Distribution Properties of Induced Subgraphs of Trees

Gerd Baron and Michael Drmota

Technical University of Vienna
Department of Discrete Mathematics
Technical University of Vienna
Wiedner Hauptstraße 8–10/118
A-1040 Vienna, Austria

Abstract. In this paper it is shown that the number of induced subgraphs (the set of edges is induced by the set of nodes) of trees of size n satisfy a central limit theorem and that multivariate asymptotic expansions can be obtained. In the case of planted plane trees, N-ary trees, and non-planar rooted labelled trees explicit formulae can be given. Furthermore the average size of the largest component of induced subgraphs in trees of size n is evaluated asymptotically.

1. Introduction

Let $G = (V(G), E(G))$ be a graph with nodes $V(G)$ and edges $E(G)$. A subgraph $H = (V(H), E(H))$ is called *induced subgraph* if $(x, y) \in E(H)$ if and only if $x, y \in V(H)$ and $(x, y) \in E(G)$. We will discuss distribution properties (such as the average size, the average number of components in general and of given size, the average size of the largest component) in *simply generated families of trees*. These kind of families of trees has been introduced and widely discussed by A. Meir and J.W. Moon. We will follow the description of [MM]:

Let $\mathcal{F}$ denote a family of rooted planar trees and $\mathcal{F}_n \subseteq \mathcal{F}$ the subset of $\mathcal{F}$ of trees of size $|V(T)| = n$ (i.e. with n nodes). Such a family is called simply generated family of trees if the generating function

$$Y(x) = \sum_{n\geq 1} y_n x^n \tag{1.1}$$

of the sequence $y_n = |\mathcal{F}_n|$ satisfies a functional equation of the type

$$Y = x\varphi(Y)\ , \tag{1.2}$$

where

$$\varphi(t) = 1 + \varphi_1 t + \varphi_2 t^2 + \cdots \tag{1.3}$$

is a power series with non-negative integral coefficients. In order to describe a more general situation it is also possible to use an arbitrary power series $\varphi(t)$ with non-negative coefficients and use the coefficients of $\varphi(t)$ to define weights $\omega(T)$ of rooted planar trees T such that $\omega(T) > 0$ if and only if $T \in \mathcal{F}$. Let

$D_i(T)$ denote the number of nodes of T with (out-)degree i, then the weight $\omega(T)$ can be defined by ($\varphi_0 = 1$)

$$\omega(T) = \prod_{i\geq 0} \varphi_i^{D_i(T)} . \tag{1.4}$$

Now it is easy to see that

$$\sum_{n\geq 1} \left(\sum_{|V(T)|=n} \omega(T) \right) x^n$$

is the solution of (1.2) analytic around $x = 0$. Thus

$$y_n = \sum_{|V(T)|=n} \omega(T) . \tag{1.5}$$

Furthermore by Lagrange's inversion formula y_n can be evaluated by

$$y_n = \frac{1}{n}[u^{n-1}]\varphi(u)^n, \tag{1.6}$$

where $[u^n]f(u)$ denotes the coefficient of u^n of the power series expansion of $f(u)$.

Note that a simply generated family of trees can also be described by a *combinatorial structure* $\mathcal{F}$ (see [Fl]) with a size function that counts the number of nodes. The structure satisfies the formal rule

$$\mathcal{F} = \{n\} \times \left(\{\varepsilon\} + \varphi_1 \mathcal{F} + \varphi_2 \mathcal{F}^2 + \cdots\right), \tag{1.7}$$

where n denotes a node, ε the empty structure and $+$ the disjoint union.

In the sequel we will mainly discuss three types of families of trees, planted plane trees, N-ary trees, and non-planar rooted labelled trees.

If $\varphi_i = 1$, then $\varphi(t) = 1/(1-t)$ and

$$y_n = \frac{1}{n}\binom{2n-2}{n-1}, \tag{1.8}$$

the Catalan number, is the number of general planted plane trees of size n.

The generating function of N-ary trees with n internal nodes (the out-degree is always N and leaves are not counted) satisfies the functional equation $Y = 1 + xY^N$. With $\overline{Y} = Y - 1$, we get $\overline{Y} = x(1+\overline{Y})^N$. So we can describe N-ary trees (only counting internal nodes) with $\varphi(t) = (1+t)^N$. The number of trees with n internal nodes is given by

$$y_n^{(N)} = \frac{1}{n}\binom{Nn}{n-1}. \tag{1.9}$$

If $\varphi(t) = e^t$ the solution of $Y = xe^Y$ can be interpreted as the exponentially generating function of the numbers of non-planar rooted labelled trees of size n. We have

$$y_n = \frac{\hat{y}_n}{n!} = \frac{n^{n-1}}{n!}. \tag{1.10}$$

These three families are not only of special interest but have formally an in some sense simple structure so that it is possilbe to get explicit formulae by Lagrange's inversion formula for many interesting parameters related to trees. The notion of simply generated families of trees also includes other interesting structures (such as Motzkin trees: $\varphi(t) = 1 + t + t^2$ or N-ary trees, where all nodes are counted: $\varphi(t) = 1 + t^N$) but they are much more difficult to handle and it is not possible to get simple explicit expressions. Nevertheless the asymptotic properties can be derived for a wide class of power series $\varphi(t)$ by the same method.

In section 2 we will prove functional equations for the generating functions of the numbers of induced subgraphs with components of given size. In section 3 we will use these functional equations to get explicit formulae for planted plane trees, for N-ary trees with n internal nodes, and for non-planar rooted labelled trees. In section 4 we will prove that those numbers satisfy a central limit theorem and give multivariate asymptotic expansions. The last section is devoted to the average size of the largest component of induced subgraphs.

2. Functional Equations

For any planar rooted tree T let $a_{klm,j}(T)$ $(k, l, m_i, j \geq 0, m = (m_1, m_2, \ldots))$ be the number of induced subgraphs H of T of size k, with l edges, and with m_i components of size i $(i \geq 1)$ such that the root $r = r(T)$ is contained in a component of H of size j ($j = 0$ means that $r \notin V(H)$) and set

$$a_{nklm,j} = \sum_{|V(T)|=n} \omega(T) a_{klm,j}(T). \tag{2.1}$$

The aim of this section is to establish functional equations for the generating functions

$$A_j(x, y, z, v) = \sum_{n,k,l,m} a_{nklm,j} x^n y^k z^l v^m \tag{2.2}$$

($v = (v_1, v_2, \ldots)$, $v^m = v_1^{m_1} v_2^{m_2} \cdots$) and for

$$C(x, y, z, v) = \sum_{j \geq 0} A_j(x, y, z, v) \tag{2.3}$$

that is the generating function for the total numbers c_{nklm} of induced subgraphs with k nodes, l edges, and m_i components of size i ($i \geq 1$) in trees of size n.

Note that $a_{nklm,j}$ or c_{nklm} can only be different from zero if

$$\sum_{j\geq 1} j m_j = k \qquad \text{and} \qquad \sum_{j\geq 1} (j-1) m_j = l \tag{2.4}$$

and that $k - l = \sum_{j\geq 1} m_j$ is the number of components.

Lemma 1. *If a simply generated family of trees is characterized by $\varphi(t)$, then we have for $A_j = A_j(x, y, z, v)$*

$$A_0 = x\varphi\left(\sum_{j\geq 0} A_j\right) \tag{2.5}$$

$$\sum_{j\geq 1} \frac{1}{v_j} A_j = xy\varphi\left(A_0 + z\sum_{j\geq 1} \frac{1}{v_j} A_j\right) \tag{2.6}$$

$$A_j = (xy)^j z^{j-1} v_j \frac{1}{j}[t^{j-1}]\varphi(A_0 + t)^j \qquad (j \geq 1). \tag{2.7}$$

Proof: In order to prove (2.5)–(2.7) remember that a simply generated family of trees can be interpreted as a combinatorial structure and satisfies the formal rule (1.7). So the combinatorial structure $\mathcal{A}_0$ of induced subgraphs not containing the root can be described as

$$\mathcal{A}_0 = \{n\} \times \left(\{\varepsilon\} + \varphi_1 \mathcal{C} + \varphi_2 \mathcal{C}^2 + \cdots\right), \tag{2.8}$$

where $\mathcal{C}$ denotes the combinatorial structure of all induced subgraphs. This gives (2.5).

Now let $\mathcal{B}$ be the combinatorial structure of induced subgraphs H of trees T where the root $r(T)$ is contained in $V(H)$ and the size function does not count the component that contains the root. Then we have

$$\begin{aligned} \mathcal{B} = \{n\} \times \{n_H\} \times \Big(\{\varepsilon\} &+ \varphi_1 \left(\mathcal{A}_0 + \{e_H\} \times \mathcal{B}\right) + \\ &\varphi_2 \left(\mathcal{A}_0 + \{e_H\} \times \mathcal{B}\right)^2 + \cdots\Big), \end{aligned} \tag{2.9}$$

where n denotes a node in T, n_H a node in H, and e_H an edge in H. Furthermore the generating function corresponding to $\mathcal{B}$ is given by

$$B(x, y, z, v) = \sum_{j\geq 1} \frac{1}{v_j} A_j(x, y, z, v). \tag{2.10}$$

This gives (2.6).

In order to prove (2.7) consider the combinatorial structure $\mathcal{D}$ of induced subgraphs H that contain the root of T and where the size function measures the size of the component that contains the root and counts the number of nodes in T not contained in this component, the number of nodes in H not contained in this component, the number of edges of H not contained in this component and so on. Then we have

$$\mathcal{D} = \{n_R\} \times \left(\{\varepsilon\} + \varphi_1\left(\mathcal{A}_0 + \mathcal{D}\right) + \varphi_2\left(\mathcal{A}_0 + \mathcal{D}\right)^2 + \cdots\right), \tag{2.11}$$

(where n_R denotes a node in the component containing the root) and the generating function corresponding to $\mathcal{D}$ can be written as

$$D(u,x,y,z,v) = \sum_{j\geq 1} A_j(x,y,z,v)(xy)^{-j} z^{-j+1} v_j^{-1} u^j. \tag{2.12}$$

Since (2.11) implies $D = u\varphi(A_0 + D)$ we get by (2.12) and Lagrange's inversion formula

$$A_j\,(xy)^{-j} z^{-j+1} v_j^{-1} u^j = \frac{1}{j}[t^{j-1}]\varphi(A_0 + t)^j. \tag{2.13}$$

Thus the proof of Lemma 1 is finished.

Remark 1: If we set $v_i = 1$ $(i \geq 1)$ we only count the total number of induced subgraphs with k nodes and l edges as it has been done in [Ba]. Using the notation of [Ba] we have

$$\begin{aligned} A(x,y,z) &= A_0(x,y,z,1) \\ B(x,y,z) &= \sum_{j\geq 1} A_j(x,y,z,1) \end{aligned} \tag{2.14}$$

and therefore by (2.5) and (2.6) the functional equations

$$\begin{aligned} A &= x\varphi(A + B) \\ B &= xy\varphi(A + zB) \end{aligned} \tag{2.15}$$

which are used in [Ba] to get explicit formulae for the total numbers of induced subgraphs with k nodes and l edges in trees of size n, where $\varphi(t) = 1/(1-t)$, $\varphi(t) = (1+t)^N$, and $\varphi(t) = e^t$.

Luckily it is possible to simplify the infinite system of functional equations for the formal power series A_j to one functional equation for C. We will use both representations, the infinite system to get explicit formulae and the single equation for the asymptotic analysis.

Theorem 1. *If a simply generated family of trees is characterized by* $\varphi(t)$, *then the generating function* $C = C(x, y, z, v)$ *satisfies the functional equation*

$$C = x\varphi(C) + xy\varphi\Big(zC + (1-z)x\varphi(C) + \sum_{j\geq 1}(1-v_j)(xyz)^j \frac{1}{j}[t^{j-1}]\varphi(x\varphi(C)+t)^j\Big) - \sum_{j\geq 1}(1-v_j)(xy)^j z^{j-1}\frac{1}{j}[t^{j-1}]\varphi(x\varphi(C)+t)^j . \quad (2.16)$$

Proof: Set

$$D = A_0 + z\sum_{j\geq 1}\frac{1}{v_j}A_j. \quad (2.17)$$

Then

$$D = zC + (1-z)A_0 + z\sum_{j\geq 1}\left(\frac{1}{v_j}-1\right)A_j \quad (2.18)$$

and we have by (2.6) that

$$\begin{aligned}\sum_{j\geq 1}\frac{1}{v_j}A_j &= \frac{D-A_0}{z}\\ &= C - A_0 + \sum_{j\geq 1}\left(\frac{1}{v_j}-1\right)A_j\\ &= xy\varphi\left(zC + (1-z)A_0 + z\sum_{j\geq 1}\left(\frac{1}{v_j}-1\right)A_j\right).\end{aligned} \quad (2.19)$$

Inserting the expressions (2.7) for A_j and $x\varphi(C)$ for A_0 we immediately get (2.16).

Remark 2: If we set $v_i = 1$ $(i \geq 1)$ we get a reduced functional equation for $C = C(x,y,z,1) = A(x,y,z) + B(x,y,z)$:

$$C = x\varphi(C) + xy\varphi(zC + (1-z)x\varphi(C)) . \quad (2.20)$$

3. Explicit Formulae

To determine explicit formulae we use the methods explained in [**Ba**] where it is proved that the tricks only work for the binomial family $\varphi(t) = \varphi_1(t) = (1+at)^N$ and for the exponential family $\varphi(t) = \varphi_2(t) = e^{at}$ of functions. We are interested in $\varphi(t) = e^t, \varphi(t) = (1+t)^N$ (the N-ary trees), and $\varphi(t) = (1-t)^{-1}$ (the planted plane trees). The formulae for the last ones can be deduced from the N-ary case by substituting -1 for N and multiplying by $(-1)^n$. So we have to discuss the functions $\varphi(t) = e^t$ and $\varphi(t) = (1+t)^N$ and will prove the following explicit formulae for the numbers c_{nklm} defined in (2.3).

Theorem 2. *If $\varphi(t) = 1/(1-t)$ then we have*

$$c_{nklm} = \frac{1}{k-l}\binom{2n-2}{n-k}\binom{n-l-2}{k-l-1}\binom{k-l}{m}P^m, \tag{3.1}$$

where $\binom{k-l}{m}$ denotes the multinomial coefficient $(k-l)!/\prod_{j\geq 1} m_j!, P^m = \prod_{j\geq 1} P_j^{m_j}$, and

$$P_j = \frac{1}{j}\binom{2j-2}{j-1}, \tag{3.2}$$

the Catalan numbers. For $\varphi(t) = (1+t)^N$ we have

$$c_{nklm} = \frac{1}{k-l}\binom{Nn-k+1}{n-k}\binom{N(n-k)}{k-l-1}\binom{k-l}{m}P^m, \tag{3.3}$$

where

$$P_j = \frac{1}{j}\binom{Nj}{j-1}, \tag{3.4}$$

and for $\varphi(t) = e^t$ we get

$$c_{nklm} = \frac{\hat{c}_{nklm}}{n!} = \frac{1}{k-l}\frac{n^{n-k}(n-k)^{k-l-1}}{(n-k)!(k-l-1)!}\binom{k-l}{m}P^m \tag{3.5}$$

with

$$P_j = \frac{j^{j-1}}{j!}. \tag{3.6}$$

Proof: We have $\varphi(v+w) = \varphi(v)\varphi\left(\frac{w}{h(v)}\right)$ with $h(v) = 1$ for $\varphi(t) = e^t$ and $h(v) = 1+v$ for $\varphi(t) = (1+t)^N$. So we will consider both cases as one as long as possible.

(2.6) now reads as

$$\begin{aligned} A_j &= (xy)^j z^{j-1} v_j \varphi(A_0)^j [t^{j-1}] \frac{1}{j}\varphi\left(\frac{t}{h(A_0)}\right)^j \\ &= (xy)^j z^{j-1} v_j \frac{\varphi(A_0)^j}{h(A_0)^{j-1}} P_j. \end{aligned} \tag{3.7}$$

With $A_j = yv_jB_j$ and $B = \sum_{j\geq 1} B_j$ we get from (2.4) - (2.6) the system

$$A_0 = x\varphi\left(A_0 + \sum_{j\geq 1} v_j y B_j\right) = x\varphi(A_0)\varphi\left(\frac{\sum v_j y B_j}{h(A_0)}\right) \tag{3.8}$$

$$B = x\varphi(A_0 + yzB) = x\varphi(A_0)\varphi\left(\frac{yzB}{h(A_0)}\right) \tag{3.9}$$

$$B_j = x^j (yz)^{j-1} \frac{\varphi(A_0)^j}{h(A_0)^{j-1}} P_j. \tag{3.10}$$

Using the tricks in [Ba] i.e. expanding $C(x,y,z,v) = C^*(u_1,u_2,u_3,v)$ with $u_1 = x$, $u_2 = xy$, $u_3 = xyz$ and determining the total number c_{nklm} as the coefficient of $u_1^{n-k}u_2^{k-l}u_3^l v^m$ we substitute $q = B/h(A_0)$, $r = yq$, and $s = yzq$. So we get from (3.9)

$$B = x\varphi(A_0)\varphi(s) \tag{3.11}$$

and from (3.10)

$$B_j = \left(\frac{B}{\varphi(s)}\right)^j \frac{(yz)^{j-1}}{h(A_0)^{j-1}}P_j = \frac{P_j}{\varphi(s)^j}s^{j-1}B. \tag{3.12}$$

With $S_j = v_jP_js^{j-1}/\varphi(s)^j$, $U_j = jS_j$, $S = \sum_{j\geq 1} S_j$, and $U = \sum_{j\geq 1} U_j$ we get from (3.8)

$$A_0 = x\varphi(A_0)\varphi(rS). \tag{3.13}$$

(3.13) divided by (3.11) implies $A_0/h(A_0) = qR = p$ with $R = \varphi(rS)/\varphi(s)$.

Preparing for later use we remark that for given $m = (m_1, m_2, \ldots)$ we have (2.4), i.e.$\sum jm_j = k$, $\sum(j-1)m_j = l$, $\sum m_j = k - l$, and

$$[v^m]S^{k-l} = \binom{k-l}{m}P^m\frac{s^l}{\varphi(s)^k} \tag{3.14}$$

$$[v^m]US^{k-l-1} = \frac{k}{k-l}[v^m]S^{k-l}. \tag{3.15}$$

For application of Lagrange's inversion formula we have to calculate derivatives with respect to q by remembering that q appears in r and s, too: $r'q = r$, $s'q = s$.

If we fix $N = 1$ for $\varphi(t) = e^t$ we can work with $\frac{d\varphi(t)}{dt} = N\frac{\varphi(t)}{h(t)}$ in both cases. Using this we get

$$S_j'q = U_j\left(1 - \frac{N}{h(s)}s\right) - S_j \tag{3.16}$$

and therefore

$$S + S'q = \left(1 - \frac{N}{h(s)}s\right)U \tag{3.17}$$

$$\begin{aligned} p' &= (qR)' = R + R'q \\ &= R\left(1 - \frac{N}{h(s)}s\right)\left(1 + \frac{N}{h(rS)}rU.\right) \end{aligned} \tag{3.18}$$

Since $1 + t\frac{dh(t)}{dt} = h(t)$ we get from $A_0 = ph(A_0)$

$$A_0' = p'h(A_0)^2. \tag{3.19}$$

Using (3.13) and $A_0 = ph(A_0) = qRh(A_0) = q\frac{\varphi(rS)}{\varphi(s)}h(A_0)$ we get

$$x = \frac{q}{g} \quad \text{with} \quad g = \frac{\varphi(A_0)\varphi(s)}{h(A_0)}. \tag{3.20}$$

The wanted number c_{nklm} is now

$$\frac{1}{n}[q^{n-k}r^{k-l}s^l v^m]\left(qA_0' + r\left(\sum v_j B_j\right)'\right)g^n. \tag{3.21}$$

Since $\sum v_j B_j = \sum S_j B = SB = Sqh(A_0)$ with (3.16)

$$\left(\sum v_j B_j\right)' = h(A_0)(S + qS') + Sq\frac{dh}{dt}(A_0)A_0' \tag{3.22}$$

and therefore with (3.18) and (3.19)

$$\begin{aligned} qA_0' + r\left(\sum v_j B_j\right)' = h(A_0)rU\left(1 - \frac{N}{h(s)}s\right) + \\ + qRh(A_0)^2\left(1 - \frac{N}{h(s)}s\right)(h(rS) + NrU). \end{aligned} \tag{3.23}$$

So we have to determine

$$[q^{n-k}r^{k-l}s^l v^m]\frac{1}{n}\left(1 - \frac{N}{h(s)}s\right)\varphi(s)^n Q \tag{3.24}$$

with

$$Q = rU\frac{\varphi(A_0)^n}{h(A_0)^{n-1}} + qR(h(rS) + NrU)\frac{\varphi(A_0)^n}{h(A_0)^{n-2}}, \tag{3.25}$$

where $A_0/h(A_0) = qR$.

Now we have to split the two cases. For $\varphi(t) = e^t$ we have $h(t) = 1$, $N = 1$, $\varphi(A_0) = e^p = e^{qR}$, and $R = e^{rS}e^{-s}$. Hence (3.23), (3.25) read

$$[q^{n-k}r^{k-l}s^l v^m]\frac{1-s}{n}e^{ns}\left(rUe^{nqR} + qRe^{nqR}(1 + rU)\right) \tag{3.26}$$

leading to $[q^{n-k}]$ equals

$$\frac{n^{n-k-2}}{(n-k)!}(1-s)e^{ks}\left(rUe^{(n-k)rS}(2n-k) + (n-k)e^{(n-k)rS}\right) \tag{3.27}$$

and therefore $[r^{k-l}]$ equals

$$\frac{n^{n-k-2}(n-k)^{k-l-1}}{(n-k)!(k-l)!}(1-s)e^{ks}\left((2n-k)(k-l)US^{k-l-1} + (n-k)^2 S^{k-l}\right). \tag{3.28}$$

Determining now $[v^m]$ using (3.14), (3.15) we get

$$\frac{n^{n-k}(n-k)^{k-l-1}}{(n-k)!(k-l)!}\binom{k-l}{m}p^m \tag{3.29}$$

times the coefficient $[s^l]s^l(1-s)e^{ks}$. Hence (3.5) with $P_j = j^{j-1}/j!$ gives the wanted coefficient c_{nklm}.

For $\varphi(t) = (1+t)^N$ we have $h(t) = 1+t$, $\frac{A_0}{1+A_0} = p$, $1+A_0 = 1/(1-p)$, and $\frac{\varphi(A_0)}{1+A_0} = (1-p)^{-(N-1)}$. The function in (3.24) reads

$$\frac{1}{n}\frac{1-(N-1)s}{1+s}(1+s)^{Nn}\left(\frac{rU}{(1-p)^{(N-1)n+1}} + q\frac{R(1+rS+NrU)}{(1-p)^{(N-1)n+2}}\right). \tag{3.30}$$

Therefore the coefficient $[q^{n-k}]$ equals

$$\frac{1}{n}\frac{1-(N-1)s}{1+s}(1+s)^{Nk}\left(\binom{Nn-k}{n-k}rU(1+rS)^{N(n-k)} + \binom{Nn-k}{n-k-1}(1+r(S+NU))(1+rS)^{N(n-k)}\right). \tag{3.31}$$

From this we get that $[r^{k-l}]$ equals

$$\frac{1}{n}\frac{1-(N-1)s}{1+s}(1+s)^{Nk}\overline{Q} \tag{3.32}$$

with

$$\overline{Q} = \binom{Nn-k}{n-k}\binom{N(n-k)}{k-l-1}US^{k-l-1} + \binom{Nn-k}{n-k-1}\binom{N(n-k)}{k-l}S^{k-l} + \binom{Nn-k}{n-k-1}\binom{N(n-k)}{k-l-1}\left(S^{k-l} + NUS^{k-l-1}\right). \tag{3.33}$$

Using (3.14), (3.15) we get after simplification (3.3) and from this the numbers for planted plane trees (3.1).

Remark 3: Note that it is also possible but rather technical to get explicit expressions for the numbers c_{nklm_I} defined in (4.16). Explicit formulae for c_{nkl} $(I = \emptyset)$ are given in [Ba].

4. Asymptotic Results

Let c_{nk} $(k = (k_1, \ldots, k_M))$ be non-negative numbers such that $c_n = \sum_k c_{nk} < \infty$. Then we say that c_{nk} satisfy a limit theorem if discrete random vectors $X_n = (X_{n1}, \ldots, X_{nM})$ with

$$\mathrm{P}[X_n = k] = \frac{c_{nk}}{c_n} \tag{4.1}$$

have a special limit distribution. For example, if these random vectors X_n are asymptotically normal with mean $\mu_n = \mathbf{E}X_n \sim n\mu$ and covariance matrix $\Sigma_n = \mathrm{Cov}X_n \sim n\Sigma$ (for some fixed μ, Σ) we say that c_{nk} satisfy a central limit theorem. The following theorems show that such a situation is quite natural in combinatorial enumeration problems where functional equations of the type $C = G(C, x, z)$ are involved.

We will use the following notation. Let $c = f_1(z), x = f_2(z)$ be a solution of the system of equations

$$\begin{aligned} c &= G(c, x, z) \\ 1 &= G_c(c, x, z), \end{aligned} \tag{4.2}$$

where $z = (z_1, \dots, z_M)$ is interpreted as a parameter. Then we define a vector $\mu(z) = (\mu_i(z))_{i=1,\dots,M}$ by

$$\mu_i(z) = \frac{z_i G_{z_i}}{x G_x}(f_1(z), f_2(z), z) \tag{4.3}$$

and a matrix $\Sigma(z) = (\sigma_{ij})_{i,j=1,\dots,M}$ by

$$\begin{aligned} \sigma_{ii}(z) &= \mu_i(z)^2 + \mu_i(z) \\ &+ \frac{z_i^2}{x G_x^3 G_{cc}} \Big[G_x^2 (G_{cc} G_{z_i z_i} - G_{cz_i}^2) - 2 G_x G_{z_i} (G_{cc} G_{x z_i} - G_{cx} G_{cz_i}) \\ &+ G_{z_i}^2 (G_{cc} G_{xx} - G_{cx}^2) \Big] (f_1(z), f_2(z), z) \end{aligned} \tag{4.4}$$

and by

$$\begin{aligned} \sigma_{ij}(z) &= \mu_i(z) \mu_j(z) \\ &+ \frac{z_i z_j}{x G_x^3 G_{cc}} \Big[G_x^2 (G_{cc} G_{z_i z_j} - G_{cz_i} G_{cz_j}) \\ &- G_x G_{z_i} (G_{cc} G_{x z_j} - G_{cx} G_{cz_j}) - G_x G_{z_j} (G_{cc} G_{x z_i} - G_{cx} G_{cz_i}) \\ &+ G_{z_i} G_{z_j} (G_{cc} G_{xx} - G_{cx}^2) \Big] (f_1(z), f_2(z), z) \end{aligned} \tag{4.5}$$

for $i \neq j$.

In many situations it will be more convenient to solve

$$\begin{aligned} c &= G(c, x, z) \\ 1 &= G_c(c, x, z) \\ \rho_i &= \frac{z_i G_{z_i}(c, x, z)}{x G_x(c, x, z)} \quad (1 \le i \le M) \end{aligned} \tag{4.6}$$

instead of (4.2). Let (c_ρ, x_ρ, z_ρ) be the solution of (4.6) with parameter $\rho = (\rho_1, \dots, \rho_M)$. Then $\mu(z_\rho) = \rho$ and of course $f_1(z_\rho) = c_\rho$ and $f_2(z_\rho) = x_\rho$.

If we set $\psi_n(z) = \sum_k c_{nk} z^k$, then let

$$d = \gcd\{n - l > 0 : \psi_n(z) \not\equiv 0\}, \tag{4.7}$$

where $l = \min\{m > 0 : \psi_m(z) \not\equiv 0\}$.

If $Z \subseteq (0, \infty)^M$, then define a set $R(Z, \varepsilon)$ by

$$R(Z,\varepsilon) = \left\{ z \in \mathbf{C}^M : |z| = (|z_1|, \ldots, |z_M|) \in Z, \max_{1 \le i \le M} \arg(z_i) < \varepsilon \right\}. \tag{4.8}$$

Theorem 3. *Let* $C(x,z) = \sum_{n,k} c_{nk} x^n z^k$ $(z^k = z_1^{k_1} \cdots z_M^{k_M})$ *be the generating function of numbers* $c_{nk} \geq 0$ *that satisfies a functional equation of the form*

$$C = G(C, x, z) = \sum_{i,j,m} g_{ijm} C^i x^j z^m \tag{4.9}$$

such that

$$\sum_m c_{ijm} z^m \geq 0 \qquad (i, j \geq 0) \tag{4.10}$$

for z in some (real) neighbourhood of $(1, \ldots, 1)$. *Suppose that there are positive solutions* $c_0 = f_1(1)$, $x_0 = f_2(1)$ *of system (4.2) (*x_0 *must be chosen to be minimal), that* $(c_0, x_0, 1)$ *is a regular point of the series expansion of* $G(c, x, z)$, *that*

$$G_{cc}(c_0, x_0, 1) G_x(c_0, x_0, 1) > 0, \tag{4.11}$$

and that $\det \Sigma(1) \neq 0$. *Then the numbers* c_{nk} $(n \equiv l \mod d)$ *satisfy a central limit theorem with*

$$\mu_n = n\mu(1) + O(1) \qquad (n \equiv l \mod d) \tag{4.12}$$

and

$$\Sigma_n = n\Sigma(1) + O(1) \qquad (n \equiv l \mod d). \tag{4.13}$$

Theorem 4. *Suppose in addition to the assumptions of Theorem 3 that there is a compact connected set* $Z \subseteq (0, \infty)^M$ *such that*

(i) $G_{cc}(f_1(z), f_2(z), z)\, G_x(f_1(z), f_2(z), z) > 0$ *for* $z \in Z$,

(ii) $\det \Sigma(z) \neq 0$ *for* $z \in Z$, *and*

(iii.a) $g_{ijm} \geq 0$ *and that there are* k_1, k_2, k_3 *with* $c_{nk_1} c_{nk_2} c_{nk_3} > 0$ *and* $\gcd\{k_{1,i}, k_{2,i}, k_{3,i}\} = 1$ *for some* $i \geq 1$ *or*

(iii.b) $c(x,z)$ *is regular for* $z \notin R(Z, \varepsilon)$, $|z| \in Z$, *and* $|x| \leq (1+\delta) f_2(|z|)$ *for some* $\varepsilon, \delta > 0$.

Then c_{nk} ($n \equiv l \mod d$) can be determined asymptotically by

$$c_{nk} = d\left[\frac{x_\rho G_x(c_\rho, x_\rho, z_\rho)}{2\pi G_{cc}(c_\rho, x_\rho, z_\rho) n^3 (2\pi n)^M \det\Sigma(z_\rho)}\right]^{\frac{1}{2}} x_\rho^{-n} z_\rho^{-k}\left(1 + O\left(n^{-\frac{1}{2}}\right)\right) \tag{4.14}$$

uniformly for $z_\rho \in Z$, where $\rho = k/n$ and c_ρ, x_ρ, z_ρ are defined in (4.6), and locally by

$$c_{nk} = c_{nj} z_{j/n}^{j-k}\left(e^{-\frac{1}{2n}(k-j)\Sigma(z_{j/n})^{-1}(k-j)'} + O\left(n^{-\frac{1}{2}}\right)\right) \tag{4.15}$$

uniformly for all k and $z_{j/n} \in Z$.

A proof of both theorems can easily be given by a combination of the ideas of [Be], [BR], and [Dr]. (Note that (iii.a) implies (iii.b).)

Now it is easy to apply these theorems to the case of induced subgraphs of trees:

Theorem 5. *Let a simply generated family of trees be characterized by a power series $\varphi(t)$ with positive radius of convergence R such that there are positive real solutions $u_0 < R, v_0 < R - u_0/2$ of the equations*

$$u_0\varphi'(u_0) = \varphi(u_0), \qquad v_0\varphi'\left(\frac{u_0}{2} + v_0\right) = \varphi\left(\frac{u_0}{2} + v_0\right). \tag{4.15}$$

Fix some finite set I of positive integers with cardinality M. Then the numbers

$$c_{nklm_I} = \sum_{m_i, i \notin I} c_{nklm} \tag{4.16}$$

satisfy a central limit theorem and multivariate asymptotic expansion can be obtained as formulated in Theorem 4. The average number of nodes of induced subsets in trees of size n is given by $n/2 + O(1)$, the average number of edges by $n/4 + O(1)$, and the average number of components of size j by $n\mu_j + O_j(1)$ with

$$\mu_j = \frac{\frac{1}{j}[t^{j-1}]\varphi\left(\frac{u_0}{2} + t\right)^j}{2^{j+1}\varphi'(u_0)^{j-1}\varphi(u_0)}. \tag{4.17}$$

Furthermore the covariance matrix related to the number of nodes and edges in induced subgraphs is given by $n\Sigma + O(1)$, where

$$\Sigma = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{5}{16} + \frac{1}{16}\frac{\varphi(u_0)\varphi''(u_0)}{\varphi'(u_0)^2} \end{pmatrix}. \tag{4.18}$$

So we have in general ($I = \emptyset$)

$$\frac{c_{nkl}}{y_n} = \frac{2^{n+2}}{\pi n\left(1 + \frac{\varphi(u_0)\varphi''(u_0)}{\varphi'(u_0)^2}\right)^{\frac{1}{2}}}\left(e^{-\frac{1}{2n}(k-\frac{n}{2}, l-\frac{n}{4})\Sigma^{-1}(k-\frac{n}{2}, l-\frac{n}{4})'} + O\left(n^{-\frac{1}{2}}\right)\right) \tag{4.19}$$

where y_n is defined in (1.5).

Remark 4: It is remarkable that the average number of nodes and edges is almost independent of $\varphi(t)$. (A similar statement holds for the covariance matrix.) Only if $\varphi_i \in \{0, 1\}$ it is easy to see that the expected number of nodes is exactly $n/2$. Since the number of components equals $k - l$, the average number of components is given by $n/4 + O(1)$, that is almost independent of $\varphi(t)$, too. Thus an average component consists of $2 + O(1/n)$ nodes. These results correspond with the identities

$$\sum_{j\geq 1} \mu_j = \frac{1}{4} \qquad \text{and} \qquad \sum_{j\geq 1} j\mu_j = \frac{1}{2}. \tag{4.20}$$

Namely, if we set $p_j = 4\mu_j$, then a random variable Y with $\mathrm{P}[Y = j] = p_j$ has $\mathrm{E}Y = 2$. These identities can be proven by Lagrange's inversion formula applied to $y = x\varphi(u_0/2 + y)$, where $y(1/(2\varphi'(u_0))) = u_0/2$. Furthermore one can obtain an explicit expression for the variance

$$\mathrm{V}Y = \sum_{j\geq 1} (j-2)^2 p_j = 2\left(1 + \frac{\varphi(u_0)\varphi''(u_0)}{\varphi'(u_0)^2}\right). \tag{4.21}$$

Remark 5: μ_j can be evaluated explicitely for $\varphi(t) = 1/(1-t)$:

$$\mu_j = \frac{1}{j}\binom{2j-2}{j-1}\frac{2^{j-2}}{3^{2j-1}}, \tag{4.22}$$

for $\varphi(t) = (1+t)^N$:

$$\mu_j = \frac{1}{j}\binom{Nj}{j-1}\frac{(2N-1)^{(N-1)j+1}}{4(2N)^{Nj}} \tag{4.23}$$

and for $\varphi(t) = e^t$:

$$\mu_j = \frac{j^{j-1}}{j!\,2^{j+1}e^{j/2}}. \tag{4.24}$$

In general it is easy to get explicit expressions for μ_j for small values (e.g. for $j \leq 3$):

$$\begin{aligned} \mu_1 &= \frac{\varphi\left(\frac{u_0}{2}\right)}{4\varphi(u_0)}, \qquad \mu_2 = \frac{\varphi\left(\frac{u_0}{2}\right)\varphi'\left(\frac{u_0}{2}\right)}{8\varphi(u_0)\varphi'(u_0)}, \\ \mu_3 &= \frac{\varphi\left(\frac{u_0}{2}\right)\varphi'\left(\frac{u_0}{2}\right)^2 + \frac{1}{2}\varphi\left(\frac{u_0}{2}\right)^2\varphi''\left(\frac{u_0}{2}\right)}{16\varphi(u_0)\varphi'(u_0)^2}. \end{aligned} \tag{4.25}$$

For large values we can use Darboux's method to get an asymptotic expansion of the Taylor coefficients of $y(x)$ of the solution of
$y = x\varphi(u_0/2 + y)$ to obtain

$$\mu_j = \frac{1}{2u_0}\left[\frac{\varphi\left(\frac{u_0}{2}+v_0\right)}{2\pi\varphi''\left(\frac{u_0}{2}+v_0\right)}\right]^{\frac{1}{2}}\left[\frac{\varphi\left(\frac{u_0}{2}+v_0\right)}{2\varphi'(u_0)}\right]^{j} j^{-3/2}\left(1+O\left(\frac{1}{j}\right)\right). \tag{4.26}$$

Proof of Theorem 5: First we use Theorem 4 for the generating function $A_0(x, y, z, v)$, where $v_i = 1$ for $i \notin I$. By (2.5) we have

$$\begin{aligned} A_0 = x\varphi\Big(A_0 &+ \sum_{j\in I}(xy)^j z^{j-1} v_j \frac{1}{j}[t^{j-1}]\varphi(A_0+t)^j \\ &+ \sum_{j\notin I}(xy)^j z^{j-1}\frac{1}{j}[t^{j-1}]\varphi(A_0+t)^j\Big). \end{aligned} \tag{4.27}$$

It satisfies all assumptions for $l = 1$ and $d = \gcd\{m > 1 : \varphi_m \neq 0\}$. So $A_0(x, y, z, v)$ is analytic and bounded in a reagion $(y, z, v) \notin R(Y \times Z \times V, \varepsilon)$, (see (4.8)), $|x| \le (1+\delta) f_2(|y|, |z|, |v|)$ for $(|y|, |z|, |v|) \in Y \times Z \times V$. But now

$$\begin{aligned} C = A_0 &+ \sum_{j\in I}(xy)^j z^{j-1} v_j \frac{1}{j}[t^{j-1}]\varphi(A_0+t)^j \\ &+ \sum_{j\notin I}(xy)^j z^{j-1}\frac{1}{j}[t^{j-1}]\varphi(A_0+t)^j \end{aligned} \tag{4.28}$$

and so $C = C(x, y, z, v)$ is analytic and bounded in the same reagion. Thus we get asymptotic normality and multivariate asymptotic expansions. It should be noted that $f_1(1,1,1) = u_0$ and that $f_2(1,1,1) = 1/(2\varphi'(u_0))$, where u_0 is the positive solution of $u_0\varphi'(u_0) = \varphi(u_0)$. So it is easy to derive $\mu(1,1,1)$ and $\Sigma(1,1,1)$.

5. Expected Value of the Largest Component

Let $a_{L,j}(T)$ be the number of induced subgraphs H of T such that all components of H are of size $\le L$ and that the root of T is contained in a component of size j. Set

$$a_{nL,j} = \sum_{|V(T)|=n} \omega(T) a_{L,j}(T) \tag{5.1}$$

and

$$A_{L,j}(x) = \sum_n a_{nL,j} x^n, \quad C_L(x) = \sum_{j=0}^{L} A_{L,j}(x). \tag{5.2}$$

As in section 2 we can prove

Lemma 2. *If a simply generated family of trees is characterized by $\varphi(t)$, then*

$$A_{L,0}(x) = x\varphi\left(\sum_{j=1}^{L} A_{L,j}(x)\right) \tag{5.3}$$

$$A_{L,j}(x) = x^j \frac{1}{j}[t^{j-1}]\varphi(A_{L,0}(x)+t)^j \tag{5.4}$$

$$C_L(x) = x\varphi(C_L(x)) + \sum_{j=0}^{L} x^j \frac{1}{j}[t^{j-1}]\varphi(x\varphi(x\varphi(C_L(x)) + t)^j . \tag{5.5}$$

Let

$$H_L(x,c) = \sum_{j=1}^{L} x^j \frac{1}{j}[t^{j-1}]\varphi(x\varphi(c)+t)^j \tag{5.6}$$

$(1 \le L \le \infty)$ and $y_{nL} = [x^n]C_L(x)$, where $C_L(x)$ is the solution of

$$C_L = x\varphi(C_L) + H_L(x, C_L). \tag{5.7}$$

Then

$$F_n(x) = \frac{y_{nL}}{y_{n\infty}} \qquad (L \le x < L+1) \tag{5.8}$$

is the distribution function of the largest component of induced subgraphs in trees of size n.

Remark 6: Note that $y_{n\infty}$ counts all subsets of trees of size n. So $C_\infty(x)$ satisfies the functional equation $C_\infty = 2x\varphi(C_\infty)$. Therefore we have

$$\begin{aligned} H_\infty(x, C_\infty(x)) &= \sum_{j\ge1} x^j \frac{1}{j}[t^{j-1}]\varphi(x\varphi(C_\infty(x)) + t)^j \\ &= x\varphi(C_\infty(x)) = \frac{1}{2}C_\infty(x). \end{aligned} \tag{5.9}$$

This relation can also be checked by using Lagrange's inversion formula, for $H_\infty(x,c)$ is the solution of $H_\infty = x\varphi(x\varphi(c) + H_\infty)$. Set $c = C_\infty(x)$. Then by (5.7) $H_\infty(x, C_\infty(x)) = x\varphi(C_\infty(x)) = \frac{1}{2}C_\infty(x)$.

Discussing the asymptotic expansion of y_{nL} we will prove the following theorem:

Theorem 6. *Let a simply generated family of trees be characterized by a power series $\varphi(t)$ with positive radius of convergence R such that there are positive real solutions $u_0 < R$, $v_0 < R - u_0/2$ of the equations*

$$u_0\varphi'(u_0) = \varphi(u_0), \qquad v_0\varphi'\left(\frac{u_0}{2} + v_0\right) = \varphi\left(\frac{u_0}{2} + v_0\right). \tag{5.10}$$

Then the expected size of the largest component of induced subgraphs of trees of size n is given by

$$\frac{\log n}{\log \frac{2\varphi'(u_0)}{\varphi'(u_0/2+v_0)}} + O(\log\log n). \tag{5.11}$$

and the variance can be estimated by $O(\log n \log\log n)$.

Theorem 6 is an easy consequence of

Lemma 3. *The numbers y_{nL} ($n \equiv 1 \mod d$) can be expanded asymptotically by*

$$y_{nL} = C_L x_L^{-n} n^{-\frac{3}{2}} \left(1 + O(n^{-1})\right) \tag{5.12}$$

uniformly for $L \geq L_0$ and $n \geq n_0$, where

$$C_L = \left[\frac{\varphi(u_0)}{2\pi\varphi''(u_0)}\right]^{\frac{1}{2}} + O(\alpha^L) \tag{5.13}$$

$$x_L = \frac{1}{2\varphi'(u_0)} + \overline{C}\frac{\alpha^L}{L^{3/2}} \left(1 + O(L^{-1})\right) \tag{5.14}$$

with

$$\alpha = \frac{\varphi'\left(\frac{u_0}{2} + v_0\right)}{2\varphi'(u_0)} \tag{5.15}$$

and

$$\overline{C} = \left[\frac{\varphi'\left(\frac{u_0}{2} + v_0\right)}{2\pi\varphi''\left(\frac{u_0}{2} + v_0\right)}\right]^{\frac{1}{2}} \frac{\varphi'\left(\frac{u_0}{2} + v_0\right)}{4\varphi(u_0)\left(2\varphi'(u_0) - \varphi'\left(\frac{u_0}{2} + v_0\right)\right)}. \tag{5.16}$$

Proof: Using Darboux's method it is easy to see that

$$y_{nL} = d\left[\frac{x_L\left(\varphi(c_L) + \frac{\partial}{\partial x}H_L(x_L, c_L)\right)}{2\pi\left(x_L\varphi''(c_L) + \frac{\partial^2}{\partial c^2}H_L(x_L, c_L)\right)}\right]^{\frac{1}{2}} x_L^{-n} n^{-\frac{3}{2}} \left(1 + O_L(n^{-1})\right) \tag{5.17}$$

for $n \equiv 1 \mod d$, where c_L, x_L are the solutions of

$$\begin{aligned} c &= x\varphi(c) + H_L(x, c) \\ 1 &= x\varphi'(c) + \frac{\partial}{\partial c}H_L(x, c). \end{aligned} \tag{5.18}$$

Set $c_L = u_0 + s_L$ and $x_L = x_0 + r_L$, where $x_0 = 1/(2\varphi'(u_0))$. Then it can be shown that (5.14) and

$$s_L = \overline{\overline{C}}\alpha^L L^{-\frac{1}{2}} \left(1 + O(L^{-1})\right) \tag{5.19}$$

hold with

$$\overline{\overline{C}} = \left[\frac{\varphi'\left(\frac{u_0}{2}+v_0\right)}{2\pi\varphi''\left(\frac{u_0}{2}+v_0\right)}\right]^{\frac{1}{2}} \frac{\varphi'(u_0)\varphi'\left(\frac{u_0}{2}+v_0\right)}{4\varphi''(u_0)\left(2\varphi'(u_0)-\varphi'\left(\frac{u_0}{2}+v_0\right)\right)}. \tag{5.20}$$

The first equation of (5.18) can be rewritten as

$$\begin{aligned} u_0 + s_L = \frac{u_0}{2} + \frac{s_L}{2} + \varphi(u_0) r_L + O\left(r_L^2 + s_L^2\right) \\ + H_\infty(x_L, s_L) + \left(H_L(x_L, s_L) - H_\infty(x_L, s_L)\right). \end{aligned} \tag{5.21}$$

But

$$H_\infty(x_L, s_L) = \frac{u_0}{2} + \frac{s_L}{2} + 3\varphi(u_0) r_L + O\left(r_L^2 + s_L^2\right) \tag{5.22}$$

and it is possible to prove the expansion

$$H_L(x_L, s_L) - H_\infty(x_L, s_L) = -\overline{C}_L \alpha_L^{L+1} L^{-\frac{3}{2}} \left(1 + O(L^{-1})\right), \tag{5.23}$$

where

$$\alpha_L = x_L \varphi'\left(x_L \varphi(c_L) + v\left(x_L \varphi(c_L)\right)\right), \tag{5.24}$$

$$\overline{C}_L = \frac{\left[\frac{\varphi(x_L\varphi(c_L)+v(x_L\varphi(c_L)))}{2\pi\varphi''(x_L\varphi(c_L)+v(x_L\varphi(c_L)))}\right]^{\frac{1}{2}}}{1 - x_L\varphi'\left(x_L\varphi(c_L) + v\left(x_L\varphi(c_L)\right)\right)}, \tag{5.25}$$

and $v(u)$ is defined by $v(u)\varphi'(u+v(u)) = \varphi(u+v(u))$.

Similarly the second equation of (5.18) can be simplified to

$$\begin{aligned} 0 = 4\left(\varphi'(u_0) + u_0\varphi''(u_0)\right) r_L + 4 x_0 \varphi''(u_0) s_L + O(r_L^2 + s_L^2) \\ - x_L \varphi'(c_L) \overline{\overline{C}}_L \alpha_L^{L+1} L^{-\frac{1}{2}} \left(1 + O(L^{-1})\right). \end{aligned} \tag{5.26}$$

So it is easy to see that (5.14) and (5.19) are correct.

Now it must be shown that the O_L-constants in (5.17) can be chosen uniformly for $L \geq L_0$. By [FO] it sufficies to ensure that the solution of $c = x\varphi(c) + H_L(x, c)$ can be expanded as

$$c = C_L(x) = c_{L0} - c_{L1}(x_L - x)^{\frac{1}{2}} - c_{L2}(x_L - x) + O\left((x_L - x)^{\frac{3}{2}}\right) \tag{5.27}$$

for $x \in \Delta_L = \{z \in \mathbf{C} : |z| < (1+\delta_1)x_L, \arg(z - x_L) > \delta_2\}$ $(\delta_1, \delta_2 > 0)$ and that this O-constant can be chosen uniformly for $L \geq L_0$. Since c_L and x_L converge to u_0 and x_0 it is not difficult to show that (5.27) holds in some neighbourhood $K(x_L, \varepsilon) = \{z \in \mathbf{C} : |z - x_L| < \varepsilon, \arg(z - x_L) \neq 0\}$. You only have to use Taylor's theorem for $c = x\varphi(c) + H_L(x, c)$. But $\Delta_L \setminus K(x_L, \varepsilon)$

is contained in a compact region for $L \geq L_0$, where $C_L(x)$ converges uniformly to $C_\infty(x)$. So we are done.

Proof of Theorem 6: We split the sum

$$\sum_{l=1}^{n} l(F_n(l) - F_n(l-1)) \tag{5.28}$$

into four parts:

$$\begin{aligned}
S_1 &: & 1 \leq l &< \frac{1}{-\log\alpha}\left(\log n - \frac{5}{2}\log\log n\right) \\
S_2 &: \frac{1}{-\log\alpha}\left(\log n - \frac{5}{2}\log\log n\right) & \leq l &< \frac{1}{-\log\alpha}\log n \\
S_3 &: & \frac{1}{-\log\alpha}\log n \leq l &< \frac{3}{-\log\alpha}\log n \\
S_4 &: & \frac{3}{-\log\alpha}\log n \leq l &\leq n.
\end{aligned}$$

Then we can estimate S_1, S_3, S_4 by

$$\begin{aligned}
S_1 &\ll \log n F_n\left(\frac{1}{-\log\alpha}\left(\log n - \frac{5}{2}\log\log n\right)\right) \ll \frac{\log n}{n} \\
S_3 &\ll \log n\left(1 - F_n\left(\frac{\log n}{-\log\alpha}\right)\right) \ll (\log n)^{-\frac{1}{2}} \\
S_4 &\ll n\left(1 - F_n\left(\frac{3\log n}{-\log\alpha}\right)\right) \ll n^{-2}
\end{aligned} \tag{5.29}$$

and expand S_2 to

$$S_2 = \frac{\log n}{-\log\alpha} + O(\log\log n). \tag{5.30}$$

Simlilary we get

$$\sum_{l=1}^{n} l^2 (F_n(l) - F_n(l-1)) = \left(\frac{\log n}{-\log\alpha} + O(\log\log n)\right)^2 + o(1). \tag{5.31}$$

This proves the estimate for the variance.

6. Appendix: An Interpretation for $\varphi(u_0)\varphi''(u_0)/\varphi'(u_0)^2$

We will show that $\varphi(u_0)\varphi''(u_0)/\varphi'(u_0)^2$, a constant that seems to be significant for many asymptotic properties in tree enumeration problems, can be interpreted as the variance of the outdegree of nodes in trees in a simply generated family of trees.

Let

$$c_p(T) = \begin{cases} 1 & \text{if} \quad D_i(T) = p_i \quad (i \geq 0) \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

$(p = (p_0, p_1, \ldots))$,

$$c_{np} = \sum_{|V(T)|=n} \omega(T) c_p(T), \tag{6.2}$$

and

$$C(x, w) = \sum_{n,p} c_{np} x^n w^p. \tag{6.3}$$

Then it is easy to see that

$$C = C(x, w) = x \sum_{n \geq 0} \varphi_n v_n C^n, \tag{6.4}$$

so that the average number of nodes with out-degree i equals $nm_i + O_i(1)$, where

$$m_i = \frac{\varphi_i u_0^i}{\varphi(u_0)}. \tag{6.5}$$

Let Z be a random variable with $\mathrm{P}[Z = i] = m_i$. Then it is easy to calculate that

$$\mathbf{E}Z = 1 \quad \text{and} \quad \mathbf{V}Z = \frac{\varphi(u_0)\varphi''(u_0)}{\varphi'(u_0)^2}. \tag{6.6}$$

References

[Ba] Baron, G., *Enumeration of Induced Subgraphs of Families of Trees*, Ars Combinatoria **20**-A (1985), 77-82.

[Be] Bender, E.A., *Central and Local Limit Theorems Applied to Asymptotic Enumeration*, J. Comb. Th. (A) **15** (1973), 91–111.

[BR] Bender, E.A., and L.B. Richmond, *Central and Local Limit Theorems Applied to Asymptotic Enumeration II: Multivariate Generating Functions*, J. Comb. Th. (A) **34** (1983), 255–265.

[Dr] Drmota, M., *Asymptotic Distributions and a Multivariate Darboux Method in Enumeration Problems*, submitted.

[Fl] Flajolet, Ph., *Mathematical Methods in the Analysis of Algorithms and Data Structures*, INRIA, Research Report 400, 1985, to appear in: A Graduate Course in Computer Science, Computer Science Press, 1987.

[FO] Flajolet, Ph., and A. Odlyzko, *Singularity Anlalysis of Generating Functions*, SIAM J. Disc. Math. 3 (1990), 216–240.

[MM] Meir, A., and J.W. Moon, *On Maximal Independent Sets of Nodes in Trees*, J. Graph Th. 12 (1988), 265–283.