the interest of obtaining reliable numerical results. Height is a characteristic of a rooted tree, the length in lines of the longest path from the root, while the *diameter* of a (free) tree is the length of the longest path joining two endpoints. The most general enumerations given are for the case where a fixed number of the points of the trees are labeled with distinct labels, which puts in one setting the classical contrast of all points alike and all points unlike. The numerical tables given extend to trees with 20 points, all alike, and to trees with 10 points, all unlike.

1. Introduction

The enumeration of (mathematical) trees and rooted trees has been described, along with its historical background, in my book [1], which also contains an extensive table of the corresponding numbers. This enumeration may be refined by introducing a new variable, the *diameter* for trees, and the *height* for rooted trees. These terms will be defined shortly. The refinement, for trees with all points alike, already appears in the extensive survey of tree enumerations by Harary and Prins [2]. Here, in the first place, the results of Harary and Prins are simplified and elaborated in the interest of obtaining reliable numerical answers. Next, they are extended to the case where some of the points of the trees are labeled with distinct labels, thus providing a bridge between the cases of all points alike and all points unlike as in the similar studies without regard for height or diameter reported in [1] and [3]. Finally the case of all points unlike is examined with the detail necessary for numerical answers.

The numerical tables extend to trees with 20 points for all points alike and to trees with 10 points for all points unlike.

For the convenience of the reader, a brief summary of the definitions forming the background of the paper follows. A *linear graph* is a collection of two kinds of entities, points and lines; the lines are pairs of points and taken together describe the connections of the graph. For enumeration purposes, lines in parallel, joining the same two points, and *slings*, lines joining points to themselves, are usually banned. A *tree* is a connected linear graph with no cycles. A *rooted tree* is a tree with one distinguished point, the root. A *planted tree* is a rooted tree whose root is an endpoint (incident to a single line). Two trees are *isomorphic* if there is a one-to-one correspondence of their point sets which preserves adjacency; two rooted trees are isomorphic if they are isomorphic as trees and the correspondence maps one root on the other. Points or lines carried into each other by an isomorphism are called *similar*. Two labeled trees are isomorphic if they are isomorphic with labels removed and the elements labeled (points or lines) either remain unchanged or are changed to similar elements. All enumerations are of nonisomorphic trees

A *path* joining points a_1 and a_n in a tree is a collection of lines of the form a_1a_2, a_2a_3, $\cdots$, $a_{n-1}a_n$ where the points a_1 to a_n are distinct; there is unique path between each pair of points of a tree. The *length* of a path is the number of lines it contains. The *diameter* of a tree (*spread* is an alternative term perhaps closer to nature) is the length of the longest path joining two of its (end) points. The *height* of a rooted tree is the length of the longest

*Bell Telephone Laboratories, New York, N. Y.

path joining the root and another end point [4]. Fig. 1 shows all trees with six points classified by diameter; Fig. 2 shows all rooted trees with five points classified by height.

The enumeration proceeds by finding relations for enumerating generating functions, which for brevity are called enumerators. The variables of enumerators are indeterminates or tags whose powers identify the coefficients, which are numbers of the various kinds of tree in question associated with the powers; e.g., if the association is through number of points, the generating function is said to be the enumerator by number of points.

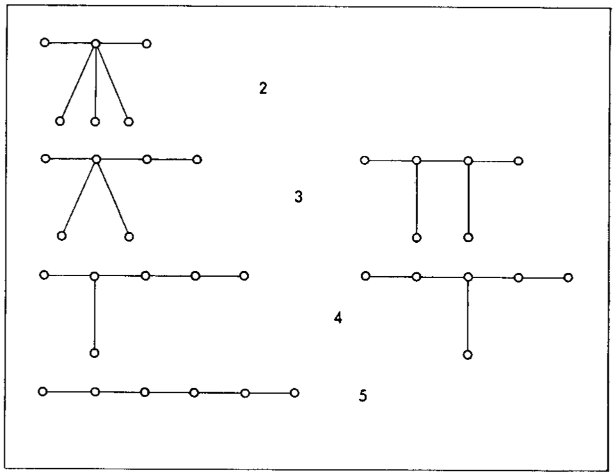

Figure 1 **Trees with six points by diameter.**

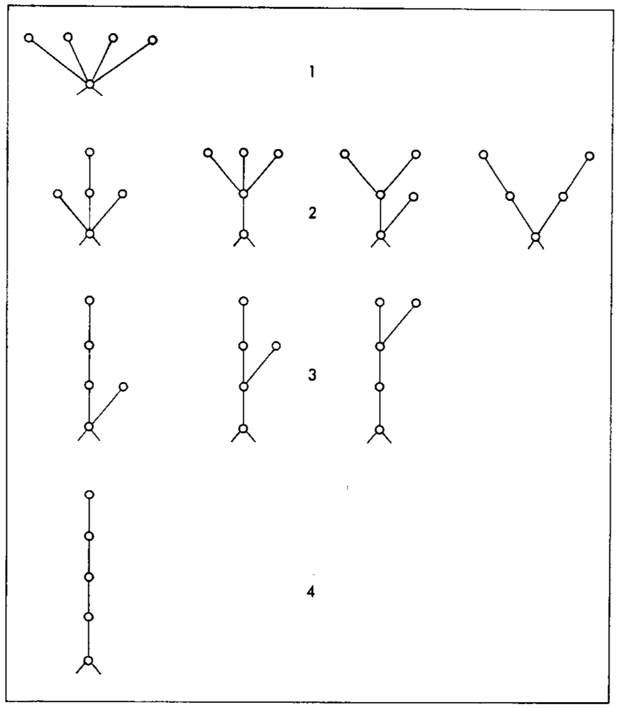

Figure 2 **Rooted trees with five points by height.**

2. Rooted trees by height

Write r_{ph} for the number of rooted trees with p points and height h, s_{ph} for the similar number with height at most h, and

$$r_h(x) = \sum r_{ph} x^p$$
$$s_h(x) = \sum s_{ph} x^p$$

for the corresponding enumerators, by number of points The rooted trees of height at most $h + 1$ with n branches at the root are a collection of n *planted* trees joined at the root; each planted tree is of height at most $h + 1$ and is in one-to-one correspondence with a rooted tree of height at most h. Hence by a procedure entirely similar to that for rooted trees with height ignored [1]

$$s_{h+1}(x) = x \exp [s_h(x) + s_h(x^2)/2 + \cdots + s_h(x^k)/k + \cdots], \tag{1}$$

a relation given in other notation by Harary and Prins [2]. This has an alternate form, similar to the Cayley form for rooted trees and similarly derivable, namely

$$s_{h+1}(x) = x/(1 - x)^{s_{1h}}(1 - x^2)^{s_{2h}} \cdots (1 - x^k)^{s_{kh}} \cdots \tag{1a}$$

The derivation of (1a) from (1) is as follows:

$$\begin{aligned} \log s_{h+1}(x) &= \log x + s_h(x) + s_h(x^2)/2 + \cdots \\ &= \log x + \sum s_{kh}(x^k + x^{2k}/2 + \cdots \\ &= \log x + \sum s_{kh} \log (1 - x^k)^{-1}. \end{aligned}$$

Equation (1a) is a natural form for computation because of its resemblance to the generating function for partitions; indeed, since $s_{p1} = r_{p1} = 1$, $p > 0$, the instance $h = 1$ of (1a) is

$$\begin{aligned} s_2(x) &= x/(1 - x)(1 - x^2) \cdots (1 - x^k) \cdots \\ &= x\pi(x) = \sum \pi_n x^{n+1}, \end{aligned}$$

with $\pi(x)$ the enumerator of partitions; then $s_{n2} = \pi_{n-1}$, $r_{n2} = \pi_{n-2} - 1$, $n > 2$, with π_n the number of partitions of n.

Also, since

$$s_{ph} = r_{p1} + r_{p2} + \cdots + r_{ph}$$
$$r_{ph} = s_{ph} - s_{p,h-1}$$

two additional forms of (1) are

$$s_{h+1}(x) = s_h(x)/(1 - x)^{r_{1h}} \cdots (1 - x^k)^{r_{kh}} \cdots$$
$$= s_h(x) \exp [r_h(x) + r_h(x^2)/2 + \cdots]. \quad (1b)$$

Note that $r_{kh} = 0, \quad k < h + 1$.

Expanding the denominator on the right of the first of (1b) leads in the first instance to

$$r_{p+1,h+1} = \sum_{i=1} r_i r_{p+1-i,h}, \qquad p < 2h + 3, \quad (2)$$

since $s_{ph} = r_p, \quad p < h + 2$, with r_p the number of rooted trees with p points. More generally, it is found that

$$r_{2h+k,h} = \sum_{i=1} r_i r_{2h+k-i,h-1} - f_k(h), \quad (3)$$

with $f_k(h) = 0,\ k < 2,\ f_2(h) = h,\ f_3(h) = h^2 + 2h - 1,\ f_4(h) = (h + 1)^3 - 2\binom{h}{3} - 6, \quad h > 1.$

It should also be noticed that (1a) may be rewritten

$$s_{h+1}(x)\sigma_{h+1}(x) = x, \quad (1c)$$

with

$$\sigma_{h+1}(x) = (1 - x)^{s_{1h}}(1 - x^2)^{s_{2h}} \cdots (1 - x^k)^{s_{kh}} \cdots$$
$$= \sum \sigma_{n,h+1} x^n.$$

Computation of the polynomials $\sigma_h(x)$ is a valuable alternative to direct use of (1a) or (1b), since the numbers s_{nh} are determinable by the system of recurrences obtained from (1c) by equating coefficients of x^n, namely

$$\sum_{i=1} s_{i,h+1}\sigma_{n-i,h+1} = \delta_{n1},$$

with δ_{n1} a Kronecker delta.

Finally, the following particular results, which may be obtained either by (2) or direct enumeration, may be noted:

$$r_{h+1,h} = 1$$

$$r_{h+2,h} = h$$

$$r_{h+3,h} = \binom{h+1}{2} + h - 1$$

$$r_{h+4,h} = \binom{h+2}{3} + 2\binom{h}{2} + 2(h - 2)$$

$$r_{h+5,h} = \binom{h+3}{4} + 3\binom{h+1}{3} + 5\binom{h-1}{2}$$
$$+ 7h - 15, \qquad h > 2. \quad (4)$$

Table 1 gives the numbers r_{ph} for $p = 2(1)20$, that is, for $p = 2, 3, \cdots, 20$, and all values of h.

3. Trees by diameter

In similar fashion to the procedure above, write t_{pd} for the number of trees with p points and diameter d, u_{pd} for the corresponding number with diameter at most d, and $t_d(x)$, $u_d(x)$ for the enumerators by number of points.

Following Harary and Prins [2], these numbers may be related to r_{pd} and s_{pd} by considering separately trees with center and trees with bicenter. For the latter, a tree with diameter $2d + 1$ clearly has a bicenter and indeed may be represented as the line containing the bicenter terminated by two like rooted trees each of height d; hence using Polya's theorem [1],

$$t_{2d+1}(x) = [r_d^2(x) + r_d(x^2)]/2. \quad (5)$$

The trees with center are of diameter $2d$ if at least

Table 1 **The number of rooted trees with p points and height h.**

h/p	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2		1	2	4	6	10	14	21	29	41	55	76	100	134	175	230	296	384	489
3			1	3	8	18	38	76	147	277	509	924	1648	2912	5088	8823	15170	25935	44042
4				1	4	13	36	93	225	528	1198	2666	5815	12517	26587	55933	116564	241151	495417
5					1	5	19	61	180	498	1323	3405	8557	21103	51248	122898	291579	685562	1599209
6						1	6	26	94	308	941	2744	7722	21166	56809	149971	390517	1005491	2564164
7							1	7	34	136	487	1615	5079	15349	45009	128899	362266	1002681	2740448
8								1	8	43	188	728	2593	8706	27961	86802	262348	776126	2256418
9									1	9	53	251	1043	3961	14102	47816	156129	494769	1530583
10										1	10	64	326	1445	5819	21858	77878	266265	880883
11											1	11	76	414	1948	8282	32695	121963	435168
12												1	12	89	516	2567	11481	47481	184903
13													1	13	103	633	3318	15564	67249
14														1	14	118	766	4218	20697
15															1	15	134	916	5285
16																1	16	151	1084
17																	1	17	169
18																		1	18
19																			1

two of the branches at the center are planted trees of height d; the remaining branches at the center are planted trees of height less than d. As usual they are enumerated by Polya's theorem. The form given by Harary and Prins [2], Eq. (8_2) is, in present notation,

$$t_{2d}(x) = s_{d-1}(x) \sum_{m=2} S_m[r_{d-1}(x), \cdots, r_{d-1}(x^m)], \qquad d > 1, \tag{6}$$

with S_m $(t_1, t_2, \cdots, t_m)$ the (multivariable) cycle index of the symmetric group. Noting (Eq. (4.3a) in [1]) that

$$1 + t_1 + \sum_{m=2} S_m(t_1, \cdots, t_m) = \exp(t_1 + t_2/2 + t_3/3 + \cdots)$$

and using the second form of (1b) leads to the simpler form

$$t_{2d}(x) = r_d(x) - r_{d-1}(x)s_{d-1}(x), \qquad d > 1. \tag{6a}$$

The initial cases for $t_{2d}(x)$ are $t_0(x) = x$, $t_2(x) = x^3/(1-x)$; note that $t_1(x) = x^2$ satisfies (5).

For the variable $u_d(x)$ note first that $u_0(x) = x$, $u_1(x) = x + x^2$, $u_2(x) = x/(1-x) = s_1(x)$. Then, first

$$\begin{aligned} u_{2d}(x) &= u_{2d-2}(x) + t_{2d-1}(x) + t_{2d}(x) \\ &= u_{2d-2}(x) + r_d(x) - [s^2_{d-1}(x) - s_{d-1}(x^2) \\ &\qquad - s^2_{d-2}(x) + s_{d-2}(x^2)]/2, \end{aligned} \tag{7}$$

and by iteration

$$u_{2d}(x) = s_d(x) - [s^2_{d-1}(x) - s_{d-1}(x^2)]/2, \tag{8}$$

which holds for all non-negative values of d if $s_{-1}(x) = 0$, since $s^2_0(x) - s_0(x^2) = 0$. Next

$$\begin{aligned} u_{2d+1}(x) &= u_{2d}(x) + t_{2d+1}(x) \\ &= s_d(x) - [s^2_{d-1}(x) - r^2_d(x) - s_d(x^2)]/2 \\ &= s_d(x) - [2s_d(x)s_{d-1}(x) \\ &\qquad - s^2_d(x) - s_d(x^2)]/2. \end{aligned} \tag{9}$$

Note that for increasing d, $u_d(x)$ approaches $t(x)$, the enumerator of trees by number of points while $s_d(x)$ approaches $r(x)$, the corresponding enumerator of rooted trees, and both (8) and (9) yield Otter's formula ([1], p. 137), namely

$$t(x) = r(x) - [r^2(x) - r(x^2)]/2.$$

Note also that both (8) and (9) may be used to check numerical results obtained from (5) and (6a).

Further numerical checks may be obtained from independent determinations of $t_{p,p-j}$, $j = 1, 2, \cdots$. These are made by adding to the diameter, a line with $p - j + 1$ points, $j - 1$ lines at all point positions which do not increase the diameter, and in all combinations which preserve the tree. There is one equivalence operation: the interchange of points k and $p - j + 2 - k$ on the diameter.

It is clear that $t_{p,p-1} = 1$, the tree consisting of $p - 1$ lines in tandem. Next

$$t_{2q,2q-2} = t_{2q+1,2q-1} = q - 1, \tag{10}$$

for the tree is formed by adding to the diameter a single line in all positions on one side of its center.

For $j = 3$, the investigation is more elaborate. At points 2 to $p - 3$ either one line or two lines in parallel may be added. At points 3 to $p - 4$ two lines in series may also be added. Hence the enumerator for additions to points 2 and $p - 3$ is $1 + x + x^2$, while for points 3 to $p - 4$ it is $(1 + x + x^2)(1 + y)$, x being the indicator of a single line, y that of two

Table 2 **The number of trees with p points and diameter d.**

d/p	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3		1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9
4			1	2	5	8	14	21	32	45	65	88	121	161	215	280	367	471
5				1	2	7	14	32	58	110	187	322	519	839	1302	2015	3032	4542
6					1	3	11	29	74	167	367	755	1515	2931	5551	10263	18677	33409
7						1	3	14	42	128	334	850	2010	4625	10201	21990	46108	94912
8							1	4	19	66	219	645	1813	4802	12265	30198	72396	169231
9								1	4	23	88	328	1065	3303	9583	26793	71986	188092
10									1	5	29	123	487	1717	5706	17843	53522	154520
11										1	5	34	156	675	2557	9124	30462	97387
12											1	6	41	204	928	3745	14109	49833
13												1	6	47	250	1223	5233	20863
14													1	7	55	313	1600	7218
15														1	7	62	374	2034
16															1	8	71	454
17																1	8	79
18																	1	9
19																		1

lines in series. The cycle index, taking into account the interchange of points on the diameter, is

$$(t_1^{2m} + t_2^m)/2, \qquad n = 2m$$

or

$$(t_1^{2m+1} + t_1 t_2^m)/2, \qquad n = 2m + 1,$$

where n is the number of positions where lines may be added, t_1 and t_2 are the indicators of cycles of lengths 1 and 2.

Thus for $p = 2q$, using Polya's theorem, the enumerator for all additions is

$$e_{2q}(x, y) = [(1 + x + x^2)^{2q-4}(1 + y)^{2q-6} + (1 + x^2 + x^4)^{q-2}(1 + y^2)^{q-3}]/2,$$

and for $p = 2q + 1$, the enumerator is

$$e_{2q+1}(x, y) = e_{2q}(x, y)(1 + x + x^2)(1 + y).$$

The numbers in question are the sums of coefficients of x^2 and y in these enumerators, which turn out to be

$$t_{2q,2q-3} = q^2 - 2q - 1$$
$$t_{2q+1,2q-2} = q^2 - q - 1. \tag{11}$$

For $j = 4$, the procedure is similar, the chief difference being occasioned by the possibility of adding either of the two rooted trees with four points at positions 4 to $p - 5$; the results are

$$t_{2q,2q-4} = (2q^3 - 6q^2 - 5q + 12)/3, \quad q > 3$$
$$t_{2q+1,2q-3} = (2q^3 - 3q^2 - 11q + 6)/3, \quad q > 3 \tag{12}$$

For $j = 5$ the results are

$$t_{2q,2q-5} = \frac{1}{2}\left[\binom{2q-3}{4} + \binom{q-2}{2}\right] + 2q^3 - 14q^2 + 24q - 10, \qquad q > 5$$

$$t_{2q+1,2q-4} = \frac{1}{2}\left[\binom{2q-2}{4} + \binom{q-1}{2}\right] + 2q^3 - 11q^2 + 12q + 1, \qquad q > 4.$$

For $j = 6$, the results are too elaborate for complete statement, but it may be noticed that, with

$$\Delta t_{p,p-6} = t_{p+1,p-5} - t_{p,p-6},$$
$$\Delta^3 t_{2q,2q-6} = -4q - 16, \qquad q > 6$$
$$\Delta^3 t_{2q+1,2q-5} = 2q^2 + 7q + 13, \qquad q > 6. \tag{13}$$

Table 2 gives the numbers t_{pd} for $p = 2(1)20$ and all values of d.

4. Point-labeled trees and rooted trees

The modifications in the enumerations occasioned by labeling points with a fixed number of distinct labels are quite similar to those employed for trees and rooted trees with diameter and height ignored. These are described in [1] and [3]; briefly, a new variable is put into the enumerators to account for the number of points labeled, and a slight modification of the theorem of Polya gives the enumerator relationships.

For the rooted trees, if s_{phm} is the number of rooted trees with p points, height at most h, and m points labeled, the modified enumerator is

$$s_h(x, y) = \sum_{p=1} x^p \sum_{m=0}^{p} s_{phm} y^m/m!$$

and the modification of (1) is

$$s_{h+1}(x, y) = x(1 + y) \exp [s_h(x, y) + s_h(x^2)/2 + \cdots + s_h(x^k)/k + \cdots] \tag{14}$$

with $s_h(x) = s_h(x, 0)$. Combining this with (1) shows that

$$s_{h+1}(x, y) \exp s_h(x) = (1 + y)s_{h+1}(x) \exp s_h(x, y). \tag{15}$$

With similar notation for point-labeled trees, the results are

$$t_{2d}(x, y) = r_d(x, y) - s_{d-1}(x, y)r_{d-1}(x, y), \qquad d > 1 \tag{16}$$

$$t_{2d+1}(x, y) = [r_d^2(x, y) + r_d(x^2, 0)]/2 \tag{17}$$

$$u_{2d}(x, y) = s_d(x, y) - [s_{d-1}^2(x, y) - s_{d-1}(x^2, 0)]/2 \tag{18}$$

$$u_{2d+1}(x, y) = s_d(x, y) - [2s_d(x, y)s_{d-1}(x, y) - s_d^2(x, y) - s_d(x^2, 0)]/2. \tag{19}$$

For concreteness and verifications note that

$$s_0(x, y) = r_0(x, y) = t_0(x, y) = x(1 + y)$$
$$s_1(x, y) = x(1 + y)e^{xy}/(1 - x) = \sum_{p=1} x^p(1 + y)(1 + y + y^2/2! + \cdots + y^{p-1}/(p - 1)!)$$
$$r_1(x, y) = x(1 + y)(e^{xy} - 1 + x)/(1 - x) = \sum_{p=2} x^p(1 + y)(1 + y + y^2/2! + \cdots + y^{p-1}/(p - 1)!)$$
$$t_1(x, y) = x^2(1 + y + y^2/2!).$$

5. Trees and rooted trees with all points labeled

This is the special case $m = p$ of the results in Section 4 and is the classical contrast to the case of

Table 3 **The numbers R_{ph} of rooted trees with p labeled points and height h.**

h/p	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
2		6	36	200	1170	7392	50568	372528	2936070
3			24	300	3360	38850	475776	6231960	87530400
4				120	2520	43680	757680	13747104	264181680
5					720	22680	551040	12836880	305605440
6						5040	221760	7136640	212133600
7							40320	2358720	96768000
8								362880	27216000
9									3628800

Table 4 **The numbers T_{pd} of trees with p labeled points and diameter d.**

d/p	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
3		12	60	210	630	1736	4536	11430
4			60	720	6090	47040	363384	2913120
5				360	7560	112560	1496880	19207440
6					2520	80640	1829520	36892800
7						20160	907200	28274400
8							181440	10886400
9								1814400

all points alike. The enumerators for this case are easily obtained by changing the y variable to $z = xy$ and passing to the limit for $x = 0$, since, e.g., by abuse of notation

$$s_h(x, z) = \sum_{p=1} x^p \sum_{m=0} s_{p+m,h,m} z^m/m! = \sum_{p=1} x^p s_p(z; h),$$

with the latter a definition; $s_0(z; h)$ is the enumerator, in this case an exponential generating function, of the rooted trees with height at most h and all points labeled (note that the root in effect has a double label).

From Eq. (14) it follows that

$$s_0(z; h + 1) = z \exp s_0(z; h). \quad (20)$$

Following the conventions of the Blissard calculus, [5] with $S^p(h) = S_p(h) = S_{ph} = s_{p,h,p}$, Eq. (20) may be rewritten

$$\exp zS(h + 1) = z \exp [\exp zS(h)]. \quad (20a)$$

Then differentiation with respect to z and equating coefficients of powers of z leads to

$$(p - 1)S_p(h + 1) = \sum_{k=0}^{p} k\binom{p}{k} S_{p-k}(h + 1)S_k(h). \quad (21)$$

This recurrence served for computation of all numbers S_{ph} and from them by differencing of the corresponding numbers R_{ph}, which are given in Table 3 for $p = 2(1)10$ and all values of h. Note that $S_p(h) = S_{ph} = p^{p-1}$, $h > p - 1$, a result corresponding to the curious identity ($0^0 = 0$),

$$(p - 1)p^{p-1} = \sum_{k=0}^{p} \binom{p}{k}(p - k)^{p-k-1}k^k.$$

Note also for verifications the following results obtained by direct enumeration:

$$R_{p,p-1} = p!$$

$$R_{p,p-2} = p!(2p - 5)/2$$

$$R_{p,p-3} = p!(3p^2 - 15p + 10)/6, \qquad p > 4$$

$$R_{p,p-4} = p!(4p^3 - 30p^2 + 40p + 3)/24, \qquad p > 6.$$

The numbers T_{pd} of trees with p labeled points and diameter d are obtained as follows Write the enumerator as

$$t_0(z; d) = \sum T_{pd} z^p/p!;$$

then (16) and (17) lead to

$$t_0(z; 2d) = r_0(z; d) - s_0(z; d - 1)r_0(z; d - 1) \quad (22)$$

$$t_0(z; 2d + 1) = r_0^2(z; d)/2. \quad (23)$$

These numbers are shown in Table 4 for $p = 3(1)10$ and all values of d. Note for verifications that

$$T_{p,p-1} = p!/2$$

$$T_{p,p-2} = p!(p - 4)/2$$

$$T_{p,p-3} = p!(6p^2 - 48p + 67)/24, \qquad p > 7$$

and

$$T_p = \sum T_{pd} = p^{p-2}.$$

It is also interesting to observe that the ratio $T_{p,p-k}/p!t_{p,p-k}$ $k = 2, 3$, approaches unity with increasing p: in other words, for large p, almost all labelings of such trees are distinct.

References and footnotes

1. J. Riordan, *An Introduction to Combinatorial Analysis*, Ch. 6, Wiley, New York, 1958.
2. F. Harary and G. Prins, *Acta Math.*, **101**, 141-162 (1959).
3. J. Riordan, *Acta Math.*, **97**, 211-225 (1957).

4. The corresponding term used by Harary and Prins is *root-diameter*. A precedent for the current usage appears in MacMahon's *Combinatory Analysis* (Volume I, p. 181); the trees he describes, "of altitude k and p terminal knots," may be regarded as rooted trees with each path from the root to an end point of length k.
5. See Ref. 1, Ch. 2.

Received Feb. 15, 1960