JOURNAL OF COMBINATORIAL THEORY, Series B 27, 109–121 (1979)

On the Number of Trees in a Random Forest

E. M. PALMER* AND A. J. SCHWENK†

Department of Mathematics, Michigan State University, East Lansing, Michigan 48824

Communicated by the Editors

Received August 24, 1976

The analytic methods of Pólya, as reported in [1, 6] are used to determine the asymptotic behavior of the expected number of (unlabeled) trees in a random forest of order p. Our results can be expressed in terms of $\eta = .338321856899208...$, the radius of convergence of $t(x)$ which is the ordinary generating function for trees. We have found that the expected number of trees in a random forest approaches $1 + \sum_{k=1}^{\infty} t(\eta^k) = 1.755510...$ and the form of this result is the same for other species of trees.

INTRODUCTION

The problem of estimating the number of trees in a large, random labeled forest was treated in Moon's book *Counting Labeled Trees* [3, p. 29]. It was found that the average number of labeled trees in all labeled forests of p points approaches 3/2 as a limit as p increases. We have investigated the same question for unlabeled trees and have found that in this case the average number of trees also approaches a constant, namely $1.755510\cdots$ This average an be expressed in terms of the ordinary generating function $t(x)$ for trees and its radius of convergence η. We use the notation and terminology of the book *Graphical Enumeration* [1] and the analytic methods of Pólya as reported in [1, 6].

Let F_p be the number of forests of order p and let $F_{p,n}$ be the number of these consisting of exactly n trees. Then the expected number of trees in a random forest is $\sum_n nF_{p,n}/F_p$. It is the asymptotic behavior of this quotient that we will determine. We begin by focusing our attention on the denominator.

* Work supported in part by a grant from the National Science Foundation.

† Work supported in part by grants from the National Science Foundation and the United States Naval Academy.

0095-8956/79/050109–13$02.00/0

Counting Forests

The counting series $F(x)$ for forests is defined by

$$F(x) = \sum_{p=1}^{\infty} F_p x^p. \tag{1}$$

As demonstrated in the book, *Graphical Enumeration* [1, p. 58], Pólya's enumeration theorem can be applied to express $F(x)$ in terms of the series $t(x)$ which counts trees and the cycle index $Z(S_n)$ of the symmetric group of degree n:

$$1 + F(x) = \sum_{n=0}^{\infty} Z(S_n, t(x)). \tag{2}$$

Formula (2) is conveniently altered by substitution in the identity (3.1.1) of [1]:

$$1 + F(x) = \exp \sum_{k=1}^{\infty} t(x^k)/k. \tag{3}$$

This formula (3) is, of course, just the usual relation between the generating function for graphs whose components are specified. Compare the relation (4.2.3) in [1] which expresses the generating function for graphs in terms of connected graphs.

For computational purposes an explicit formula for F_p can be obtained by differentiating (3) and equating coefficients. For neatness, we set $F_0 = 1$ but we emphasize that the empty forest is not admitted even temporarily (see Figure 1 of [2]). The formula for F_p can then be stated in terms of the number t_d of trees of order d:

$$F_p = \frac{1}{p} \sum_{k=1}^{p} F_{p-k} \sum_{d|k} dt_d . \tag{4}$$

The number t_d of trees of order d has been computed from Otter's formulas by the second author through $d = 44$ and reported [6] through $d = 36$.

To establish the asymptotic behavior of F_p we first review Otter's formula [4] for trees as reported in [1, p. 214]. Following Pólya's approach [5], Otter proved that

$$t_p \sim \frac{b_1^3}{4(\pi)^{1/2}} \eta^{3/2} \frac{\eta^{-p}}{p^{5/2}}, \tag{5}$$

where $\eta = .3383219...$ is the radius of convergence of the power series $t(x)$ and $(b_1^3/4\,(\pi)^{1/2})\,\eta^{3/2} = .5349485\cdots$. This was accomplished by showing that $0 < \eta < 1$, that $x = \eta$ is the unique singularity of $t(x)$ on the circle of convergence $|x| = \eta$, and that η is a branch point of order 2 for the continuation of $t(x)$. Thus he could expand $t(x)$ as follows (see formula (9.5.30) of [1, p. 213]):

$$t(x) = a_0 - a_1(\eta - x)^{1/2} + a_2(\eta - x)^{2/2} + a_3(\eta - x)^{3/2} + \cdots. \tag{6}$$

From this expression and the relation between $t(x)$ and the generating function $T(x)$ for rooted trees, he was able to show that $a_1 = 0$ and $a_3 = b_1{}^3/3$, where b_1 had already been determined from a similar treatment of $T(x)$. The details of the computation of b_1 and hence a_3 are found in the exposition of Otter's work in [1, Chapter 9.5]. Then the asymptotic estimate (5) for t_p follows from Pólya's lemma [5, p. 240 or 6, p. 367] presented below, together with the observation that $\Gamma(-3/2) = 4(\pi)^{1/2}/3$.

LEMMA (Pólya). *Let the power series*

$$f(x) = c_0 + c_1 x + c_2 x^2 + \cdots \tag{7}$$

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

$$f(x) = (1 - x/\rho)^{-s} g(x) + (1 - x/\rho)^{-t} h(x), \tag{8}$$

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

$$c_n \sim \frac{g(\rho)}{\Gamma(s)} n^{s-1} \rho^{-n}. \tag{9}$$

To apply the lemma to $t(x)$, we first note that $t(x)$ can be expressed as

$$\begin{aligned} t(x) = {} & \frac{b_1{}^3}{3}(\eta - x)^{3/2}\{1 + d_5(\eta - x) + d_7(\eta - x)^2 + \cdots\} \\ & + \{a_0 + a_2(\eta - x)^1 + a_4(\eta - x)^2 + \cdots\}. \end{aligned} \tag{10}$$

Therefore we can also write

$$t(x) = (1 - x/\eta)^{3/2} g(x) + h(x) \tag{11}$$

where

$$g(x) = \frac{b_1{}^3 \eta^{3/2}}{3}\{1 + d_5(\eta - x) + d_7(\eta - x)^2 + \cdots\} \tag{12}$$

and

$$h(x) = a_0 + a_2(\eta - x)^1 + a_4(\eta - x)^2 + \cdots. \tag{13}$$

The two series $g(x)$ and $h(x)$ are analytic at $x = \eta$ and so the lemma can be applied with $t = 0$, $s = -3/2$ and $g(\eta) = b_1{}^3\eta^{3/2}/3$ to obtain (5).

A similar procedure can be applied to the series $1 + F(x)$ for forests to obtain the next theorem.

THEOREM 1. *The asymptotic behavior of the number F_p of forests of order p is given by*

$$F_p \sim t_p(1 + F(\eta)). \tag{14}$$

Proof. Our goal is to apply Pólya's lemma to the generating function $1 + F(x)$ for forests, so first we must investigate its behavior at $x = \eta$. We begin by rewriting (3) as

$$1 + F(x) = \exp t(x) \exp \sum_{k=2}^{\infty} t(x^k)/k. \tag{15}$$

Next we note that the second factor on the right side of (15) is analytic at $x = \eta$. To establish this fact it is sufficient to show that $\sum_{k=2}^{\infty} t(x^k)/k$ converges for real x between 0 and $\eta^{1/2}$. This is quickly accomplished by the following inequalities:

$$\begin{aligned} \frac{1}{2} t(x^2) \leqslant \sum_{k=2}^{\infty} t(x^k)/k &= \sum_{k=2}^{\infty} \frac{1}{k} \sum_{p=1}^{\infty} t_p x^{kp} \\ &\leqslant \sum_{p=1}^{\infty} t_p \sum_{k=2}^{\infty} (x^p)^k \\ &\leqslant \sum_{p=1}^{\infty} t_p (x^2)^p (1 - x^p)^{-1} \\ &\leqslant (1 - x)^{-1} t(x^2). \end{aligned}$$

Therefore, the infinite sum above has the same radius of convergence as $t(x^2)$, namely $\eta^{1/2}$. The fact that $1 + F(x)$ has radius of convergence η with $x = \eta$ as the sole singularity on its circle of convergence now follows from the analogous fact for $t(x)$.

Now we turn our attention to $\exp t(x)$.

From (11) it follows that

$$\exp t(x) = \exp h(x) \exp\{(1 - x/\eta)^{3/2} g(x)\} \tag{16}$$

where $h(x)$ and $g(x)$ are both analytic at $x = \eta$, $h(\eta) = a_0 = t(\eta)$ and $g(\eta) =$

$b_1{}^3\eta^{3/2}/3$. On expanding the second factor on the right side of (16) in a Maclaurin series and rearranging the terms we have

$$\exp t(x) = \exp h(x) \left\{ g(x)(1 - x/\eta)^{3/2} \sum_{k=0}^{\infty} [(1 - x/\eta)^3 g(x)^2]^k/(2k + 1)! \right.$$

$$\left. + \sum_{k=0}^{\infty} [(1 - x/\eta)^3 g(x)^2]^k/(2k)! \right\}. \tag{17}$$

Therefore we can simplify our notation and conclude that

$$\exp t(x) = (1 - x/\eta)^{3/2} g(x) e^{h(x)} d(x) + a(x) \tag{18}$$

where $d(x)$ and $a(x)$ are both analytic at $x = \eta$, $d(\eta) = 1$, and $a(\eta) = \exp h(\eta) = \exp t(\eta)$.

Now it follows from (15) and (18) that

$$1 + F(x) = (1 - x/\eta)^{3/2} q(x) + b(x) \tag{19}$$

where $q(x)$ and $b(x)$ are analytic at $x = \eta$ and

$$q(\eta) = g(\eta) \exp \sum_{k=1}^{\infty} t(\eta^k)/k. \tag{20}$$

and

$$b(x) = a(x) \exp \sum_{k=2}^{\infty} t(x^k)/k. \tag{21}$$

Formula (20) has the compact form

$$q(\eta) = g(\eta)(1 + F(\eta)) \tag{22}$$

on substitution with (3). And for future reference, note also that

$$b(\eta) = 1 + F(\eta). \tag{23}$$

Finally, on applying Pólya's lemma we have F_p in terms of t_p and $F(\eta)$.

Counting Trees in Forests

To count forests with both points and number of trees as enumeration parameters, we simply introduce another variable in formulas (1), (2) and (3). Formula (1) becomes

$$1 + F(x, y) = \sum_{p,n=1}^{\infty} F_{p,n} x^p y^n \tag{24}$$

where $F_{p,n}$ is the number of forests with exactly n trees. Then Pólya's enumeration theorem can be applied to express $F(x, y)$ in terms of $yt(x)$, the series which counts forests with exactly one tree in each:

$$1 + F(x, y) = \sum_{n=0}^{\infty} Z(S_n, yt(x)). \tag{25}$$

The multi-variable form of the identity (3.1.1) of [1] is then used to obtain

$$1 + F(x, y) = \exp \sum_{k=1}^{\infty} y^k t(x^k)/k. \tag{26}$$

Now we note that on differentiating formula (24) with respect to y and setting $y = 1$ we have

$$\left.\frac{\partial F(x, y)}{\partial y}\right|_{y=1} = \sum_{p=1}^{\infty} \left(\sum_{n=1}^{p} nF_{p,n} \right) x^p, \tag{27}$$

the generating function which has as the coefficient of x^p the total number of trees in all forests of order p. To simplify our notation we denote this series by $F_y(x, 1)$. Therefore in differentiating (26) we have $F_y(x, 1)$ in explicit form:

$$F_y(x, 1) = \left\{ \exp \sum_{k=1}^{\infty} t(x^k)/k \right\} \sum_{k=1}^{\infty} t(x^k). \tag{28}$$

Now substitution from (3) gives

$$F_y(x, 1) = \{1 + F(x)\} \sum_{k=1}^{\infty} t(x^k). \tag{29}$$

The next formula for the coefficients of $F_y(x, 1)$ is quickly derived:

$$\sum_{n=1}^{p} nF_{p,n} = \sum_{k=1}^{p} F_{p-k} \left(\sum_{d \mid k} t_d \right). \tag{30}$$

The behavior of (30) as p increases is obtained in the next theorem.

THEOREM 2. *The asymptotic behavior of the number* $\sum_{n=1}^{p} nF_{p,n}$ *of trees in all forests of order* p *is given by*

$$\sum_{n=1}^{p} nF_{p,n} \sim t_p(1 + F(\eta)) \left\{ 1 + \sum_{k=1}^{\infty} t(\eta^k) \right\}. \tag{31}$$

Proof. It follows from (29) and our discussion at the beginning of the proof of Theorem 1 that $F_y(x, 1)$ also has η as its radius of convergence and $x = \eta$ in the only singularity on the circle of convergence. On substituting the expressions (11) for $t(x)$ and (19) for $1 + F(x)$ in (29) and simplifying, we have

$$F_y(x, 1) = (1 - x/\eta)^{3/2} \left\{ q(x) \left[h(x) + \sum_{k=2}^{\infty} t(x^k) \right] + g(x)\, b(x) \right\}$$

$$+ (1 - x/\eta)^3 g(x)\, q(x) + b(x) \left[h(x) + \sum_{k=2}^{\infty} t(x^k) \right]. \qquad (32)$$

Now the values of $q(x)$, $h(x)$, $g(x)$ and $b(x)$ at $x = \eta$ can be collected to evaluate the coefficient of $(1 - x/\eta)^{3/2}$ in (32) at $x = \eta$:

$$q(\eta) \left\{ h(\eta) + \sum_{k=2}^{\infty} t(\eta^k) \right\} + g(\eta)\, b(\eta)$$

$$= g(\eta)(1 + F(\eta)) \left[1 + \sum_{k=1}^{\infty} t(\eta^k) \right].$$

Then on applying Pólya's lemma to $F_y(x, 1)$ we have (31).

The Average Number of Trees in a Random Forest

An immediate consequence of the two theorems is the following corollary which provides an estimate for the average number of trees in a large random forest.

COROLLARY. *The asymptotic behavior of the expected number of trees in a random forest is given by*

$$\sum_{n=1}^{p} nF_{p,n}/F_p \sim 1 + \sum_{k=1}^{\infty} t(\eta^k). \qquad (34)$$

The first eight coefficients of $t(x)$ are sufficient to calculate the first five digits of the right side of (34) and we find that it begins $1.7555\cdots$.

Formulas (4) and (30) have been used to calculate the entries in Table I which includes the average number of trees per forest of order p.

We have also investigated the same questions for forests of rooted trees and planted trees. Formulas (4) and (30) are exactly the same except for interpretation. The number T_p of rooted trees of order p was calculated [6] for $p \leqslant 44$ and T_p is also the number of planted trees of order $p + 1$. Note also that deletion of the root in a tree leaves a rooted forest, so that the number of forests of rooted trees with p points is just T_{p+1}. Tables II and III contain the relevant data for rooted and planted forests.

TABLE I
Forests of Trees

p	t_p	F_p	$\Sigma\, nF_{p,n}$	Average
1	1	1	1	1
2	1	2	3	1.5
3	1	3	6	2
4	2	6	13	2.16667
5	3	10	24	2.4
6	6	20	49	2.45
7	11	37	93	2.51351
8	23	76	190	2.5
9	47	153	381	2.4902
10	106	329	803	2.44073
11	235	710	1703	2.39859
12	551	1601	3755	2.34541
13	1301	3658	8401	2.29661
14	3159	8599	19338	2.24887
15	7741	20514	45275	2.20703
16	19320	49905	108229	2.1687
17	48629	122963	26204	2.13563
18	123867	307199	647083	2.1064
19	317955	775529	1613941	2.08108
20	823065	1977878	4072198	2.05887
21	2144505	5086638	10374138	2.03949
22	5623756	13184156	26663390	2.02238
23	14828074	34402932	69056163	2.00728
24	39299897	90328674	180098668.	1.99382
25	104636890	238474986.	472604314.	1.98178
26	279793450.	632775648.	1247159936.	1.97094
27	751065460.	1686705630.	3307845730.	1.96113
28	2023443032.	4514955632.	8814122981.	1.95221
29	5469566585.	12132227370.	23585720703.	1.94406
30	14830871802.	32717113805.	63359160443.	1.93658
31	40330829030.	88519867048.	$1.7081554171E+11$	1.92969
32	$1.0997241022E+11$	$2.402356753E+11$	$4.6204925016E+11$	1.92332
33	$3.0062886248E+11$	$6.53843295E+11$	$1.2536852527E+12$	1.91741
34	$8.2377963172E+11$	$1.7843008656E+12$	$3.4114299618E+12$	1.91191
35	$2.2623663437E+12$	$4.881427413E+12$	$9.3078478603E+12$	1.90679
36	$6.2263060372E+12$	$1.3385770765E+13$	$2.5459661335E+13$	1.90199
37	$1.7169677491E+13$	$3.6787224247E+13$	$6.9803793987E+13$	1.8975
38	$4.7436313524E+13$	$1.0131005004E+14$	$1.9180827015E+14$	1.89328
39	$1.3129054378E+14$	$2.7954989467E+14$	$5.2815539543E+14$	1.88931
40	$3.6399025778E+14$	$7.7280707782E+14$	$1.4571734141E+15$	1.88556
41	$1.0107480767E+15$	$2.1401546924E+15$	$4.0278115401E+15$	1.88202
42	$2.8109864835E+15$	$5.9366613173E+15$	$1.1153020082E+16$	1.87867
43	$7.8289862215E+15$	$1.6494004573E+16$	$3.0934399422E+16$	1.87549
44	$2.1835027913E+16$	$4.5894756724E+16$	$8.5937004292E+16$	1.87248

TABLE II
Forests of Rooted Trees

p	T_p	F_p	$\Sigma\, nF_{p,n}$	Average
1	1	1	1	1
2	1	2	3	1.5
3	2	4	7	1.75
4	4	9	17	1.88889
5	9	20	39	1.95
6	20	48	96	2
7	48	115	232	2.01739
8	115	286	583	2.03846
9	286	719	1474	2.05007
10	719	1842	3797	2.06135
11	1842	4766	9864	2.06966
12	4766	12486	25947	2.07809
13	12486	32973	68738	2.08468
14	32973	87811	183612	2.09099
15	87811	235381	493471	2.09648
16	235381	634847	1334143	2.10152
17	634847	1721159	3624800	2.10602
18	1721159	468676	9893860	2.11016
19	468676	12826228	27113492	2.11391
20	12826228	35221832	74577187	2.11736
21	35221832	97055181	205806860.	2.12051
22	97055181	268282855.	569678759.	2.12343
23	268282855.	743724984.	1581243203.	2.12611
24	743724984.	2067174645.	4400193551.	2.1286
25	2067174645.	5759636510.	12273287277.	2.13091
26	5759636510.	16083734329.	34307646762.	2.13306
27	16083734329.	45007066269.	96093291818.	2.13507
28	45007066269.	1.2618655431$E + 11$	2.696540049$E + 11$	2.13695
29	1.2618655431$E + 11$	3.544268476$E + 11$	7.5801431209$E + 11$	2.1387
30	3.544268476$E + 11$	9.97171513$E + 11$	2.134300171$E + 12$	2.14035
31	9.97171513$E + 11$	2.8099343527$E + 12$	6.0186133956$E + 12$	2.14191
32	2.8099343527$E + 12$	7.9298197844$E + 12$	1.6996511299$E + 13$	2.14337
33	7.9298197844$E + 12$	2.2409533674$E + 13$	4.8062747591$E + 13$	2.14475
34	2.2409533674$E + 13$	6.3411730258$E + 13$	1.3608466446$E + 14$	2.14605
35	6.3411730258$E + 13$	1.7965593044$E + 14$	3.8577199147$E + 14$	2.14728
36	1.7965593044$E + 14$	5.0958804981$E + 14$	1.0948251932$E + 15$	2.14845
37	5.0958804981$E + 14$	1.4470233846$E + 15$	3.1104653829$E + 15$	2.14956
38	1.4470233846$E + 15$	4.1132541199$E + 15$	8.8460299877$E + 15$	2.15062
39	4.1132541199$E + 15$	1.170378008$E + 16$	2.5182082762$E + 16$	2.15162
40	1.170378008$E + 16$	3.3333125878$E + 16$	7.1752079388$E + 16$	2.15258
41	3.3333125878$E + 16$	9.5020085894$E + 16$	2.0462458967$E + 17$	2.15349
42	9.5020085894$E + 16$	2.7109773717$E + 17$	5.8404170531$E + 17$	2.15436
43	2.7109773717$E + 17$	7.7408802343$E + 17$	1.6683073006$E + 18$	2.15519
44	7.7408802343$E + 17$	2.2120392457$E + 18$	4.7681276168$E + 18$	2.15599

TABLE III
Forests of Planted Trees

p	T_{p-1}	F_p	$\Sigma\, nF_{p,n}$	Average
1	0	0	0	0
2	1	1	1	1
3	1	1	1	1
4	2	3	4	1.33333
5	4	5	6	1.2
6	9	13	18	1.38462
7	20	27	35	1.2963
8	48	68	93	1.36765
9	115	160	214	1.3375
10	286	404	549	1.35891
11	719	1010	1362	1.34851
12	1842	2604	3534	1.35714
13	4766	6726	9102	1.35326
14	12486	17661	23951	1.35615
15	32973	46628	63192	1.35524
16	87811	124287	168561	1.35622
17	235381	333162	451764	1.35599
18	634847	898921	1219290	1.35639
19	1721159	2437254	3305783	1.35636
20	4688676	6640537	9008027	1.35652
21	12826228	18166568	24643538	1.35653
22	35221832	49890419	67681372	1.3566
23	97055181	137478389.	186504925.	1.35661
24	268282855.	380031868.	515566016.	1.35664
25	743724984.	1053517588.	1429246490.	1.35664
26	2067174645.	2928246650.	3972598378.	1.35665
27	5759636510.	8158727139.	11068477743.	1.35664
28	16083734329.	22782938271.	30908170493.	1.35664
29	45007066269.	63752461474.	86488245455.	1.35663
30	1.2618655431$E + 11$	1.7874001451$E + 11$	2.4248115991$E + 11$	1.35661
31	3.544268476$E + 11$	5.0202656579$E + 11$	6.8104878438$E + 11$	1.3566
32	9.97171513$E + 11$	1.4124098942$E + 12$	1.916051726$E + 12$	1.35658
33	2.8099343527$E + 12$	3.9799473781$E + 12$	5.39906262$E + 12$	1.35657
34	7.9298197844$E + 12$	1.1231414222$E + 13$	1.52359599$E + 13$	1.35655
35	2.2409533674$E + 13$	3.1739032973$E + 13$	4.3054962476$E + 13$	1.35653
36	6.3411730258$E + 13$	8.9809052506$E + 13$	1.2182703769$E + 14$	1.35651
37	1.7965593044$E + 14$	2.5443781772$E + 14$	3.4514310652$E + 14$	1.35649
38	5.0958804981$E + 14$	7.2168711875$E + 14$	9.7894984871$E + 14$	1.35647
39	1.4470233846$E + 15$	2.0492500044$E + 15$	2.7797157244$E + 15$	1.35646
40	4.1132541199$E + 15$	5.8249836893$E + 15$	7.9012195814$E + 15$	1.35644
41	1.170378008$E + 16$	1.6573919634$E + 16$	2.2481157702$E + 16$	1.35642
42	3.3333125878$E + 16$	4.720251853$E + 16$	6.4025459588$E + 16$	1.3564
43	9.5020085894$E + 16$	1.3455342753$E + 17$	1.8250571297$E + 17$	1.35638
44	2.7109773717$E + 17$	3.8388013022$E + 17$	5.2068082317$E + 17$	1.35636

Otter showed [4] that the series $T(x)$ for rooted trees can be expanded as

$$T(x) = 1 - b_1(\eta - x)^{1/2} + b_2(\eta - x)^{2/2} + b_3(\eta - x)^{3/2} + \cdots. \tag{35}$$

We hasten to point out that the constants b_1 in (35) and (5) are the same (see also [1, p. 212]).

Therefore $T(x)$ can be expressed in the proper form for Pólya's lemma:

$$T(x) = (1 - x/\eta)^{1/2} g_1(x) + h_1(x) \tag{36}$$

where $g_1(\eta) = -b_1\eta^{1/2}$ and $h_1(\eta) = 1$. The asymptotic behavior follows with the observation that $\Gamma(-1/2) = -2(\pi)^{1/2}$.

The generating function for planed trees is just $xT(x)$ and on multiplying the right side of (35) by $\eta - (\eta - x)$, it too can be brought into an expansion of the proper form.

On carrying out the details for the expected values for rooted forests and planted forests, we again arrive at a conclusion of the form of (34). Moreover, this result holds in general whenever the generating function for a class of trees can be expanded as in (8). In particular, the expected number of rooted trees in a large random forest is $1 + \sum_{k=1}^{\infty} T(\eta^k) = 2.191837...$ and for forests of planted trees the number is $1 + \sum_{k=1}^{\infty} \eta^k T(\eta^k) = 1.355131\cdots$.

The actual calculations of these averages are simplified by certain rearrangements of the terms in the series. Since the series for trees and planted trees depend on the series $T(x)$ for rooted trees, it is sufficient to indicate the approach for $T(x)$ alone. By interchanging the order of summation and using $T(\eta) = 1$, the average for rooted trees can be written as

$$1 + \sum_{k=1}^{\infty} T(\eta^k) = 2 + \sum_{p=1}^{\infty} T_p \eta^{2p}/(1 - \eta^p). \tag{37}$$

The numbers T_p of rooted trees are bounded above by the Catalan numbers (see [1, p. 209]). These bounds can be used to estimate the truncation error when the right side of (37) is terminated at $p = m$. We find that this error is less than $(4\eta^2)^{m+1}/(m + 1)$ for all $m \geqslant 0$ and so, for example, with $m = 5$, the truncation error is already less than $2 \cdot 10^{-1}$.

A routine analysis of the rounding error also shows that if η is estimated with s significant digits, we can obtain $s - 1$ significant digits in our result, provided m is sufficiently large.

Note that there are fewer planted trees per large, random forest than ordinary trees, even though for large p there are always more planted trees than ordinary trees of order p. This difference is accounted for by noting that a single point is a tree but not a planted tree. Therefore $t(x)$ begins with x

itself and this term contributes to the estimate $1.755510\cdots$ for trees the important sum $\eta + \eta^2 + \cdots = \eta/(1 - \eta)$ which is more than enough to make up the difference between $1.755510\cdots$ and $1.355131\cdots$.

Probability of Exactly k Trees in a Forest

A related question asks for the probability that a large random forest has exactly k trees. The generating function for forests with exactly k rooted trees is $Z(S_k; T(x), T(x^2),...)$. The desired probability for a forest on p points is the coefficient of x^p in this cycle index divided by the total number of p-point forests. Now in [6] we demonstrated that the coefficient in x^p in the cycle index is asymptotic to $T_p \cdot Z(S_{k-1}; T(\eta), T(\eta^2),...)$. Since $F_p \sim T_p (1 + F(\eta))$, we conclude that as p increases, the desired probability that a forest on p points has exactly k trees approaches the limiting value of

$$\frac{Z(S_{k-1}\,; T(\eta), T(\eta^2),...)}{1 + F(\eta)}. \tag{38}$$

The same analysis applies to other varieties of forests, so long as $T(x)$ is replaced by the appropriate counting series, for example, $t(x)$ for unrooted forests and $xT(x)$ for planted forests.

The identity

$$kZ(S_k\,; s_1\,, s_2\,,...) = \sum_{i=1}^{k} s_i Z(S_{k-i}\,; s_1\,, s_2\,,...) \tag{39}$$

(see [1, p. 36]) permits the cycle indices in (38) to be evaluated recursively. Thus, we have computed the limiting probabilities in Table IV.

In conclusion, we observe that it has been shown [7] that for large k

$$Z(S_k\,; T(\eta), T(\eta^2),...) \sim c\eta^k \tag{40}$$

where the constant c is given by

$$c = \prod_{i=1}^{\infty} (1 - \eta^i)^{-T_i+1} = 7.758\,160\,291. \tag{41}$$

A similar analysis for unrooted threes provides that for a large k

$$Z(S_k; t(\eta), t(\eta^2),...) \sim d\eta^k \tag{42}$$

with

$$d = \prod_{i=1}^{\infty} (1 - \eta^i)^{-t_i+1} = 2.129\,384\,514. \tag{43}$$

TABLE IV

Limiting Probability of Exactly k Trees in a Forest

k	forest of trees	forest of rooted trees	forest of planted trees
1	.522 841 424	.338 321 857	.707 218 415
2	.295 794 369	.338 321 857	.239 267 448
3	.117 521 272	.191 403 541	.045 796 637
4	.041 950 320	.083 180 075	.006 733 378
5	.014 451 944	.031 622 733	.000 866 050
6	.004 919 432	.011 256 322	.00 104 297
7	.001 667 802	.003 887 755	.000 012 187
8	.000 564 649	.001 325 758	.000 001 406
9	.000 191 079	.000 449 832	.000 000 161
10	.000 064 651	.000 152 344	.000 000 018
large k	$3.290\,743\,438\,\eta^k$	$7.758\,160\,291\,\eta^k$	$49.934\,941\,04\,\eta^{2k}$

The planted problem is solved by the identity

$$Z(S_k;\, \eta T(\eta),\, \eta^2 T(\eta^2),\ldots) = \eta^k Z(S_k\,;\, T(\eta),\, T(\eta^2),\ldots). \tag{44}$$

Expression (38) can now be estimated to obtain the final line of Table IV.

References

1. F. Harary and E. M. Palmer, "Graphical Enumeraction," Academic Press, New York, 1973.
2. F. Harary and R. C. Read, Is the null-graph a pointless concept?, *in* "Graphs and Combinatorics" (R. A. Bari and F. Harary, Eds.), pp. 37–44, Springer–Verlag, Berlin, 1974.
3. J. W. Moon, "Counting labeled trees, *in* "Canad. Math. Congress, Montreal, 1970," p. 29.
4. R. Otter, The number of trees, *Ann. of Math.* **49** (1948), 583–599.
5. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68** (1973), 145–254.
6. R. W. Robinson and A. J. Schwenk, The distribution of degrees in a large random tree, *Discrete Math.* **12** (1975), 359–372.
7. A. J. Schwenk, An asymptotic evaluation of the cycle index of a symmetric group, *Discrete Math.* **18** (1977), 71–78.