On edge-weighted recursive trees and inversions in random permutations.

M. Kuba and A. Panholzer*
Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104
1040 Wien, Austria,
Markus.Kuba@tuwien.ac.at, Alois.Panholzer@tuwien.ac.at

Corresponding Author: M. Kuba

May 30, 2006

Abstract

We introduce random recursive trees, where deterministically weights are attached to the edges according to the labeling of the trees. We will give a bijection between recursive trees and permutations, which relates the arising edge-weights in recursive trees with inversions of the corresponding permutations. Using this bijection we obtain exact and limiting distribution results for the number of permutation of size n, where exactly m elements have j inversions. Furthermore we analyze the distribution of the sum of labels of the elements, which have exactly j inversions, where we can identify Dickman's infinitely distribution as the limit law. Moreover we give a distributional analysis of weighted depths and weighted distances in edge-weighted recursive trees.

Keywords: Recursive tree, permutation, inversion, Dickman distribution
2000 Mathematics Subject Classification 05C05.

1 Introduction

There are several well-known bijections between recursive trees and permutations. For example, a rotation correspondence is given in [2], which immediately characterizes the distribution of the root degree, the number of leaves, etc., in a random recursive tree. Here we state a natural, but new bijection (best to our knowledge), which maps inversions in random permutations of $\{1, 2, \dots, n-1\}$ to suitably defined weights on the edges of recursive trees with n nodes. Using this bijection we are able to relate parameters in permutations, as, e.g., the number of inversions, with parameters in recursive trees.

A rooted non-plane size-n tree labeled with distinct integers $1, 2, \dots, n$ is a recursive tree if the node labeled 1 is distinguished as the root, and, for each $2 \le k \le n$, the labels of the nodes on the unique path from the root to the node labeled k form an increasing sequence. Every size-n recursive tree can be obtained uniquely by attaching node n to one of the $n-1$ nodes in a recursive tree of size $n-1$. This immediately shows that the number T_n of recursive trees of size n is given by $(n-1)!$, for $n \ge 1$. Throughout this paper we assume as the model of randomness the random tree model, which means that all $(n-1)!$ recursive trees of size n are considered to appear equally likely. We speak then about random recursive trees. Equivalently one may describe random recursive trees via the following tree evolution process, which generates random recursive

*This work was supported by the Austrian Science Foundation FWF, grant S9608.

trees of arbitrary size n. At step 1 the process starts with the root labeled by 1. At step $i+1$ the node with label $i+1$ is attached to any previous node v of the already grown tree T of size i with probability $p_i(v) = 1/i$.

Due to this simple growth rule random recursive trees have been introduced as a probability model in several areas. E.g., they are used to model the spread of epidemics [16], to aid in the construction of the family trees of preserved copies of ancient manuscripts [17], or to model chain letter and pyramid schemes [6]. Furthermore they are used to model the stochastic growth of networks [3]. See also the survey paper [15].

Let T denote a recursive tree of size $|T|$. Throughout this paper we always consider edge-weighted recursive trees, where every edge $e \in E = E(T)$ of the tree will be weighted deterministically as follows. If the edge $e = (j,k)$ is adjacent to the nodes j and k, then we define the weight w_e of the edge e as $w_e := |k-j|$. The aim of this paper is to study several parameters of random recursive trees appearing in connection with this edge-weights.

We will analyze the random variable $S_n := \sum_{e\in E_n} w_e$ counting the sum of all edge-weights and the random variable $S_{n,j} = \sum_{e\in E_n} \mathbb{1}_{\{w_e=j\}}$ counting the number of edge-weights w_e with $w_e = j$ in a random recursive tree of size n. By using the bijection between recursive trees and permutations already mentioned it turns out that the r. v. S_n essentially counts the number inversions and the r. v. $S_{n,j}$ is given by the number of elements with exactly $j-1$ inversions in a random permutation of size $n-1$.

We also study the r. v. $W_{n,j} := \sum_{e=(e_1,e_2)\in E_n} e_2 \mathbb{1}_{\{w_e=j\}}$, which counts the sum of the labels of nodes attached via edges of weight j, for the full range of $j = j(n)$ growing with the size n of a random recursive tree. We show that the limiting distribution of $W_{n,j}$ depends on the growth of $j = j(n)$, where Dickman's infinitely divisible distribution appears in the limit law.

Furthermore we analyze the random variable G_n counting the number of absent edge-weights of the set $\{1,2,\dots,n-1\}$ in a random recursive tree of size n. We can show a Gaussian limit law for the suitably normalized and centered r. v. G_n. Moreover we consider the random variable M_n, which gives the maximum edge-weight appearing in a random recursive tree of size n. Due to the correspondence with inversion statistics in random permutations we obtain a Rayleigh distribution as limit law of the suitably normalized and scaled r. v. M_n.

In a rooted tree the depth of node v, also called the level of node v, is measured by the number of edges lying on the unique path from the root to node v. Here we consider a generalization of the depth for edge-weighted trees. Let $X_{n,j} := \sum_{e\in E} w_e \mathbb{1}(A_e)$ denote the random variable counting the edge-weighted depth of node j in a random recursive tree of size n, where A_e denotes the event that edge e is on the unique path from the root to node j. Further we denote by $\mathfrak{X}_{n,j}$ the random variable counting the edge-weighted distance between nodes j and n in a random recursive tree of size n, measured by the sum of the edge-weights on the unique path from j to n. The r. v. $X_{n,j}$ can be trivially characterized, whereas the characterization of $\mathfrak{X}_{n,j}$ leads to a discrete limit law. Using the distribution law of $\mathfrak{X}_{n,j}$ we are also able to obtain the distribution of the label of the root of the spanning tree of two randomly selected nodes in a random recursive tree of size n.

When speaking about inversions in permutations we will here always think about "right inversions", i.e., inversions caused by elements to the right. For a permutation $\sigma = (\sigma_1 \dots \sigma_n)$ of $\{1,2\dots,n\}$ and $1 \le k \le n$ we call the number of elements in σ to the right of k, which are smaller than k, the element inversions $i_k = i_k(\sigma)$ of k. Hence the inversion table of $\sigma = (\sigma_1 \dots \sigma_n)$ is given by $(i_1, i_2, \dots, i_n)$, with the restrictions $0 \le i_k \le k-1$, for $1 \le k \le n$.

Throughout this paper we denote by $X \overset{(d)}{=} Y$ the equality in distribution of the random variables X and Y, and with $X_n \xrightarrow{(d)} X$ the weak convergence, i. e., the convergence in distribution, of the sequence of random variables X_n to a random variable X. For independent random variables X and Y we denote the sum of X and Y by $X \oplus Y$, whereas for not necessarily independent random variables X and Y we write $X+Y$. We will denote by $\left[{n \atop k}\right]$ the signless Stirling numbers of the first kind and by $\left\langle{n \atop k}\right\rangle$ the Eulerian numbers. Furthermore we use the Iverson bracket-notation: for a statement A we have $[\![A]\!] = 1$ if A is true and $[\![A]\!] = 0$ otherwise. Moreover we denote by $\{j <_c k\}$ the event that node j is attached to node k (i.e., j is a child of k) for a given tree T.

In Section 2 we give the bijection between recursive trees and inversions in permutations and study the random variables S_n, G_n and M_n. Section 3 is devoted to the analysis of specific edge weights $S_{n,j}$ and $W_{n,j}$, whereas Section 4 is devoted to the analysis of edge-weighted depths and distances, $X_{n,j}$ and $\mathfrak{X}_{n,j}$.

2 Edge-weights and inversions in random permutations

2.1 A bijection between recursive trees and permutations

We present the following bijection between recursive trees of size n and permutations of $\{1, \dots, n-1\}$, which turns out to be appropriate when studying parameters in edge-weighted recursive trees.

Bijection 1. *Consider a recursive tree T of size n and its edge set E. We enumerate the $n-1$ edges of E by $e_2, e_3, \dots, e_n$, where e_k, with $k \geq 2$, is defined as the edge $e_k = (j, k)$ connecting j and k, with $1 \leq j \leq k-1$. The edge e_k is uniquely defined, since every node $k \geq 2$ in a recursive tree is attached to exactly one node j with $1 \leq j \leq k-1$. We define now the numbers $q_k := w_{e_k} = k - j$ as the edge-weight of edge e_k and consider the edge-weight table $(q_2, q_3, \dots, q_n)$ of T. Of course, it holds $1 \leq q_k \leq k-1$, for $2 \leq k \leq n$. If we define numbers $i_k := q_{k+1} - 1$, for $1 \leq k \leq n-1$, then it holds $0 \leq i_k \leq k-1$ and the array $(i_1, i_2, \dots, i_{n-1})$ corresponds to the inversion table of a permutation σ of $\{1, 2, \dots, n-1\}$, which uniquely determines σ. To construct a recursive tree T of size n from a given permutation σ of $\{1, \dots, n\}$ with inversion table $(i_1, \dots, i_{n-1})$ one starts with 1 as the root of T and attaches successively node k, with $2 \leq k \leq n$, to node $j = k - 1 - i_{k-1}$, which leads to an edge-weight table $(q_2, \dots, q_n)$ with $q_k = k - j = i_{k-1} + 1$, for $2 \leq k \leq n$.*

2.2 The sum of edge-weights

Theorem 1. *The distribution of the random variable S_n, counting the sum of the edge-weights in a random recursive tree of size n, is given by*

$$S_n \stackrel{(d)}{=} \xi_{n-1} \oplus (n-1) \stackrel{(d)}{=} \mathrm{U}_1 \oplus \mathrm{U}_2 \oplus \cdots \oplus \mathrm{U}_{n-1}, \tag{1}$$

where the random variable ξ_n counts the number of inversions of a random permutation of size n, and U_k denotes a uniform distribution on the set $\{1, 2, \dots, k\}$.

Proof. The theorem is an immediate consequence of Bijection 1, since for a given recursive tree of size n with edge-weight table $(q_2, \dots, q_n)$ and inversion table $(i_1, \dots, i_{n-1})$ of the corresponding permutation of size $n-1$ we always have $\sum_{2 \leq k \leq n} q_k = n - 1 + \sum_{1 \leq k \leq n-1} i_{k-1}$. Furthermore for a random recursive tree of size n the edge-weights q_k are uniformly distributed on $\{1, \dots, k-1\}$ independent of the edge-weights q_l, $l \neq k$, since k is attached to one of the nodes $1 \leq j \leq k-1$ at random. □

Using Theorem 1 and the central limit theorem for the number of inversions ξ_n of a random permutation of size n, see [19], we obtain the following corollary.

Corollary 1. *The properly scaled and shifted random variable S_n converges, for $n \to \infty$, in distribution to a standard normal distributed r. v.:*

$$S_n^* := \frac{S_n - \mathbb{E}(S_n)}{\mathbb{V}(S_n)} \xrightarrow{(d)} \mathcal{N}(0, 1), \tag{2}$$

where $\mathcal{N}(0, 1)$ denotes the standard normal distribution.

2.3 The maximal edge-weight

Let M_n denote the random variable, which gives the maximal edge-weight in a random recursive tree of size n. As an immediate consequence of Bijection 1 we obtain the following result.

Theorem 2. *The distribution of the r. v. M_{n+1} is for $n \geq 1$ given as follows.*

$$M_{n+1} \overset{(d)}{=} \eta_n \oplus 1 \overset{(d)}{=} \max\{\mathrm{U}_1, \mathrm{U}_2, \dots \mathrm{U}_n\}, \tag{3}$$

where the random variable η_n gives the maximal entry in the inversion table of a random permutation of size n, and U_k denotes a uniform distribution on the set $\{1, 2, \dots, k\}$ and all r. v. U_k are mutually independent.

We want to remark that, due to Theorem 2, M_{n+1} also counts the number of passes that are required to sort a random permutation of size n by the sorting algorithm bubble sort. Using results from [9] and [14] we obtain distributional results for M_{n+1}.

Corollary 2. *The exact distribution of M_n is given as follows.*

$$\mathbb{P}\{M_{n+1} \leq m\} = \frac{m!\, m^{n-m}}{n!}, \quad \text{for } 1 \leq m \leq n, \quad \text{and} \quad n \geq 1.$$

The limiting distribution of the suitably scaled and shifted r. v. M_n can be characterized as follows:

$$M_n^* := \frac{M_n - n}{\sqrt{n}} \xrightarrow{(d)} -X,$$

where X is a Rayleigh distributed random variable with density function $f(x) = xe^{-\frac{x^2}{2}}$, for $x \geq 0$, and $f(x) = 0$, otherwise.

2.4 The number of absent edge-weights

Let G_n denote the random variable counting the number of elements of the set $\{1, 2, \dots, n-1\}$, which do not appear as an edge-weight in a random recursive tree of size n. Using Bijection 1 we immediately get the following result.

Theorem 3. *G_n is distributed as the number of elements of the set $\{0, 1, \dots, n-2\}$, which are not appearing in the inversion table of a random permutation of size $n-1$. Furthermore the following distributional equation holds:*

$$G_n \overset{(d)}{=} |\{1, 2, \dots, n-1\} \setminus \{\mathrm{U}_1, \mathrm{U}_2, \dots, \mathrm{U}_{n-1}\}|, \tag{4}$$

where U_k denotes a uniform distribution on $\{1, 2, \dots, k\}$ and all random variables are mutually independent.

We will study now the distribution of G_n, where it will turn out that it is slightly more convenient when considering G_{n+1}. Obviously we have $\mathbb{P}\{G_{n+1} = 0\} = \mathbb{P}\{G_{n+1} = n-1\} = 1/n!$, for $n \geq 1$, which corresponds to a star-like tree: $\{k <_c 1\}$ for $1 \leq k \leq n$, or a chain: $\{k+1 <_c k\}$ for $1 \leq k \leq n$. Furthermore it holds that $\mathbb{P}\{G_{n+1} = n\} = 0$, for $n \geq 1$. When distinguishing whether the weight of the edge adjacent to node $n+1$ is occurring also amongst the edge-weights of the remaining edges or not, we can set up easily the following recurrence for G_{n+1}:

$$\mathbb{P}\{G_{n+1} = m\} = \frac{n-m}{n}\mathbb{P}\{G_n = m-1\} + \frac{m+1}{n}\mathbb{P}\{G_n = m\}, \quad \text{for } 1 \leq m \leq n-1 \quad \text{and} \quad n \geq 2. \tag{5}$$

We introduce now, for $0 \leq m \leq n$ and $n \geq 1$, the numbers $a_{n,m} := \mathbb{P}\{G_{n+1} = m\}T_{n+1} = \mathbb{P}\{G_{n+1} = m\}n!$. We obtain then the margin values $a_{n,0} = 1$ and $a_{n,n} = 0$, for $n \geq 1$, and for $1 \leq m \leq n-1$ and $n \geq 2$ the recurrence

$$a_{n,m} = (n-m)a_{n-1,m-1} + (m+1)a_{n-1,m}. \tag{6}$$

But this is exactly the recurrence for the Eulerian numbers $\left\langle {n \atop m} \right\rangle$ counting the number of permutations of size n with exactly m *ascents*:

$$\left\langle {n \atop m} \right\rangle = (n-m)\left\langle {n-1 \atop m-1} \right\rangle + (m+1)\left\langle {n-1 \atop m} \right\rangle. \tag{7}$$

Note that an ascent in the permutation $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ is defined as a position j, with $1 \le j \le n-1$, such that $\sigma_j < \sigma_{j+1}$. This immediately leads to the following result.

Theorem 4. *The distribution of the random variable G_{n+1} is for $n \ge 1$ given as follows:*

$$\mathbb{P}\{G_{n+1} = m\} = \frac{\left\langle {n \atop m} \right\rangle}{n!}, \quad \text{for } 0 \le m \le n, \tag{8}$$

where $\left\langle {n \atop m} \right\rangle$ denote the Eulerian numbers.

Using Theorem 4 and the central limit theorem for the Eulerian numbers [1] we obtain the limit distribution of G_n.

Corollary 3. *The properly scaled and shifted random variable G_n converges, for $n \to \infty$, in distribution to a standard normal distributed r. v.:*

$$G_n^* := \frac{G_n - \mathbb{E}(G_n)}{\mathbb{V}(G_n)} \xrightarrow{(d)} \mathcal{N}(0,1), \tag{9}$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

Of course the previous considerations lead to the following corollary, which seems to be new best to our knowledge.

Corollary 4. *The number of permutations of $\{1,2,\dots,n\}$ with exactly m ascents and the number of permutations of $\{1,2,\dots,n\}$ with exactly m absent values in its inversion table (i.e., exactly m elements of the set $\{0,1,\dots,n-1\}$, which are not appearing in its inversion table) coincide and are given by the Eulerian numbers $\left\langle {n \atop m} \right\rangle$.*

Remark 1. We want to point out that only the *number* of permutations coincide in Corollary 4. It is not true in general that a permutation with exactly m ascents also has m absent values in its inversion table. Consider, e.g., the permutation $\sigma = (2,4,1,3)$ with 2 ascents, but when considering its inversion table $(0,1,0,2)$ we see that only one value, namely 3, is absent.

3 The distribution of specific edge-weights

3.1 The number of edge-weights of a given size

Let us denote by E_n the set of edges of an edge-weighted random recursive tree of size n. We will study the random variable $S_{n,j} := \sum_{e \in E_n} \mathbb{1}_{\{w_e = j\}}$, which counts the number of edges with weight j in a size-n random recursive tree. This can be done by showing relations to the node degree of specified nodes in random recursive trees.

At first we give the following immediate consequence of Bijection 1.

Proposition 1.

$$S_{n,j} \stackrel{(d)}{=} I_{n-1,j-1}, \quad \text{for } n > j \ge 1, \tag{10}$$

where $I_{n,j}$ denotes the r. v., which counts the number of elements in a random permutation of size n with exactly j inversions.

The next theorem gives a relation between the distribution of edges with a given weight and the out-degree of specified nodes in recursive trees and characterizes the distribution appearing.

Theorem 5. *The random variable $S_{n,j}$ satisfies the following distribution law.*

$$S_{n,j} \stackrel{(d)}{=} Z_{n,j} \stackrel{(d)}{=} B_j \oplus \cdots \oplus B_{n-1}, \tag{11}$$

where the random variable $Z_{n,j}$ gives the out-degree (i.e., the number of children, or the the number of attached nodes) of node j in a random recursive tree of size n, and B_k denotes a Bernoulli-distributed r. v. $\mathrm{Be}(\frac{1}{k})$, *i.e.,* $\mathbb{P}\{B_k = 1\} = 1/k$ *and* $\mathbb{P}\{B_k = 0\} = 1 - 1/k$.

Proof. We obtain the following distributional equation of $S_{n,j}$:

$$S_{n,j} \stackrel{(d)}{=} \bigoplus_{k=j+1}^{n} \mathbb{1}_{\{k <_c k-j\}}, \tag{12}$$

where $\{k <_c i\}$ denotes the event that node k is attached to node i (i.e., k is a child of i). These indicators are mutually independent for recursive trees and by definition given via $\mathbb{P}\{k <_c i\} = 1/(k-1)$, for $1 \le i \le k-1$. Thus $S_{n,j} \stackrel{(d)}{=} \bigoplus_{k=j}^{n-1} \mathrm{Be}(\frac{1}{k})$.

To establish the connection to the out-degree $Z_{n,j}$ of node j in a size-n random recursive tree we use the following distributional equation:

$$Z_{n,j} \stackrel{(d)}{=} \bigoplus_{k=j+1}^{n} \mathbb{1}_{\{k <_c j\}}. \tag{13}$$

Since $\mathbb{P}\{k <_c j\} = 1/(k-1)$, for $1 \le j \le k-1$, we also get $S_{n,j} \stackrel{(d)}{=} Z_{n,j}$. □

Since the distribution of the out-degree of specified nodes in recursive trees $Z_{n,j}$ has been studied in [11] for the whole range $1 \le j < n$ one immediately obtains corresponding results for $S_{n,j}$ and $I_{n,j}$ also, which are given next.

Corollary 5. *The distribution of $Z_{n,j}$, $S_{n,j}$ and $I_{n,j}$ are given as follows:*

$$\mathbb{P}\{S_{n,j} = m\} = \mathbb{P}\{Z_{n,j} = m\} = \mathbb{P}\{I_{n-1,j-1} = m\} = \frac{1}{\binom{n-1}{j-1}} \sum_{k=m}^{n-j} \binom{n-k-2}{j-2} \frac{\left[{k \atop m}\right]}{k!}, \quad \text{for } 1 \le j < n. \tag{14}$$

The limiting distribution behavior of the random variable $X_{n,j}$, which stands for $Z_{n,j}$, $S_{n,j}$ or $I_{n,j}$, is, for $n \to \infty$ and depending on the growth of j, given as follows.

- *The region for j small: $j = o(n)$. The suitably scaled and shifted r. v. $X_{n,j}$ is asymptotically Gaussian distributed,*

$$X_{n,j}^* := \frac{X_{n,j} - (\log n - \log j)}{\sqrt{\log n - \log j}} \xrightarrow{(d)} \mathcal{N}(0,1). \tag{15}$$

- *The central region for j: $j \to \infty$ such that $j = \mu n$, with $0 < \mu < 1$. The random variable $X_{n,j}$ is asymptotically Poisson distributed* $\mathrm{Poisson}(\lambda)$ *with parameter $\lambda = -\log \mu$.*

$$X_{n,j} \xrightarrow{(d)} X_\mu, \quad \text{with} \quad \mathbb{P}\{X_\mu = m\} = \frac{\mu(-\log \mu)^m}{m!}. \tag{16}$$

- *The region for j large: $l := n - j = o(n)$. $\mathbb{P}\{X_{n,j} = 0\} \to 1$.*

We also give the following corollary.

Corollary 6. *The number $I_{n,j,m}$ of permutations of $\{1,2,\dots,n\}$, where exactly m elements have j inversions is given by*

$$I_{n,j,m} = j!(n-j)! \sum_{k=m}^{n-j} \binom{n-k-1}{j-1} \frac{\left[{k \atop m}\right]}{k!}. \tag{17}$$

$I_{n,j,m}$ is for j and m fixed and $n \to \infty$ asymptotically given by

$$I_{n,j,m} \sim \frac{j\sqrt{2\pi}}{m!} \frac{n^{n-\frac{3}{2}} \log^m(n)}{e^n}, \tag{18}$$

whereas for $l := n-j$ and m fixed and $n \to \infty$ by

$$I_{n,j,m} \sim \frac{\sqrt{2\pi}}{m!(l-m)!} \frac{n^{n-m}}{e^n}. \tag{19}$$

Remark 2. Note that, by using the correspondence between recursive trees and permutations as given by Bijection 1 and results stated in [12] about the node degrees in recursive trees, one can get even a refinement of the results presented. It is possible to obtain a closed formula for $I_{n,J,M}$, which gives the number of permutations of $\{1,2,\dots,n\}$, where exactly m_i elements have j_i inversions, with $1 \le i \le r$, for vectors $J = (j_1,\dots,j_r)$ and $M = (m_1,\dots,m_r)$.

3.2 The sum of the labels of nodes attached via edges of a given weight

We study now the random variable $W_{n,j} := \sum_{e=(e_1,e_2)\in E_n} e_2 \mathbb{1}_{\{w_e=j\}}$, which counts the sum of the labels of nodes attached via edges of weight j, where again E_n denotes the set of edges of a random recursive tree of size n. In order to get a direct correspondence with inversions we study a variant, i.e., the r. v. $\mathcal{W}_{n,j} := \sum_{e=(e_1,e_2)\in E_n} (e_2-1)\mathbb{1}_{\{w_e=j\}}$, and obtain as a consequence of Bijection 1 the following proposition.

Proposition 2. *The random variable $\mathcal{W}_{n,j}$ counts the sum of the values of the elements in a random permutation of $\{1,2,\dots,n-1\}$, which have exactly $j-1$ inversions.*

We immediately obtain due to considerations analogous to the proof of Theorem 5 the following distribution law of $\mathcal{W}_{n,j}$.

Theorem 6. *The distribution of the r. v. $\mathcal{W}_{n,j}$ is given as follows.*

$$\mathcal{W}_{n,j} \stackrel{(d)}{=} \mathcal{B}_j \oplus \cdots \oplus \mathcal{B}_{n-1}, \tag{20}$$

where $\mathcal{B}_k$ denotes a scaled Bernoulli distributed r. v. $k\,\mathrm{Be}(\frac{1}{k})$, i.e., $\mathbb{P}\{\mathcal{B}_k = k\} = 1/k$ and $\mathbb{P}\{\mathcal{B}_k = 0\} = 1-1/k$.

The next theorem characterizes the limiting distribution of $\mathcal{W}_{n,j}$ for the three phases appearing depending on the growth of $j = j(n)$.

Theorem 7. *The limiting distribution behaviour of the r. v. $\mathcal{W}_{n,j}$ is, for $n \to \infty$ and depending on the growth of j, given as follows.*

- *The region for j small: $j = o(n)$. The limiting distribution of the suitably scaled r. v. $\mathcal{W}_{n,j}$ is Dickman's infinitely divisible distribution,*

$$\frac{\mathcal{W}_{n,j}}{n} \xrightarrow{(d)} W, \quad \text{with} \quad \mathbb{P}\{W < x\} = e^{-\gamma} \int_0^x \rho(v)\,dv, \quad \text{for } x > 0, \tag{21}$$

where $\rho(v)$ denotes the Dickman function and $\gamma \approx 0.57721$ is the Euler-Mascheroni constant.

- *The central region for j: $j \to \infty$ such that $j = \mu n$, with $0 < \mu < 1$. The suitably scaled r. v. $\mathcal{W}_{n,j}$ converges in distribution to a r. v. W_μ, whose distribution can be characterized via its Laplace transform,*

$$\frac{\mathcal{W}_{n,j}}{n} \xrightarrow{(d)} W_\mu, \quad \text{with} \quad \psi_\mu(t) := \mathbb{E}(e^{-tW_\mu}) = \exp\Big(\int_{\mu t}^{t} \frac{e^{-v}-1}{v} dv\Big). \tag{22}$$

  Furthermore the distribution of W_μ is infinitely divisible.

- *The region for j large: $l := n - j = o(n)$.*

$$\frac{\mathcal{W}_{n,j}}{n} \xrightarrow{(d)} 0. \tag{23}$$

For an overview of the Dickman distribution, the Dickman function and an extensive list of combinatorial objects, which lead to the Dickman distribution, we refer to [7].

Proof. The proof of the case $j = 1$ already appeared in [7] and an extension of the arguments appearing there leads to the theorem presented. As a consequence of Theorem 6 the Laplace transform $\psi_{n,j}(t) := \mathbb{E}(e^{-t\frac{\mathcal{W}_{n,j}}{n}})$ of $\mathcal{W}_{n,j}/n$ is given by

$$\psi_{n,j}(t) = \prod_{k=j}^{n-1} \Big(1 + \frac{e^{\frac{-tk}{n}} - 1}{k}\Big). \tag{24}$$

First we consider the case $j = o(n)$. Using Curtiss' theorem [5] it suffices to show that the Laplace transform $\psi_{n,j}(t)$ converges pointwise in a neighborhood of $t = 0$ to the Laplace transform of the Dickman distribution, i.e.,

$$\lim_{n\to\infty} \psi_{n,j}(t) = \int_0^t \frac{e^{-v}-1}{v} dv. \tag{25}$$

We write equation (24) as

$$\psi_{n,j}(t) = \exp\Big(\sum_{k=j}^{n-1} \log\big(1 + \frac{e^{-\frac{tk}{n}} - 1}{k}\big)\Big). \tag{26}$$

By using the expansion $\log(1+x) = \sum_{k\ge1}(-x)^l/l$ we get further

$$\psi_{n,j}(t) = \exp\Big(\sum_{k=j}^{n-1} \frac{e^{-\frac{tk}{n}} - 1}{k} + R_{n,j}(t)\Big), \tag{27}$$

and by standard estimates we obtain for the remainder term

$$R_{n,j}(t) = \sum_{l\ge2} \frac{(-1)^l}{l} \sum_{k=j}^{n-1} \frac{(e^{-\frac{tk}{n}} - 1)^l}{k^l} = \mathcal{O}(|t^2|/n), \tag{28}$$

uniformly for all j, $1 \le j \le n-1$.

By an application of the Euler-MacLaurin summation formula and the estimate (28) we get the following expansion of the sum appearing in (27):

$$\sum_{k=j}^{n-1} \frac{e^{-\frac{tk}{n}} - 1}{k} = \int_{\frac{jt}{n}}^{\frac{t(n-1)}{n}} \frac{e^{-v}-1}{v} dv + \mathcal{O}(|t^2|/n). \tag{29}$$

This leads to the stated result for $j = o(n)$.

For the case $j = \mu n$, with $0 < \mu < 1$, we proceed as before and get the following expansion of the sum appearing in (27):

$$\sum_{k=j}^{n-1} \frac{e^{-\frac{tk}{n}} - 1}{k} = \int_{\mu t}^{\frac{t(n-1)}{n}} \frac{e^{-v} - 1}{v} dv + \mathcal{O}(|t^2|/n), \qquad (30)$$

which again proves the stated result. The infinitely divisibility follows by using the characterization of Chow and Teicher [4], p. 420.

The remaining case $l := n - j = o(n)$ is proven fully analogous. □

Remark 3. It can be seen easily that the limit law of $W_{n,j}$ is the same as the limit law of $\mathcal{W}_{n,j}$, and thus Theorem 7 also holds for $W_{n,j}$.

4 Weighted depths and distances

4.1 The edge-weighted depth of nodes and distance between nodes

A weighted recursive tree readily provides generalizations to depths and distances by adding the edge-weights on the connecting path between two specified nodes instead of simple counting the number of edges. We start with a simple observation leading to the (degenerate) distribution of the edge-weighted depth $X_{n,j}$, which is given by the sum of the weights of the edges lying on the unique path from the root to node j in a random recursive tree of size n.

Proposition 3. *The r. v. $X_{n,j}$ satisfies the following distribution.*

$$X_{n,j} \stackrel{(d)}{=} X_{j,j} \stackrel{(d)}{=} j - 1. \qquad (31)$$

Proof. Consider a recursive tree of size $n \geq j$ and let us assume that the connecting path from the root 1 to node j is given by the node sequence $1 = v_0, v_1, \dots, v_r = j$. Then the sum of the edge-weights on this path is given by $\sum_{1 \leq i \leq r}(v_i - v_{i-1}) = v_r - v_0 = j - 1$, which proves this proposition. □

Next we are going to study the edge-weighted distance between arbitrary nodes $j_1 \leq j_2$ in random recursive trees of size n, where it is sufficient to consider only the case $j_2 = n$, since nodes larger than j_2 do not have an influence on this parameter (this is a consequence of the description of random recursive trees via a tree evolution process). To do this we introduce the r. v. $\mathfrak{X}_{n,j}$, which gives the sum of the weights of the edges lying on the unique path from node j to node n in a random recursive tree of size n. The next proposition reduces the analysis of $\mathfrak{X}_{n,j}$ to the special instance $n = j + 1$.

Proposition 4.

$$\mathfrak{X}_{n,j} \stackrel{(d)}{=} (n - j - 1) \oplus \mathfrak{X}_{j+1,j}. \qquad (32)$$

Proof. A combinatorial argument is the following. The edge-weighted distance between the nodes n and j in a size-n recursive tree depends only on the label of the root of the spanning tree of n and j. Assume that node k, with $1 \leq k \leq n$, is the root of the spanning tree of the nodes n and j. Then the edge-weighted distance is given by $n + j - 2k$, regardless of the nodes lying on the path from k to n. Now merge all nodes on the path from k to n with labels larger than j into one node and delete all other nodes larger than j except n itself. Think of the new node as node $j + 1$. Thus to obtain the edge-weighted distance between n and j we only have to add $n - j - 1$ to the edge-weighted distance between nodes j and $j + 1$, which proves the stated result.

A more probabilistic argument is also given next.

$$\begin{aligned}\mathbb{P}\{\mathfrak{X}_{n,j} = m\} &= \mathbb{P}\{\mathfrak{X}_{n,j} = m | n <_c n-1\}\mathbb{P}\{n <_c n-1\} + \mathbb{P}\{\mathfrak{X}_{n,j} = m | n \not<_c n-1\}\mathbb{P}\{n \not<_c n-1\} \\ &= \mathbb{P}\{\mathfrak{X}_{n-1,j} = m-1\}\mathbb{P}\{n <_c n-1\} + \mathbb{P}\{\mathfrak{X}_{n-1,j} = m-1\}\mathbb{P}\{n \not<_c n-1\}\end{aligned}$$

$$= \mathbb{P}\{\mathfrak{X}_{n-1,j} = m-1\}. \qquad (33)$$

Iterating this argument leads to the stated result. □

Remark 4. Proposition 3 and Proposition 4 are valid for a larger class of trees, i.e., for the family of so called "grown simple increasing trees", see [18, 11] for a definition of this tree family.

The distribution of $\mathfrak{X}_{j+1,j}$ is characterized next.

Theorem 8. *The r. v. $\mathfrak{X}_{j+1,j}$ satisfies the following distribution law.*

$$\mathbb{P}\{\mathfrak{X}_{j+1,j} = 2m-1\} = \begin{cases} \frac{1}{j}, & m = 1, \\ \frac{1}{(j+2-m)(j+1-m)}, & 1 < m \le j. \end{cases} \qquad (34)$$

The expectation and the variance of $\mathfrak{X}_{j+1,j}$ are given by the following exact formulæ.

$$\mathbb{E}(\mathfrak{X}_{j+1,j}) = 2j - 2H_j + 1, \quad \mathbb{V}(\mathfrak{X}_{j+1,j}) = 8j - 4H_j^2 - 4H_j. \qquad (35)$$

The limiting distribution of the suitably shifted r. v. $\mathfrak{X}_{j+1,j}$ is given as follows.

$$X_{j+1,j} - 2j - 1 \xrightarrow{(d)} -2R, \quad \text{with} \quad \mathbb{P}\{R = m\} = \frac{1}{m(m+1)}, \quad \text{for } m \in \mathbb{N}. \qquad (36)$$

Thus the integer moments of R do not exist.

Proof. We start with the remark that, as a consequence of the proof of Proposition 4, $\mathfrak{X}_{j+1,j}$ can have only values $1, 3, 5, \dots, 2j-3, 2j-1$.

Conditioning on the node ℓ, where $j+1$ is attached, gives then

$$\begin{aligned}
\mathbb{P}\{\mathfrak{X}_{j+1,j} = 2m-1\} &= [\![m=1]\!]\mathbb{P}\{j+1 <_c j\} + \sum_{\ell=1}^{j-1} \mathbb{P}\{\mathfrak{X}_{j+1,j} = 2m-1 | j+1 <_c \ell\}\mathbb{P}\{j+1 <_c \ell\} \\
&= \frac{[\![m=1]\!]}{j} + \sum_{\ell=1}^{j-1} \mathbb{P}\{\mathfrak{X}_{j,\ell} = 2m-1-(j+1-\ell) | j+1 <_c \ell\}\frac{1}{j} \\
&= \frac{[\![m=1]\!]}{j} + \sum_{\ell=1}^{j-1} \mathbb{P}\{\mathfrak{X}_{j,\ell} = 2m-1-(j+1-\ell)\}\frac{1}{j} \\
&= \frac{[\![m=1]\!]}{j} + \sum_{\ell=1}^{j-1} \mathbb{P}\{\mathfrak{X}_{\ell+1,\ell} = 2m-1-(2j-2\ell)\}\frac{1}{j}. \qquad (37)
\end{aligned}$$

Taking differences leads then to

$$j\mathbb{P}\{\mathfrak{X}_{j+1,j} = 2m-1\} - (j-1)\mathbb{P}\{\mathfrak{X}_{j,j-1} = 2m-3\} = [\![m=1]\!] - [\![m=2]\!] + \mathbb{P}\{\mathfrak{X}_{j,j-1} = 2m-3\},$$

and after normalization we obtain the recurrence

$$\mathbb{P}\{\mathfrak{X}_{j+1,j} = 2m-1\} = \mathbb{P}\{\mathfrak{X}_{j,j-1} = 2m-3\} + \frac{[\![m=1]\!] - [\![m=2]\!]}{j}. \qquad (38)$$

Iterating (38) gives

$$\mathbb{P}\{\mathfrak{X}_{j+1,j} = 2m-1\} = \sum_{k=1}^{j-1} \frac{[\![m=k]\!] - [\![m=k+1]\!]}{j+1-k} + \mathbb{P}\{X_{2,1} = 2m-1-2(j-1)\}, \qquad (39)$$

and after simplifying the expression we get the stated distribution law. The formulæ for the expectation and the variance are obtained by easy summation.

To obtain the limiting distribution result we simply observe that

$$\mathbb{P}\{\mathfrak{X}_{j+1,j} - 2j - 1 = -2m\} = \begin{cases} \frac{1}{m(m+1)}, & 1 \le m < j, \\ \frac{1}{j}, & m = j. \end{cases} \qquad (40)$$

□

4.2 The root of the spanning tree of consecutive nodes

Let us denote by $R_{j+1,j}$ the random variable, which counts the label of the root of the spanning tree of the nodes j and $j+1$ in a random recursive tree of size $j+1$. Using our results for $\mathfrak{X}_{j+1,j}$ as obtained in Subsection 4.1 we immediately obtain the following theorem.

Theorem 9. *The r. v. $R_{j+1,j}$ satisfies the following distribution law.*

$$\mathbb{P}\{R_{j+1,j} = m\} = \begin{cases} \frac{1}{m(m+1)}, & 1 \le m < j, \\ \frac{1}{j}, & m = j. \end{cases} \tag{41}$$

The expectation and the variance of $R_{j+1,j}$ are given as follows.

$$\mathbb{E}(R_{j+1,j}) = H_j, \quad \mathbb{V}(\mathfrak{X}_{j+1,j}) = 2j - 2H_j. \tag{42}$$

Furthermore $R_{j+1,j}$ converges in distribution to a r. v. R, without convergence of any integer moment,

$$R_{j+1,j} \xrightarrow{(d)} R, \quad \text{with} \quad \mathbb{P}\{R = m\} = \frac{1}{m(m+1)}, \quad \text{for } m \in \mathbb{N}, \tag{43}$$

Proof. As remarked in the proof of Proposition 4 the weighted distance between nodes n and j is given by the smallest label on the connecting path of n and j and thus by the root of the spanning tree. We obtain that the weighted distance is $n + j - 2m$ if and only if node m is the root of the spanning tree. Thus $\mathbb{P}\{R_{j+1,j} = m\} = \mathbb{P}\{\mathfrak{X}_{j+1,j} = 2(j + 1 - m) - 1\}$, and the exact distribution of $R_{j+1,j}$ is characterized, which immediately also leads to the limiting distribution result. Again the formulæ for the expectation and the variance are obtained easily. □

Remark 5. Note that curiously $R_{j+1,j}$ is distributed as the size of the subtree rooted at a randomly chosen node in a random recursive tree of size j, [10].

4.3 The root of the spanning tree of two randomly chosen nodes

We study also the random variable $Y_j^{[R]}$, which gives the label of the root of the spanning tree of two randomly chosen nodes in a random recursive tree of size j.

Corollary 7.

$$\mathbb{P}\{Y_{j+1}^{[R]} = m\} = \frac{1}{m(m+1)} + \frac{2}{jm(m+1)} - \frac{1}{j(j+1)}, \quad \text{for } 1 \le m \le j. \tag{44}$$

The random variable $Y_j^{[R]}$ converges in distribution to a r. v. Y, without convergence of any integer moment.

$$Y_j^{[R]} \xrightarrow{(d)} Y, \quad \text{with} \quad \mathbb{P}\{Y = m\} = \frac{1}{m(m+1)}, \quad \text{for } m \in \mathbb{N}. \tag{45}$$

Proof. Since two nodes in a recursive tree of size $j+1$ are chosen at random, any pair (i, l), with $1 \le i < l \le j+1$, is selected with probability $\frac{2}{j(j+1)}$. Thus we obtain

$$\mathbb{P}\{Y_{j+1}^{[R]} = m\} = \sum_{m \le i < l \le j+1} \frac{2}{j(j+1)} \mathbb{P}\{\mathfrak{X}_{l,i} = i + l - 2m\}$$

$$= \sum_{m \le i < l \le j+1} \frac{2}{j(j+1)} \mathbb{P}\{\mathfrak{X}_{i+1,i} = 2i - 2m + 1\} = \frac{2}{j(j+1)} \sum_{i=m}^{j} (j + 1 - i) \mathbb{P}\{\mathfrak{X}_{i+1,i} = 2i - 2m + 1\}$$

$$= \frac{2}{j(j+1)} \Big(\frac{j+1-m}{m} + \sum_{i=m+1}^{j} (j + 1 - i) \mathbb{P}\{\mathfrak{X}_{i+1,i} = 2i - 2m + 1\} \Big). \tag{46}$$

Since $\mathbb{P}\{\mathfrak{X}_{i+1,i} = 2i - 2m + 1\} = \frac{1}{m(m+1)}$, for $1 \le m < i$, due to Theorem 8, we obtain the stated exact distribution result by summation and the limiting distribution result easily follows from that. □

5 Conclusion

We have analyzed several parameters in edge-weighted random recursive trees by establishing relations to inversions in random permutations. It would be interesting to study the behavior of the parameters considered for other families of labeled trees like plane-oriented recursive trees or binary increasing trees, but it seems that an analysis becomes more involved than for recursive trees.

References

[1] E. A. Bender, Central and local limit theorems applied to asymptotic enumeration, *J. Combinatorial Theory Ser. A* 15, 91–111, 1973.

[2] F. Bergeron, P. Flajolet and B. Salvy, Varieties of Increasing Trees, *Lecture Notes in Computer Science* 581, 24–48, 1992.

[3] D. Y. C. Chan, B. D. Hughes, A. S. Leong and W. J. Reed. Stochastically evolving networks, *Physical Review E* 68, 066124, 24 pages, 2003.

[4] Y. S. Chow and H. Teicher, *Probability theory*, Springer Verlag, Ney York, 1978.

[5] J. H. Curtiss, A note on the theory of moment generating functions, *Annals of Math. Stat.* 13-4, 430-433, 1942.

[6] J. L. Gastwirth and P. K. Bhattacharya. Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable, *Operations Research* 32, 527-536, 1984.

[7] H. -K. Hwang and T. -H. Tsai, Quickselect and Dickman function, *Combinatorics, Probability and Computing*, 11:4, 353-371, 2002.

[8] S. Janson, Asymptotic degree distribution in random recursive trees, *Random Structures Algorithms* 26, 69–83, 2005.

[9] D. Knuth, *The art of computer programming, Vol. 3: Sorting and Searching*, Addison-Wesley, Reading, 1973.

[10] M. Kuba, *Analysis of node isolation procedures and label-based parameters in tree structures.* PhD thesis, TU Wien, Wien, 2006.

[11] M. Kuba and A. Panholzer, On the degree distribution of the nodes in increasing trees, submitted, 2005.
available at `http://info.tuwien.ac.at/panholzer/nodedegree3.pdf`

[12] M. Kuba and A. Panholzer, On the distribution of the node degree and related parameters in recursive trees, preprint.

[13] M. Loève, *Probability Theory I*, 4th Edition, Springer-Verlag, New York, 1977.

[14] H. Mahmoud, *Sorting. A distribution theory*, Wiley, New York, 2000.

[15] H. Mahmoud and R. Smythe, A Survey of Recursive Trees, *Theoretical Probability and Mathematical Statistics* 51, 1–37, 1995.

[16] A. Meir and J. W. Moon. Cutting down recursive trees, *Mathematical Biosciences* 21, 173-181, 1974.

[17] D. Najock and C. C. Heyde. On the number of terminal vertices in certain random trees with an application to stemma construction in philology, *Journal of Applied Probability* 19, 675-680, 1982.

[18] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, accepted for publication in *Random Structures & Algorithms*.
available at `http://info.tuwien.ac.at/panholzer/levelinc3.pdf`

[19] V. N. Sachkov, *Probabilistic Methods in Combinatorial Analysis*, Cambridge University Press, New York, NY, 1997.

[20] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences.
Published electronically at http://www.research.att.com/ njas/sequences/.