

# CUTTING DOWN RANDOM TREES

A. MEIR and J. W. MOON

(Received 20 January 1969)

Communicated by G. B. Preston

## 1. Introduction

Let  $T_n$  denote a tree with  $n (\geq 2)$  labelled points: we assume  $T_n$  is rooted at a given point  $x$ , say the point labelled 1 (see [3] for definitions not given here). If we remove some edge  $e$  of  $T_n$ , then  $T_n$  falls into two subtrees one of which, say  $T_k$ , contains the root  $x$ . If  $k \geq 2$  we can remove some edge of  $T_k$  and obtain an even smaller subtree of  $T_n$  that contains  $x$ . If we repeat this process we will eventually obtain the subtree consisting of  $x$  itself. Let  $\lambda = \lambda(T_n)$  denote the number of edges removed from  $T_n$  before the root  $x$  is isolated. Our main object here is to determine the expected value  $\mu(n)$  and variance  $\sigma^2(n)$  of  $\lambda(T_n)$  under the assumptions (1)  $T_n$  is chosen at random from the set of  $n^{n-2}$  trees with  $n$  labelled points that are rooted at point  $x$ , and (2) at each stage the edge removed is chosen at random from the edges of the remaining subtree containing  $x$ . It follows from our results that  $\mu(n) \sim (\frac{1}{2}\pi n)^{\frac{1}{2}}$  and  $\sigma^2(n) \sim (2 - \frac{1}{2}\pi)n$  as  $n$  tends to infinity. We also consider the corresponding problem for forests of rooted trees and for trees in which the degree of the root is specified. We are indebted to Professor Alistair Lachlan for suggesting the original problem to us.

## 2. Some identities

In what follows we shall make use of certain properties of the function

$$Y = \sum_{n=1}^{\infty} n^{n-2} \frac{x^n}{(n-1)!};$$

it can be shown that

$$(1) \quad Y = xe^Y$$

and that

$$(2) \quad Y^h = \sum_{n=h}^{\infty} hn^{n-h-1} \frac{x^n}{(n-h)!},$$

for  $h = 1, 2, \dots$ . These relations can be established by Lagrange's inversion formula (see [4; p. 301] or [5]) or by purely combinatorial arguments (see [2]). It follows from (1) that

$$(3) \quad Y' = Y/x(1-Y).$$

### 3. The mean and variance of $\lambda(T_n)$

If  $0 \leq l \leq n-1$ , let  $P(n, l)$  denote the probability that  $\lambda(T_n)$  equals  $l$  under the assumptions stated earlier. We adapt the convention that  $P(n, 0)$  equals one if  $n = 1$  and zero otherwise.

Suppose we remove one of the  $n-1$  edges of a random tree  $T_n$  and obtain a subtree  $T_k$  containing the root  $x$ . There are  $\binom{n-1}{k-1}$  possible choices for the  $k-1$  points of  $T_k$  other than  $x$  and, having chosen these points, there are  $k^{k-2}$  possible trees  $T_k$ . There are  $(n-k)^{n-k-2}$  trees that could be formed on the remaining  $n-k$  points and the edge we removed could have joined any of the  $k$  points of  $T_k$  to any of the remaining  $n-k$  points. Consequently,

$$(4) \quad P(n, l) = \frac{1}{(n-1)n^{n-2}} \sum_{k=l}^{n-1} \binom{n-1}{k-1} P(k, l-1)k^{k-1}(n-k)^{n-k-1}$$

for  $1 \leq l \leq n-1$ .

If  $\mu(n)$  denotes the expected value of  $\lambda$ , then

$$(5) \quad \begin{aligned} \mu(n) &= \sum_{l=1}^{n-1} lP(n, l) \\ &= \frac{1}{(n-1)n^{n-2}} \sum_{k=1}^{n-1} \binom{n-1}{k-1} k^{k-1}(n-k)^{n-k-1} \{\mu(k) + 1\} \end{aligned}$$

by equation (4), for  $n = 2, 3, \dots$  ( $\mu(1) = 0$ , by definition). Let

$$M = \sum_{n=2}^{\infty} \mu(n)n^{n-2} \frac{x^n}{(n-1)!}.$$

(Since  $\mu(n) \leq n-1$ , this certainly converges if  $|x| < e^{-1}$ , as do all other generating functions we consider here.) If we multiply relation (5) by  $(n-1)n^{n-2}x^n/(n-1)!$  and sum over  $n$ , we obtain the relation

$$xM' - M = xM'Y + xY'Y$$

between  $M$  and  $Y$ . This may be rewritten, using (3), as

$$(M/Y)' = Y'(1-Y)^{-1}.$$

We conclude, therefore, that

$$(6) \quad M = -Y \ln(1-Y) = \sum_{j=1}^{\infty} \frac{1}{j} Y^{j+1}.$$

(The constant of integration must be zero since  $\mu(1) = 0$ .) When we equate the coefficients of  $x^n$  in (6), using (2), we obtain the following formula for  $\mu(n)$ . (In what follows we adopt the convention that  $(m)_0 = 1$  and  $(m)_r = m(m-1) \cdots (m-r+1)$  for positive integers  $r$ .)

THEOREM 1. If  $n \geq 2$ , then

$$\mu(n) = \sum_{j=1}^{n-1} \frac{j+1}{j} \cdot \frac{(n-1)_j}{n^j}.$$

Instead of determining the variance of  $\lambda$  directly, we determine  $\tau(n)$ , the expected value of  $\lambda(\lambda-1)$ . The variance  $\sigma^2(n)$  of  $\lambda$  is then given by the formula  $\sigma^2(n) = \tau(n) + \mu(n) - \mu^2(n)$ . It follows from equation (4) that

$$\begin{aligned} \tau(n) &= \sum_{l=1}^{n-1} l(l-1)P(n, l) \\ (7) \quad &= \frac{1}{(n-1)n^{n-2}} \sum_{k=1}^{n-1} \binom{n-1}{k-1} k^{k-1}(n-k)^{n-k-1} \{\tau(k) + 2\mu(k)\}, \end{aligned}$$

for  $n = 2, 3, \dots$  ( $\tau(1) = 0$ , by definition). If we let

$$S = \sum_{n=2}^{\infty} \tau(n)n^{n-2} \frac{x^n}{(n-1)!},$$

then (7) implies that

$$xS' - S = xS'Y + 2xM'Y.$$

This may be rewritten, using (3) and (6), as

$$(S/Y)' = 2Y'\{Y(1-Y)^{-2} - \ln(1-Y) \cdot (1-Y)^{-1}\}.$$

Consequently,

$$\begin{aligned} S &= Y\{2Y(1-Y)^{-1} + 2\ln(1-Y) + \ln^2(1-Y)\} \\ &= 2 \sum_{j=2}^{\infty} \left(1 - \frac{1}{j} + \frac{\alpha_j}{j}\right) Y^{j+1}, \end{aligned}$$

where  $\alpha_1 = 0$  and

$$\alpha_j = \sum_{i=1}^{j-1} \frac{1}{i}$$

for  $j = 2, 3, \dots$ . If we equate the coefficients of  $x^n$  in (8) we obtain the following formula for  $\tau(n)$ .

THEOREM 2. If  $n \geq 2$ , then

$$\tau(n) = 2 \sum_{j=1}^{n-1} \left(1 - \frac{1}{j} + \frac{\alpha_j}{j}\right) (j+1) \frac{(n-1)_j}{n^j}.$$

We shall discuss the asymptotic behaviour of  $\mu(n)$  and  $\sigma^2(n)$  in § 6.

We remark that the recurrence relation (4) can also be used to express the generating functions

$$P_l = \sum_{n=l+1}^{\infty} P(n, l)n^{n-2} \frac{x^{n-1}}{(n-1)!},$$

$l = 1, 2, \dots$ , in terms of  $Y$ . For example,

$$P_1 = (Y - \frac{1}{2}Y^2),$$

$$P_2 = (Y^2 - \frac{5}{6}Y^3 + \frac{1}{8}Y^4),$$

and

$$P_3 = (Y^3 - \frac{13}{12}Y^4 + \frac{7}{24}Y^5 - \frac{1}{48}Y^6),$$

from which it follows, using (2), that

$$P(n, 1) = (n-1)^{n-3}/n^{n-2},$$

$$P(n, 2) = (5n+1)(n-2)(n-1)^{n-5}/2n^{n-2},$$

and

$$P(n, 3) = (103n^2 + 73n + 4)(n-3)(n-2)(n-1)^{n-7}/24n^{n-2}.$$

These generating function relations, however, seem to be too complicated in general to be particularly useful by themselves.

#### 4. Cutting down random forests

If  $1 \leq t \leq n$ , then there are  $tn^{n-t-1}$  forests  $F_n$  of  $t$  trees with a total of  $n$  labelled points such that  $t$  given points, say the points labelled  $1, 2, \dots, t$ , belong to different trees (see [1] or [5]). We may consider these  $t$  points as roots of the trees in  $F_n$ . Let  $\mu(n, t)$  and  $\sigma^2(n, t)$  denote the mean and variance of the number  $\lambda$  of edges that must be removed from a random forest  $F_n$  of  $t$  rooted trees before isolating the  $t$  roots, where at each stage the edge removed is chosen at random from the edges of the remaining subtrees containing the  $t$  roots.

The argument used to establish equation (5) can easily be extended to show that

$$\mu(n, t) = \frac{1}{(n-t)tn^{n-t-1}} \sum_{k=t+1}^{n-1} \binom{n-t}{k-t} tk^{k-t}(n-k)^{n-k-1} \{\mu(k, t) + 1\}$$

for  $1 \leq t \leq n-1$  ( $\mu(n, t) = 0$  otherwise). If we let

$$M_t = \sum_{n=t+1}^{\infty} \mu(n, t)tn^{n-t-1} \frac{x^n}{(n-1)!},$$

then it follows from this recurrence relation that

$$xM'_t - tM_t = xM'_t Y + x(Y')' \cdot Y,$$

or equivalently, that

$$(M_t/Y^t)' = tY'(1-Y)^{-1}.$$

Therefore,

$$M_t = -tY^t \ln(1-Y) = t \sum_{j=1}^{\infty} \frac{1}{j} Y^{j+t}$$

This implies the following generalization of theorem 1.

**THEOREM 3.** If  $1 \leq t \leq n-1$ , then

$$\mu(n, t) = \sum_{j=1}^{n-t} \frac{j+t}{j} \cdot \frac{(n-t)_j}{n^j}.$$

Let  $\tau(n, t)$  denote the expected value of  $\lambda(\lambda-1)$  for forests  $F_n$  of  $t$  rooted trees so that  $\sigma^2(n, t) = \tau(n, t) + \mu(n, t) - \mu^2(n, t)$ . The argument that led to equation (7) can be extended to show that

$$\tau(n, t) = \frac{1}{(n-t)tn^{n-t-1}} \sum_{k=t+1}^{n-1} \binom{n-t}{k-t} tk^{k-t}(n-k)^{n-k-1} \{\tau(k, t) + 2\mu(k, t)\}$$

for  $1 \leq t \leq n-1$ . If we let

$$S_t = \sum_{n=t+1}^{\infty} \tau(n, t)tn^{n-t-1} \frac{x^n}{(n-t)!},$$

then it follows from this recurrence relation that

$$xS'_t - tS_t = xS'_t Y + 2xM'_t Y,$$

or equivalently, that

$$(S_t/Y^t)' = 2tY'\{Y(1-Y)^{-2} - t \ln(1-Y) \cdot (1-Y)^{-1}\}.$$

Therefore,

$$\begin{aligned} S_t &= tY^t\{2Y(1-Y)^{-1} + 2 \ln(1-Y) + t \ln^2(1-Y)\} \\ &= 2t \sum_{j=2}^{\infty} \left(1 - \frac{1}{j} + \frac{t\alpha_j}{j}\right) Y^{j+t}. \end{aligned}$$

This implies the following generalization of theorem 2.

**THEOREM 4.** If  $1 \leq t \leq n-1$ , then

$$\tau(n, t) = 2 \sum_{j=1}^{n-t} \left(1 - \frac{1}{j} + \frac{t\alpha_j}{j}\right) (j+t) \frac{(n-t)_j}{n^j}.$$

Notice that  $M_t = tY^{t-1}M_1$  and  $S_t = tY^{t-1}S_1$ ; thus one can express the numbers  $\mu(n, t)$  and  $\tau(n, t)$  in terms of the numbers  $\mu(m) = \mu(m, 1)$  and  $\tau(m) = \tau(m, 1)$ , for  $2 \leq m \leq n-t+1$ .

### 5. Cutting down trees whose roots have specified degrees

If  $1 \leq t \leq n-1$ , then there are  $\binom{n-2}{t-1}(n-1)^{n-t-1}$  trees  $T_n$  in which the point  $x$  is joined to exactly  $t$  other points (see [2]); let  $D(n, t)$  and  $V(n, t)$  denote the mean and variance of the number  $\lambda$  of edges that must be removed from such a tree before isolating the root  $x$ .

It is not difficult to show that

$$\begin{aligned}
 D(n, t) = & \frac{1}{\binom{n-2}{t-1} (n-1)^{n-t}} \left\{ \sum_{k=t+1}^{n-1} \binom{n-1}{k-1} \binom{k-2}{t-1} (k-1)^{k-t} (n-k)^{n-k-1} (D(k, t) + 1) \right. \\
 (9) \quad & \left. + \sum_{k=t}^{n-1} \binom{n-1}{k-1} \binom{k-2}{t-2} (k-1)^{k-t} (n-k)^{n-k-1} (D(k, t-1) + 1) \right\}
 \end{aligned}$$

if  $1 \leq t \leq n-1$  (otherwise  $D(n, t) = 0$ ). The argument is essentially the same as that used to establish equation (5) except for one added complication. We obtain a subtree  $T_k$  containing the root  $x$  by removing some edge  $e$  of the random tree  $T_n$  (in which  $x$  is joined to  $t$  other points). If  $e$  is not incident with  $x$ , then  $x$  is joined to  $t$  other points of  $T_k$  and  $e$  joins one of the  $k-1$  points of  $T_k$  other than  $x$  to one of the  $n-k$  points not in  $T_k$ ; but if  $e$  is incident with  $x$ , then  $x$  is joined only to  $t-1$  other points of  $T_k$  and  $e$  joins  $x$  to one of the  $n-k$  points not in  $T_k$ . The two sums on the right hand side of equation (9) correspond to these two possibilities.

If we let

$$D_t = \sum_{n=t+1}^{\infty} D(n, t) \binom{n-2}{t-1} (n-1)^{n-t-1} \frac{x^{n-1}}{(n-1)!},$$

then equation (9) implies that

$$xD'_t = xD'_t Y + D_{t-1} Y + x(Y^t)' Y/t! + Y^{t-1} \cdot Y/(t-1)!,$$

or equivalently, that

$$D'_t = \frac{1}{(t-1)!} Y^{t-1} \cdot Y'(1-Y)^{-1} + D_{t-1} Y'$$

for  $t = 1, 2, \dots$ . Since  $D_t(0) = 0$  for all  $t$  and  $D_0 = 0$ , by definition, it follows by induction that

$$D_t = \frac{-1}{(t-1)!} Y^{t-1} \ln(1-Y) = \frac{1}{(t-1)!} \sum_{j=1}^{\infty} \frac{1}{j} Y^{j+t-1}$$

This implies the following result.

**THEOREM 5.** *If  $0 \leq t \leq n-1$ , then*

$$D(n+1, t+1) = \sum_{j=1}^{n-t} \frac{j+t}{j} \cdot \frac{(n-1-t)_{j-1}}{n^{j-1}}$$

**COROLLARY 5.1.** *If  $1 \leq t \leq n-1$ , then*

$$\mu(n, t) = \left(1 - \frac{t}{n}\right) D(n+1, t+1).$$

This corollary follows from theorems 3 and 5. As  $t$  increases, for fixed values of  $n$ ,  $D(n+1, t+1)$  increases to  $n$  while  $\mu(n, t)$  eventually decreases to 1.

Other recurrence relations can also be written. For example, it is not difficult to see that

$$t! \binom{n-2}{t-1} (n-1)^{n-t-1} D(n, t) = \sum \binom{n-1}{a_1, \dots, a_t} a_1^{a_1-1} \dots a_t^{a_t-1} \{D(a_1+1, 1) + \dots + D(a_t+1, 1)\},$$

where the sum is over all solutions in positive integers to the equation  $a_1 + \dots + a_t = n-1$ .

Let  $U(n, t)$  denote the expected value of  $\lambda(\lambda-1)$  for trees  $T_n$  in which the root  $x$  is joined to exactly  $t$  other points so that  $V(n, t) = U(n, t) + D(n, t) - D^2(n, t)$ . It can be shown, by essentially the same arguments as have been used before, that

$$U(n, t) = \frac{1}{\binom{n-2}{t-1} (n-1)^{n-t}} \left\{ \sum_{k=t+1}^{n-1} \binom{n-1}{k-1} \binom{k-2}{t-1} (k-1)^{k-t} (n-k)^{n-k-1} (U(k, t) + 2D(k, t)) + \sum_{k=t}^{n-1} \binom{n-1}{k-1} \binom{k-2}{t-2} (k-1)^{k-t} (n-k)^{n-k-1} (U(k, t-1) + 2D(k, t-1)) \right\}$$

for  $1 \leq t \leq n-1$  (otherwise  $U(n, t) = 0$ ). If we let

$$U_t = \sum_{n=t+1}^{\infty} U(n, t) \binom{n-2}{t-1} (n-1)^{n-t-1} \frac{x^{n-1}}{(n-1)!},$$

then it follows from this recurrence relation that

$$xU'_t = xU'_t Y + 2xD'_t Y + U_{t-1} Y + 2D_{t-1} Y,$$

or equivalently, that

$$U'_t = \frac{2}{(t-1)!} Y^t \cdot Y'(1-Y)^{-2} - \frac{2}{(t-2)!} Y^{t-2} Y' \ln(1-Y) \cdot (1-Y)^{-1} + U_{t-1} Y'$$

for  $t = 1, 2, \dots$ . Since  $U_t(0) = 0$  for all  $t$  and  $U_0 = 0$ , by definition, it follows by induction that

$$U_t = \frac{1}{(t-1)!} \{2Y^t(1-Y)^{-1} + 2Y^{t-1} \ln(1-Y) + (t-1)Y^{t-2} \ln^2(1-Y)\} = \frac{2}{(t-1)!} \sum_{j=0}^{\infty} \left(1 - \frac{1}{j+1} + \frac{(t-1)\alpha_{j+2}}{j+2}\right) Y^{j+t}.$$

This implies the following result.

**THEOREM 6.** *If  $1 \leq t \leq n-1$ , then*

$$U(n, t) = 2 \sum_{j=0}^{n-1-t} \left( 1 - \frac{1}{j+1} + \frac{(t-1)\alpha_{j+2}}{j+2} \right) (j+t) \frac{(n-1-t)_j}{(n-1)^j}$$

**COROLLARY 6.1.** *If  $1 \leq t \leq n-1$ , then*

$$\tau(n, t) = (n-t)(n-t-1)n^{-2}U(n+1, t+2) - 2 \sum_{j=2}^{n-t} \alpha_{j-1} \frac{j+t}{j} \frac{(n-t)_j}{n^j}.$$

This corollary follows from theorems 4 and 6.

Notice that

$$(t-1)! \cdot D_t = Y^{t-1} D_1 \text{ and } (t-1)! \cdot U_t = Y^{t-1} U_1 + (t-1)Y^{t-2} \ln^2(1-Y);$$

thus one can express the numbers  $D(n, t)$  and  $U(n, t)$  in terms of the numbers  $D(m, 1)$  and  $U(m, 1)$  for  $2 \leq m \leq n-t+1$ .

### 6. Asymptotic results

We now determine the asymptotic behaviour of  $\mu(n, t)$  and  $D(n, t)$  for large  $n$  provided that  $t$  is not too large. Some of the following inequalities will be valid only when  $n$  is larger than some absolute constant but we will not repeat this qualification every time.

Let  $0 \leq t \leq n-1$ ; then

$$D(n+1, t+1) = \sum_{k=0}^{n-t-1} \left( 1 + \frac{t}{k+1} \right) c_k,$$

where

$$c_k = \prod_{j=1}^k \left( 1 - \frac{t+j}{n} \right).$$

Since  $1-x < e^{-x}$  for  $x > 0$ , it follows that for all  $k$

$$c_k \leq e^{-k^2/2n};$$

in particular, if  $k \geq K = 2(n \log n)^{\frac{1}{2}}$ , then

$$c_k \leq \frac{1}{n^2}.$$

Therefore,

$$\sum_{k=0}^{n-t-1} c_k \leq 1 + n^{\frac{1}{2}} \int_0^\infty e^{-x^2/2} dx = 1 + (\frac{1}{2}\pi n)^{\frac{1}{2}}$$

and

$$\sum_{k=0}^{n-t-1} \frac{t}{k+1} c_k \leq t \sum_{k=0}^{[K]-1} \frac{1}{k+1} + \frac{t}{n^2} \cdot \frac{n}{K} \leq \frac{1}{2}t(\log n + \log \log n + 4).$$

Consequently, if  $n$  is sufficiently large and  $0 \leq t \leq n-1$ , then

$$(10) \quad D(n+1, t+1) \leq (\frac{1}{2}\pi n)^{\frac{1}{2}} + \frac{1}{2}t(\log n + \log \log n + 4) + 1.$$

On the other hand,  $1-x \geq e^{-x/(1-x)}$  if  $0 < x < 1$ ; thus if

$$t+k \leq (n \log n)^{\frac{1}{2}} = N,$$

then

$$c_k \geq e^{-(t+k+1)^2/2(n-N)}.$$

Hence,

$$\sum_{k=0}^{[N]-t} c_k \geq (n-N)^{\frac{1}{2}} \int_{\alpha}^{\beta} e^{-x^2/2} dx$$

where  $\alpha = (t+1)/(n-N)^{\frac{1}{2}}$  and  $\beta = N/(n-N)^{\frac{1}{2}}$ . Now

$$\int_{\beta}^{\infty} e^{-x^2/2} dx \leq \beta^{-1} \int_{\beta}^{\infty} x e^{-x^2/2} dx = \beta^{-1} e^{-\beta^2/2} \leq (n \log n)^{-\frac{1}{2}},$$

$$\int_0^{\alpha} e^{-x^2/2} dx \leq \alpha,$$

and  $(1-z)^{\frac{1}{2}} > 1 - \frac{3}{4}z$  for small  $z$ ; therefore,

$$\begin{aligned} \sum_{k=0}^{[N]-t} c_k &\geq (n-N)^{\frac{1}{2}} \{ (\frac{1}{2}\pi)^{\frac{1}{2}} - (n \log n)^{-\frac{1}{2}} - (t+1)/(n-N)^{\frac{1}{2}} \} \\ &\geq (\frac{1}{2}\pi n)^{\frac{1}{2}} - \frac{3}{4}(\frac{1}{2}\pi \log n)^{\frac{1}{2}} - (\log n)^{-\frac{1}{2}} - (t+1) \\ &\geq (\frac{1}{2}\pi n)^{\frac{1}{2}} - (\log n)^{\frac{1}{2}} - t. \end{aligned}$$

If  $t+k+1 \leq 2(n/\log n)^{\frac{1}{2}} = M$ , then

$$c_k \geq e^{-M^2/2(n-M)} \geq 1 - 3/\log n;$$

so if  $t \leq \frac{1}{2}M$ , then

$$\begin{aligned} \sum_{k=0}^{[\frac{1}{2}M]} \frac{t}{k+1} c_k &\geq t(1 - 3/\log n) \sum_{k=0}^{[\frac{1}{2}M]} \frac{1}{k+1} \\ &\geq \frac{1}{2}t(\log n - \log \log n - 3). \end{aligned}$$

Consequently, if  $n$  is sufficiently large and  $0 \leq t \leq (n/\log n)^{\frac{1}{2}}$ , then

$$(11) \quad D(n+1, t+1) \geq (\frac{1}{2}\pi n)^{\frac{1}{2}} + \frac{1}{2}t(\log n - \log \log n - 5) - (\log n)^{\frac{1}{2}}.$$

When we combine inequalities (10) and (11) we obtain the following result.

**THEOREM 7.** If  $0 \leq i \leq (n/\log n)^{\frac{1}{2}}$ , then

$$D(n+1, t+1) = (\frac{1}{2}\pi n)^{\frac{1}{2}} + \frac{1}{2}t(\log n + O(\log \log n)) + O((\log n)^{\frac{1}{2}})$$

as  $n \rightarrow \infty$ .

It follows from corollary 5.1 that if  $1 \leq t \leq (n/\log n)^{\frac{1}{2}}$ , the conclusion of theorem 7 remains valid if  $D(n+1, t+1)$  is replaced by  $\mu(n, t)$ . Thus, for example,

$$D(n+1, 1) = (\frac{1}{2}\pi n)^{\frac{1}{2}} + O((\log n)^{\frac{1}{2}}),$$

$$D(n+1, 2) = (\frac{1}{2}\pi n)^{\frac{1}{2}} + \frac{1}{2} \log n + O((\log n)^{\frac{1}{2}}),$$

and

$$\mu(n) = (\frac{1}{2}\pi n)^{\frac{1}{2}} + \frac{1}{2} \log n + O((\log n)^{\frac{1}{2}}).$$

The average number of points joined to the root  $x$  in a random tree  $T_n$  is

$$2(n-1)/n = 2(1-n^{-1})$$

so perhaps it is not too surprising that  $\mu(n)$  is asymptotically equal to  $D(n, 2)$ .

We now determine the asymptotic behaviour of the variance  $\sigma^2(n)$  of  $\lambda$  for ordinary rooted trees  $T_n$ . If we use the identity

$$\sum_{k=1}^n k \frac{\binom{n}{k}}{n^k} = n,$$

then it follows from theorems 1 and 2 that

$$\tau(n) = 2(n-1) - 2\mu(n) + 2 \sum_{k=1}^{n-1} \alpha_k \frac{k+1}{k} \frac{(n-1)_k}{n^k}.$$

If  $s_n$  denotes the last sum, then

$$s_n = \sum_{k=1}^{n-1} \{\log k + O(1)\} \frac{k+1}{k} \frac{(n-1)_k}{n^k}$$

$$= \{\frac{1}{2} \log n + O(1)\} \mu(n) + \sum_{k=1}^{n-1} \log(kn^{-\frac{1}{2}}) \cdot \frac{(k+1)}{k} \frac{(n-1)_k}{n^k}$$

$$= \{\frac{1}{2} \log n + O(1)\} \mu(n) + O(n^{\frac{1}{2}}),$$

since  $(k+1)/k \leq 2$  and  $(n-1)_k/n^k < \exp(-k^2/2n)$  and so, if divided by  $n^{\frac{1}{2}}$ , the sum is bounded by an approximate Riemann sum for

$$\int_0^{\sqrt{n}} \log x \cdot e^{-x^2/2} dx.$$

Consequently,

$$\sigma^2(n) = \tau(n) + \mu(n) - \mu^2(n)$$

$$= 2n + \{\log n + O(1)\} \mu(n) - \mu^2(n) + O(n^{\frac{1}{2}}).$$

This and the estimate for  $\mu(n)$  given by theorem 7 imply the following result.

**THEOREM 8.**  $\sigma^2(n) = (2 - \frac{1}{2}\pi)n + O((n \log n)^{\frac{1}{2}})$  as  $n \rightarrow \infty$ .

A similar argument, the details of which we omit, shows that if  $t \leq n^{\frac{1}{2}}/\log n$  then

$$V(n + 1, t + 1) = (2 - \frac{1}{2}\pi)n + O(tn^{\frac{1}{2}} \log \log n) + O((n \log n)^{\frac{1}{2}})$$

and if  $t \leq (n/\log n)^{\frac{1}{2}}$  then

$$V(n + 1, t + 1) = (2 - \frac{1}{2}\pi)n + o(n)$$

as  $n \rightarrow \infty$ .

The above asymptotic formulae hold if  $V(n + 1, t + 1)$  is replaced by  $\sigma^2(n, t)$ .

### 7. Numerical results

The entries in the following tables were computed by Mr. J. Hubert.

TABLE 1: Values of  $\mu(n, t)$

$t/n$	2	3	4	5	6	7	8	9	10
1	1	1.6667	2.1875	2.624	3.0046	3.3451	3.6551	3.9409	4.2072
2		1	1.75	2.36	2.8796	3.3357	3.7444	4.1163	4.4586
3			1	1.8	2.4722	3.0554	3.5728	4.0394	4.4656
4				1	1.8333	2.5510	3.1836	3.7508	4.2662
5					1	1.8571	2.6094	3.2812	3.8894
6						1	1.875	2.6543	3.358
7							1	1.8889	2.69
8								1	1.9
9									1

TABLE 2: Values of  $D(n, t)$

$t/n$	2	3	4	5	6	7	8	9	10
1	1	1.5	1.8889	2.2188	2.5104	2.7747	3.0181	3.2450	3.4583
2		2	2.5	2.9167	3.28	3.6056	3.9026	4.1772	4.4336
3			3	3.5	3.9333	4.3194	4.6700	4.9925	5.2923
4				4	4.5	4.9444	5.3469	5.7164	6.0591
5					5	5.5	5.9524	6.3672	6.7514
6						6	6.5	6.9583	7.3827
7							7	7.5	7.9630
8								8	8.5
9									9

### References

[1] A. Cayley, 'A theorem on trees', *Quart. J. Math.* 23 (1889), 376–378. Collected papers, Cambridge, 13 (1897), 26–28.  
 [2] L. E. Clarke, 'On Cayley's formula for counting trees', *J. London Math. Soc.* 33 (1958), 471–475.

- [3] O. Ore, *Theory of Graphs*, (American Math. Soc. Colloq. Publi. 38, Providence, 1962).
- [4] G. Pólya and G. Szegő, *Aufgaben und Lehrsätze aus der Analysis*, Bd. 1 (Springer-Verlag, Berlin, Aufg. 3, 1964).
- [5] A. Rényi, 'Some remarks on the theory of trees', *Publi. Math. Inst. Hung. Acad. Sci.* 4(1959), 73–85.

The University of Alberta  
Edmonton, Alberta  
Canada