SIMPLY GENERATED TREES, CONDITIONED GALTON–WATSON TREES, RANDOM ALLOCATIONS AND CONDENSATION: EXTENDED ABSTRACT

SVANTE JANSON

1. Simply generated trees and Galton–Watson trees

We suppose that we are given a fixed *weight sequence* $\mathbf{w} = (w_k)_{k\geqslant 0}$ of non-negative real numbers. We then define the *weight* of a finite rooted and ordered (a.k.a. plane) tree T by

$$w(T) := \prod_{v\in T} w_{d^+(v)}, \tag{1.1}$$

taking the product over all nodes v in T, where $d^+(v)$ is the outdegree of v. Trees with such weights are called *simply generated trees* and were introduced by Meir and Moon [24]. We let $\mathcal{T}_n$ be the random simply generated tree obtained by picking a tree with n nodes at random with probability proportional to its weight. (To avoid trivialities, we assume that $w_0 > 0$ and that there exists some $k \geqslant 2$ with $w_k > 0$. We consider only n such that there exists some tree with n vertices and positive weight.)

One particularly important case is when $\sum_{k=0}^{\infty} w_k = 1$, so the weight sequence (w_k) is a probability distribution on $\mathbb{Z}_{\geqslant 0}$. (We then say that (w_k) is a *probability weight sequence.*) In this case we let ξ be a random variable with the corresponding distribution: $\mathbb{P}(\xi = k) = w_k$. It is easily seen that the simply generated random tree $\mathcal{T}_n$ equals the *conditioned Galton–Watson tree* with offspring distribution ξ, i.e., the random Galton–Watson tree defined by ξ conditioned on having exactly n vertices.

One of the reasons for the interest in these trees is that many kinds of random trees occuring in various applications (random ordered trees, unordered trees, binary trees, ...) can be seen as simply generated random trees and conditioned Galton–Watson trees, see e.g. Aldous [3, 4], Devroye [9] and Drmota [10].

It is easily seen that if $a, b > 0$ and we change w_k to

$$\widetilde{w}_k := ab^k w_k, \tag{1.2}$$

Date: 16 March 2012.

2000 *Mathematics Subject Classification.* 60C50; 05C05, 60F05, 60J80.

This is an extended abstract of [16], where further results, examples and proofs are given.

then the distribution of $\mathcal{T}_n$ is not changed. In other words, the new weight sequence $(\widetilde{w}_k)$ defines the same simply generated random trees $\mathcal{T}_n$ as (w_k). (This is essentially due to Kennedy [19], who did not consider trees but showed the corresponding result for Galton–Watson processes. See also Aldous [4].) We say that weight sequence (w_k) and $(\widetilde{w}_k)$ related by (1.2) (for some $a, b > 0$) are *equivalent*.

In many cases it is possible to change the weight (w_k) to an equivalent probability weight sequence; in this case $\mathcal{T}_n$ can thus be seen as a conditioned Galton–Watson tree. Moreover, in many cases this can be done such that the resulting probability distribution has mean 1. In such cases it thus suffices to consider the case of a probability weight sequence with mean $\mathbb{E}\,\xi = 1$; then $\mathcal{T}_n$ is a conditional critical Galton–Watson tree. It turns out that this is a nice and natural setting, with many known results proved by many different authors. We extend here some of these results to the general case, including cases where no equivalent probability weight sequence exists.

2. Notation

We consider a fixed weight sequence $\mathbf{w} = (w_k)_{k\geqslant 0}$; we let

$$\Phi(z) := \sum_{k=0}^{\infty} w_k z^k \tag{2.1}$$

be the generating function of the given weight sequence, and let $\rho \in [0,\infty]$ be its radius of convergence.

We further define, for t such that $\Phi(t) < \infty$,

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} = \frac{\sum_{k=0}^{\infty} k w_k t^k}{\sum_{k=0}^{\infty} w_k t^k}; \tag{2.2}$$

moreover, if $\Phi(\rho) = \infty$, we define $\Psi(\rho) := \lim_{t\nearrow\rho} \Psi(t) \leqslant \infty$.

We define

$$\nu := \Psi(\rho). \tag{2.3}$$

In particular, if $\Phi(\rho) < \infty$, then

$$\nu = \frac{\rho\Phi'(\rho)}{\Phi(\rho)} \leqslant \infty. \tag{2.4}$$

Then $\nu = 0 \iff \rho = 0$, and if $\rho > 0$, then

$$\nu := \Psi(\rho) = \lim_{t\nearrow\rho} \Psi(t) = \sup_{0\leqslant t<\rho} \Psi(t) \in (0,\infty]. \tag{2.5}$$

If $\rho > 0$, then ν is the supremum of the means of all probability weight sequences equivalent to (w_k).

We define $\mathbb{N}_0 = \mathbb{Z}_{\geqslant 0} := \{0,1,2,\dots\}$, $\mathbb{N}_1 = \mathbb{Z}_{>0} := \{1,2,\dots\}$, $\overline{\mathbb{N}}_0 := \mathbb{N}_0 \cup \{\infty\}$ and $\overline{\mathbb{N}}_1 := \mathbb{N}_1 \cup \{\infty\}$.

3. Main result for simply generated random trees

Our main limit theorem for simply generated random trees is the following. The case when $\nu \geqslant 1$ and $\sigma^2 < \infty$ was shown implicitly by Kennedy [19] (who considered Galton–Watson processes and not trees), and explicitly by Aldous and Pitman [5], see also Grimmett [14], Kolchin [21], Kesten [20] and Aldous [4]. Special cases with $0 < \nu < 1$ and $\nu = 0$ are given by Jonsson and Stefánsson [18] and Janson, Jonsson and Stefánsson [17], respectively.

The limit (in distribution) in the theorem is for a topology where convergence means convergence of outdegree for any fixed node; it thus really means local convergence close to the root. (It is for this purpose convenient to regard the trees as subtrees of the infinite Ulam–Harris tree.) See [16] for details.

Theorem 3.1. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be any weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 2$.*

(i) *If $\nu \geqslant 1$, let τ be the unique number in $[0, \rho]$ such that $\Psi(\tau) = 1$.*
(ii) *If $\nu < 1$, let $\tau := \rho$.*

In both cases, $0 \leqslant \tau < \infty$ and $0 < \Phi(\tau) < \infty$. Let

$$\pi_k := \frac{\tau^k w_k}{\Phi(\tau)}, \qquad k \geqslant 0; \tag{3.1}$$

then $(\pi_k)_{k\geqslant 0}$ is a probability distribution, with expectation

$$\mu = \Psi(\tau) = \min(\nu, 1) \leqslant 1 \tag{3.2}$$

and variance $\sigma^2 = \tau\Psi'(\tau) \leqslant \infty$. Let $\widehat{\mathcal{T}}$ be the infinite modified Galton–Watson tree constructed in Section 4 below for the distribution $(\pi_k)_{k\geqslant 0}$. Then $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ as $n \to \infty$.

Remark 3.2. In case (ii), there is no $\tau \geqslant 0$ with $\Psi(\tau) = 1$. Hence the definition of τ can also be expressed as follows, recalling $\Psi(t) := t\Phi'(t)/\Phi(t)$ from (2.2): τ is the unique number in $[0, \rho]$ such that

$$\tau\Phi'(\tau) = \Phi(\tau), \tag{3.3}$$

if there exists any such τ; otherwise $\tau := \rho$. (Equation (3.3) is used in many papers to define τ, in the case $\nu \geqslant 1$.)

Remark 3.3. If $0 < t < \rho$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\Phi(t)}{t}\right) = \frac{t\Phi'(t) - \Phi(t)}{t^2} = \frac{\Phi(t)}{t^2}\bigl(\Psi(t) - 1\bigr).$$

Since $\Psi(t)$ is increasing, it follows that $\Phi(t)/t$ decreases on $[0, \tau]$ and increases on $[\tau, \rho]$, so τ can, alternatively, be characterised as the (unique) minimum point in $[0, \rho]$ of the convex function $\Phi(t)/t$, cf. e.g. Minami [26] and Jonsson and Stefánsson [18]. Consequently,

$$\frac{\Phi(\tau)}{\tau} = \inf_{0\leqslant t\leqslant\rho} \frac{\Phi(t)}{t} = \inf_{0\leqslant t<\infty} \frac{\Phi(t)}{t}. \tag{3.4}$$

(This holds also when $\rho = 0$, trivially, since then $\Phi(t)/t = \infty$ for every $t \geqslant 0$.)

Remark 3.4. If we replace (w_k) by an equivalent weight sequence $(\widetilde{w}_k)$, we obtain the same distribution (π_k).

Remark 3.5. If $\rho > 0$, then $\tau > 0$ and the distribution (π_k) is a probability weight sequence equivalent to (w_k). There are other equivalent probability weight sequences, but Theorem 3.1 shows that (π_k) has a special role and therefore is a canonical choice of a weight sequence in its equivalence class. Furthermore, (π_k) is the unique probability distribution with mean 1 that is equivalent to (w_k), if any such distribution exists. If no such distribution exists but $\rho > 0$, then (π_k) is the probability distribution equivalent to (w_k) that has the maximal mean.

Remark 3.6. When $\nu \geqslant 1$, the quantity σ^2 is a natural parameter of the weight sequence (w_k), which frequently occurs in asymptotic results. In this case a calculation yields also the formula [4]

$$\sigma^2 = \tau\Psi'(\tau) = \frac{\tau^2\Phi''(\tau)}{\Phi(\tau)}. \tag{3.5}$$

4. The infinite limit tree

Let $(\pi_k)_{k\geqslant 0}$ be a probability distribution on $\mathbb{N}_0$ and let ξ be a random variable on $\mathbb{N}_0$ with distribution $(\pi_k)_{k=0}^{\infty}$:

$$\mathbb{P}(\xi = k) = \pi_k, \qquad k = 0, 1, 2, \dots \tag{4.1}$$

We assume that the expectation $\mu := \mathbb{E}\,\xi = \sum_k k\pi_k \leqslant 1$ (the subcritical or critical case). We then define (based on Kesten [20] and Jonsson and Stefánsson [18]) a modified Galton–Watson tree $\widehat{\mathcal{T}}$ as follows: There are two types of nodes: *normal* and *special*, with the root being special. Normal nodes have offspring (outdegree) according to independent copies of ξ, while special nodes have offspring according to independent copies of $\widehat{\xi}$, where

$$\mathbb{P}(\widehat{\xi} = k) := \begin{cases} k\pi_k, & k = 0, 1, 2, \dots, \\ 1 - \mu, & k = \infty. \end{cases} \tag{4.2}$$

(Note that this is a probability distribution on $\overline{\mathbb{N}}_1$.) Moreover, all children of a normal node are normal; when a special node gets an infinite number of children, all are normal; when a special node gets a finite number of children, one of its children is selected uniformly at random and is special, while all other children are normal.

Since each special node has at most one special child, the special nodes form a path from the root; we call this path the *spine* of $\widehat{\mathcal{T}}$. We distinguish two different cases:

(T1) If $\mu = 1$ (the critical case), then $\widehat{\xi} < \infty$ a.s. so each special node has a special child and the spine is an infinite path. Each outdegree $d^+(v)$ in $\widehat{\mathcal{T}}$ is finite, so the tree is infinite but locally finite.

In this case, the distribution of $\widehat{\xi}$ in (4.2) is the *size-biased* distribution of ξ, and $\widehat{\mathcal{T}}$ is the size-biased Galton–Watson tree defined by Kesten [20], see also Aldous [4], Aldous and Pitman [5] and Lyons, Pemantle and Peres [23]. The underlying size-biased Galton–Watson process is the same as the *Q-process* studied in Athreya and Ney [6, Section I.14], which is an instance of Doob's h-transform. (See Lyons, Pemantle and Peres [23] for further related constructions in other contexts and Geiger and Kauffmann [13] for a generalization.)

An alternative construction of the random tree $\widehat{\mathcal{T}}$ is to start with the spine (an infinite path from the root) and then at each node in the spine attach further branches; the number of branches at each node in the spine is a copy of $\widehat{\xi} - 1$ and each branch is a copy of the Galton–Watson tree $\mathcal{T}$ with offspring distributed as ξ; furthermore, at a node where k new branches are attached, the number of them attached to the left of the spine is uniformly distributed on $\{0, \dots, k\}$. (All random choices are independent.) Since the critical Galton–Watson tree $\mathcal{T}$ is a.s. finite, it follows that $\widehat{\mathcal{T}}$ a.s. has exactly one infinite path from the root, viz. the spine.

(T2) If $\mu < 1$ (the subcritical case), then a special node has with probability $1 - \mu$ no special child. Hence, the spine is a.s. finite and the number L of nodes in the spine has a (shifted) geometric distribution $\mathrm{Ge}(1 - \mu)$,

$$\mathbb{P}(L = \ell) = (1 - \mu)\mu^{\ell - 1}, \qquad \ell = 1, 2, \dots. \tag{4.3}$$

The tree $\widehat{\mathcal{T}}$ has a.s. exactly one node with infinite outdegree, viz. the top of the spine. $\widehat{\mathcal{T}}$ has a.s. no infinite path.

In this case, an alternative construction of $\widehat{\mathcal{T}}$ is to start with a spine of random length L, where L has the geometric distribution (4.3). We attach as in (T1) further branches that are independent copies of the Galton–Watson tree $\mathcal{T}$; at the top of the spine we attach an infinite number of branches and at all other nodes in the spine the number we attach is a copy of $\xi^* - 1$ where $\xi^* \overset{\mathrm{d}}{=} (\widehat{\xi} \mid \widehat{\xi} < \infty)$ has the size-biased distribution $\mathbb{P}(\xi^* = k) = k\pi_k/\mu$. The spine thus ends with an explosion producing an infinite number of branches, and this is the only node with an infinite degree. This is the construction by Jonsson and Stefánsson [18].

Example 4.1. In the extreme case $\mu = 0$, or equivalently $\xi = 0$ a.s., i.e., $\pi_0 = 1$ and $\pi_k = 0$ for $k \geqslant 1$, (4.2) shows that $\widehat{\xi} = \infty$ a.s. Hence, every normal node has no child and is thus a leaf, while every special node has an infinite number of children, all normal. Consequently, the root is the only special node, the spine consists of the root only (i.e., its length $L = 1$), and the tree $\widehat{\mathcal{T}}$ consists of the root with an infinite number of leaves attached to it, i.e., $\widehat{\mathcal{T}}$ is an infinite star. (This is also given directly by the alternative construction in (T2) above.) In contrast, $\mathcal{T}$ consists of the root only, so $|\mathcal{T}| = 1$. In this case there is no randomness in $\mathcal{T}$ or $\widehat{\mathcal{T}}$.

Remark 4.2. In case (T1), if we remove the spine, we obtain a random forest that can be regarded as coming from a Galton–Watson process with immigration, where the immigration is described by an i.i.d. sequence of random variables with the distribution of $\widehat{\xi} - 1$, see Lyons, Pemantle and Peres [23]. (In the Poisson case, Grimmett [14] gave a slightly different description of $\widehat{\mathcal{T}}$ using a Galton–Watson process with immigration.)

In case (T2), we can do the same, but now the immigration is different: at a random (geometric) time, there is an infinite immigration, and after that there is no more immigration at all.

Remark 4.3. Some related modifications of Galton–Watson trees having a finite spine have been considered previously, see Sagitov and Serra [29], Addario-Berry, Devroye and Janson [1], Geiger [12] and Aldous [2]. Kurtz, Lyons, Pemantle and Peres [22] and Chassaing and Durhuus [8] have constructed related trees with infinite spines using multi-type Galton–Watson processss.

Remark 4.4. In case (T1), the random variable $\widehat{\xi}$ is a.s. finite and has mean

$$\mathbb{E}\,\widehat{\xi} = \sum_{k=0}^{\infty} k\,\mathbb{P}(\widehat{\xi} = k) = \sum_{k=0}^{\infty} k^2 \pi_k = \mathbb{E}\,\xi^2 = \sigma^2 + 1, \tag{4.4}$$

where $\sigma^2 := \operatorname{Var}\xi \leqslant \infty$. In case (T2), we have $\mathbb{P}(\widehat{\xi} = \infty) > 0$ and thus $\mathbb{E}\,\widehat{\xi} = \infty$. This suggests that in results that are known in the critical case (T1), and where σ^2 appears as a parameter, the correct generalization of σ^2 to the subcritical case (T2) is not $\operatorname{Var}\xi$ but $\mathbb{E}\,\widehat{\xi} - 1 = \infty$.

5. Three different types of weights

Although Theorem 3.1 has only two cases, it makes sense to treat the case $\rho = 0$ separately. We thus have the following three (mutually exclusive) cases for the weight sequence (w_k):

I. $\nu \geqslant 1$. Then $0 < \tau < \infty$ and $\tau \leqslant \rho \leqslant \infty$. The weight sequence (w_k) is equivalent to (π_k), which is a probability distribution with mean $\mu = \Psi(\tau) = 1$ and probability generating function $\sum_{k=0}^{\infty} \pi_k z^k$ with radius of convergence $\rho/\tau \geqslant 1$.

II. $0 < \nu < 1$. Then $0 < \tau = \rho < \infty$. The weight sequence (w_k) is equivalent to (π_k), which is a probability distribution with mean $\mu = \Psi(\tau) = \nu < 1$ and probability generating function $\sum_{k=0}^{\infty} \pi_k z^k$ with radius of convergence $\rho/\tau = 1$.

III. $\nu = 0$. Then $\tau = \rho = 0$, and (w_k) is not equivalent to any probability distribution.

If we consider the modified Galton–Watson tree in Theorem 3.1, then III is the case discussed in Example 4.1; excluding this case, I and II are the same as (T1) and (T2) in Section 4.

We can reformulate the partition into three cases in more probabilistic terms. We say that ξ, or the distribution (p_k), has *some finite exponential moment* if $\mathbb{E}\, R^{\xi} < \infty$ for some $R > 1$, or equivalently, $\mathbb{E}\, e^{r\xi} < \infty$ for some $r > 0$; this is equivalent to the probability generating function $\sum_{k=0}^{\infty} p_k z^k$ having radius of convergence strictly larger than 1. The cases I–III can then be described as follows:

I. $\nu \geqslant 1$. Then (w_k) is equivalent to a probability distribution with mean $\mu = 1$ (with or without some exponential moment). Moreover, (π_k) in (3.1) is the unique such distribution.
II. $0 < \nu < 1$. Then (w_k) is equivalent to a probability distribution with mean $\mu < 1$ and no finite exponential moment. Moreover, (π_k) in (3.1) is the unique such distribution.
III. $\nu = 0$. Then (w_k) is not equivalent to any probability distribution.

Case I may be further subdivided. From an analytic point of view, it is natural to split I into two subcases:

Ia. $\nu > 1$; equivalently, $0 < \tau < \rho \leqslant \infty$. The weight sequence (w_k) is equivalent to a probability distribution with mean $\mu = 1$ and some finite exponential moment. (Then (π_k) is the unique such distribution.) (This case is called *generic* in [11] and [18].)
Ib. $\nu = 1$; then $0 < \tau = \rho < \infty$. The weight sequence (w_k) is equivalent to a probability distribution with mean $\mu = 1$ and no finite exponential moment. (Then (π_k) is the unique such distribution.)

Case Ia is convenient when using analytic methods, since it says that the point τ is strictly inside the domain of convergence of Φ, which is convenient for methods involving contour integrations in the complex plane. (See e.g. Drmota [10] for several such results of different types.) For that reason, many papers using such methods consider only case Ia. However, it has repeatedly turned out, for many different problems, that results proved by such methods often hold assuming only that we are in case I with finite variance of (π_k). (In fact, as shown in [15], it is at least sometimes possible to use complex analytic methods also in this case.) Consequently, it is often more important to partition case I into the following two cases:

Iα. $\nu \geqslant 1$ and (π_k) has variance $\sigma^2 < \infty$. In other words, (w_k) is equivalent to a probability distribution (π_k) with mean $\mu = 1$ and finite variance σ^2.
Iβ. $\nu = 1$ and (π_k) has variance $\sigma^2 = \infty$. In other words, (w_k) is equivalent to a probability distribution with mean $\mu = 1$ and infinite variance.

Note that Ia is a subcase of Iα, since a finite exponential moment implies that the second moment is finite.

By Theorem 3.1 we have, if $\nu \geqslant 1$, case Iα when $\Psi'(\tau) < \infty$ and Iβ when $\Psi'(\tau) = \infty$; equivalently, by (3.5), Iα is the case $\nu \geqslant 1$ and $\Phi''(\tau) < \infty$.

Remark 5.1. We have seen that except in case III, we may without loss of generality assume that the weight (w_k) is a probability weight sequence. If this distribution is critical, i.e. has mean 1, we are in case I with $\pi_k = w_k$, so we do not have to change the weights.

If the distribution (w_k) is supercritical, then $\nu > 1$ and we are in case Ia; we can change to an equivalent critical probability weight. Hence we never have to consider supercritical weights.

If the distribution (w_k) is subcritical, we can only say that we are in case I or II. We can often change to an equivalent critical probability weight, but not always.

6. Node degrees

Theorem 3.1 implies the following result for the degree $d^+_{\mathcal{T}_n}(o)$ of the root o.

Theorem 6.1. *Let $(w_k)_{k\geqslant 0}$ and $(\pi_k)_{k\geqslant 0}$ be as in Theorem 3.1. Then, as $n \to \infty$,*

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = d) \to d\pi_d, \quad d \geqslant 0. \tag{6.1}$$

Consequently, regarding $d^+_{\mathcal{T}_n}(o)$ as a random number in $\overline{\mathbb{N}}_0$,

$$d^+_{\mathcal{T}_n}(o) \xrightarrow{\mathrm{d}} \widehat{\xi}, \tag{6.2}$$

where $\widehat{\xi}$ is a random variable in $\overline{\mathbb{N}}_0$ with the distribution given in (4.2).

Note that the sum $\sum_0^\infty d\pi_d = \mu$ of the limiting probabilities in (6.1) may be less than 1; in that case we do not have convergence to a proper finite random variable, which is why we regard $d^+_{\mathcal{T}_n}(o)$ as a random number in $\overline{\mathbb{N}}_0$.

If we instead take a random node, we obtain a different limit distribution, viz. (π_k). (When $\nu > 1$, this was proved by Otter [27], see also Minami [26].)

Theorem 6.2. *Let $(w_k)_{k\geqslant 0}$ and $(\pi_k)_{k\geqslant 0}$ be as in Theorem 3.1, and let v be a uniformly random node in $\mathcal{T}_n$. Then, as $n \to \infty$,*

$$\mathbb{P}(d^+_{\mathcal{T}_n}(v) = d) \to \pi_d, \quad d \geqslant 0. \tag{6.3}$$

7. The maximum degree

Results for the maximum degree are more complicated, and we discuss the different cases separately. We denote the maximum outdegree in the tree $\mathcal{T}_n$ by $Y_{(1)}$.

Case Ia: $\nu > 1$. In this case $0 < \tau < \rho \leqslant \infty$, and we have a logarithmic bound due to Meir and Moon [25]:

$$Y_{(1)} \leqslant \frac{1}{\log(\rho/\tau)} \log n + o_{\mathrm{p}}(\log n); \tag{7.1}$$

if further $w_k^{1/k} \to 1/\rho$ as $k \to \infty$, then

$$\frac{Y_{(1)}}{\log n} \xrightarrow{\mathrm{p}} \frac{1}{\log(\rho/\tau)}. \tag{7.2}$$

In particular, if $\rho = \infty$, then $Y_{(1)} = o_{\mathrm{p}}(\log n)$.

Moreover, if $w_{k+1}/w_k \to a > 0$ as $k \to \infty$, then $Y_{(1)} = k(n) + O_{\mathrm{p}}(1)$ for some deterministic sequence $k(n)$, so $Y_{(1)}$ is essentially concentrated in an interval of length $O(1)$. The distribution of $Y_{(1)}$ is asymptotically given by a discretised Gumbel distribution, but different subsequences may have different limits and no limit distribution exists.

Similarly, if $w_{k+1}/w_k \to 0$, then $Y_{(1)} \in \{k(n), k(n)+1\}$ so $Y_{(1)}$ is concentrated on at most two values, and often (but not always) on a single value.

Case Iα: $\nu \geqslant 1$ and $\sigma^2 < \infty$. The maximum outdegree $Y_{(1)}$ is asymptotically distributed as the maximum $\xi_{(1)}$ of n i.i.d. copies of ξ; this holds in the strong sense that the total variation distance tends to 0. Since $\mathbb{E}\,\xi^2 < \infty$, this implies in particular

$$Y_{(1)} = o_{\mathrm{p}}(n^{1/2}). \tag{7.3}$$

Case Iβ: $\nu \geqslant 1$ and $\sigma^2 = \infty$. We have

$$Y_{(1)} = o_{\mathrm{p}}(n), \tag{7.4}$$

and this is (more or less) best possible.

Case II: $0 < \nu < 1$. In this case, if further (w_k) satisfies an asymptotic power-law $w_k \sim ck^{-\beta}$ as $k \to \infty$, then Jonsson and Stefánsson [18] showed that

$$Y_{(1)} = (1-\nu)n + o_{\mathrm{p}}(n), \tag{7.5}$$

while the second largest node degree $Y_{(2)} = o_{\mathrm{p}}(n)$. However, if the weight sequence is more irregular, this is no longer always true; it is possible (at least along a subsequence) that $Y_{(1)} = o_{\mathrm{p}}(n)$, which can be seen as incomplete condensation; it is also possible (at least along a subsequence) that $Y_{(2)}$ too is of order n, meaning condensation to two or more giant nodes.

Case III: $\nu = \rho = 0$. This is similar to case II. In some regular cases we have (7.5), which now says $Y_{(1)} = n + o_{\mathrm{p}}(n)$, and then necessarily $Y_{(2)} = o_{\mathrm{p}}(n)$, but there are exceptions in other cases with an irregular weight sequence.

8. Balls-in-boxes

The *balls-in-boxes* model is a model for random allocation of m (unlabelled) balls in n (labelled) boxes. The set of possible allocations is thus

$$\mathcal{B}_{m,n} := \Bigl\{(y_1,\dots,y_n)\in\mathbb{N}_0^n : \sum_{i=1}^n y_i = m\Bigr\}, \tag{8.1}$$

where y_i counts the number of balls in box i. We suppose again that $\mathbf{w} = (w_k)_{k=0}^\infty$ is a fixed weight sequence, and we define the weight of an allocation $\mathbf{y} = (y_1,\dots,y_n)$ as

$$w(\mathbf{y}) := \prod_{i=1}^n w_{y_i}. \tag{8.2}$$

Given m and n, we choose a random allocation $B_{m,n}$ with probability proportional to its weight, see e.g. Bialas, Burda and Johnston [7]. We can replace the weight sequence by an equivalent weight sequence for the balls-in-boxes model just as we did for the random trees above.

Example 8.1 (probability weights)**.** In the special case when (w_k) is a probability weight sequence, let $\xi_1,\xi_2,\dots$ be i.i.d. random variables with the distribution (w_k). Then, $B_{m,n}$ has the same distribution as $(\xi_1,\dots,\xi_n)$ conditioned on $\sum_{i=1}^n \xi_i = m$. (This construction of a random allocation $B_{m,n}$ is used by Kolchin [21] and there called the *general scheme of allocation.*)

The connection to random trees is that if T is a tree with $|T| = n$, then its degree sequence (in depth-first order, say) is an allocation in $\mathcal{B}_{n-1,n}$, with the same weight as the tree. Moreover, a converse holds by the following lemma, see e.g. Takács [30], Wendel [31], Pitman [28].

Lemma 8.2. *If $(d_1,\dots,d_n)\in\mathcal{B}_{n-1,n}$, then exactly one of the n cyclic shifts of $(d_1,\dots,d_n)$ is the degree sequence $\Lambda(T)$ of a tree $T\in\mathfrak{T}_n$.*

Other examples of random allocations are different types of random forests with a given number of components, with each component regarded as a box, and each vertex as a ball, see [16].

Our main result for trees is a consequence of the following result for balls-in-boxes, where $\omega := \sup\{i : w_i > 0\} \leqslant \infty$ and $N_k(B_{m,n})$ is the number of boxes with exactly k balls in the allocation $B_{m,n}$.

Theorem 8.3. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be any weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 1$. Suppose that $n\to\infty$ and $m = m(n)$ with $m/n \to \lambda$ with $0 \leqslant \lambda < \omega$.*

(i) *If $\lambda \leqslant \nu$, let τ be the unique number in $[0,\rho]$ such that $\Psi(\tau) = \lambda$.*
(ii) *If $\lambda > \nu$, let $\tau := \rho$.*

In both cases, $0 \leqslant \tau < \infty$ and $0 < \Phi(\tau) < \infty$. Let

$$\pi_k := \frac{w_k\tau^k}{\Phi(\tau)}, \qquad k \geqslant 0. \tag{8.3}$$

Then $(\pi_k)_{k\geqslant 0}$ is a probability distribution, with expectation

$$\mu = \Psi(\tau) = \min(\lambda, \nu) \tag{8.4}$$

and variance $\sigma^2 = \tau\Psi'(\tau) \leqslant \infty$. Moreover, for every $k \geqslant 0$,

$$N_k(B_{m,n})/n \xrightarrow{\mathrm{p}} \pi_k. \tag{8.5}$$

If we regard the weight sequence $\mathbf{w}$ as fixed and vary λ (i.e., vary $m(n)$), we see that if $0 < \nu < \infty$, there is a phase transition at $\lambda = \nu$.

Note that τ and π_k in Theorem 3.1 are the same as in Theorem 8.3 with $\lambda = 1$. (By Lemma 8.2, the random trees correspond to $m = n - 1$ and thus $\lambda = 1$.)

9. Condensation

By (8.5), there are roughly $n\pi_k$ boxes with k balls in a random allocation $B_{m,n}$. Summing this approximation over all k we would get n boxes (as we should) with a total of $n\sum_{k=0}^{\infty} k\pi_k = n\mu$ balls. However, the total number of balls is $m \approx n\lambda$, so in the case $\lambda > \nu$, (8.4) shows that about $n(\lambda - \mu) = n(\lambda - \nu)$ balls are missing. Where are they?

The explanation is that the sums $\sum_{k=0}^{\infty} kN_k(B_{m,n})/n = m$ are not uniformly summable, and we cannot take the limit inside the summation sign. The "missing balls" appear in one or several boxes with very many balls, but these "giant" boxes are not seen in the limit (8.5) for fixed k. In physical terminology, this can be regarded as condensation of part of the mass (= balls).

The simplest case is that there is a single giant box with $\approx (\lambda - \nu)n$ balls. This happens in the important case of a power-law weight sequence: $w_k \sim ck^{-\beta}$ as $k \to \infty$ for some $c > 0$, as shown by Jonsson and Stefánsson [18]; however, there are also other possibilities when the weight sequence is less regular.

Recall that for simply generated random trees, which as said above correspond to balls-in-boxes with $\lambda = 1$, Theorem 3.1 too shows a related form of condensation when $\nu < \lambda = 1$ (since then $\mu < 1$ by (3.2)); in this case the condensation appears as a node of infinite degree in the random limit tree $\widehat{\mathcal{T}}$ of type ($\mathsf{T2}$).

Acknowledgement. This research was started during a visit to NORDITA, Stockholm, during the program *Random Geometry and Applications*, 2010. I thank the participants, in particular Zdzisław Burda, Bergfinnur Durhuus, Thordur Jonsson and Sigurður Stefánsson, for stimulating discussions.

References

[1] L. Addario-Berry, L. Devroye & S. Janson, Sub-Gaussian tail bounds for the width and height of conditioned Galton–Watson trees. *Ann. Probab.*, to appear. `arXiv:1011.4121`.

[2] D. Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** (1991), no. 2, 228–266.

[3] D. Aldous, The continuum random tree I. *Ann. Probab.* **19** (1991), no. 1, 1–28.

[4] D. Aldous, The continuum random tree II: an overview. *Stochastic Analysis (Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser. 167, Cambridge Univ. Press, Cambridge, 1991.

[5] D. Aldous & J. Pitman, Tree-valued Markov chains derived from Galton–Watson processes. *Ann. Inst. H. Poincaré Probab. Statist.* **34** (1998), no. 5, 637–686.

[6] K. B. Athreya & P. E. Ney, *Branching Processes.* Springer-Verlag, Berlin, 1972.

[7] P. Bialas, Z. Burda & D. Johnston, Condensation in the backgammon model. *Nuclear Physics* **493** (1997), 505–516.

[8] P. Chassaing & B. Durhuus, Local limit of labeled trees and expected volume growth in a random quadrangulation. *Ann. Probab.* **34** (2006), no. 3, 879–917.

[9] L. Devroye, Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer, Berlin, 1998, pp. 249–314.

[10] M. Drmota, *Random Trees*, Springer, Vienna, 2009.

[11] B. Durhuus, T. Jonsson & J. F. Wheater, The spectral dimension of generic trees. *J. Stat. Phys.* **128** (2007), 1237–1260.

[12] J. Geiger, Elementary new proofs of classical limit theorems for Galton–Watson processes. *J. Appl. Probab.* **36** (1999), no. 2, 301–309.

[13] J. Geiger & L. Kauffmann, The shape of large Galton–Watson trees with possibly infinite variance. *Random Struct. Alg.* **25** (2004), no. 3, 311–335.

[14] G. R. Grimmett, Random labelled trees and their branching networks. *J. Austral. Math. Soc. Ser. A* **30** (1980/81), no. 2, 229–237.

[15] S. Janson, Random cutting and records in deterministic and random trees. *Random Struct. Alg.* **29** (2006), no. 2, 139–179.

[16] S. Janson, Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys* **9** (2012), 103–252.

[17] S. Janson, T. Jonsson & S. Ö. Stefánsson, Random trees with super-exponential branching weights. *J. Phys. A: Math. Theor.* **44** (2011), 485002.

[18] T. Jonsson & S. Ö. Stefánsson, Condensation in nongeneric trees. *J. Stat. Phys.* **142** (2011), no. 2, 277–313.

[19] D. P. Kennedy, The Galton–Watson process conditioned on the total progeny. *J. Appl. Probab.* **12** (1975), 800–806.

[20] H. Kesten, Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.* **22** (1986), no. 4, 425–487.

[21] V. F. Kolchin, *Random Mappings*. Nauka, Moscow, 1984 (Russian). English transl.: Optimization Software, New York, 1986.

[22] T. Kurtz, R. Lyons, R. Pemantle & Y. Peres, A conceptual proof of the Kesten–Stigum Theorem for multi-type branching processes. *Classical and Modern Branching Processes (Minneapolis, MN, 1994)*, IMA Vol. Math. Appl., 84, Springer, New York, 1997, pp. 181–185.

[23] R. Lyons, R. Pemantle & Y. Peres, Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.* **23** (1995), no. 3, 1125–1138.

[24] A. Meir & J. W. Moon, On the altitude of nodes in random trees. *Canad. J. Math.*, **30** (1978), 997–1015.

[25] A. Meir & J. W. Moon, On the maximum out-degree in random trees. *Australas. J. Combin.* **2** (1990), 147–156.

[26] N. Minami, On the number of vertices with a given degree in a Galton–Watson tree. *Adv. Appl. Probab.* **37** (2005), no. 1, 229–264.

[27] R. Otter, The multiplicative process. *Ann. Math. Statistics* **20** (1949), 206–224.

[28] J. Pitman, Enumerations of trees and forests related to branching processes and random walks. *Microsurveys in Discrete Probability (Princeton, NJ, 1997)*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 41, Amer. Math. Soc., Providence, RI, 1998, pp. 163–180.

[29] S. Sagitov & M. C. Serra, Multitype Bienaymé–Galton–Watson processes escaping extinction. *Adv. Appl. Probab.* **41** (2009), no. 1, 225–246.

[30] L. Takács, A generalization of the ballot problem and its application in the theory of queues. *J. Amer. Statist. Assoc.* **57** (1962), 327–337.

[31] J. G. Wendel, Left-continuous random walk and the Lagrange expansion. *Amer. Math. Monthly* **82** (1975), 494–499.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www2.math.uu.se/~svante/`