ON THE HEIGHT OF TREES

A. RÉNYI and G. SZEKERES [1]

(Received 12 October 1965)

1. Introduction

In this note we shall deal with the enumeration of labelled trees of given order and given height over a selected point.

An undirected graph is called a tree if it is connected and contains no cycle. If we select any two vertices P and Q of a tree T, there is evidently a uniquely determined path in T leading from P to Q. We shall call the length of this path (i.e. the number of edges in the path) the distance of P and Q in T and denote it by $d_T(P, Q)$. If a vertex P is distinguished as the root of T, we define the height of T over P as the length of the longest path in T starting from P; thus if $h_P(T)$ denotes the height of T over the root P, we have

$$h_P(T) = \max_{Q \in T} d_T(P, Q). \tag{1.1}$$

Let us consider the set $\mathscr{T}_n$ of all possible trees with n given labelled vertices $P_1, P_2, \cdots, P_n$. According to a classical result of Cayley [1] if t_n denotes the number of elements of $\mathscr{T}_n$, we have

$$t_n = n^{n-2} \tag{1.2}$$

Let $t_n(k)$ denote the number of those trees $T \in \mathscr{T}_n$ for which $h_{P_1}(T) \leqq k$. Clearly

$$t_1(0) = 1, \quad t_n(0) = 0 \quad \text{for } n > 1 \tag{1.3}$$

and

$$t_n(k) = t_n \quad \text{for} \quad k \geqq n-1. \tag{1.4}$$

J. Riordan [2] has shown that the enumerator

$$G_k(x) = \sum_{n=1}^{\infty} \frac{t_n(k)}{(n-1)!} x^n \qquad (k = 0, 1, \cdots) \tag{1.5}$$

[1] Part of this work was done at the fifth Summer Research Institute of the Australian Mathematical Society in Canberra. The first author is greatly indebted to the Australian National University for its generous support of his visit to the Institute.

satisfies the recursion formula

$$G_{k+1}(x) = x \exp G_k(x) \qquad (k = 0, 1, \cdots) \tag{1.6}$$

with

$$G_0(x) = x; \tag{1.7}$$

the latter follows from (1.3) and (1.5).

From the recursion formula (1.6) one can determine $t_n(k)$ for any k and n $(0 \leqq k \leqq n-1)$. For instance

$$G_1(x) = xe^x, \quad G_2(x) = xe^{xe^x}, \quad \text{etc.}, \tag{1.8}$$

and thus

$$t_n(1) = 1 \qquad (n = 1, 2, \cdots) \tag{1.9}$$

$$t_n(2) = \sum_{m=0}^{n-1} \binom{n-1}{m} m^{n-m-1} \qquad (n = 1, 2, \cdots) \tag{1.10}$$

and generally for $k \geqq 1$

$$t_n(k) = \sum_{\substack{m_1+\cdots+m_k=n-1 \\ m_i \geqq 0}} \frac{(n-1)!}{m_1!m_2!\cdots m_k!} m_1^{m_2} m_2^{m_3} \cdots m_{k-1}^{m_k} \qquad (i = 1, 2, \cdots, k). \tag{1.11}$$

In these formulae 0^0 always means 1.

In view of (1.2) and (1.4) one has

$$\lim_{k\to\infty} G_k(x) = \sum_{n=1}^{\infty} \frac{n^{n-2}}{(n-1)!} x^n \tag{1.12}$$

provided that the series on the right of (1.12) is convergent. But the series

$$y = \sum_{n=1}^{\infty} \frac{n^{n-2}}{(n-1)!} x^n \tag{1.13}$$

converges for $|x| \leqq 1/e$ and represents the inverse function of

$$x = ye^{-y}. \tag{1.14}$$

This equation also follows from (1.6) and (1.12).

Riordan [2] obtained the formula (1.6) as a special case of a more general result on enumerators of trees. In § 2 we shall give direct proofs of (1.6) and (1.11).

In § 3 we shall investigate the asymptotic distribution of

$$d_n(k) = t_n(k) - t_n(k-1), \tag{1.15}$$

i.e. the number of trees $T \in \mathscr{T}_n$ having exact height k over P_1. Let us mention that if $D(T)$ denotes the diameter of T (i.e. the length of the longest path in T) one has evidently

$$\tag{1.16} \tfrac{1}{2}D(T) \leqq \min_i h_{P_i}(T) \leqq h_{P_i}(T) \leqq \max_i h_{P_i}(T) \leqq D(T).$$

Thus the study of the distribution of $h_{P_i}(T)$ for $T \in \mathscr{T}_n$ gives us also some information on the distribution of $D(T)$.

Our thanks are due to F. Harary and J. W. Moon for calling our attention to the paper [2] of Riordan.

2. Proof of the recursion formula

To prove (1.6) we start from the formula

$$\tag{2.1} t_n(k) = \sum_{p=1}^{n-1} \binom{n-1}{p} \sum_{m_1+\cdots+m_p=n-1} \frac{(n-1-p)!}{(m_1-1)!\cdots(m_p-1)!} t_{m_1}(k-1)\cdots t_{m_p}(k-1).$$

(2.1) can be proved as follows: Let E denote the set of those points of $T \in \mathscr{T}_n$ which are directly connected (i.e. connected by an edge) with P_1. If p is the number of elements of E then $1 \leqq p \leqq n-1$ and denoting these points by $Q_1, \cdots, Q_p$, the points Q_i can be selected in $\binom{n-1}{p}$ ways. All the remaining $n-1-p$ points $P_m (P_m \neq Q_j, 1 \leqq j \leqq p; P_m \neq P_1)$ can be classified into p classes which are defined as follows: P_m lies in the j-th class $(j = 1, 2, \cdots, p)$ if the unique path from P_1 to P_m goes through Q_j. Clearly if the j-th class contains m_j-1 points, these points together with Q_j form a tree of order m_j and height $\leqq k-1$ over the basic point Q_j. Thus (2.1) follows.

Multiplication of (2.1) by $x^n/(n-1)!$ and summation for $n = 1, 2, \cdots$ leads immediately to (1.6). (1.11) can be deduced from (1.6) by using several times the power series of the exponential function. It can also be proved directly as follows:

Let $T \in \mathscr{T}_n$ be a tree the height of which over the basic point P_1 is $\leqq k$. Then all points of T different from P_1 can be classified into k classes, the j-th class $\mathscr{C}_j$ consisting of those points whose distance from P_1 is equal to j $(1 \leqq j \leqq k)$. Let m_j denote the number of points in the class $\mathscr{C}_j$ $(1 \leqq j \leqq k)$; then

$$\sum_{j=1}^{k} m_j = n-1.$$

If the numbers m_j are fixed, the distribution of the $n-1$ points in the classes $\mathscr{C}_j$ can be carried out in $(n-1)!/m_1! \cdots m_k!$ ways. Now evidently each point in the class $\mathscr{C}_1$ is directly connected with P_1, each point in $\mathscr{C}_2$ is directly connected with some point in $\mathscr{C}_1$ etc., each point in $\mathscr{C}_k$ is directly connected with some point of $\mathscr{C}_{k-1}$. As the connections can be established in $m_1^{m_2} m_2^{m_3} \cdots m_{k-1}^{m_k}$ different ways and by choosing these connections the tree T is completely determined, (1.11) follows.

For $d_n(k) = t_n(k) - t_n(k-1)$ the proof of (1.11) gives

$$(2.2) \qquad d_n(k) = \sum_{\substack{m_1+\cdots+m_k=n-1 \\ m_i \geq 1}} \frac{(n-1)!}{m_1! \cdots m_k!} m_1^{m_2} m_2^{m_3} \cdots m_{k-1}^{m_k}, \qquad (i = 1, \cdots, k).$$

If $d_n(k, m)$ denotes the number of trees $T \in \mathscr{T}_n$ for which $h_{P_1}(T) = k$ and in which there are exactly m points connected with P_1 by an edge then (2.2) gives

$$(2.3) \qquad d_n(k, m) = \frac{1}{m!} \sum_{\substack{m_2+\cdots+m_k=n-1-m \\ m_i \geq 1}} \frac{(n-1)!}{m_2! \cdots m_k!} m^{m_2} \cdots m_{k-1}^{m_k}.$$

From here the following recursion formula can be deduced:

$$(2.4) \qquad d_n(k, m) = \binom{n-1}{m} \sum_{p=1}^{n-1-m} m^p d_{n-m}(k-1, p).$$

Similarly if $t_n(k, m)$ is the number of those trees $T \in \mathscr{T}_n$ which have height $\leq k$ over P_1 and in which the number of points having distance k from P_1 equals m, then

$$(2.5) \quad t_n(k, m) = \frac{1}{m!} \sum_{m_1+\cdots+m_{k-1}=n-1-m} \frac{(n-1)!}{m_1! \cdots m_{k-1}!} m_1^{m_2} \cdots m_{k-2}^{m_{k-1}} m_{k-1}^{m},$$

and

$$(2.6) \qquad t_n(k, m) = \binom{n-1}{m} \sum_{p=1}^{n-1-m} p^m t_{n-m}(k-1, p).$$

Thus putting

$$(2.7) \qquad F_k(x, z) = \sum_{n=1}^{\infty} \sum_{m=0}^{n-1} \frac{t_n(k, m)}{(n-1)!} x^{n-m} z^m$$

we have the recursion formula

$$(2.8) \qquad F_k(x, z) = F_{k-1}(x, xe^z)$$

with $F_0(x, z) = z$. We obtain

$$F_1(x, z) = xe^z, \quad F_2(x, z) = xe^{xe^z}, \text{ etc.,}$$

hence

$$(2.9) \qquad F_{k+1}(x, z) = x \exp F_k(x, z)$$

further

$$(2.10) \qquad F_k(x, x) = F_{k+1}(x, 0) = G_k(x).$$

https://doi.org/10.1017/S1446788700004432 Published online by Cambridge University Press

3. The asymptotic distribution of $d_n(k)$

We consider now the asymptotic distribution of $d_n(k)$ when n and k are large. We shall make use of the generating function

$$G_k(x)-G_{k-1}(x) = \sum_{n=1}^{\infty} \frac{d_n(k)}{(n-1)!} x^n \tag{3.1}$$

where

$$G_0(x) = x, \quad G_k(x) = x \exp G_{k-1}(x) \qquad (k = 1, 2, \cdots) \tag{3.2}$$

by (1.5) and (1.15). From (3.2) it is seen that $G_k(z)-G_{k-1}(z)$ is an entire function and hence

$$\frac{d_n(k)}{(n-1)!} = \frac{1}{2\pi i}\int_{C+} \frac{G_k(z)-G_{k-1}(z)}{z^{n+1}}\, dz \tag{3.3}$$

where C is any circular path with centre 0. For the radius of C we shall take $r = e^{-1}$; this is the largest positive value of r for which the sequence $G_k(r)$, $k = 1, 2, \cdots$ tends to a finite limit, namely $\lim_{k\to\infty} G_k(e^{-1}) = 1$. Moreover if k is of order $\sqrt{n}$ which is the case of principal interest then the point e^{-1} lies very close to a saddle point of the integrand.

As in (2.9), write for a fixed complex number ζ

$$F_0(\zeta, z) = z, \tag{3.4}$$

$$F_k(\zeta, z) = \zeta \exp F_{k-1}(\zeta, z) \qquad (k = 1, 2, \cdots). \tag{3.5}$$

Thus $F_k(\zeta, z)$ is the k-th iterate of

$$F(\zeta, z) = \zeta e^z \quad (\zeta \text{ fixed}) \tag{3.6}$$

and

$$G_k(\zeta) = F_k(\zeta, \zeta) = F_{k+1}(\zeta, 0), \tag{3.7}$$

as in (2.10).

In the particular case of $\zeta = e^{-1}$, $z = 1$ is a fixed point with multiplier 1 of the function [2] $F(e^{-1}, z) = e^{z-1}$; in fact $F(e^{-1}, 1) = 1$, $F'(e^{-1}, 1) = 1$.

The sequence

$$\gamma_k = F_k(e^{-1}, e^{-1}) = G_k(e^{-1}), \qquad k = 0, 1, 2, \cdots \tag{3.8}$$

satisfies

$$\gamma_k = \exp(\gamma_{k-1}-1), \qquad k = 1, 2, \cdots \tag{3.9}$$

and has an asymptotic expansion

[2] $z = a$ is a fixed-point with multiplier μ of the function $F(z)$ if $F(a) = a$ and $F'(a) = \mu$. If $|\mu| < 1$, the fixed-point is called attractive. (Fatou [4], p. 186).

https://doi.org/10.1017/S1446788700004432 Published online by Cambridge University Press

$$\gamma_k \cong 1 - \frac{2}{k} + \tfrac{2}{3}\frac{\log k}{k^2} + \frac{c}{k^2} + \cdots \qquad (k \to \infty) \tag{3.10}$$

where c is a certain constant (see e.g. [3], Lemma 3, p. 247). For all other values of ζ on the circle

$$\zeta = e^{-1+it}, \quad -\pi \leqq t \leqq \pi,$$

$F(\zeta, z)$ has an attractive fixed-point [2] $\omega = u+iv$ with multiplier ω ($|\omega| < 1$), given by the equation

$$\omega = \zeta\, e^{\omega} = e^{-1+it+\omega}. \tag{3.11}$$

These fixed-points lie on the curve

$$u^2+v^2 = e^{2(u-1)}, \quad u \leqq 1 \tag{3.12}$$

with

$$\tan\,(v+t) = v/u = u^{-1}\,(e^{2(u-1)}-u^2)^{\frac{1}{2}}. \tag{3.13}$$

Thus to each $\zeta = e^{-1+it}$ there corresponds a unique $\omega = \omega(\zeta) = u+iv$ on the curve (3.12). In the neighbourhood of $t = 0$, i.e. of $u = 1$, $v = 0$, the curve of fixed-points has a double point and satisfies an expansion

$$\begin{aligned} u &= 1-\sqrt{t}+0\cdot t+at\sqrt{t}+\cdots, \\ v &= \sqrt{t}-\tfrac{2}{3}t+bt\sqrt{t}+\cdots \qquad \text{if } t > 0 \\ &= -\sqrt{-t}-\tfrac{2}{3}t-bt\sqrt{-t}+\cdots \text{ if } t < 0. \end{aligned} \tag{3.14}$$

The fixed-points ω on (3.12) are attractive for all z on the circle $|z| = e^{-1}$; in fact

$$|F(\zeta, z)| = |\zeta e^z| < 1 \tag{3.15}$$

for all $|\zeta| = e^{-1}$, Re $z < 1$ and the functions

$$F_k(\zeta, z), \ |\zeta| = e^{-1}, \qquad k = 0, 1, 2, \cdots$$

form a normal family on the half plane Re $z < 1$.

By a more refined argument one can show that if we set

$$D_k(\zeta) = G_k(\zeta)-\omega(\zeta) \tag{3.16}$$

then $D_k(\zeta)/\omega(\zeta)^k$ is uniformly bounded for all $\zeta = e^{-1+it}$, $-\pi \leqq t \leqq \pi$ and $k = 0, 1, 2, \cdots$, i.e.

$$|D_k(\zeta)| < A|\omega(\zeta)|^k \tag{3.17}$$

for a suitable positive constant A. For we have

$$D_k(\zeta) = \omega(\zeta)[\exp D_{k-1}(\zeta)-1], \qquad k = 1, 2, \cdots \tag{3.18}$$

by (3.2), (3.11) and (3.16), and the sequence behaves very nearly like the sequence D_k^* given by the recursion

$$D_k^* = \omega D_{k-1}^*/(1-\tfrac{1}{2}D_{k-1}^*). \tag{3.19}$$

For this sequence the statement can be verified by direct calculation since

$$D_k^* = -\frac{2\omega^k}{1+\omega+\cdots+\omega^{k-1}-a}, \quad a = 1/D_0^*. \tag{3.20}$$

We omit details.

For $|t| < ((\log^2 k)/k)^2$ the sequence $G_k(\zeta)$, $k = 0, 1, 2, \cdots$ has a uniform asymptotic expansion

$$\begin{aligned} G_k(e^{-1+it}) &\cong 1-\frac{2}{k}\,\tau\cot\tau+\tfrac{2}{3}\,\frac{\tau^2}{\sin^2\tau}\,\frac{\log k}{k^2} \\ &\quad+\frac{1}{k^2}\left[c\,\frac{\tau^2}{\sin^2\tau}+\tfrac{2}{3}\tau^2\left(2+\log\frac{\sin\tau}{\tau}\right)\right]+\cdots \end{aligned} \tag{3.21}$$

where

$$it = 2\tau^2/k^2 \tag{3.22}$$

and c is the constant in (3.10). The expansion becomes (3.10) for $\tau = 0$ and can be verified formally by setting

$$G_k(e^{-1+it}) \cong 1-\frac{\theta_1(\tau)}{k}+\theta_2(\tau)\,\frac{\log k}{k^2}+\frac{\theta_3(\tau)}{k^2}+\cdots \tag{3.23}$$

and using (3.2).

We obtain (for fixed t) by (3.22)

$$\begin{aligned} G_{k+1}(e^{-1+it}) &\cong 1-\frac{\theta_1(\tau+\tau/k)}{k+1}+\theta_2(\tau+\tau/k)\,\frac{\log(k+1)}{(k+1)^2} \\ &\quad+\theta_3(\tau+\tau/k)\,\frac{1}{(k+1)^2}+\cdots \\ &\cong 1-\theta_1/k+\theta_1/k^2-\theta_1/k^3-\tau\theta_1'/k^2 \\ &\quad+\tau\theta_1'/k^3-\tau^2\theta_1''/2k^3+\theta_2\log k/k^2 \\ &\quad-2\theta_2\log k/k^3+\theta_2/k^3+\tau\theta_2'\log k/k^3 \\ &\quad+\theta_3/k^2-2\theta_3/k^3+\tau\theta_3'/k^3+\cdots, \end{aligned}$$

$$\begin{aligned} \exp(G_k-1+2\tau^2/k^2) &\cong \exp\left[-\frac{\theta_1}{k}+\theta_2\,\frac{\log k}{k^2}+\frac{\theta_3}{k^2}+\frac{2\tau^2}{k^2}+\cdots\right] \\ &\cong 1-\theta_1/k+\theta_2\log k/k^2+(\theta_3+2\tau^2)/k^2 \\ &\quad+\tfrac{1}{2}\theta_1^2/k^2-\theta_1\theta_2\log k/k^3-\theta_1\theta_3/k^3 \\ &\quad-2\theta_1\tau^2/k^3-\tfrac{1}{6}\theta_1^3/k^3+\cdots \end{aligned}$$

where values of the function θ_i and their derivatives are taken at τ. These two expressions are equal by (3.2), hence comparing coefficients

$$\theta_1-\tau\theta_1' = \tfrac{1}{2}\theta_1^2+2\tau^2, \tag{3.24}$$

$$\tau\theta_2'-2\theta_2 = -\theta_1\theta_2, \tag{3.25}$$

$$-\theta_1+\tau\theta_1'-\tfrac{1}{2}\tau^2\theta_1''+\theta_2-2\theta_3+\tau\theta_3' = -\theta_1\theta_3-2\tau^2\theta_1-\tfrac{1}{6}\theta_1^3. \tag{3.26}$$

With the initial conditions $\theta_1(0) = 2$, $\theta_2(0) = \frac{2}{3}$, $\theta_3(0) = c$, obtained from (3.10), the equations (3.24)—(3.26) give

$$\theta_1(\tau) = 2\tau \cot \tau, \quad \theta_2(\tau) = \tfrac{2}{3}\frac{\tau^2}{\sin^2 \tau},$$

$$\theta_3(\tau) = c\,\frac{\tau^2}{\sin^2 \tau}+\tfrac{2}{3}\tau^2\left(2+\log\frac{\sin \tau}{\tau}\right), \quad \text{i.e. } (3.21).$$

Thus we only have to prove the existence of an expansion of the form (3.23). This can be achieved by a step by step method such as the one used in [3] for the proof of general expansions of the type (3.11); we omit details. Actually we only need the expansion in the weaker form

$$G_k(e^{-1+it}) = 1-\frac{2}{k}\,\tau \cot \tau+O\,(k^{-1-2\delta}) \tag{3.27}$$

for some $\delta > 0$ when $|t| \leqq ((\log^2 k)/k)^2$.

We then get from (3.2), since

$$2\tau \cot \tau = O\,(k\sqrt{|t|}) = O\,(\log^2 k),$$

$$\begin{aligned} G_k(e^{-1+it})-G_{k-1}(e^{-1+it}) &= \exp\,(-1+it+G_{k-1})-G_{k-1} \\ &= it+\tfrac{1}{2}(G_{k-1}-1)^2+O(k^{-3+\delta}) \\ &= 2\tau^2(1+\cot^2 \tau)/k^2+O(k^{-2-\delta}), \end{aligned}$$

$$G_k(e^{-1+it})-G_{k-1}(e^{-1+it}) = \frac{2}{k^2}\frac{\tau^2}{\sin^2 \tau}+O(k^{-2-\delta}) \tag{3.28}$$

for $|t| \leqq ((\log^2 k)/k)^2$.

Now from (3.17)

$$\begin{aligned} G_k(e^{-1+it})-G_{k-1}(e^{-1+it}) &= D_k(e^{-1+it})-D_{k-1}(e^{-1+it}) \\ &= O\left(\left(1-\frac{\log^2 k}{k}\right)^k\right) = O(e^{-\log^2 k}) \end{aligned}$$

for $|t| \geqq ((\log^2 k)/k)^2$, by (3.14), hence by (3.3) and (3.28)

https://doi.org/10.1017/S1446788700004432 Published online by Cambridge University Press

$$(3.29)\qquad \begin{aligned}\frac{d_n(k)}{(n-1)!} &= \frac{1}{2\pi i}\frac{2e^n}{k^2}\left(\int_{|t|\leq((\log^2 k)/k)^2}\frac{\tau^2}{\sin^2\tau}e^{-nit}idt+O(k^{-2-\delta})\right)\\ &\cong \frac{1}{2\pi i}\frac{8e^n}{k^4}\int_\Gamma \frac{\tau^3}{\sin^2\tau}e^{-\beta\tau^2}d\tau\end{aligned}$$

by (3.22), where

$$(3.30)\qquad \beta = 2n/k^2$$

and Γ is the path

$$\begin{aligned}\tau &= (i-1)u, \quad -\tfrac{1}{2}\log^2 k \leqq u \leqq 0\\ \tau &= (i+1)u, \quad 0 \leqq u \leqq \tfrac{1}{2}\log^2 k.\end{aligned}$$

Hence

$$\frac{d_n(k)}{(n-1)!} \cong \frac{8e^n}{k^4}\sum_{p=1}^{\infty}\operatorname*{res}_{\tau=p\pi}\left(\frac{\tau^3}{\sin^2\tau}e^{-\beta\tau^2}\right).$$

By using Stirling's formula we obtain from here

$$(3.31)\qquad p_n(k) = \frac{d_n(k)}{n^{n-2}} \cong 2\left(\frac{2\pi}{n}\right)^{\frac{1}{2}}\beta^2\sum_{p=1}^{\infty}(2p^4\pi^4\beta-3p^2\pi^2)e^{-\beta\pi^2p^2}$$

for large n and k where β is given by (3.30).

This is the required asymptotic probability distribution. Note that

$$\begin{aligned}\sum_{k=1}^{n-1}p_n(k) &\sim 2\left(\frac{2\pi}{n}\right)^{\frac{1}{2}}\int_0^{n-1}\beta^2\sum_{p=1}^{\infty}(2p^4\pi^4\beta-3p^2\pi^2)e^{-\beta\pi^2p^2}dk\\ &\cong 2\pi^{\frac{1}{2}}\int_0^{\infty}\sum_{p=1}^{\infty}\beta^{\frac{1}{2}}(-3p^2\pi^2+2p^4\pi^4\beta)e^{-\beta\pi^2p^2}d\beta\\ &= 2\pi^{\frac{1}{2}}\lim_{\epsilon=0}\left[-\sum_{p=1}^{\infty}2p^2\pi^2\beta^{\frac{3}{2}}e^{-\beta\pi^2p^2}\right]_\epsilon^\infty\\ &= \lim_{\beta\to 0}4\pi^{\frac{5}{2}}\beta^{\frac{3}{2}}\sum_{p=1}^{\infty}p^2e^{-\beta\pi^2p^2}\\ &= 4\pi^{-\frac{1}{2}}\int_0^\infty u^2e^{-u^2}du = 1,\end{aligned}$$

as required.

The maximum of the distribution curve is reached when in (3.31), $(d/d\beta)p_n(k) = 0$, i.e.

$$\sum_{p=1}^{\infty}(9p^4\pi^4\beta^2-6p^2\pi^2\beta-2p^6\pi^6\beta^3)e^{-\beta\pi^2p^2} = 0.$$

Numerically β (max) $= 0.373138525$ and

$$(3.32)\qquad k\,(\max) = 2.31515436\sqrt{n}.$$

4. Conclusion

The result of the previous section can be stated as follows:

Let $\mathscr{H}_n$ be the height over P_1 of a labelled random tree of order n i.e. of a tree selected at random from the set of n^{n-2} elements of $\mathscr{T}_n$ with uniform probability distribution. Then

$$\lim_{n\to\infty} P\left(\frac{\mathscr{H}_n}{\sqrt{2n}} < x\right) = H(x) \tag{4.1}$$

where

$$H(x) = 4x^{-3}\pi^{\frac{5}{2}} \sum_{p=1}^{\infty} p^2 e^{-\pi^2 p^2/x^2}. \tag{4.2}$$

This can be transformed (e.g. by means of Poisson's formula) to the form

$$H(x) = \sum_{v=-\infty}^{\infty} e^{-v^2x^2}(1-2v^2x^2) \tag{4.3}$$

whence

$$h(x) = H'(x) = 4x\sum_{v=1}^{\infty} v^2(2v^2x^2-3)e^{-v^2x^2}. \tag{4.4}$$

From (4.4) we can calculate all moments of the distribution function $H(x)$:

$$\begin{aligned} M_s &= \int_0^\infty x^s h(x)dx \\ &= 2\Gamma(\tfrac{1}{2}s+1)(s-1)\zeta(s) \qquad (s > 1) \end{aligned} \tag{4.5}$$

where $\zeta(s) = \sum_{m=1}^{\infty} 1/m^s$. In particular we obtain for M_1, since

$$\lim_{s\to 1} (s-1)\zeta(s) = 1,$$

$$M_1 = \sqrt{\pi}. \tag{4.6}$$

Hence the expectation value of $\mathscr{H}_n$ is

$$E(\mathscr{H}_n) \sim \sqrt{2n\pi} = 2.50663\sqrt{n} \tag{4.7}$$

and the variance is

$$D^2 = M_2 - M_1^2 = \frac{\pi(\pi-3)}{3}. \tag{4.8}$$

References

[1] Cayley, A., 'A theorem on trees', *Quart. J. of Pure and Applied Math.* 23 (1889), 376–378.

[2] Riordan, J., 'The enumeration of trees by height and diameter', *IBM Journal of Research and Development* 4 (1960), 473–478.

[3] Szekeres, G., 'Regular iteration of real and complex functions', *Acta Math.* 100 (1958), 103–258.
[4] Fatou, M. P., 'Sur les équations fonctionnelles', *Bull. Soc. Math. France* 47 (1919), 161–271.

Mathematical Institute of the Academy, Budapest
and
University of New South Wales