ON THE DISTRIBUTION OF THE NUMBER OF VERTICES IN LAYERS OF RANDOM TREES

LAJOS TAKÁCS
Case Western Reserve University
Cleveland, Ohio

ABSTRACT

Denote by $S_n$ the set of all distinct rooted trees with $n$ labeled vertices. A tree is chosen at random in the set $S_n$, assuming that all the possible $n^{n-1}$ choices are equally probable. Define $r_n(m)$ as the number of vertices in layer $m$, that is, the number of vertices at a distance $m$ from the root of the tree. The distance of a vertex from the root is the number of edges in the path from the vertex to the root. This paper is concerned with the distribution and the moments of $r_n(m)$ and their asymptotic behavior in the case where $m = \lfloor 2\alpha \sqrt{n} \rfloor$, $0 < \alpha < \infty$ and $n \to \infty$. In addition, more random trees, branching processes, the Bernoulli excursion and the Brownian excursion are also considered.

Key words: Random trees, Branching processes, Bernoulli excursion, Brownian excursion, Local times, Limit theorems.

AMS (MOS) subject classifications: 60F05, 05C05, 60J55, 60J65, 60J80.

1. INTRODUCTION

In 1889, A. Cayley [3] observed that the number of distinct trees with $n$ labeled vertices is $n^{n-2}$. Since then various proofs have been found for Cayley’s formula. For a simple proof see L. Takács [23]. The number of distinct rooted trees with $n$ labeled vertices is

$$R_n = n^{n-1}$$

for $n = 1, 2, \ldots$. Since among the $n$ vertices we can choose a root in $n$ ways, (1) immediately follows from Cayley’s formula.

The number of vertices in layer $m$ in a rooted tree is the number of vertices at a distance $m$ from the root. The distance of a vertex from the root is the number of edges in the path from the vertex to the root.
Let $S_n$ be the set of all distinct rooted trees with $n$ labeled vertices and denote by $t_n(j, m), j = 0, 1, \ldots, n - m$, the number of trees in $S_n$ having $j$ vertices at a distance $m$ from the root. Let us choose a tree at random in the set $S_n$, assuming that all the possible $n^{n-1}$ choices are equally probable. Define $\tau_n(m)$ as the number of vertices in layer $m$, that is, the number of vertices at a distance $m$ from the root of the tree chosen at random. If all the possible trees in $S_n$ are equally probable, then

$$P\{\tau_n(m) = j\} = t_n(j, m)/n^{n-1}$$

for $j = 0, 1, \ldots, n - m$.

In this paper we are concerned with the distribution and the moments of $\tau_n(m)$ and their asymptotic behavior in the case where $m = [2\alpha \sqrt{n}], 0 < \alpha < \infty$ and $n \to \infty$. The results derived for $\tau_n(m)$ are extended to other random trees, branching processes, the Bernoulli excursion and the Brownian excursion.

2. AUXILIARY THEOREMS

Let us define the generating functions

$$g_n(z, m) = \sum_{j=0}^{n-m} t_n(j, m)z^j$$

and

$$G_m(z, w) = \sum_{n=1}^{\infty} g_n(z, m)w^n/n!$$

for $n \geq 1$ and $m \geq 0$. If $|z| \leq 1$ and $|w| \leq 1/e$, then (4) is convergent.

**Lemma 1:** If $|w| \leq 1/e$, then the equation

$$ye^{-y} = w$$

has exactly one root in the unit disk $|y| \leq 1$ and

$$y^r = [y(w)]^r = r \sum_{n=r}^{\infty} \frac{n^{n-r}w^n}{n!(n-r)!}$$

for $|w| \leq 1/e$ and $r = 1, 2, \ldots$.

**Proof:** By Rouché's theorem it follows that (5) has exactly one root in the unit disk $|y| \leq 1$ and we obtain (6) by Lagrange's expansion. For $r = 1$ the expansion (6) was already known to L. Euler [7].
Lemma 2: If \( m \geq 1, \ |z| \leq 1 \) and \( |w| \leq 1/e \), then
\[
G_m(z, w) = w e^{G_{m-1}(z, w)}
\] (7)
where \( G_0(z, w) = z y(w) \), and \( y = y(w) \) is given by (6) with \( r = 1 \).

Proof: If we take into consideration that the degree of the root of a tree may be \( k = 0, 1, 2, \ldots \), then we obtain that
\[
G_m(z, w) = w + w \sum_{k=1}^{\infty} \frac{[G_{m-1}(z, w)]^k}{k!} = w e^{G_{m-1}(z, w)}
\] (8)
for \( m = 1, 2, \ldots \) and obviously
\[
G_0(z, w) = z \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} w^n = z y(w)
\] (9)
for \( |w| \leq 1/e \). Equation (7) appears also in A. Meir and J.W. Moon [19] and in A.M. Odlyzko and H.S. Wilf [20].

3. THE MOMENTS OF \( \tau_n(m) \)

The following theorem has been found by V.E. Stepanov [21]. In what follows we shall give a simple proof for it.

Theorem 1: If \( 0 < \alpha < \infty \), then
\[
\lim_{n \to \infty} E \left\{ \left( \frac{2\tau_n((2\alpha \sqrt{n})^1)}{\sqrt{n}} \right)^r \right\} = \mu_r(\alpha)
\] (10)
e exists for \( r = 0, 1, 2, \ldots \). We have \( \mu_0(\alpha) = 1, \mu_1(\alpha) = 4 \alpha e^{-2\alpha^2} \), and
\[
\mu_r(\alpha) = 2^r + 1 r! \alpha^r \int_0^{r-1} (1 + x)e^{-2\alpha^2(1 + x)^2} g_{r-1}(x) dx
\] (11)
for \( r \geq 2 \), where
\[
g_{r-1}(x) = \sum_{j=0}^{[x]} (-1)^j \binom{r-1}{j} (x-j)\frac{r-2}{(r-2)!}
\] (12)
for \( r \geq 2 \) and \( x \geq 0 \).

Proof: Let us define
\[
B_r(w, m) = \frac{1}{r!} \left( \frac{\partial^r G_m(z, w)}{\partial z^r} \right)_{z=1} = \sum_{n=1}^{\infty} E \left\{ \binom{\tau_n(m)}{r} \right\} \frac{n^{n-1} w^n}{n!}
\] (13)
for $r \geq 0, m \geq 0$, and $|w| \leq 1/e$.

By forming the derivative of (7) with respect to $z$ we obtain

$$\frac{\partial G_m(z, w)}{\partial z} = G_m(z, w) \frac{\partial G_{m-1}(z, w)}{\partial z}$$

(14)

for $m \geq 1$. Hence

$$B_1(w, m) = B_0(w, m)B_1(w, m - 1)$$

(15)

for $m \geq 1$. Since

$$B_0(w, m) = y(w)$$

(16)

for $m \geq 0$, by (15) we obtain that

$$B_1(w, m) = [y(w)]^{m + 1}$$

(17)

for $m \geq 0$, and thus by (6)

$$E\{\tau_n(m)\} = \gamma_n(m) = (m + 1) \left(\frac{n}{m + 1}\right) \left(\frac{m + 1}{n + 1}\right).$$

(18)

If $r \geq 2$, and $m \geq 1$, then the $(r-1)$st derivative of (14) with respect to $z$ at $z = 1$ yields

$$r[B_r(w, m) - y(w)B_r(w, m - 1)] = \sum_{j=1}^{r-1} (r - j)B_j(w, m)B_{r-j}(w, m - 1),$$

(19)

whence for the determination of $B_r(w, m), (r = 2, 3, \ldots)$, we get the following recurrence formula:

$$rB_r(w, m) = \sum_{j=1}^{r-1} (r - j) \sum_{0 \leq i < m} [y(w)]^{m - i - 1}B_j(w, i + 1)B_{r-j}(w, i).$$

(20)

If $r = 2$ in (20), then by (17)

$$B_2(w, m) = \frac{1}{2} \sum_{0 \leq i < m} [y(w)]^{m + i + 2}$$

(21)

and thus by (6)

$$E\left\{\left(\frac{\tau_n(m)}{2}\right)\right\} = \frac{1}{2} \sum_{0 \leq i < m} \gamma_n(m + i + 1)$$

$$= \frac{n!}{2n^{n-1}} \left(\frac{n^{n-m-2}}{(n-m-2)!} - \frac{n^{n-2m-2}}{(n-2m-2)!}\right).$$

(22)

If $r = 3$ in (20), then by (17) and (21)
\[ B_3(w, m) = \frac{1}{2} \sum_{0 \leq i < j < m} [y(w)]^{m+i+j+3} + \frac{1}{6} \sum_{0 \leq i = j < m} [y(w)]^{m+i+j+3} \]  \hspace{1cm} (23)

and hence

\[ E \left\{ \left( \frac{\tau_n(m)}{3} \right) \right\} = \frac{1}{2} \sum_{0 \leq i < j < m} \gamma_n(m+i+j+2) + \frac{1}{6} \sum_{0 \leq i = j < m} \gamma_n(m+i+j+2). \]  \hspace{1cm} (24)

By continuing this procedure we obtain that for \( r \geq 2, \)

\[ B_r(w, m) = \frac{(r-1)!}{2^{r-1}} \sum_{0 \leq i_1 < i_2 < \ldots < i_{r-1} < m} [y(w)]^{m+i_1+\ldots+i_{r-1}+r} \]  \hspace{1cm} (25)

where the neglected terms are constant multiples of sums similar to the one displayed, except that in these sums \( i_1, i_2, \ldots, i_{r-1} \) are not distinct; for at least one \( \nu = 2, \ldots, r-1 \) we have \( i_{\nu-1} = i_{\nu}. \) Formula (25) can be proved by mathematical induction. If we suppose that (25) is true for \( B_2(w, m), \ldots, B_{r-1}(w, m) \) where \( r = 3, 4, \ldots, \) then by (20) it follows that (25) is true for \( B_r(w, m) \) too. Accordingly, (25) is true for every \( r \geq 2. \)

It is easy to prove that

\[ | \gamma_n(m) - me^{-m^2/(2n)} | < 4/3 \]  \hspace{1cm} (26)

for \( 0 \leq m < n. \) If \( r = 1 \) and \( m = \lfloor 2\alpha \sqrt{n} \rfloor, \) then by (18) we obtain that

\[ E\{\tau_n(m)\} = \gamma_n(m) \sim 2\alpha \sqrt{n} e^{-2\alpha^2} \]  \hspace{1cm} (27)

as \( n \to \infty, \) or

\[ \lim_{n \to \infty} 2E\{\tau_n(m)\}/\sqrt{n} = 4\alpha e^{-2\alpha^2}. \]  \hspace{1cm} (28)

This proves (10) for \( r = 1. \) If \( r \geq 2, \) \( m = \lfloor 2\alpha \sqrt{n} \rfloor, 0 < \alpha < \infty \) and \( n \to \infty, \) then by (25)

\[ E\left\{ \left( \frac{\tau_n(m)}{r} \right) \right\} = \frac{(r-1)!}{2^{r-1}} \sum_{0 \leq i_1 < i_2 < \ldots < i_{r-1} < m} \gamma_n(m+i_1+\ldots+i_{r-1}+r-1) + \ldots \]  \hspace{1cm} (29)

where the neglected terms are of smaller order than the displayed one. If \( r \geq 1, \) \( m = \lfloor 2\alpha \sqrt{n} \rfloor, 0 < \alpha < \infty \) and \( n \to \infty, \) then

\[ E\{[\tau_n(m)]^r\} \sim r! E\left\{ \left( \frac{\tau_n(m)}{r} \right) \right\}, \]  \hspace{1cm} (30)

and by (26) and (29) we obtain that

\[ \lim_{n \to \infty} 2^r E\{[\tau_n(m)]^r\}/n^{r/2} = \mu_r(\alpha) \]  \hspace{1cm} (31)
exists and

\[ \mu_r(\alpha) = (r - 1)!\alpha_r \]

\[ \int \cdots \int_{0 < x_1 < \ldots < x_{r-1} < 1} (1 + x_1 + \ldots + x_{r-1}) e^{-2\alpha^2(1 + x_1 + \ldots + x_{r-1})^2} dx_1 \ldots dx_{r-1} \] (32)

\[ = \alpha_r \int_0^1 \cdots \int_0^1 (1 + x_1 + \ldots + x_{r-1}) e^{-2\alpha^2(1 + x_1 + \ldots + x_{r-1})^2} dx_1 \ldots dx_{r-1} \]

for \( r \geq 2 \), where \( \alpha_r = 2^{r+1}r!\alpha^r \). We can write also that

\[ \mu_r(\alpha) = 2^{r+1}r!\alpha^r \int_0^{r-1} (1 + x) e^{-2\alpha^2(1 + x)^2} g_{r-1}(x) dx \] (33)

for \( r \geq 2 \) where \( g_{r-1}(x) \) is the density function of \( \xi_1 + \xi_2 + \ldots + \xi_{r-1} \) where \( \xi_1, \xi_2, \ldots, \xi_{r-1} \) are independent random variables each having a uniform distribution over the interval (0,1). For the density function \( g_{r-1}(x) \), formula (12) has been found by P.S. Laplace [14], pp. 256-257. For a simple proof of (12) see L. Takács [22].

We note that

\[ \mu_2(\alpha) = 4(e^{-2\alpha^2} - e^{-8\alpha^2}) \] (34)

and

\[ \mu_3(\alpha) = 12\sqrt{2\pi}[2\Phi(4\alpha) - \Phi(2\alpha) - \Phi(6\alpha)] \] (35)

where

\[ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du \] (36)

is the normal distribution function.

4. THE ASYMPTOTIC DISTRIBUTION OF \( \tau_n(m) \)

The asymptotic distribution of \( \tau_n(m) \) has been found by V.E. Stepanov [21] in a different form.
On the Distribution of the Number of Vertices

Theorem 1: If $0 < \alpha < \infty$, then

$$\lim_{n \to \infty} P\left\{ \frac{2\tau_n([2\alpha\sqrt{n}])}{\sqrt{n}} \leq x \right\} = G_\alpha(x)$$

(37)

for $x > 0$ where $G_\alpha(x)$ is the distribution function of a nonnegative random variable and is given by

$$G_\alpha(x) = 1 - 2 \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} \binom{j-1}{k} e^{-\frac{(x+2\alpha j)^2}{2}} (-x)^k H_k + (x+2\alpha j) / k!$$

(38)

for $x \geq 0$ where $H_0(x), H_1(x), \ldots$ are the Hermite polynomials defined by

$$H_n(x) = n! \sum_{j=0}^{[n/2]} \frac{(-1)^j x^{n-2j}}{2^j j!(n-2j)!}.$$

(39)

We have

$$G_\alpha(0) = 1 - 2 \sum_{j=1}^{\infty} (4\alpha^2 j^2 - 1) e^{-2\alpha^2 j^2}$$

(40)

and

$$\frac{dG_\alpha(x)}{dx} = 2 \sum_{j=1}^{\infty} \sum_{k=1}^{j} \frac{j}{k!} \frac{j!}{k!} e^{-\frac{(x+2\alpha j)^2}{2}} (-x)^{k-1} H_k + (x+2\alpha j) / (k-1)!$$

(41)

if $x > 0$.

Proof: Since

$$ue^{-u^2} \leq (2e)^{-1/2} < 1/2$$

(42)

if $u \geq 0$, it follows from (11) that

$$\mu_r(\alpha)/r! < (2\alpha)^r / \alpha^0$$

(43)

for $r \geq 2$. Accordingly, there exists one and only one distribution function $G_\alpha(x)$ such that $G_\alpha(x) = 0$ for $x < 0$ and

$$\int_{-\infty}^{\infty} x^r dG_\alpha(x) = \mu_r(\alpha)$$

(44)

for $r \geq 0$. By the moment convergence theorem of M. Fréchet and J. Shohat [8] it follows from (10) that (37) holds in every continuity point of $G_\alpha(x)$. If $|s| < 1/(2\alpha)$, then the Laplace-Stieltjes transform

$$\Psi_\alpha(s) = \int_{-\infty}^{\infty} e^{-sx} dG_\alpha(x)$$

(45)
can be expressed as
\[ \Psi_\alpha(s) = \sum_{r=0}^{\infty} (-1)^r \mu_r(\alpha) s^r / r!. \] (46)

By (11) we obtain that
\[ \Psi_\alpha(s) = 1 + 2 \sum_{k=1}^{\infty} \frac{(2\alpha s)^k}{(k-1)!} \int_k^\infty (1 - 4\alpha^2 u^2)(u - k)^{k-1} e^{-2\alpha^2 u^2 - 2\alpha(u - k)s} du \] (47)

for \( |s| < 1/(2\alpha) \). Hence (38) and (41) follow by inversion.

5. VARIOUS EXTENSIONS

By using the same method which we used in proving Theorems 1 and 2 we can demonstrate that the distribution function \( G_\alpha(x) \) appears also in the solutions of various other problems in probability theory. Apparently, the interesting interrelation among these problems has not been noticed before, and \( G_\alpha(x) \) has appeared in various disguises. Here are some examples.

(i) Random trees. Denote by \( T_{n+1} \) the set of distinct rooted ordered trees with \( n + 1 \) unlabeled vertices. There are
\[ C_n = \binom{2n}{n} \frac{1}{n+1} \] (48)
distinct trees in \( T_{n+1} \). This follows from the obvious recurrence formula
\[ C_n = \sum_{i=1}^{n} C_{i-1} C_{n-i} \] (49)
for \( n = 1, 2, \ldots \) where \( C_0 = 1 \). In (48) \( C_n \) is the \( n \)th Catalan number. Let us choose a tree at random, assuming that all the possible \( C_n \) trees are equally probable. Denote by \( \tau_{n+1}(m) \) the number of vertices at a distance \( m \) from the root of a tree chosen at random. If \( 0 < \alpha < \infty \), then we have
\[ \lim_{n \to \infty} P \left\{ \frac{2\tau_{n+1}(\lfloor \alpha \sqrt{2n} \rfloor)}{\sqrt{2n}} \leq x \right\} = G_\alpha(x) \] (50)
for \( x > 0 \).

Denote by \( T_{2n+2}^* \) the set of distinct planted trivalent trees with \( 2n + 2 \) unlabeled vertices. A planted tree is rooted at an end vertex. In a trivalent tree every vertex has degree 3 except the end vertices which have degree 1. In 1859, A. Cayley [2] demonstrated that there
are $C_n$ distinct trees in $T_{2n+2}^*$ where $C_n$ is given by (48). Let us choose a tree at random in $T_{2n+2}^*$ assuming that all the possible $C_n$ choices are equally probable. Denote by $\tau_{2n+2}(m)$ the number of vertices at a distance $m$ from the root of a tree chosen at random. If $0 < \alpha < \infty$, then we have

$$
\lim_{n \to \infty} P \left\{ \frac{2\tau_{2n+2}\left(\alpha\sqrt{8n}\right)}{\sqrt{2n}} \leq x \right\} = G_\alpha(x)
$$

for $x > 0$.

(ii) Branching processes. Let us suppose that in a population initially we have a progenitor and in each generation each individual reproduces, independently of the others, and has probability $p_j$, ($j = 0, 1, \ldots$), of giving rise to $j$ descendants in the following generation. Denote by $\xi(m)$, ($m = 0, 1, \ldots$), the number of individuals in the $m$th generation; $\xi(0) = 1$. Define

$$
\rho = \sum_{m \geq 0} \xi(m),
$$

that is, $\rho$ is the total number of individuals (total progeny) in the process (possibly $\rho = \infty$). Let

$$
f(z) = \sum_{j=0}^{\infty} p_j z^j,
$$

and

$$
gcd\{j: p_j > 0\} = d.
$$

If $f(1) = 1$, $f'(1) = 1$, $f''(1) = \sigma^2$ where $0 < \sigma < \infty$, $f^{(r)}(1) < \infty$ for $r \geq 2$, and $0 < \alpha < \infty$, then

$$
\lim_{n \to \infty} P \left\{ \frac{2\xi\left(\alpha\sqrt{nd}/\sigma\right)}{\sigma\sqrt{nd}} \leq x | \rho = nd + 1 \right\} = G_\alpha(x)
$$

for $x > 0$ where $G_\alpha(x)$ is defined by (38).

If $p_j = e^{-1}/j!$ for $j = 0, 1, 2, \ldots$, then $\sigma^2 = 1$ and $d = 1$ and (55) reduces to (37). If $p_j = 1/2^{j+1}$ for $j = 0, 1, 2, \ldots$, then $\sigma^2 = 2$ and $d = 1$ and (55) reduces to (50). If $p_0 = p_2 = 1/2$ and $p_j = 0$ otherwise, then $\sigma^2 = 1$ and $d = 2$ and (55) reduces to (51).

The limit distribution (55) has already been determined by D.P. Kennedy [12] in a different form. By his results we can conclude that
for $x > 0$ and

$$
\int_0^\infty \int_0^\infty e^{-su - uw} f(u,v) du dv = \left\{ \frac{\sinh(\sqrt{2w})}{\sqrt{2w}} + s \left( \frac{\sinh\left(\frac{w}{2}\right)}{\sqrt{w/2}} \right) \right\}^{-1}
$$

for $\text{Re}(s) \geq 0$ and $\text{Re}(w) \geq 0$.

(iii) Bernoulli excursion. Let us arrange $n$ white balls and $n$ black balls in a row in such a way that for every $i = 1, 2, \ldots, 2n$ among the first $i$ balls there are at least as many white balls as black. The total number of such arrangements is given by the $n$th Catalan number $C_n$, defined by (48). Let us suppose that all the possible $C_n$ sequences are equally probable and choose a sequence at random. We associate a random walk with the random sequence chosen by assuming that a particle starts at time $t = 0$ at the origin of the $z$-axis and in the time interval $(i-1,i], i = 1, 2, \ldots, 2n$, it moves with a unit velocity to the right or to the left according to whether the $i$th ball in the row is white or black respectively. Denote by $x = \eta^+_n(t)$ the position of the particle at time $2nt$ where $0 \leq t \leq 1$. The process $\{\eta^+_n(t), 0 \leq t \leq 1\}$ is called a Bernoulli excursion. Denote by $2\tau^+_n(m)(m = 1, 2, \ldots, n)$ the number of crossings of the sample function of the process $\{\eta^+_n(t), 0 \leq t \leq 1\}$ through the line $x = m - 1/2$. In other words, $\tau^+_n(m)/n$ is the total time spent in the interval $(m-1,m)$ by the process $\{\eta^+_n(t), 0 \leq t \leq 1\}$. If $0 < \alpha < \infty$, then

$$
\lim_{n \to \infty} P \left\{ \frac{2\tau^+_n(\sqrt{n})}{\sqrt{2n}} \leq x \right\} = G_\alpha(x)
$$

for $x > 0$. Since $\tau^+_n(m)$ has exactly the same distribution as $\tau^+_{n+1}(m)$ in (50), the two results, (50) and (58), imply each other.

(iv) Brownian excursion. The process $\{\eta^+_n(t)/\sqrt{2n}, 0 \leq t \leq 1\}$, where $\eta^+_n(t)$ is defined under (iii), converges weakly to the Brownian excursion $\{\eta^+(t), 0 \leq t \leq 1\}$. For the definition and properties of the Brownian excursion we refer to P. Lévy [15], [16], K. Itô and H.P. McKean, Jr. [11] and K.L. Chung [4]. For the process $\{\eta^+(t), 0 \leq t \leq 1\}$ define $\tau^+(\alpha)$ as the local time at the level $\alpha$ for $\alpha \geq 0$. From (58) we can conclude that

$$
P(\tau^+(\alpha) \leq x) = G_\alpha(x)
$$

for $x > 0$, and also
for \( r = 0, 1, 2, \ldots \) where \( \mu_r(\alpha) \) is defined by (10).

The distribution function (59) has attracted considerable interest. In the articles by R.K. Getoor and M.J. Sharpe [9], J.W. Cohen and G. Hooghiemstra [5], G. Louchard [17], [18], E. Csáki and S.G. Mohanty [6], and Ph. Biane and M. Yor [1], \( P\{\tau^+(\alpha) \leq x\} \) is expressed in the form of a complex integral. F.B. Knight [13] and G. Hooghiemstra [10] expressed \( P\{\tau^+(\alpha) \leq x\} \) in explicit forms, but their formulas are hardly suitable for numerical calculations. We can easily produce tables and graphs for \( G_\alpha(x) \) and \( G'_\alpha(x) \) by using formulas (38) and (41) and the remarkable program MATHEMATICA by S. Wolfram [24].

REFERENCES


Submit your manuscripts at http://www.hindawi.com