Scan A6677 ✓
– A6682

RWR

Bin? Phylogenetic
Trees

add to all segs

ENUMERATION OF BINARY PHYLOGENETIC TREES

L.R. FOULDS AND R.W. ROBINSON*

Evolutionary trees of biology are represented by a special class of labelled trees, termed phylogenetic *trees. These are characterised by having disjoint subsets of the labelling set assigned to the points of a tree, in such a way that no point of degree less than 3 is assigned an empty set of labels. By a* binary *tree is meant one in which every point has degree 1 or 3. The exact and asymptotic numbers of binary phylogenetic trees are determined under the presence or absence of two additional conditions on the labelling. The optional constraints studied require nonempty label sets to be singletons, and that only endpoints be labelled.*

1. INTRODUCTION

Biologists often generate phylogenetic (evolutionary) trees from protein sequence data; see [2],[3],[4] and [8]. Determination of a common ancestor, or root, is based on separate criteria. It is common for the directed rooted tree so created to be binary. The trees considered here are binary trees, in which every point has degree 1 or 3. An evolutionary tree is labelled with the names of known species. If two species are not distinguished by the protein sequences under study then they are assigned to the same point in the tree. Conversely, it is often convenient to hypothesize common ancestors for whom no direct evidence is known, and these become points to which no name at all is assigned. Of course, a hypothetical point with no name would not have degree 1, because including such a point would serve no purpose in explaining the biological evidence.

As a mathematical model of an evolutionary tree containing n known species we take a binary tree labelled by an assignment of the label set {1,...,n} to points so that every endpoint receives at least one label. Such a tree is called a *binary phylogenetic tree*. The number n of labels is termed the *magnitude* of the phylogenetic tree, and the number of points in it is termed the *order*. A *planted* binary tree is one with a distinguished endpoint called the *root*. This corresponds to the common ancestor in an evolutionary tree. Since the root is already distinguished we never assign it a label.

The exact and asymptotic numbers of phylogenetic trees with given magnitude, along with the average and variance of their orders, were determined in [5]. This

* The second author is grateful for the support of the Australian Research Grants Committee for the project "Numerical Implementation of Unlabelled Graph Counting Algorithms", under which research and computing for this paper were performed.

was also accomplished under the restriction to 1-1 labellings. In the present paper we perform the same analysis for binary phylogenetic trees, but in addition, study the effect of restricting the labelling to endpoints. This gives a total of four cases to study. The calculations are facilitated by the particularly simple form of the most restricted case, and the fact that the other three can then be obtained by algebraic transformations of the exponential generating functions.

2. LABELLING RESTRICTED TO ENDPOINTS AND 1-1

The simplest case is the most restricted, being binary phylogenetic trees with 1-1 labelling and no interior points labelled. It is a standard result that the number of such trees with n endpoints is $\prod_{i=0}^{n-3} (2i+1)$. Direct combinatorial proofs of this fact appear in [2,p.241] and [3,p.28]; a more elaborate proof involving generating functions is given in [8,pp.51-52]. In [9,p.72] it is pointed out that this is a corollary of Prüfer's proof of Cayley's result that there are n^{n-2} labelled trees. Here the labelling is 1-1 onto all the points, and the proof gives a 1-1 correspondence between trees and sequences of length n-2 from the label set. It was noted by Prüfer [10] that each point of degree d is represented in its sequence exactly d-1 times. Thus if 1,...,n are reserved for labelling the endpoints of a binary tree and n+1,...,2n-2 for labelling the interior points, then there are $(2n-4)!/2^{n-2}$ sequences in which the latter occur exactly twice each. We simply divide by (n-2)! for the labellings on the interior points, and find that the number with just endpoints labelled is

$$(2n-4)!/(n-2)!2^{n-2} = 1.3 \cdot \ldots \cdot (2n-5),$$

as claimed.

If we denote by T_n the number of binary trees of magnitude n in this case we have, by the above discussion, $T_1 = 0$, $T_2 = 1$, and

$$T_n = \prod_{i=0}^{n-3} (2i+1) \tag{2.1}$$

for $n \geq 3$. It is clear that in a binary tree the number of endpoints exceeds the number of interior points by exactly 2. Thus each binary tree of magnitude n has order 2n-2. We denote by R_n the sum of the orders p of the binary trees of magnitude n, so that

$$R_n = (2n-2)T_n \tag{2.2}$$

for $n \geq 1$. Also, denote by S_n the sum of p(p-1), which gives

$$S_n = (2n-2)(2n-3)T_n \tag{2.3}$$

for $n \geq 1$. In general, R_n and S_n are needed in order to determine the average μ_n

and variance ν_n of the order for the trees of magnitude n. The relations for $n \geq 1$ are:

$$\mu_n = R_n/T_n ,$$

$$\nu_n = S_n/T_n + \mu_n - \mu_n^2 . \qquad (2.4)$$

Of course in this case $\mu_n = 2n-2$ and $\nu_n = 0$, but R_n and S_n will be useful for the three cases considered later.

Let P_n denote the number of planted binary trees of magnitude n. Since the root is not labelled this means there are exactly n+1 endpoints in such a planted tree. If the root point is labelled with the number n+1 and then unrooted, the result is an ordinary binary tree of magnitude n+1. This process gives a 1-1 correspondence, so that

$$P_n = T_{n+1} \qquad (2.5)$$

for $n \geq 1$.

In order to obtain asymptotic estimates we need the exponential generating functions T(x), R(x), S(x) and P(x) for these four sequences. From (2.1) and the binomial theorem we have

$$T(x) = -\frac{1}{3} + x + \frac{1}{3}(1-2x)^{3/2}. \qquad (2.6)$$

By Stirling's formula it can be seen that the coefficient of x^n in $(1-x)^{-s}$ is

$$\frac{n^{s-1}}{\Gamma(s)} (1 + \frac{s(s-1)}{2n} + O(\frac{1}{n^2})) \qquad (2.7)$$

as long as $s \neq 0,-1,-2,\ldots$ Thus

$$T_n = \frac{n!2^n}{4\pi^{1/2}n^{5/2}} (1 + \frac{15}{8n} + O(\frac{1}{n^2})). \qquad (2.8)$$

Similarly one has

$$P(x) = 1 - (1-2x)^{1/2}, \qquad (2.9)$$

and

$$P_n = \frac{n!2^n}{2\pi^{1/2}n^{3/2}} (1 + \frac{3}{8n} + O(\frac{1}{n^2})). \qquad (2.10)$$

From (2.2) it follows that

$$R(x) = 2xT'(x) - 2T(x),$$

so from (2.6) we have

$$R(x) = \frac{2}{3} - \frac{2}{3}(1+x)(1-2x)^{1/2}. \qquad (2.11)$$

Likewise, (2.3) gives

$$S(x) = 4x^2T''(x) - 6xT'(x) + 6T(x),$$

and combining with (2.6) yields

$$S(x) = -2 + 2(1-x)(1-2x)^{-1/2}. \qquad (2.12)$$

3. LABELLING RESTRICTED TO ENDPOINTS

This case differs from the previous one in allowing more than one label to be assigned to an endpoint. Interior points are still not allowed labels. Thus whereas the exponential generating function for labelling a single endpoint was x, it is now e^x-1. This is because an endpoint may receive 1,2,3,... labels. For any $k \geq 1$ there is just one way to assign k labels to a point. Interleaving of label sets is accounted for in multiplying exponential generating functions together; see [6,Chapter 1] for an account of the uses of exponential generating functions in labelled enumeration. Thus $T(e^x-1)$ is the exponential generating function by magnitude for binary phylogenetic trees in which interior points are not labelled. Similarly $R(e^x-1)$, $S(e^x-1)$ and $P(e^x-1)$ give exponential generating functions for the sum of the order p, the sum of p(p-1) and the number of planted trees, respectively. We denote these generating functions by $\tilde{T}(x)$, $\tilde{R}(x)$, $\tilde{S}(x)$ and $\tilde{P}(x)$.

Substitition of e^x-1 for x in (2.6), (2.11), (2.12) and (2.9) gives

$$\begin{aligned} \tilde{T}(x) &= e^x - \frac{4}{3} + \frac{1}{3}(3-2e^x)^{3/2}, \\ \tilde{R}(x) &= \frac{2}{3} - \frac{2}{3}e^x(3-2e^x)^{1/2}, \\ \tilde{S}(x) &= -2 + (4-2e^x)(3-2e^x)^{-1/2} \\ \tilde{P}(x) &= 1 - (3-2e^x)^{1/2}. \end{aligned} \qquad (3.1)$$

Recurrence relations for $\tilde{T}_n$, $\tilde{R}_n$, $\tilde{S}_n$ and $\tilde{P}_n$ could be deduced directly from these equations. However for numerical purposes it is easier to start with the simple expression for T_n, R_n, S_n and P_n provided in Section 2. From the fact that $(e^x-1)^k/k!$ is the exponential generating function

$$\sum_{n=0}^{\infty} S(n,k)x^n/n!$$

for Stirling numbers of the second kind, it follows that

$$\tilde{T}_n = \sum_{k=2}^{n} S(n,k)T_k \qquad (3.2)$$

for $n \geq 2$. The analogous equations hold for $\tilde{R}_n$, $\tilde{S}_n$ and $\tilde{P}_n$. In the latter case the sum must include $k = 1$, and the result is also valid for $n = 1$. Stirling numbers are readily calculated from the recurrence

$$S(n+1,k) = S(n,k-1) + kS(n,k),$$

which holds for $k \geq 1$, and from the boundary conditions $S(0,0) = 1$, $S(0,k) = 0$ if $k > 0$ and $S(n,0) = 0$ if $n > 0$.

Roughly speaking, the asymptotic behaviour of $\tilde{T}_n$ is determined by the radius of convergence $\tilde{\rho}$ of the exponential generating function $\tilde{T}(x)$. From (3.1) it is evident that $\tilde{\rho} = \ln 3/2$, that $\tilde{\rho}$ is also the radius of convergence of $\tilde{R}(x)$, $\tilde{S}(x)$ and $\tilde{P}(x)$, and that in each case the point $x = \tilde{\rho}$ is the sole singularity on the circle of convergence. It is then classical (see [1,Theorem 4] or [7,p.489]) that an expansion of the generating function in powers of $(1-x/\tilde{\rho})^{\frac{1}{2}}$ can be used in conjunction with (2.7) to determine the precise asymptotic growth rate of the coefficients. The first two odd powers are sufficient to give the nth coefficient with a factor of $(1+O(\frac{1}{n^2}))$.

Because $e^{\tilde{\rho}} = 3/2$ we have

$$3 - 2e^x = 3\tilde{\rho}(1-x/\tilde{\rho})(1 - \frac{\tilde{\rho}}{2}(1-x/\tilde{\rho}) \pm \ldots),$$

and so from (3.1)

$$\tilde{T}(x) = 3^{1/2}\tilde{\rho}^{3/2}(1-x/\tilde{\rho})^{3/2} - \frac{3^{3/2}\tilde{\rho}^{5/2}}{4}(1-x/\tilde{\rho})^{5/2} \pm \ldots$$

Summing the contributions of these two terms according to (2.7) yields

$$\tilde{T}_n = \frac{3^{3/2}\tilde{\rho}^{3/2}}{4\pi^{1/2}} \cdot \frac{n!}{n^{5/2}\tilde{\rho}^n}\left(1 + \frac{15(1+\tilde{\rho})}{8n} + O(\frac{1}{n^2})\right) . \qquad (3.3)$$

In the same fashion, the other three expressions in (3.1) can be expanded, with the following results:

$$\tilde{R}_n = \frac{3^{1/2}\tilde{\rho}^{1/2}}{2\pi^{1/2}} \cdot \frac{n!}{n^{3/2}\tilde{\rho}^n}\left(1 + \frac{3(1+5\tilde{\rho})}{8n} + O(\frac{1}{n^2})\right); \qquad (3.4)$$

$$\tilde{S}_n = \frac{3^{-1/2}\tilde{\rho}^{-1/2}}{\pi^{1/2}} \quad \frac{n!}{n^{1/2}\tilde{\rho}^n}\left(1 - \frac{1+13\tilde{\rho}}{8n} + O(\frac{1}{n^2})\right); \qquad (3.5)$$

$$\tilde{P}_n = \frac{3^{1/2}\tilde{\rho}^{1/2}}{2\pi^{1/2}} \quad \frac{n!}{n^{3/2}\tilde{\rho}^n}\left(1 + \frac{3(1+\tilde{\rho})}{8n} + O(\frac{1}{n^2})\right). \qquad (3.6)$$

The mean $\tilde{\mu}_n$ and variance $\tilde{\nu}_n$ of the number of points in the binary trees of magnitude n in this case are given by the obvious analogue of (2.4)

$$\tilde{\mu}_n = \tilde{R}_n/\tilde{T}_n,$$

$$\tilde{\nu}_n = (\tilde{S}_n/\tilde{T}_n) + \tilde{\mu}_n - \tilde{\mu}_n^2 . \qquad (3.7)$$

Thus (3.3), (3.4) and (3.5) can be immediately applied, resulting in

$$\tilde{\mu}_n = \frac{2n}{3\tilde{\rho}}(1 - \frac{3}{2n} + O(\frac{1}{n^2}))$$

$$\tilde{\nu}_n = \frac{4(1-2\tilde{\rho})n}{9\tilde{\rho}^2}(1 + O(\frac{1}{n})). \qquad (3.8)$$

4. 1-1 LABELLING

This case differs from the first case in allowing interior points to be labelled. Let $\overline{T}_n$ denote the number of trees of magnitude n under this labelling convention. Likewise, let $\overline{R}_n$ and $\overline{S}_n$ denote the totals of the order p and of p(p-1) respectively, over these $\overline{T}_n$ trees. Finally, let $\overline{P}_n$ be the number of planted trees of magnitude n. As usual, we denote the exponential generating functions of these four sequences by $\overline{T}(x)$, $\overline{R}(x)$, $\overline{S}(x)$ and $\overline{P}(x)$.

Since labelling is optional for interior points, and at most one label can be assigned to each, the exponential generating function of the labelling possibilities for an interior point is $1 + x$. For a tree with n endpoints there are $n - 2$ interior points. Each endpoint is labelled, so labelling possibilities for an endpoint has x as its exponential generating function. Thus each 1-1 labelled basic tree with magnitude n and only endpoints labelled gives rise to a number of compatible versions in which interior points may be labelled, and these have $x^n(1+x)^{n-2}$ as exponential generating function. Summing over all T_n basic trees and then over all $n \geq 2$, this gives

$$\overline{T}(x) = T(x+x^2)/(1+x)^2. \qquad (4.1)$$

It is now easy to obtain a recurrence for $\overline{T}_n$. Putting the equation in the form

$$\overline{T}(x) = -2x\overline{T}(x) - x^2\overline{T}(x) + \overline{T}(x+x^2)$$

and comparing coefficients of $x^n/n!$ yields

$$\overline{T}_n = -2n\overline{T}_{n-1} - n(n-1)\overline{T}_{n-2} + \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k}\binom{n-k}{k} k! T_{n-k} \qquad (4.2)$$

for $n \geq 2$. Here $T_0 = 0$ and $T_1 = 0$ are needed as boundary conditions. Exactly the same transformation gives $\overline{R}(x)$ from $R(x)$ and $\overline{S}(x)$ from $S(x)$, so recurrence relations analagous to (4.2) are valid for $\bar{R}_n$ and $\bar{S}_n$.

In a planted tree the root is an endpoint which is not labelled, so with n labelled endpoints there are n - 1 interior points which might be labelled. This gives

$$\bar{P}(x) = P(x+x^2)/(1+x), \tag{4.3}$$

and

$$\bar{P}_n = -n\bar{P}_{n-1} + \sum_{k=0}^{\lfloor n/2\rfloor} \binom{n}{k}\binom{n-k}{k} k!P_{n-k} \tag{4.4}$$

for $n \geq 2$ with $\bar{P}_1 = 1$.

Explicit expressions for the exponential generating functions can be found at once from (2.6), (2.9), (2.11) and (2.12):

$$\begin{aligned}
\bar{P}(x) &= \frac{1-(1-2x-2x^2)^{1/2}}{1+x};\\
\bar{T}(x) &= \frac{-1+3x+3x^2+(1-2x-2x^2)^{3/2}}{3(1+x)^2};\\
\bar{R}(x) &= \frac{2-2(1+x+x^2)(1-2x-2x^2)^{1/2}}{3(1+x)^2};\\
\bar{S}(x) &= \frac{-2+2(1-x-x^2)(1-2x-2x^2)^{-1/2}}{(1+x)^2}.
\end{aligned} \tag{4.5}$$

In each case the radius of convergence is $\bar{\rho} = (\sqrt{3}-1)/2$, and $x = \bar{\rho}$ is the sole singularity on the circle of convergence. We have $1+\bar{\rho} = 1/2\bar{\rho}$, so that

$$1 + x = \frac{1}{2\bar{\rho}}(1-2\bar{\rho}(\bar{\rho}-x))$$

and

$$1 - 2x - 2x^2 = 2\sqrt{3}(\bar{\rho}-x)(1-\frac{1}{\sqrt{3}}(\bar{\rho}-x)).$$

Substituting into (4.5), one finds the first two odd powers of $(1-x/\bar{\rho})^{\frac{1}{2}}$. Finally, (2.7) is applied, to give the following asymptotic estimates:

$$\begin{aligned}
\bar{P}_n &= \frac{2^{1/2}3^{1/4}\bar{\rho}^{-3/2}}{\pi^{1/2}} \frac{n!}{n^{3/2}\bar{\rho}^{n}}\left(1+\frac{11\sqrt{3}-18}{8n}+O(\frac{1}{n^2})\right);\\
\bar{T}_n &= \frac{2^{1/2}3^{3/4}(2-\sqrt{3})\bar{\rho}^{-3/2}}{\pi^{1/2}} \frac{n!}{n^{5/2}\bar{\rho}^{n}}\left(1+\frac{5(7\sqrt{3}-10)}{8n}+O(\frac{1}{n^2})\right);\\
\bar{R}_n &= \frac{2^{1/2}3^{1/4}(2-\sqrt{3})\bar{\rho}^{-1/2}}{\pi^{1/2}} \frac{n!}{n^{3/2}\bar{\rho}^{n}}\left(1+\frac{19\sqrt{3}-30}{8n}+O(\frac{1}{n^2})\right);\\
\bar{S}_n &= \frac{2^{1/2}3^{1/4}(2-\sqrt{3})\bar{\rho}^{-1/2}}{\pi^{1/2}} \frac{n!}{n^{1/2}\bar{\rho}^{n}}\left(1-\frac{90-37\sqrt{3}}{24n}+O(\frac{1}{n^2})\right).
\end{aligned} \tag{4.6}$$

The mean $\bar{\mu}_n$ and the variance $\bar{\nu}_n$ of the order for trees of magnitude n are found from $\bar{T}_n$, $\bar{R}_n$ and $\bar{S}_n$ just as in (3.7) for the previous case. Asymptotic estimates then follow from (4.6);

$$\bar{\mu}_n = \frac{n}{\sqrt{3}\,\bar{\rho}}\left(1 - \frac{4\sqrt{3}-5}{2n} + O\left(\frac{1}{n^2}\right)\right),$$

and (4.7)

$$\bar{\nu}_n = \frac{(2\sqrt{3}-3)n}{9\bar{\rho}^2}\left(1 + O\left(\frac{1}{n}\right)\right).$$

The recurrence relations (4.2) and (4.3) for $\overline{T}_n$ and $\overline{P}_n$ require $O(n^2)$ arithmetic operations to compute the values up to n, even given that T_k is available already for $k \leq n$. Improved recurrences can be obtained directly from (4.5) by differentiating the explicit expressions for the generating functions, simplifying and then comparing coefficients of $x^n/n!$. In this way one finds:

$$\overline{P}_n = (n-3)\overline{P}_{n-1} + (n-1)(4n-9)\overline{P}_{n-2} + 2(n-1)(n-2)(n-3)\overline{P}_{n-3} \tag{4.8}$$

for $n \geq 3$, with $\overline{P}_0 = 0$, $\overline{P}_1 = 1$ and $\overline{P}_2 = 1$;

$$\overline{T}_n = \overline{P}_{n-1} + 2(n-1)\overline{P}_{n-2} - (n+1)\overline{T}_{n-1} \tag{4.9}$$

for $n \geq 3$, with $\overline{T}_2 = 1$;

$$\overline{R}_n = -n\overline{R}_{n-1} + \tfrac{2}{3}(n(n-1)\overline{P}_{n-2} + n\overline{P}_{n-1} + \overline{P}_n) \tag{4.10}$$

for $n \geq 3$, with $R_2 = 2$;

$$\overline{S}_n = n\overline{S}_{n-1} + 4n(n-1)\overline{S}_{n-2} + 2n(n-1)(n-2)\overline{S}_{n-3} - 2(\overline{P}_n - n\overline{P}_{n-1} - n(n-1)\overline{P}_{n-2}) \tag{4.11}$$

for $n \geq 3$, with $\overline{S}_2 = 2$. These relations only require $O(n)$ arithmetic operations in order to calculate values of $\overline{P}_k$, $\overline{T}_k$, $\overline{R}_k$ or $\overline{S}_k$ for $k \leq n$.

5. UNRESTRICTED LABELLING

The final case allows all binary labelled trees, including the possibility of multiple labels and labels for interior points. As for any phylogenetic trees, it is still the case that each endpoint must be assigned at least one label. This differs from the previous case only in allowing multiple labels, so the relation of this section to the previous section is exactly the same as the relation of Section 3 to Section 2. We denote the number of trees of magnitude n by $\hat{T}_n$, and the number of planted trees by $\hat{P}_n$. Similarly, the sum of the order p and the sum of p(p-1) for magnitude n trees are denoted $\hat{R}_n$ and $\hat{S}_n$. The exponential generating functions are $\hat{T}(x)$, $\hat{P}(x)$, $\hat{R}(x)$ and $\hat{S}(x)$ respectively. These are obtained from $\overline{T}(x)$, $\overline{P}(x)$, $\overline{R}(x)$ and $\overline{S}(x)$ by replacing x with $e^x - 1$. Thus the exact numbers are related by

$$\hat{T}_n = \sum_{k=2}^{n} S(n,k)\bar{T}_k \tag{5.1}$$

for $n \geq 2$, which is similar to (3.2). Of course $\hat{R}_n$, $\hat{S}_n$ and $\hat{P}_n$ are calculated analogously from the corresponding numbers determined in the previous section. In the case of $\hat{P}_n$ the sum starts at $k = 1$ and the result is valid for $n = 1$.

To obtain the exponential generating functions explicitly one need only substitute $e^x - 1$ for x in (4.5). The results are:

$$\begin{aligned}
\hat{P}(x) &= e^{-x} - e^{-x}(1+2e^x-2e^{2x})^{1/2};\\
\hat{T}(x) &= 1 - e^{-x} - \tfrac{1}{3}e^{-2x} + \tfrac{1}{3}e^{-2x}(1+2e^x-2e^{2x})^{3/2};\\
\hat{R}(x) &= \tfrac{2}{3}e^{-2x} - \tfrac{2}{3}e^{-2x}(1-e^x+e^{2x})(1+2e^x-2e^{2x})^{1/2};\\
\hat{S}(x) &= -2e^{-2x} + 2e^{-2x}(1+e^x-e^{2x})(1+2e^x-2e^{2x})^{-1/2}.
\end{aligned} \tag{5.2}$$

In each of these generating functions the radius of convergence is $\hat{\rho} = \ln((\sqrt{3}+1)/2)$, and $x = \hat{\rho}$ is the only singularity on the circle of convergence. As in the previous three sections we expand the generating functions in terms of $(1-x/\hat{\rho})^{\frac{1}{2}}$, and apply (2.7) to the first two odd powers. The asymptotic estimates so obtained are:

$$\begin{aligned}
\hat{P}_n &= \frac{(\sqrt{3}-1)(3+\sqrt{3})^{1/2}\hat{\rho}^{1/2}}{2\pi^{1/2}} \cdot \frac{n!}{n^{3/2}\hat{\rho}^n}\left(1 + \frac{3-\hat{\rho}(6-\sqrt{3})}{8n} + O\left(\frac{1}{n^2}\right)\right);\\
\hat{T}_n &= \frac{3^{1/2}(3-\sqrt{3})^{1/2}\hat{\rho}^{3/2}}{2^{1/2}\pi^{1/2}} \cdot \frac{n!}{n^{5/2}\hat{\rho}^n}\left(1 + \frac{5(3-\hat{\rho}(2-\sqrt{3}))}{8n} + O\left(\frac{1}{n^2}\right)\right);\\
\hat{R}_n &= \frac{(2-\sqrt{3})(3+\sqrt{3})^{1/2}\hat{\rho}^{1/2}}{\pi^{1/2}} \cdot \frac{n!}{n^{3/2}\hat{\rho}^n}\left(1 + \frac{3+\hat{\rho}(5\sqrt{3}-6)}{8n} + O\left(\frac{1}{n^2}\right)\right);\\
\hat{S}_n &= \frac{2(2-\sqrt{3})(3+\sqrt{3})^{-1/2}\hat{\rho}^{-1/2}}{\pi^{1/2}} \cdot \frac{n!}{n^{1/2}\hat{\rho}^n}\left(1 - \frac{3+\hat{\rho}(66+13\sqrt{3})}{24n} + O\left(\frac{1}{n^2}\right)\right).
\end{aligned} \tag{5.3}$$

Now the mean $\hat{\mu}_n$ and the variance $\hat{\sigma}_n$ of the order for trees of magnitude n depend on $\hat{T}_n$, $\hat{R}_n$ and $\hat{S}_n$ as in (3.7). From the asymptotic estimates above one then calculates

$$\hat{\mu}_n = \frac{2(2-\sqrt{3})n}{\hat{\rho}(3-\sqrt{3})}\left(1 - \frac{3-\hat{\rho}}{2n} + O\left(\frac{1}{n^2}\right)\right),$$

and (5.4)

$$\hat{\sigma}_n = \frac{2n}{9\hat{\rho}^2}\left(6 - (3+\hat{\rho})\sqrt{3}\right)\left(1 + O\left(\frac{1}{n}\right)\right).$$

6. NUMERICAL RESULTS

The values of P_n, $\tilde{P}_n$, $\bar{P}_n$ and $\hat{P}_n$ for $1 \leq n \leq 10$ and $n = 15,20,25,30,35$ and 40 are presented in Table 1. The corresponding values of T_n, $\tilde{T}_n$, $\bar{T}_n$ and $\hat{T}_n$ are presented in Table 2, those of μ_n, $\tilde{\mu}_n$, $\bar{\mu}_n$ and $\hat{\mu}_n$ in Table 3, and those of $\tilde{\nu}_n$, $\bar{\nu}_n$ and $\hat{\nu}_n$ in Table 4. The full range of values for $1 \leq n \leq 40$ in all cases is available from the second author. The calculations are based on the equations giving recurrences for P_n, T_n, R_n, S_n, $\tilde{P}_n$, $\tilde{T}_n$, $\tilde{R}_n$, $\tilde{S}_n$, $\bar{P}_n$, $\bar{T}_n$, $\bar{R}_n$, $\bar{S}_n$, $\hat{P}_n$, $\hat{T}_n$, $\hat{R}_n$ and $\hat{S}_n$ in the previous four sections. Then μ_n and ν_n are computed from T_n, R_n and S_n by (2.4), and $\tilde{\mu}_n$, $\tilde{\nu}_n$, $\bar{\mu}_n$, $\bar{\nu}_n$, $\hat{\mu}_n$ and $\hat{\nu}_n$ are found in the same way. Since $\nu_n = 0$ for all n, those values are omitted.

Asymptotic estimates for all of these quantities are derived in the preceeding four sections. In Table 5 the relative errors of the estimates are presented. If E is an estimate for the quantity Q, we define the relative error to be (E-Q)/Q. The estimates for P_n, T_n, μ_n, $\tilde{P}_n$, $\tilde{T}_n$, $\tilde{\mu}_n$, $\bar{P}_n$, $\bar{T}_n$, $\bar{\mu}_n$, $\hat{P}_n$, $\hat{T}_n$ and $\hat{\mu}_n$ are to second order in 1/n, so that the relative errors are $O(1/n^2)$. The estimates for $\hat{\nu}_n$, $\bar{\nu}_n$ and $\tilde{\nu}_n$ are only to first order in 1/n, since the leading terms added out exactly, and so the relative errors are O(1/n). Since the estimate for μ_n is exact the relative error is always zero, and those values have been omitted.

The computations were programmed on a PDP11/45 by A. Nymeyer while employed under an A.R.G.C. grant. Multiple precision integer arithmetic was employed for the exact results, so no errors should have been introduced by arithmetic operations in the course of the computations.

n	P_n	$\tilde{P}_n$	$\bar{P}_n$	$\hat{P}_n$
1	1	1	1	1
2	1	2	1	2
3	3	7	6	10
4	15	41	39	83

Table 1. Exact Numbers of Planted Trees.

105	5
346	
390	
946	
945	6
3797	
4815	
13772	
10395	7
51157	
73080	
2 44315	
1 35135	8
8 16356	
13 04415	
51 13208	
20 27025	9
150 50581	
268 47450	
1233 42166	
344 59425	10
3147 26117	
6255 28575	
33695 68817	
21345 80466 76875	15
5 52134 63465 43307	
18 60930 96068 88000	
222 49860 76835 28550	
82 00794 53263 78915 59375	20
6030 06160 20371 20901 60876	
34162 88904 24238 71173 19375	
9 07848 89714 77640 85142 98308	
1 19256 81927 74434 12353 99076 40625	25
249 60937 77924 41570 77429 36703 12521	
2369 38660 81672 49563 88051 41829 06250	
1 40020 31727 81284 38254 77005 19138 11921	
4951 79769 00801 98183 90136 61171 60891 40625	30
29 52055 00537 50105 69454 33962 29379 21916 46717	
468 78078 22176 09468 90311 78233 86499 38771 09375	
61622 23725 36992 56577 22418 48868 47524 62475 40292	
488 96013 03686 63401 54392 27834 73071 78464 62136 71875	35
8 30555 12190 54585 07530 50112 37947 61629 76250 03010 43432	
220 45043 83215 09719 97185 12351 30927 75473 42274 65921 87500	
64469 94776 42821 80627 78241 11507 01198 15512 97655 72710 58200	
100 98473 64737 86927 09053 02433 22159 25040 62302 66320 27246 09375	40
4 88849 59488 48468 12925 66057 13518 13962 35628 62289 55846 36023 59701	
216 76354 20265 56596 78464 48363 20829 15417 44296 91092 26936 17246 09375	
1 41043 29642 04081 01696 47860 11232 21536 87652 57986 44515 59972 16529 95683	

Table 1 concluded

n	T_n	$\tilde{T}_n$	$\bar{T}_n$	$\hat{T}_n$
2	1	1	1	1
3	1	4	1	4
4	3	16	7	20
5	15	85	45	155
6	105	646	465	1716
7	945	6664	5775	24654
8	10395	86731	88515	4 34155
9	1 35135	13 54630	15 88545	90 43990
10	20 27025	246 07816	328 52925	2174 57456
15	790 58535 80625	26178 52548 56584	62467 65138 36375	8 98476 50088 10504

Table 2. Exact Numbers of Free Trees

2 21643 09547 66997 71875	20
205 49784 36233 45183 15851	
842 45483 44899 82007 04375	
26682 69789 26535 31333 74235	
2537 37913 35626 25794 76576 09375	25
6 64178 57351 59954 30396 13789 93560	
46 17076 71760 51374 73244 85886 40625	
3235 11311 52182 42235 34626 34054 77330	
86 87364 36856 17511 99826 95810 02822 65625	30
64436 62064 76134 50310 45988 54233 29473 09166	
7 55042 77813 17552 31463 00849 57120 96622 65625	
1172 90718 32815 25159 49765 43242 31672 39084 02156	
7 29791 23935 62140 32155 10863 20493 60872 60628 90625	35
15368 80705 80797 59840 80975 60737 16744 65780 96218 14259	
3 02580 48379 86006 84198 80394 40559 05889 76656 22199 21875	
1043 30147 98740 01154 14781 09291 78957 78854 81627 17151 95629	
1 31149 00840 75154 89727 96135 49638 43182 34575 35926 23730 46875	40
7850 74372 68277 04558 58143 14560 59876 66156 89466 40712 27350 48676	
2 59200 36605 50234 64023 20968 12167 26967 31842 95195 74464 69511 71875	
1985 12629 86757 57145 96126 26837 39795 87614 49844 23808 78544 49936 30160	

Table 2 concluded

μ_n	$\tilde{\mu}_n$	$\bar{\mu}_n$	$\hat{\mu}_n$	n
2	2.00000 00000 00000	2.00000 00000 00000	2.00000 00000 00000	2
4	2.50000 00000 00000	4.00000 00000 00000	2.50000 00000 00000	3
6	3.50000 00000 00000	4.85714 28571 42857	3.60000 00000 00000	4
8	5.05882 35294 11765	6.66666 66666 66667	4.96774 19354 83871	5
10	6.87616 09907 12074	8.06451 61290 32258	6.34149 18414 91841	6
12	8.67647 05882 35294	9.67272 72727 27273	7.70017 03577 51278	7
14	10.4124 93802 67724	11.2099 64412 81139	9.05386 78582 53389	8
16	12.1104 72970 47902	12.7804 87804 87805	10.4063 90321 08616	9
18	13.7896 80969 65615	14.3440 24162 23213	11.7588 57383 11957	10
28	22.0911 28106 38671	22.1987 85618 08196	18.5248 31560 70087	15
38	30.3436 73372 50291	30.0699 42364 27419	25.2950 68606 39104	20
48	38.5816 66700 12733	37.9475 81912 66192	32.0672 91540 57386	25
58	46.8133 11991 00284	45.8283 62637 25108	38.8405 35372 58541	30
68	55.0416 12908 97430	53.7109 00716 17330	45.6143 66223 76070	35
78	63.2679 34337 78013	61.5945 20432 18619	52.3885 64087 49734	40

Table 3. Mean Order of Trees with Fixed Magnitude

$\tilde{\nu}_n$	$\bar{\nu}_n$	$\hat{\nu}_n$	n
0.00000 00000 00000	0.00000 00000 00000	0.00000 00000 00000	2
0.75000 00000 00000	0.00000 00000 00000	0.75000 00000 00000	3
2.25000 00000 00000	0.97959 18367 34694	1.84000 00000 00000	4
3.82006 92041 52248	0.88888 88888 88888	2.54734 65140 47868	5
4.55122 73672 70844	1.67325 70239 33404	3.12114 55599 56749	6
4.82810 18289 66879	1.85652 89256 19834	3.71248 71941 46887	7
5.13028 35400 88531	2.35093 27389 47078	4.31499 16329 39918	8
5.52204 49273 73383	2.69839 38132 06426	4.92129 87937 60629	9
5.96404 81520 58989	3.10673 00739 16410	5.52838 11503 75196	10
8.38429 32248 87521	5.04591 99744 39442	8.55839 33406 31767	15
10.8925 61215 87343	6.98141 89245 42325	11.5800 79720 01440	20
13.4236 40224 09535	8.91280 60113 55968	14.5975 78794 19007	25
15.9642 03499 83919	10.8420 07039 11216	17.6128 72291 44108	30
18.5096 36929 00962	12.7699 13435 93193	20.6268 81592 54253	35
21.0579 06897 65325	14.6969 92847 02091	23.6400 81773 08085	40

Table 4. Variance of the Order of Trees with Fixed Magnitude

Quantity	10	20	30	40
P_n	-.00198255	-.000491952	-.000218101	-.000122529
$\tilde{P}_n$	-.00566099	-.00134128	-.000586665	-.000327466
$\bar{P}_n$	.00130770	.000311656	.000136469	.0000761654
$\hat{P}_n$	.000919071	.000202635	.0000861797	.0000474024
T_n	-.0290372	-.00739040	-.00330385	-.00186380
$\tilde{T}_n$	-.0716409	-.0173360	-.00765566	-.00429481
$\bar{T}_n$	-.0107192	-.00277488	-.00125092	-.000708669
$\hat{T}_n$	-.0296345	-.00734383	-.00326257	-.00183542
$\tilde{\mu}_n$	.0134911	.00244101	.000992320	.000535098
$\bar{\mu}_n$	-.00636135	-.00145205	-.000623642	-.000344765
$\hat{\mu}_n$	-.00251239	-.000600032	-.000259226	-.000143563
$\tilde{\nu}_n$	-.142979	-.0615036	-.0394795	-.0290935
$\bar{\nu}_n$	.238924	.102642	.0650247	.0475617
$\hat{\nu}_n$	.0892020	.0399797	.025646	.0188668

Table 5. Relative Error of Asymptotic Estimates

7. RELATED RESULTS

In [5] the numbers of phylogenetic trees were studied with no restriction on the degrees of the points. There it was noted that the mean and the variance of the order were both O(n). Therefore as $n \to \infty$ the distribution of orders in trees of magnitude n becomes gradually sharper as a percentage of mean value. This is also true of all four cases considered in Sections 2-5.

The methods of the present paper have been applied to other classes of trees which are relevant to the formation of phylogenetic diagrams in biology. These classes are determined by applying certain combinations of the following conditions: no points of degree 2 are allowed; no interior points are labelled; the labelling is 1-1. It is planned to present those results elsewhere.

REFERENCES

[1] E.A. Bender, Asymptotic methods in enumeration, *SIAM Rev.* 16 (1974), 485-515.

[2] L.L. Cavalli-Sforza and A.W.F. Edwards, Phylogenetic analysis. Models and estimation procedures, *Amer. J. Human Genet.* 19 (1967), 233-257 and *Evolution* 21 (1967), 550-570.

[3] J. Felsenstein, The numbers of evolutionary trees, *Syst. Zool.* 27 (1978), 27-33.

[4] L.R. Foulds, David Penny and M.D. Hendy, A Graph Theoretic Approach to the Development of Minimal Phylogenetic Trees, *J. Mol. Evol.* 13 (1979), 127-150.

[5] L.R. Foulds and R.W. Robinson, Determining the asymptotic numbers of phylogeneti trees. *Combinatorial Mathematics VII*, Lecture Notes in Mathematics 829 (Springer, Berlin, 1980), 110-126.

[6] F. Harary and E.M. Palmer, *Graphical Enumeration* (Academic Press, New York, 1973).

[7] F. Harary, R.W. Robinson and A.J. Schwenk, Twenty step algorithm for determining the asymptotic number of trees of various species, *J. Austral. Math. Soc. Ser. A* 20 (1975), 483-503.

[8] E.F. Harding, The probabilities of rooted tree-shapes generated by random bifurcation, *Adv. Appl. Prob.* 3 (1971), 44-77.

[9] J.W. Moon, Various proofs of Cayley's formula for counting trees. *A Seminar on Graph Theory* (F. Harary, ed, Holt, Rinehart and Winston, New York, 1967), 70-78.

[10] H. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Arch. Math. Phys.* 27 (1918), 742-744.

Operations Research
University of Canterbury
Christchurch
New Zealand

Department of Mathematics
University of Newcastle
New South Wales 2308