SIMPLY GENERATED TREES, CONDITIONED GALTON–WATSON TREES, RANDOM ALLOCATIONS AND CONDENSATION.

SVANTE JANSON

Abstract. We give a unified treatment of the limit, as the size tends to infinity, of simply generated random trees, including both the well-known result in the standard case of critical Galton–Watson trees and similar but less well-known results in the other cases (i.e., when no equivalent critical Galton–Watson tree exists). There is a well-defined limit in the form of an infinite random tree in all cases; for critical Galton–Watson trees this tree is locally finite but for the other cases the random limit has exactly one node of infinite degree.

The proofs use a well-known connection to a random allocation model that we call balls-in-boxes, and we prove corresponding theorems for this model.

This survey paper contains many known results from many different sources, together with some new results.

Contents

Date: 2 December 2011; revised 14 February 2012.
2000 *Mathematics Subject Classification.* 60C50; 05C05, 60F05, 60J80.

1. Introduction

The main purpose of this survey paper is to study the asymptotic shape of simply generated random trees in complete generality; this includes conditioned Galton–Watson trees as a special case, but we will also go beyond that case. Definitions are given in Section 2; here we only recall that simply generated trees are defined by a weight sequence (w_k), and that the case when the weight sequence is a probability distribution yields conditioned Galton–Watson trees.

It is well-known that in the case of a critical conditioned Galton–Watson tree, i.e., when the defining offspring distribution has expectation 1, the random tree has a limit (as the size tends to infinity); this limit is an infinite random tree, the size-biased Galton–Watson tree defined by Kesten [74], see also Aldous [4], Aldous and Pitman [6] and Lyons, Pemantle and Peres [84]. It is also well-known that this case is less special than it might seem; there is a notion of equivalent weight sequences defining the same simply generated random tree, see Section 4, and a large class of weight sequences have an equivalent probability weight sequence defining a critical conditioned Galton–Watson tree. Many probabilists, including myself, have often concentrated on this "standard" case of critical conditioned Galton–Watson trees and dismissed the remaining cases as uninteresting exceptional cases. However, some researchers, in particular mathematical physicists, have studied such cases too. Bialas and Burda [13] studied one case (Example 10.7 below) and found a phase transition as we leave the standard case; this can be interpreted as a condensation making the tree bushy with one or a few nodes of very high degree. This interesting condensation was studied further by Jonsson and Stefánsson [67], who showed that (in the power-law case), there is a limit tree of a different type, having one node of infinite degree.

We give in the present paper a unified treatment of the limit as the size tends to infinity for all simply generated trees, including both the well-known result in the standard case of critical Galton–Watson trees and the "exceptional" cases (i.e., when no equivalent probability weight sequence exists, or when such a sequence exists but not with mean 1). We will see that there is a well-defined limit in the form of an infinite random tree for any weight sequence. In the non-standard cases, this infinite random limit has exactly one node of infinite degree, so its form differs from the standard case of a critical Galton–Watson tree where all nodes in the limit tree have finite degrees, but nevertheless the trees are similar; see Sections 5 and 7 for details.

Some important notation, used throughout the paper, is introduced in Section 3, while Sections 4 and 6 contain further preliminaries. The main limit theorem for simply generated random trees is stated in Section 7, together with some other, related, limit theorems concerning node degrees and fringe subtrees. The differences between different types of weight sequences are discussed further in Section 8, and this is continued in Section 9 with a summary of the main results from Section 19 on the maximum outdegree in the random tree.

The proofs of the limit theorems for random trees use a well-known connection to a random allocation model that we call balls-in-boxes; this model exhibits a similar behaviour, with condensation in the non-classical cases, see e.g. Bialas, Burda and Johnston [14]. The model is defined in Section 11, and the relation between the models is described in Section 15. The balls-in-boxes model is interesting in its own right, and it has been used for several other applications; we give some examples from probability theory, combinatorics and statistical physics in Section 12. We therefore also develop the general theory for balls-in-boxes with arbitrary weight sequences (in the range where the mean occupancy is bounded). In particular, we give in Section 11 theorems corresponding to (and in some ways extending) our main theorems for random trees.

The limit theorems for balls-in-boxes are proved in Sections 13–14, and then these results are used to prove the limit theorems for random trees in Sections 15–17.

The remaining sections contain additional results. Section 18 gives asymptotic results for the partition functions of the models. The very long

Section 19 gives results on the largest degrees in random trees, and the largest numbers of balls in a box in the balls-in-boxes model; the section is long because there are several different cases with different types of behaviour. (See also the summary in Section 9.) In particular, we study in Section 19.6 the case when there is condensation, and investigate whether this appears as condensation to a single box (or node), or whether the condensation is distributed over several boxes (nodes); it turns out that both cases can occur. We give also, in Section 19.7, applications to the size of the largest tree in random forests. In Section 20, the condensation in random trees is discussed in further detail. Finally, some additional comments, results and open problems are given in Sections 21 and 22; Section 21 mentions briefly various other types of asymptotic results for simply generated random trees, and Section 22 discusses alternative ways to condition Galton–Watson trees.

This paper contains many known results from many different sources, together with some new results. (We believe, for example, that the theorems in Section 7 are new in the present generality.) We have tried to give relevant references, but the absence of references does not necessarily imply that a result is new.

2. Simply generated trees

2.1. **Ordered rooted trees.**

The trees that we consider are (with a few explicit exceptions) *rooted* and *ordered* (such trees are also called *plane* trees). Recall that a tree is *rooted* if one node is distinguished as the *root* o; this implies that we can arrange the nodes in a sequence of generations (or levels), where generation x consists of all nodes of distance x to the root. (Thus generation 0 is the root; generation 1 is the set of neighbours of the root, and so on.) If v is a node with $v \neq o$, then the *parent* of v is the neighbour of v on the path from v to o; thus, every node except the root has a unique parent, while the root has no parent. Conversely, for any node v, the neighbours of v that are further away from the root than v are the *children* of v. The number of children of v is the *outdegree* $d^+(v) \geqslant 0$ of v. Note that if v is in generation x, then its parent is in generation $x-1$ and its children are in generation $x+1$.

Recall further that a rooted tree is *ordered* if the children of each node are ordered in a sequence $v_1, \dots, v_d$, where $d = d^+(v) \geqslant 0$ is the outdegree of v. See e.g. Drmota [33] for more information on these and other types of trees. (The trees we consider are called *planted plane trees* in [33].) We identify trees that are isomorphic in the obvious (order preserving) way. (Formally, we can define our trees as equivalence classes. Alternatively, we may select a specific representative in each equivalence class as in Section 6.)

Remark 2.1. Some authors prefer to add an extra (phantom) node as a parent of the root; such trees are called *planted.* (An alternative version is to add only a pendant edge at the root, with no second endpoint.) There is an obvious one-to-one correspondence between trees with and without the extra node, so the difference is just a matter of formulations, but when comparing results one should be careful whether, for example, the extra node is counted or not. The extra node yields the technical advantage that also the root has indegree 1 and thus total degree $= 1 + d^+(v)$; it further gives each embedding in the plane a unique ordering of the children of every node (in clockwise order from the parent, say). Nevertheless, we find this device less natural and we will not use it in the present paper. (We use outdegrees instead of degrees and assume that an ordering of the children as above is given; then there are no problems.)

We are primarily interested in (large) finite trees, but we will also consider infinite trees, for example as limit objects in our main theorem (Theorem 7.1). The infinite trees may have nodes with infinite outdegree $d^+(v) = \infty$; in this case we assume that the children are ordered $v_1, v_2, \dots$ (i.e., the order type of the set of children is $\mathbb{N}$).

We let $\mathfrak{T}_n$ be the set of all ordered rooted trees with n nodes (including the root) and let $\mathfrak{T}_{\mathrm{f}} := \bigcup_{n=1}^{\infty} \mathfrak{T}_n$ be the set of all finite ordered rooted trees; see further Section 6.

Remark 2.2. Note that $\mathfrak{T}_n$ is a finite set. In fact, it is well-known that its size $|\mathfrak{T}_n|$ is the $(n-1)$:th Catalan number

$$C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = \frac{(2n-2)!}{n!\,(n-1)!}, \tag{2.1}$$

see e.g. [33, Section 1.2.2 and Theorem 3.2], [40, Section I.2.3] or [103, Exercise 6.19(e)], but we do not need this.

For any tree T, we let $|T|$ denote the number of nodes; we call $|T|$ the *size* of T. As is well known, for any finite tree T,

$$\sum_{v \in T} d^+(v) = |T| - 1, \tag{2.2}$$

since every node except the root is the child of exactly one node.

2.2. **Galton–Watson trees.** An important class of examples of random ordered rooted trees is given by the *Galton–Watson trees*. These are defined as the family trees of Galton–Watson processes: Given a probability distribution $(\pi_k)_{k=0}^{\infty}$ on $\mathbb{Z}_{\geqslant 0}$, or, equivalently, a random variable ξ with distribution $(\pi_k)_{k=0}^{\infty}$, we build the tree $\mathcal{T}$ recursively, starting with the root and giving each node a number of children that is an independent copy of ξ. (We call $(\pi_k)_{k=0}^{\infty}$ the *offspring distribution* of $\mathcal{T}$; we sometimes also abuse the language and call ξ the offspring distribution.) In other words, the outdegrees $d^+(v)$ are i.i.d. with the distribution $(\pi_k)_{k=0}^{\infty}$.

Recall that the Galton–Watson process is called *subcritical*, *critical* or *supercritical* as the expected number of children $\mathbb{E}\,\xi = \sum_{k=0}^{\infty} k\pi_k$ satisfies $\mathbb{E}\,\xi < 1$, $\mathbb{E}\,\xi = 1$ or $\mathbb{E}\,\xi > 1$. It is a standard basic fact of branching process theory that $\mathcal{T}$ is finite a.s. if $\mathbb{E}\,\xi \leqslant 1$ (i.e., in the subcritical and critical cases), but $\mathcal{T}$ is infinite with positive probability if $\mathbb{E}\,\xi > 1$ (the supercritical case), see e.g. Athreya and Ney [8].

The Galton–Watson trees have random sizes. We are mainly interested in random trees with a given size; we thus define $\mathcal{T}_n$ as $\mathcal{T}$ conditioned on $|\mathcal{T}| = n$. These random trees $\mathcal{T}_n$ are called *conditioned Galton–Watson trees.* By definition, $\mathcal{T}_n$ has size $|\mathcal{T}_n| = n$.

It is well-known that several important classes of random trees can be seen as conditioned Galton–Watson tree, see e.g. Aldous [4], Devroye [32], Drmota [33] and Section 10.

2.3. **Simply generated trees.** The random trees that we will study are a generalization of the Galton–Watson trees. We suppose in this paper 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 tree $T \in \mathfrak{T}_{\mathrm{f}}$ by

$$w(T) := \prod_{v\in T} w_{d^+(v)}, \tag{2.3}$$

taking the product over all nodes v in T. Trees with such weights are called *simply generated trees* and were introduced by Meir and Moon [85]. To avoid trivialities, we assume that $w_0 > 0$ and that there exists some $k \geqslant 2$ with $w_k > 0$.

We let $\mathcal{T}_n$ be the random tree obtained by picking an element of $\mathfrak{T}_n$ at random with probability proportional to its weight, i.e.,

$$\mathbb{P}(\mathcal{T}_n = T) = \frac{w(T)}{Z_n}, \qquad T \in \mathfrak{T}_n, \tag{2.4}$$

where the normalizing factor Z_n is given by

$$Z_n = Z_n(\mathbf{w}) := \sum_{T\in\mathfrak{T}_n} w(T); \tag{2.5}$$

Z_n is known as the *partition function.* This definition makes sense only when $Z_n > 0$; we tacitly consider only such n when we discuss $\mathcal{T}_n$. Our assumptions $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 2$ imply that $Z_n > 0$ for infinitely many n, see Corollary 15.6 for a more precise result. (In most applications, $w_1 > 0$, and then $Z_n > 0$ for every $n \geqslant 1$, so there is no problem at all. The archetypical example with a parity restriction is given by the random (full) binary tree, see Example 10.3, for which $Z_n > 0$ if and only if n is odd.)

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: $\mathbb{P}(\xi = k) = w_k$; we further let $\mathcal{T}$ be the random Galton–Watson tree generated by ξ. It follows directly from the definitions that for every finite tree $T \in \mathfrak{T}_{\mathrm{f}}$, $\mathbb{P}(\mathcal{T} = T) = w(T)$. Hence

$$Z_n = \mathbb{P}(|\mathcal{T}| = n) \tag{2.6}$$

and the simply generated random tree $\mathcal{T}_n$ is the same as the random Galton–Watson tree $\mathcal{T}$ conditioned on $|\mathcal{T}| = n$, i.e., it equals the conditioned Galton–Watson tree $\mathcal{T}_n$ defined above.

It is well-known, see Section 4 for details, that in many cases it is possible to change the weight sequence (w_k) to a probability weight sequence without changing the distribution of the random trees $\mathcal{T}_n$; 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. (In many papers it is further assumed that ξ has finite variance, or even a finite exponential moment. This is not needed for the main results presented here, but may be necessary for other results. See also Sections 8, 19 and 21.)

3. Notation

We consider a fixed weight sequence $\mathbf{w} = (w_k)_{k\geqslant 0}$. The *support* $\operatorname{supp}(\mathbf{w})$ of the weight sequence $\mathbf{w} = (w_k)$ is $\{k : w_k > 0\}$. We define

$$\omega = \omega(\mathbf{w}) := \sup \operatorname{supp}(\mathbf{w}) = \sup\{k : w_k > 0\} \leqslant \infty, \tag{3.1}$$

(When considering $\mathcal{T}_n$, we assume, as said above, $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 2$; this can be written $0 \in \operatorname{supp}(\mathbf{w})$ and $\omega \geqslant 2$.)

We further define (assuming that the support contains at least two points)

$$\operatorname{span}(\mathbf{w}) := \max\{d \geqslant 1 : d \mid (i-j) \text{ whenever } w_i, w_j > 0\}. \tag{3.2}$$

Since we assume $w_0 > 0$, i.e., $0 \in \operatorname{supp}(\mathbf{w})$, we can simplify this to

$$\operatorname{span}(\mathbf{w}) = \max\{d \geqslant 1 : d \mid i \text{ whenever } w_i > 0\}, \tag{3.3}$$

the greatest common divisor of $\operatorname{supp}(\mathbf{w})$.

We let

$$\Phi(z) := \sum_{k=0}^{\infty} w_k z^k \tag{3.4}$$

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

$$\rho = 1/\limsup_{k\to\infty} w_k^{1/k}. \tag{3.5}$$

$\Phi(\rho)$ is always defined, with $0 < \Phi(\rho) \leqslant \infty$. Note that (assuming $\omega > 0$) $\Phi(\infty) = \infty$; in particular, if $\rho = \infty$, then $\Phi(\rho) = \infty$. On the other hand, if $\rho < \infty$, then both $\Phi(\rho) = \infty$ and $\Phi(\rho) < \infty$ are possible. If $\rho > 0$, then $\Phi(t) \nearrow \Phi(\rho)$ as $t \nearrow \rho$ by monotone convergence.

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

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} = \frac{\sum_{k=0}^{\infty} k w_k t^k}{\sum_{k=0}^{\infty} w_k t^k}; \tag{3.6}$$

$\Psi(t)$ is thus defined and finite at least for $0 \leqslant t < \rho$, and if $\Phi(\rho) < \infty$, then $\Psi(\rho)$ is still defined by (3.6), with $\Psi(\rho) \leqslant \infty$ (note that the numerator in (3.6) may diverge in this case, but not for $0 \leqslant t < \rho$). Moreover, if $\Phi(\rho) = \infty$, we define $\Psi(\rho) := \lim_{t \nearrow \rho} \Psi(t) \leqslant \infty$. (The limit exists by Lemma 3.1(i) below, but may be infinite.)

Alternatively, (3.6) may be written

$$\Psi(e^x) = e^x \frac{\Phi'(e^x)}{\Phi(e^x)} = \frac{\mathrm{d}}{\mathrm{d}x} \log \Phi(e^x). \tag{3.7}$$

The function Ψ will play a central role in the sequel. This is mainly because of Lemma 4.2 below, which gives a probabilistic interpretation of $\Psi(t)$. Its basic properties are given by the following lemma, which is proved in Section 13.

Lemma 3.1. *Let $\mathbf{w} = (w_k)_{k=0}^\infty$ be a given weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 1$ (i.e., $\omega(\mathbf{w}) > 0$).*

(i) *If $0 < \rho \leqslant \infty$, then the function*

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} = \frac{\sum_{k=0}^\infty k w_k t^k}{\sum_{k=0}^\infty w_k t^k} \tag{3.8}$$

is finite, continuous and (strictly) increasing on $[0, \rho)$, with $\Psi(0) = 0$.

(ii) *If $0 < \rho \leqslant \infty$, then $\Psi(t) \to \Psi(\rho) \leqslant \infty$ as $t \nearrow \rho$.*

(iii) *For any ρ, Ψ is continuous $[0, \rho] \to [0, \infty]$.*

(iv) *If $\rho < \infty$ and $\Phi(\rho) = \infty$, then $\Psi(\rho) := \lim_{t \to \rho} \Psi(t) = \infty$.*

(v) *If $\rho = \infty$, then $\Psi(\rho) := \lim_{t \to \rho} \Psi(t) = \omega \leqslant \infty$.*

Consequently, if $\rho > 0$, then

$$\Psi(\rho) = \lim_{t \nearrow \rho} \Psi(t) = \sup_{0 \leqslant t < \rho} \Psi(t) \in (0, \infty]. \tag{3.9}$$

We define

$$\nu := \Psi(\rho). \tag{3.10}$$

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

$$\nu = \frac{\rho \Phi'(\rho)}{\Phi(\rho)} \leqslant \infty. \tag{3.11}$$

It follows from Lemma 3.1 that $\nu = 0 \iff \rho = 0$, and that if $\rho > 0$, then

$$\nu := \Psi(\rho) = \lim_{t \nearrow \rho} \Psi(t) = \sup_{0 \leqslant t < \rho} \Psi(t) \in (0, \infty]. \tag{3.12}$$

It follows from (3.8) that $\nu \leqslant \omega$.

Note that all these parameters depend on the weight sequence $\mathbf{w} = (w_k)$; we may occasionally write e.g. $\omega(\mathbf{w})$ and $\nu(\mathbf{w})$, but usually we for simplicity do not show $\mathbf{w}$ explicitly in the notation.

Remark 3.2. Let $\mathcal{Z}(z)$ denote the generating function $\mathcal{Z}(z) := \sum_{n=1}^\infty Z_n z^n$. Then

$$\mathcal{Z}(z) = z\Phi(\mathcal{Z}(z)), \tag{3.13}$$

as shown already by Otter [93]. This equation is the basis of much work on simply generated trees using algebraic and analytic methods, see e.g. Drmota [33], but the present paper uses different methods and we will use (3.13) only in a few minor remarks.

3.1. **More notation.** 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 $\overline{\mathbb{N}}_1 := \mathbb{N}_1 \cup \{\infty\}$.

All unspecified limits are as $n \to \infty$. Thus, $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$. We use $\xrightarrow{\mathrm{p}}$ and $\xrightarrow{\mathrm{d}}$ for convergence in probability and distribution, respectively, of random variables, and $\overset{\mathrm{d}}{=}$ for equality in distribution. We use o_{p} and O_{p} in the standard senses: $o_{\mathrm{p}}(a_n)$ is an unspecified random variable X_n such that $X_n/a_n \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$, and $O_{\mathrm{p}}(a_n)$ is a random variable X_n such that X_n/a_n is stochastically bounded (usually called *tight*). We say that some event holds *w.h.p.* (with high probability) if its probability tends to 1 as $n \to \infty$. (See further e.g. [62].)

A *coupling* of two random variables X and Y is formally a pair of random variables X' and Y' defined on a common probability space such that $X \overset{\mathrm{d}}{=} X'$ and $Y \overset{\mathrm{d}}{=} Y'$; with a slight abuse of notation we may continue to write X and Y, thus replacing the original variables with new ones having the same distributions.

We write $X_n \overset{\mathrm{d}}{\approx} X'_n$ for two sequences of random variables or vectors X_n and X'_n if there exists a coupling of X_n and X'_n with $X_n = X'_n$ w.h.p.; this is equivalent to $d_{\mathrm{TV}}(X_n, X'_n) \to 0$ as $n \to \infty$, where d_{TV} denotes the total variation distance.

We use $C_1, C_2, \dots$ to denote unimportant constants, possibly different at different occurrences.

Recall that $d^+(v) = d^+_T(v)$ always denotes the *outdegree* of a node v in a tree T. (We use the notation $d^+(v)$ rather than $d(v)$ to emphasise this.) We will not use the total degree $d(v) = 1 + d^+(v)$ (when $v \neq o$), but care should be taken when comparing with other papers.

4. Equivalent weights

If $a, b > 0$ and we change w_k to

$$\widetilde{w}_k := ab^k w_k, \tag{4.1}$$

then, for every tree $T \in \mathfrak{T}_n$, $w(T)$ is changed to, using (2.2),

$$\widetilde{w}(T) = a^n b^{\sum_v d^+(v)} w(T) = a^n b^{n-1} w(T). \tag{4.2}$$

Consequently, Z_n is changed to

$$\widetilde{Z}_n := a^n b^{n-1} Z_n, \tag{4.3}$$

and the probabilities in (2.4) are 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 [73], 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 (4.1) (for some $a, b > 0$) are *equivalent*. (This is clearly an equivalence relation on the set of weight sequences.)

Let us see how replacing (w_k) by the equivalent weight sequence $(\widetilde{w}_k)$ affects the parameters defined above. The support, span and ω are not affected at all.

The generating function $\Phi(t)$ is replaced by

$$\widetilde{\Phi}(t) := \sum_{k=0}^{\infty} \widetilde{w}_k t^k = \sum_{k=0}^{\infty} ab^k t^k = a\Phi(bt), \tag{4.4}$$

with radius of convergence $\widetilde{\rho} = \rho/b$. Further, $\Psi(t)$ is replaced by

$$\widetilde{\Psi}(t) := \frac{t\widetilde{\Phi}'(t)}{\widetilde{\Phi}(t)} = \frac{tab\Phi'(bt)}{a\Phi(bt)} = \Psi(bt). \tag{4.5}$$

Hence, if $\rho > 0$, ν is replaced by, using (3.12),

$$\widetilde{\nu} := \sup_{0 \leqslant t < \widetilde{\rho}} \widetilde{\Psi}(t) = \sup_{0 \leqslant t < \rho/b} \Psi(bt) = \sup_{0 \leqslant s < \rho} \Psi(s) = \nu;$$

if $\rho = 0$ then $\widetilde{\nu} = \widetilde{\rho} = 0 = \nu$ is trivial. In other words, ν is invariant and depends only on the equivalence class of the weight sequence.

Lemma 4.1. *There exists a probability weight sequence equivalent to* (w_k) *if and only if and only if* $\rho > 0$. *In this case, the probability weight sequences equivalent to* (w_k) *are given by*

$$p_k = \frac{t^k w_k}{\Phi(t)}, \tag{4.6}$$

for any $t > 0$ *such that* $\Phi(t) < \infty$.

Proof. The equivalent weight sequence $(\widetilde{w}_k)$ given by (4.1) is a probability distribution if and only if

$$1 = \sum_{k=0}^{\infty} \widetilde{w}_k = a \sum_{k=0}^{\infty} w_k b^k = a\Phi(b),$$

i.e., if and only if $\Phi(b) < \infty$ and $a = \Phi(b)^{-1}$. Thus, there exists a probability weight sequence equivalent to (w_k) if and only if there exists $b > 0$ with $\Phi(b) < \infty$, i.e., if and only if $\rho > 0$; in this case we can choose any such b and take $a := \Phi(b)^{-1}$, which yields (4.6) (with $t = b$). $\square$

We easily find the probability generating function and thus moments of the probability weight sequence in (4.6); we state this in a form including the trivial case $t = 0$.

Lemma 4.2. *If* $t \geqslant 0$ *and* $\Phi(t) < \infty$, *then*

$$p_k := \frac{t^k w_k}{\Phi(t)}, \qquad k \geqslant 0, \tag{4.7}$$

defines a probability weight sequence (p_k). *This probability distribution has probability generating function*

$$\Phi_t(z) := \sum_{k=0}^{\infty} p_k z^k = \frac{\Phi(tz)}{\Phi(t)}, \tag{4.8}$$

and a random variable ξ *with this distribution has expectation*

$$\mathbb{E}\,\xi = \Phi_t'(1) = \frac{t\Phi'(t)}{\Phi(t)} = \Psi(t) \tag{4.9}$$

and variance

$$\operatorname{Var}\xi = t\Psi'(t); \tag{4.10}$$

furthermore, for any $s \geqslant 0$ *and* $x \geqslant 0$,

$$\mathbb{P}(\xi \geqslant x) \leqslant e^{-sx}\frac{\Phi(e^s t)}{\Phi(t)} \leqslant e^{-sx}\frac{\Phi(e^s t)}{\Phi(0)}. \tag{4.11}$$

If $t < \rho$, then $\mathbb{E}\,\xi$ and $\operatorname{Var}\xi$ are finite. If $t = \rho$, however, $\mathbb{E}\,\xi$ and $\operatorname{Var}\xi$ may be infinite (we define $\operatorname{Var}\xi = \infty$ when $\mathbb{E}\,\xi = \infty$, but $\operatorname{Var}\xi$ may be infinite also when $\mathbb{E}\,\xi$ is finite); (4.9)–(4.10) still hold, with $\Psi'(\rho) \leqslant \infty$ defined as the limit $\lim_{s\nearrow\rho}\Psi'(s)$. The tail estimate (4.11) is interesting only when $t < \rho$, when we may choose any $s < \log(\rho/t)$ and obtain the estimate $O(e^{-sx})$.

Proof. Direct summations yield

$$\sum_{k=0}^{\infty} p_k = \frac{\sum_{k=0}^{\infty} t^k w_k}{\Phi(t)} = 1 \tag{4.12}$$

and, more generally,

$$\sum_{k=0}^{\infty} p_k z^k = \frac{\sum_{k=0}^{\infty} w_k t^k z^k}{\Phi(t)} = \frac{\Phi(tz)}{\Phi(t)}, \tag{4.13}$$

showing that (p_k) is a probability distribution with the probability generating function Φ_t given in (4.8).

The expectation $\mathbb{E}\,\xi = \Phi_t'(1)$ is evaluated by differentiating (4.8) (for $z < 1$ and then taking the limit as $z \to 1$ to avoid convergence problems if $t = \rho$), or directly from (4.7) as

$$\mathbb{E}\,\xi = \sum_{k=0}^{\infty} k p_k = \frac{\sum_{k=0}^{\infty} k w_k t^k}{\Phi(t)} = \Psi(t).$$

Similarly, the variance is given by, using (4.8) and (4.9),

$$\operatorname{Var}\xi = \Phi_t''(1) + \Phi_t'(1) - (\Phi_t'(1))^2 = \frac{t^2\Phi''(t)}{\Phi(t)} + \frac{t\Phi'(t)}{\Phi(t)} - \left(\frac{t\Phi'(t)}{\Phi(t)}\right)^2 = t\Psi'(t).$$

Alternatively,

$$t\Psi'(t) = t\frac{\mathrm{d}}{\mathrm{d}t}\frac{\sum_{k=0}^{\infty} kt^k w_k}{\Phi(t)} = \frac{\sum_{k=0}^{\infty} k^2 t^k w_k}{\Phi(t)} - \left(\frac{\sum_{k=0}^{\infty} kt^k w_k}{\Phi(t)}\right)^2$$
$$= \sum_{k=0}^{\infty} k^2 p_k - \left(\sum_{k=0}^{\infty} kp_k\right)^2 = \mathbb{E}\,\xi^2 - (\mathbb{E}\,\xi)^2 = \operatorname{Var}\xi.$$

(In the case $t = \rho$ and $\operatorname{Var}\xi = \infty$, we use this calculation for $t' < t$ and let $t' \to t$.)

Finally, by (4.8),

$$\mathbb{P}(\xi \geqslant x) \leqslant e^{-sx}\,\mathbb{E}\,e^{s\xi} = e^{-sx}\Phi_t(e^s) = e^{-sx}\frac{\Phi(e^s t)}{\Phi(t)}. \qquad \square$$

In particular, taking $t = 1$, we recover the standard facts that if (w_k) is a probability distribution, so $\Phi(1) = 1$, then it has expectation $\Phi'(1) = \Psi(1)$ and variance $\Psi'(1)$.

Remark 4.3. We see from Lemma 4.1 that the probability weight sequences equivalent to (w_k) are given by (4.6), where $t \in (0, \rho]$ when $\Psi(\rho) < \infty$ and $t \in (0, \rho)$ when $\Psi(\rho) = \infty$. By Lemma 3.1, $t \mapsto \mathbb{E}\,\xi = \Psi(t)$ is an increasing bijection $(0, \rho] \to (0, \nu]$ and $(0, \rho) \to (0, \nu)$. Hence, any equivalent probability weight sequence is uniquely determined by its expectation, and the possible expectations are $(0, \nu]$ (when $\Psi(\rho) < \infty$) or $(0, \nu)$ (when $\Psi(\rho) = \infty$).

Remark 4.4. Note that we will frequently use (4.6) to define a new probability weight sequence also if we start with a probability weight sequence (w_k). Probability distributions related in this way are called *conjugated* or *tilted*. Conjugate distributions were introduced by Cramér [27] as an important tool in large deviation theory, see e.g. [31]. The reason is essentially the same as in the present paper: by conjugating the distribution we can change its mean in a way that enables us to keep control over sums S_n.

5. A modified Galton–Watson 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 $(\pi_k)_{k=0}^{\infty}$:

$$\mathbb{P}(\xi = k) = \pi_k, \qquad k = 0, 1, 2, \ldots \tag{5.1}$$

We assume that the expectation $\mu := \mathbb{E}\,\xi = \sum_k k\pi_k \leqslant 1$ (the subcritical or critical case).

In this case, we define (based on Kesten [74] and Jonsson and Stefánsson [67]) 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

$$\mathbb{P}(\widehat{\xi}=k):=\begin{cases}k\pi_k, & k=0,1,2,\dots,\\ 1-\mu, & k=\infty.\end{cases} \tag{5.2}$$

(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.

Thus, for a special node, and any integers j,k with $1\leqslant j\leqslant k<\infty$, the probability that the node has exactly k children and that the j:th of them is special is $k\pi_k/k=\pi_k$.

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 (5.2) is the *size-biased* distribution of ξ, and $\widehat{\mathcal{T}}$ is the size-biased Galton–Watson tree defined by Kesten [74], see also Aldous [4], Aldous and Pitman [6], Lyons, Pemantle and Peres [84] and Remark 5.7 below. The underlying size-biased Galton–Watson process is the same as the *Q-process* studied in Athreya and Ney [8, Section I.14], which is an instance of Doob's h-transform. (See Lyons, Pemantle and Peres [84] for further related constructions in other contexts and Geiger and Kauffmann [45] 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 $\mathrm{Ge}(1-\mu)$,

$$\mathbb{P}(L=\ell)=(1-\mu)\mu^{\ell-1},\qquad \ell=1,2,\dots. \tag{5.3}$$

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 (5.3). 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 [67].

Example 5.1. In the extreme case $\mu = 0$, or equivalently $\xi = 0$ a.s., i.e., $\pi_0 = 1$ and $\pi_k = 0$ for $k \geqslant 1$, (5.2) 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 5.2. 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 [84]. (In the Poisson case, Grimmett [47] 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 5.3. Some related modifications of Galton–Watson trees having a finite spine have been considered previously. Sagitov and Serra [102] construct (as a limit for a certain two-type branching process) a random tree similar to the one in (T2) above (with a subcritical ξ), with a finite spine having a length with the geometric distribution (5.3); the difference is that at the top of the spine, only a finite number of Galton–Watson trees $\mathcal{T}$ are attached. (This number may be a copy of $\xi^* - 1$ as at the other points of the spine, or it may have a different distribution, see [102].) Thus there is no explosion, and the tree is finite. Another modified Galton–Watson tree is used by Addario-Berry, Devroye and Janson [1]; the proofs use a truncated version of (T1) above (with a critical ξ), where the spine has a fixed length k; at the top of the spine the special node becomes normal and reproduces normally with ξ children. Geiger [44] studied $\mathcal{T}$ conditioned on its height being at least n, see Section 22, and gave a construction of it using a spine of length n, but with more complicated rules for the branches. See also the modified trees $\widehat{\mathcal{T}}_{1n}$, $\widehat{\mathcal{T}}_{2n}$, $\widehat{\mathcal{T}}_{3n}$ in Section 20.

The invariant random sin-tree constructed by Aldous [2] in a more general situation, is for a critical Galton–Watson process another related tree; it has an infinite spine as $\widehat{\mathcal{T}}$, but differs from $\widehat{\mathcal{T}}$ in that the root has $\xi+1$ children (and thus ξ normal children) instead of $\widehat{\xi}$. In this case, it may be better to reverse the orientation of the spine and consider the spine as an infinite path $\cdots v_{-2}v_{-1}v_0$ starting at $-\infty$ (there is thus no root); we attach further branches (copies of $\mathcal{T}$) as above, with all v_i, $i<0$, special (the number of children is a copy of $\widehat{\xi}$), but the top node v_0 normal (the number of children is a copy of ξ, and all are normal).

Kurtz, Lyons, Pemantle and Peres [78] and Chassaing and Durhuus [23] have constructed related trees with infinite spines using multi-type Galton–Watson processs.

Remark 5.4. If ξ has the probability generating function $\varphi(x) := \mathbb{E}\,x^{\xi} = \sum_{k=0}^{\infty}\pi_k x^k$, then $\widehat{\xi}$ has by (5.2) the probability generating function

$$\mathbb{E}\,x^{\widehat{\xi}} = \sum_{k=0}^{\infty} k\pi_k x^k = x\varphi'(x), \tag{5.4}$$

at least for $0 \leqslant x < 1$. (Also for $\mu < 1$ when $\widehat{\xi}$ may take the value ∞.)

Remark 5.5. In case $(\mathsf{T1})$, the random variable $\widehat{\xi}$ is a.s. finite and has mean

$$\mathbb{E}\,\widehat{\xi} = \sum_{k=0}^{\infty} k\,\mathbb{P}(\widehat{\xi}=k) = \sum_{k=0}^{\infty} k^2\pi_k = \mathbb{E}\,\xi^2 = \sigma^2+1, \tag{5.5}$$

where $\sigma^2 := \operatorname{Var}\xi \leqslant \infty$. In case $(\mathsf{T2})$, 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 $(\mathsf{T1})$, and where σ^2 appears as a parameter (see e.g. Section 21), the correct generalization of σ^2 to the subcritical case $(\mathsf{T2})$ is not $\operatorname{Var}\xi$ but $\mathbb{E}\,\widehat{\xi}-1=\infty$. (See Remark 5.6 below for a simple example.) We thus define, for any distribution $(\pi_k)_{k=0}^{\infty}$ with expectation $\mu \leqslant 1$,

$$\hat{\sigma}^2 := \mathbb{E}\,\widehat{\xi} - 1 = \begin{cases} \sigma^2, & \mu = 1, \\ \infty, & \mu < 1. \end{cases} \tag{5.6}$$

Remark 5.6. Let $l_k(T)$ denote the number of nodes with distance k to the root in a rooted tree T. (This is thus the size of the k:th generation.) Trivially, $l_0(T)=1$, while $l_1(T)=d_T^+(o)$, the root degree.

It follows by the construction of $\widehat{\mathcal{T}}$ and induction that in case $(\mathsf{T1})$, using (5.5),

$$\mathbb{E}\,l_k(\widehat{\mathcal{T}}) = 1 + k(\mathbb{E}\,\widehat{\xi}-1) = k\sigma^2+1, \qquad k \geqslant 0. \tag{5.7}$$

In case $(\mathsf{T2})$, we have if $\mu > 0$ and $k \geqslant 1$ a positive probability that $L = k$ and then $l_k(\widehat{\mathcal{T}}) = \infty$. Thus $\mathbb{E}\,l_k(\widehat{\mathcal{T}}) = \infty$. Consequently, using (5.6), if

$0 < \mu \leqslant 1$, then

$$\mathbb{E}\, l_k(\widehat{\mathcal{T}}) = k\hat{\sigma}^2 + 1, \qquad k \geqslant 1. \tag{5.8}$$

However, this fails if $\mu = 0$; in that case, $l_1(\widehat{\mathcal{T}}) = \infty$ but $l_k(\widehat{\mathcal{T}}) = 0$ for $k \geqslant 2$, see Example 5.1.

Remark 5.7. As said above, in the case $\mu = 1$, the tree $\widehat{\mathcal{T}}$ is the size-biased Galton–Watson tree, see [74], [6] and [84]. For comparison, we give the definition of the latter, for an arbitrary distribution $(\pi_k)_{k\geqslant 0}$ with finite mean $\mu > 0$: Let, as above, ξ have the distribution (π_k), see (5.1), and let ξ^* have the size-biased distribution defined by

$$\mathbb{P}(\xi^* = k) = \frac{k\pi_k}{\mu}, \qquad k = 0, 1, 2, \dots \tag{5.9}$$

(Note that this is a probability distribution on $\mathbb{N}_1$.) Construct $\mathcal{T}^*$ as $\widehat{\mathcal{T}}$ above, with normal and special nodes, with the only difference that the number of children of a special node has the distribution of ξ^* in (5.9).

In the critical case $\mu = 1$, we have $\xi^* = \widehat{\xi}$ and thus $\mathcal{T}^* = \widehat{\mathcal{T}}$, but in the subcritical case $\mu < 1$, $\mathcal{T}^*$ and $\widehat{\mathcal{T}}$ are clearly different. (Note that $\mathcal{T}^*$ always is locally finite, but $\widehat{\mathcal{T}}$ is not when $\mu < 1$.) When $\mu > 1$, $\widehat{\mathcal{T}}$ is not even defined, but $\mathcal{T}^*$ is. (As remarked by Aldous and Pitman [6], in the supercritical case $\mathcal{T}^*$ has a.s. an uncountable number of infinite paths from the root, in contrast to the case $\mu \leqslant 1$ when the spine a.s. is the only one.)

$\mathcal{T}^*$ can also be constructed by the alternative construction in (T1) above starting with an infinite spine, again with the difference that $\widehat{\xi}-1$ is replaced by ξ^*-1. $\mathcal{T}^*$ can also be seen as a Galton–Watson process with immigration in the same way as in Remark 5.2.

By (5.9), the probability that a given special node in $\mathcal{T}^*$ has $k \geqslant 1$ children, with a given one of them special, is

$$\frac{1}{k}\,\mathbb{P}(\xi^* = k) = \frac{k\pi_k}{k\mu} = \frac{\pi_k}{\mu}. \tag{5.10}$$

Let T be a fixed tree of height ℓ, and let u be a node in the ℓ:th (and last) generation in T. Let $\mathcal{T}^{*(\ell)}$ denote $\mathcal{T}^*$ truncated at height ℓ. It follows from (5.10) and independence that the probability that $\mathcal{T}^{*(\ell)} = T$ and that u is special (i.e., u is the unique element of the spine at distance ℓ from the root) equals $\mu^{-\ell}\,\mathbb{P}(\mathcal{T}^{(\ell)} = T)$. Hence, summing over the $l_\ell(T)$ possible u,

$$\mathbb{P}(\mathcal{T}^{*(\ell)} = T) = \mu^{-\ell} l_\ell(T)\,\mathbb{P}(\mathcal{T}^{(\ell)} = T), \tag{5.11}$$

which explains the name size-biased Galton–Watson tree. (As an alternative, one can thus define $\mathcal{T}^*$ directly by (5.11), noting that this gives consistent distributions for $m = 1, 2, \dots$, see Kesten [74].) See further Section 22.2.

6. The Ulam–Harris tree and convergence

It is convenient, especially when discussing convergence, to regard our trees as subtrees of the infinite Ulam–Harris tree defined as follows. (See e.g. Otter [93], Harris [51, § VI.2], Neveu [91] and Kesten [74].)

Definition 6.1. The Ulam–Harris tree U_∞ is the infinite rooted tree with node set $V_\infty := \bigcup_{k=0}^{\infty} \mathbb{N}_1^k$, the set of all finite strings $i_1 \cdots i_k$ of positive integers, including the empty string $\emptyset$ which we take as the root o, and with an edge joining $i_1 \cdots i_k$ and $i_1 \cdots i_{k+1}$ for any $k \geqslant 0$ and $i_1, \dots, i_{k+1} \in \mathbb{N}_1$.

Thus every node $v = i_1 \cdots i_k$ has outdegree $d^+(v) = \infty$; the children of v are the strings $v1$, $v2$, $v3$, ..., and we let them have this order so U_∞ becomes an infinite ordered rooted tree. The parent of $i_1 \cdots i_k$ ($k > 0$) is $i_1 \cdots i_{k-1}$.

The family $\mathfrak{T}$ of ordered rooted trees can be identified with the set of all rooted subtrees T of U_∞ that have the property

$$i_1 \cdots i_k i \in V(T) \implies i_1 \cdots i_k j \in V(T) \text{ for all } j \leqslant i. \tag{6.1}$$

Equivalently, by identifying T and its node set $V(T)$, we can regard $\mathfrak{T}$ as the family of all subsets V of V_∞ that satisfy

$$\emptyset \in V, \tag{6.2}$$

$$i_1 \cdots i_{k+1} \in V \implies i_1 \cdots i_k \in V, \tag{6.3}$$

$$i_1 \cdots i_k i \in V \implies i_1 \cdots i_k j \in V \quad \text{for all} \quad j \leqslant i. \tag{6.4}$$

We let $\mathfrak{T}_{\mathrm{f}} := \{T \in \mathfrak{T} : |T| < \infty\}$ be the set of all finite ordered rooted trees and $\mathfrak{T}_n := \{T \in \mathfrak{T} : |T| = n\}$ the set of all ordered rooted trees of size n.

If $T \in \mathfrak{T}$, we let as above $d^+(v) = d_T^+(v)$ denote the outdegree of v for every $v \in V(T)$, For convenience, we also define $d^+(v) = 0$ for $v \notin V(T)$; thus $d^+(v)$ is defined for every $v \in V_\infty$, and the tree $T \in \mathfrak{T}$ is uniquely determined by the (out)degree sequence $(d_T^+(v))_{v \in V_\infty}$. It is easily seen that this gives a bijection between $\mathfrak{T}$ and the set of sequences $(d_v) \in \overline{\mathbb{N}}_0^{V_\infty}$ with the property

$$d_{i_1 \cdots i_k i} = 0 \quad \text{when} \quad i > d_{i_1 \cdots i_k}. \tag{6.5}$$

The family $\mathfrak{T}_{\mathrm{lf}}$ of locally finite trees corresponds to the subset of all such sequences with all $d_v < \infty$, and the family $\mathfrak{T}_{\mathrm{f}}$ of finite trees correspond to the subset of all such sequences (d_v) with all $d_v < \infty$ and only finitely many $d_v \neq 0$.

In this way we have $\mathfrak{T}_{\mathrm{f}} \subset \mathfrak{T}_{\mathrm{lf}} \subset \mathfrak{T} \subset \overline{\mathbb{N}}_0^{V_\infty}$; note that $\mathfrak{T}_{\mathrm{lf}} = \mathfrak{T} \cap \mathbb{N}_0^{V_\infty}$, so $\mathfrak{T}_{\mathrm{f}} \subset \mathfrak{T}_{\mathrm{lf}} \subset \mathbb{N}_0^{V_\infty}$.

We give $\overline{\mathbb{N}}_0$ the usual compact topology as the one-point compactification of the discrete space $\mathbb{N}_0$. Thus $\overline{\mathbb{N}}_0$ is a compact metric space. (One metric, among many equivalent ones, is given by the homomorphism $n \mapsto 1/(n+1)$ onto $\{1/n\}_{n=1}^{\infty} \cup \{0\} \subset \mathbb{R}$.) We give $\overline{\mathbb{N}}_0^{V_\infty}$ the product topology and its subspaces $\mathfrak{T}_{\mathrm{f}}$, $\mathfrak{T}_{\mathrm{lf}}$ and $\mathfrak{T}$ the induced topologies. Thus $\overline{\mathbb{N}}_0^{V_\infty}$ is a compact metric space, and its subspaces $\mathfrak{T}_{\mathrm{f}}$, $\mathfrak{T}_{\mathrm{lf}}$ and $\mathfrak{T}$ are metric spaces. (The precise choice of metric on these spaces is irrelevant; we will not use any explicit metric except briefly in Section 20.) Moreover, the condition (6.5) defines $\mathfrak{T}$ as a closed subset of $\overline{\mathbb{N}}_0^{V_\infty}$; thus $\mathfrak{T}$ is a compact metric space. ($\mathfrak{T}_{\mathrm{f}}$ and $\mathfrak{T}_{\mathrm{lf}}$ are not compact. In fact, it is easily seen that they are dense proper subsets of $\mathfrak{T}$. $\mathfrak{T}_{\mathrm{f}}$ is a countable discrete space.)

In other words, if T_n and T are trees in $\mathfrak{T}$, then $T_n \to T$ if and only if the outdegrees converge pointwise:

$$d^+_{T_n}(v) \to d^+_T(v) \qquad \text{for each } v \in V_\infty. \tag{6.6}$$

It is easily seen that it suffices to consider $v \in V(T)$, i.e., (6.6) is equivalent to

$$d^+_{T_n}(v) \to d^+_T(v) \qquad \text{for each } v \in V(T), \tag{6.7}$$

since (6.7) implies that if $v \notin V(T)$, then $v \notin V(T_n)$ for sufficiently large n, and thus $d^+_{T_n}(v) = 0$. (Consider the last node w in $V(T)$ on the path from the root to v and use $d^+_{T_n}(w) \to d^+_T(w)$.)

Alternatively, we may as above consider the node set $V(T)$ as a subset of V_∞ and regard $\mathfrak{T}$ as the family of all subsets of V_∞ that satisfy (6.2)–(6.4). We identify the family of all subsets of V_∞ with $\{0,1\}^{V_\infty}$, and give this family the product topology, making it into a compact metric space. (Thus, convergence means convergence of the indicator $\mathbf{1}\{v \in \cdot\}$ for each $v \in V_\infty$.) This induces a topology on $\mathfrak{T}$, where $T_n \to T$ means that, for each $v \in V_\infty$, if $v \in V(T)$, then $v \in V(T_n)$ for all large n, and, conversely, if $v \notin V(T)$, then $v \notin V(T_n)$ for all large n.

If $v = i_1 \dots i_k$ with $k > 0$, then $v \in V(T)$ if and only of $i_k \leqslant d^+_T(i_1 \dots i_{k-1})$. It follows immediately that $V(T_n) \to V(T)$ in the sense just described, if and only if (6.6) holds. The two definitions of $T_n \to T$ above are thus equivalent (for $\mathfrak{T}$, and thus also for its subsets $\mathfrak{T}_{\mathrm{f}}$ and $\mathfrak{T}_{\mathrm{lf}}$).

Furthermore, we see, e.g. from (6.6), that the convergence of trees can be described recursively: Let $T_{(j)}$ denote the j:th subtree of T, i.e., the subtree rooted at the j:th child of T, for $j = 1, \dots, d^+_T(o)$. (We consider only finite j, even when $d^+_T(o) = \infty$.) Then, $T_n \to T$ if and only if

(i) the root degrees converge: $d^+_{T_n}(o) \to d^+_T(o)$, and further,
(ii) for each $j = 1, \dots, d^+_T(o)$, $T_{n,(j)} \to T_{(j)}$.

(Note that $T_{n,(j)}$ is defined for large n, at least, by (i).)

It is important to realize that the notion of convergence used here is a local (pointwise) one, so we consider only a single v at a time, or, equivalently, a finite set of v; there is no uniformity in v required.

If T is a locally finite tree, $T \in \mathfrak{T}_{\mathrm{lf}}$, then $d^+_T(v) < \infty$ for each v, and thus (6.6) means that for each v, $d^+_{T_n}(v) = d^+(v)$ for all sufficiently large n.

Let $T^{(m)}$ denote the tree T truncated at height m, i.e., the subtree of T consisting of all nodes in generations $0, \dots, m$. If T is locally finite, then each $T^{(m)}$ is a finite tree, and it is easily seen from (6.7) that convergence to T can be characterised as follows:

Lemma 6.2. *If T is locally finite, then, for any trees $T_n \in \mathfrak{T}$,*

$$\begin{aligned} T_n \to T &\iff T_n^{(m)} \to T^{(m)} \quad \textit{for each } m \\ &\iff T_n^{(m)} = T^{(m)} \quad \textit{for each } m \textit{ and all large } n. \end{aligned}$$

(The last condition means for n larger than some $n(m)$ depending on m.) □

This notion of convergence for locally finite trees is widely used; see e.g. Otter [93] and Aldous and Pitman [6].

In general, if T is not locally finite, this characterization fails. (For example, if S_n, $1 \leqslant n \leqslant \infty$, is a star where the root has outdegree n and its children all have outdegree 0, then $S_n \to S_\infty$, but $S_n^{(m)} \neq S_\infty^{(m)}$ for all n and $m \geqslant 1$.) Instead, we have to localise also horizontally: Let $V^{[m]} := \bigcup_{k=0}^{m} \{1, \dots, m\}^k$, the subset of V_∞ consisting of strings of length at most m with all elements at most m. For a tree $T \in \mathfrak{T}$, let $T^{[m]}$ be the subtree with node set $V(T) \cap V^{[m]}$, i.e., the tree T truncated at height m and pruned so that all outdegrees are at most m. It is then easy to see from (6.6) that the following analogue and generalization of Lemma 6.2 holds:

Lemma 6.3. *For any trees $T, T_n \in \mathfrak{T}$,*

$$\begin{aligned} T_n \to T &\iff T_n^{[m]} \to T^{[m]} \quad \textit{for each } m \\ &\iff T_n^{[m]} = T^{[m]} \quad \textit{for each } m \textit{ and all large } n. \end{aligned}$$

(The last condition means for n larger than some $n(m)$ depending on m.) □

Our notion of convergence for general trees $T \in \mathfrak{T}$ was introduced in this form by Jonsson and Stefánsson [67] (where the truncation $T^{[m]}$ is called a *left ball*).

Remark 6.4. It is straightforward to obtain versions of Lemmas 6.2–6.3 for random trees T, T_n and convergence in probability or distribution. For example: For any random trees $T, T_n \in \mathfrak{T}$,

$$T_n \xrightarrow{\mathrm{d}} T \iff T_n^{[m]} \xrightarrow{\mathrm{d}} T^{[m]} \quad \text{for each } m. \tag{6.8}$$

If $T \in \mathfrak{T}_{\mathrm{lf}}$, a.s., then we also have

$$T_n \xrightarrow{\mathrm{d}} T \iff T_n^{(m)} \xrightarrow{\mathrm{d}} T^{(m)} \quad \text{for each } m, \tag{6.9}$$

see e.g. Aldous and Pitman [6]. The proofs are standard using the methods in e.g. Billingsley [15].

7. Main result for simply generated random trees

Our main result for trees is the following, proved in Section 16. The case when $\nu \geqslant 1$ and $\sigma^2 < \infty$ was shown implicitly by Kennedy [73] (who considered Galton–Watson processes and not trees), and explicitly by Aldous and Pitman [6], see also Grimmett [47], Kolchin [76], Kesten [74] and Aldous [4]. Special cases with $0 < \nu < 1$ and $\nu = 0$ are given by Jonsson and Stefánsson [67] and Janson, Jonsson and Stefánsson [64], respectively.

Theorem 7.1. *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*

$$\pi_k := \frac{\tau^k w_k}{\Phi(\tau)}, \qquad k \geqslant 0; \tag{7.1}$$

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

$$\mu = \Psi(\tau) = \min(\nu, 1) \leqslant 1 \tag{7.2}$$

and variance $\sigma^2 = \tau\Psi'(\tau) \leqslant \infty$. *Let* $\widehat{\mathcal{T}}$ *be the infinite modified Galton–Watson tree constructed in Section 5 for the distribution* $(\pi_k)_{k\geqslant 0}$. *Then* $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ *as* $n \to \infty$, *in the topology defined in Section 6.*

Furthermore, in case (i), $\mu = 1$ *(the critical case) and* $\widehat{\mathcal{T}}$ *is locally finite with an infinite spine; in case* (ii) $\mu = \nu < 1$ *(the subcritical case) and* $\widehat{\mathcal{T}}$ *has a finite spine ending with an explosion.*

Remark 7.2. Note that we can combine the two cases $\nu \geqslant 1$ and $\nu < 1$ and define, using Lemma 3.1 and with $\Psi(\rho) = \nu$,

$$\tau := \max\Bigl\{t \leqslant \rho : \Psi(t) \leqslant 1\Bigr\}. \tag{7.3}$$

Remark 7.3. In case (ii), there is no $\tau \geqslant 0$ with $\Psi(\tau) = 1$, see Lemma 3.1. Hence the definition of τ can also be expressed as follows, recalling $\Psi(t) := t\Phi'(t)/\Phi(t)$ from (3.6): τ is the unique number in $[0, \rho]$ such that

$$\tau\Phi'(\tau) = \Phi(\tau), \tag{7.4}$$

if there exists any such τ; otherwise $\tau := \rho$. (Equation (7.4) is used in many papers to define τ, in the case $\nu \geqslant 1$.)

Remark 7.4. 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 by Lemma 3.1, 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 [89] and Jonsson and Stefánsson [67]. Consequently,

$$\frac{\Phi(\tau)}{\tau} = \inf_{0\leqslant t\leqslant\rho} \frac{\Phi(t)}{t} = \inf_{0\leqslant t<\infty} \frac{\Phi(t)}{t}. \tag{7.5}$$

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

Remark 7.5. By Remark 7.4, τ is, equivalently, the (unique) maximum point in $[0,\rho]$ of $t/\Phi(t)$, which by (3.13) is the inverse function of the generating function $\mathcal{Z}(z)$. It follows easily that

$$\tau = \mathcal{Z}(\rho_{\mathcal{Z}}), \tag{7.6}$$

where $\rho_{\mathcal{Z}} = \tau/\Phi(\tau)$ is the radius of convergence of $\mathcal{Z}$; see also Corollary 18.17. Note that $0 \leqslant \rho_{\mathcal{Z}} < \infty$ and that $\rho_{\mathcal{Z}} = 0 \iff \tau = 0 \iff \rho = 0$. Otter [93] uses (7.6) as the definition of τ (by him denoted a); see also Minami [89].

Remark 7.6. When $\nu = 0$ (which is equivalent to $\rho = 0$), the limit $\widehat{\mathcal{T}}$ is the non-random infinite star in Example 5.1, so Theorem 7.1 gives $\mathcal{T}_n \xrightarrow{\mathrm{p}} \widehat{\mathcal{T}}$.

Remark 7.7. We consider briefly the cases excluded from Theorem 7.1. The case when $w_0 = 0$ is completely trivial, since then $w(T) = 0$ for every finite tree, so $\mathcal{T}_n$ is undefined. The same holds (for $n \geqslant 2$) when $w_0 > 0$ but $w_k = 0$ for all $k \geqslant 1$, i.e., when $\omega = 0$.

The case when $w_0 > 0$ and $w_1 > 0$ but $w_k = 0$ for $k \geqslant 2$, so $\omega = 1$, is also trivial. Then $w(T) = 0$ unless T is a rooted path P_n for some n. Thus $Z_n = w(P_n) = w_0 w_1^{n-1}$, and (a.s.) $\mathcal{T}_n = P_n$, which converges as $n \to \infty$ to the infinite path P_∞. We have $\nu = 1 = \omega$, but, in contrast to Theorem 7.1, $\tau = \infty$, with τ defined e.g. by (7.3). Further, interpreting (7.1) as a limit, we have $\pi_k = \delta_{k1}$, so (π_k) is the distribution concentrated at 1; thus (5.2) yields $\widehat{\xi} = 1$ a.s., so $\widehat{\mathcal{T}}$ consists of an infinite spine only, i.e. $\widehat{\mathcal{T}} = P_\infty$. Consequently, $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ holds in this case too.

Remark 7.8. If we replace (w_k) by the equivalent weight sequence $(\widetilde{w}_k)$ given by (4.1), then (7.3) and (4.5) show that τ is replaced by

$$\widetilde{\tau} := \max\{t \leqslant \widetilde{\rho} : \widetilde{\Psi}(t) \leqslant 1\} = \max\{t \leqslant \rho/b : \Psi(bt) \leqslant 1\} = \tau/b. \tag{7.7}$$

The corresponding probability weight sequence given by (7.1) thus is, using (4.4),

$$\widetilde{\pi}_k := \frac{\widetilde{\tau}^k \widetilde{w}_k}{\widetilde{\Phi}(\widetilde{\tau})} = \frac{(\tau/b)^k a b^k w_k}{a\Phi(\tau)} = \frac{\tau^k w_k}{\Phi(\tau)} = \pi_k, \tag{7.8}$$

so the distribution (π_k) is invariant and depends only on the equivalence class of (w_k).

Remark 7.9. 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, see Lemma 4.1, but Theorem 7.1 and the theorems below show that (π_k) has a special role and therefore is a canonical choice of a weight sequence in its equivalence class. Remark 4.3 shows that (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.

A heuristic motivation for this choice of probability weight sequence is that when we construct $\mathcal{T}_n$ as a Galton–Watson tree $\mathcal{T}$ conditioned on $|\mathcal{T}| = n$, it is better to condition on an event of not too small probability; in the critical case this probability decreases as $n^{-3/2}$ provided $\sigma^2 < \infty$, see [93] ($\nu > 1$) and [76, Theorem 2.3.1] ($\nu \geqslant 1$, $\sigma^2 < \infty$), and always subexponentially, but in the subcritical and supercritical cases it typically decreases exponentially fast, see Theorems 18.7 and 18.11.

As a special case of Theorem 7.1 we have the following result for the root degree $d^+_{\mathcal{T}_n}(o)$, proved in Section 15.

Theorem 7.10. *Let $(w_k)_{k\geqslant 0}$ and $(\pi_k)_{k\geqslant 0}$ be as in Theorem 7.1. Then, as $n \to \infty$,*

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = d) \to d\pi_d, \quad d \geqslant 0. \tag{7.9}$$

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

$$d^+_{\mathcal{T}_n}(o) \xrightarrow{\mathrm{d}} \widehat{\xi}, \tag{7.10}$$

where $\widehat{\xi}$ is a random variable in $\overline{\mathbb{N}}_0$ with the distribution given in (5.2).

Note that the sum $\sum_0^\infty d\pi_d = \mu$ of the limiting probabilities in (7.9) 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 $\overline{\mathbb{N}}_0$.

Theorem 7.10 describes the degree of the root. If we instead take a random node, we obtain a different limit distribution, viz. (π_k). We state two versions of this; the two results are of the types called *annealed* and *quenched* in statistical physics. In the first (annealed) version, we take a random tree $\mathcal{T}_n$ and, simultaneously, a random node v in it. In the second (quenched) version we fix a random tree $\mathcal{T}_n$ and study the distribution of outdegrees in it. (This yields a random probability distribution. Equivalently, we study the outdegree of a random node conditioned on the tree $\mathcal{T}_n$.)

Theorem 7.11. *Let $(w_k)_{k\geqslant 0}$ and $(\pi_k)_{k\geqslant 0}$ be as in Theorem 7.1.*

(i) *Let v be a uniformly random node in $\mathcal{T}_n$. Then, as $n \to \infty$,*

$$\mathbb{P}(d^+_{\mathcal{T}_n}(v) = d) \to \pi_d, \quad d \geqslant 0. \tag{7.11}$$

(ii) *Let N_d be the number of nodes in $\mathcal{T}_n$ of outdegree d. Then*

$$\frac{N_d}{n} \xrightarrow{\mathrm{p}} \pi_d, \quad d \geqslant 0. \tag{7.12}$$

The proof is given in Section 17. (When $\nu > 1$, this was proved by Otter [93], see also Minami [89].) See Section 21.2 for further results.

Instead of considering just the outdegree of a random node, i.e., its number of children, we may obtain a stronger result by considering the subtree containing its children, grandchildren and so on. (This random subtree is called a *fringe subtree* by Aldous [2].) We have an analogous result, also proved in Section 17. Cf. [2], which in particular contains (i) below in the case $\nu \geqslant 1$ and $\sigma^2 < \infty$; this was extended by Bennies and Kersting [11] to the general case $\nu \geqslant 1$. (Note that the limit distribution, i.e. the distribution of $\mathcal{T}$, is a fringe distribution in the sense of [2] only if $\mu = 1$, i.e., if and only if $\nu \geqslant 1$.)

Theorem 7.12. *Let $(w_k)_{k\geqslant 0}$ and $(\pi_k)_{k\geqslant 0}$ be as in Theorem 7.1, and let $\mathcal{T}$ be the Galton–Watson tree with offspring distribution (π_k). Further, if v is a node in $\mathcal{T}_n$, let $\mathcal{T}_{n;v}$ be the subtree rooted at v.*

(i) *Let v be a uniformly random node in $\mathcal{T}_n$. Then, $\mathcal{T}_{n;v} \xrightarrow{\mathrm{d}} \mathcal{T}$, i.e., for any fixed tree T,*
$$\mathbb{P}(\mathcal{T}_{n;v} = T) \to \mathbb{P}(\mathcal{T} = T). \tag{7.13}$$
(ii) *Let T be an ordered rooted tree and let $N_T := |\{v : \mathcal{T}_{n;v} = T\}|$ be the number of nodes in $\mathcal{T}_n$ such that the subtree rooted there equals T. Then*
$$\frac{N_T}{n} \xrightarrow{\mathrm{p}} \mathbb{P}(\mathcal{T} = T). \tag{7.14}$$

Remark 7.13. Aldous [2] considers also the tree obtained by a random re-rooting of $\mathcal{T}_n$, i.e., the tree obtained by declaring a uniformly random node v to be the root. Note that this re-rooted tree contains $\mathcal{T}_{n;v}$ as a subtree, and that, provided $v \neq o$, there is exactly one branch from the new root not in this subtree, viz. the branch starting with the original parent of v. Aldous [2] shows, at least when $\nu \geqslant 1$ and $\sigma^2 < \infty$, convergence of this randomly re-rooted tree to the random sin-tree in Remark 5.3. The limit of the re-rooted tree is thus very similar to the limit of $\mathcal{T}_n$ in Theorem 7.1, but not identical to it.

8. Three different types of weights

Although Theorem 7.1 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 7.1, then III is the case discussed in Example 5.1; excluding this case, I and II are the same as (T1) and (T2) in Section 5.

We can reformulate the partition into three cases in more probabilistic terms. If ξ is a non-negative integer valued random variable with distribution given by $p_k = \mathbb{P}(\xi = k)$, $k \geqslant 0$, then the *exponential moments* of ξ are $\mathbb{E}\, R^\xi = \sum_{k=0}^\infty p_k R^k$ for $R > 1$. (Equivalently, $\mathbb{E}\, e^{r\xi}$ for $r := \log R > 0$.) We say that ξ, or the distribution (p_k), has *some finite exponential moment* if $\mathbb{E}\, R^\xi < \infty$ for some $R > 1$; this is equivalent to the probability generating function $\sum_{k=0}^\infty p_k z^k$ having radius of convergence strictly larger than 1.

Consider again a probability distribution $(\widetilde{w}_k)$ equivalent to (w_k), with $\widetilde{w}_k = t^k w_k / \Phi(t)$ for some $t \leqslant \rho$. By Section 4, the radius of convergence of the probability generating function $\widetilde{\Phi}(z)$ of this distribution is ρ/t, cf. (4.4). Hence, the distribution $(\widetilde{w}_k)$ has some finite exponential moment if and only if $0 < t < \rho$. The cases I–III can thus 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 (7.1) 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 (7.1) 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 (π_k), which is a probability distribution with mean $\mu = 1$ and probability generating function $\sum_{k=0}^\infty \pi_k z^k$ with radius of convergence $\rho/\tau > 1$. In other words, (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.) By (7.6), the condition can also be written analytically as $\mathcal{Z}(\rho_{\mathcal{Z}}) < \rho$, a version used e.g. in [35]. (This case is called *generic* in [35] and [67].)

Ib. $\nu = 1$; then $0 < \tau = \rho < \infty$. The weight sequence (w_k) is equivalent to (π_k), which is a probability distribution with mean 1 and probability generating function $\sum_{k=0}^\infty \pi_k z^k$ with radius of convergence $\rho/\tau = 1$. In other words, (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 [33] 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, by other proofs, assuming only that we are in case I with finite variance of (π_k). (In fact, as shown in [59], it is at least sometimes possible to use complex analytic methods also in the case when $\tau = \rho$ and (π_k) has a finite second moment.) 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 second moment 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.

When $\nu \geqslant 1$, the quantity σ^2 is another natural parameter of the weight sequence (w_k), which frequently occurs in asymptotic results, see e.g. Section 21. (When $\nu < 1$, the natural analogue is ∞, see Remark 5.5.) By Theorem 7.1 (or (4.10)), $\sigma^2 = \tau\Psi'(\tau)$, so (assuming $\nu \geqslant 1$), we have case Iα when $\Psi'(\tau) < \infty$ and Iβ when $\Psi'(\tau) = \infty$. Moreover, when $\nu \geqslant 1$, then (π_k) has mean $\mu = 1$, and it follows from (4.8) that the variance σ^2 of (π_k) also is given by the formula [4]

$$\sigma^2 = \Phi''_\tau(1) + \mu - \mu^2 = \Phi''_\tau(1) = \frac{\tau^2\Phi''(\tau)}{\Phi(\tau)}. \tag{8.1}$$

Hence Iα is the case $\nu \geqslant 1$ and $\Phi''(\tau) < \infty$; equivalently, either $\nu > 1$ or $\nu = 1$ and $\Phi''(\rho) < \infty$.

Remark 8.1. 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. (Recall that by Remark 4.3, ν is the supremum of the means of the equivalent probability weight sequences.)

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.

9. The maximum degree

Theorem 7.1 studies convergence of the random tree $\mathcal{T}_n$ in the topology defined in Section 6, which really means local convergence close to the root; we have seen that the limit is of somewhat different types depending on the weight sequence, with condensation in the form of a node of infinite degree in the limit tree $\widehat{\mathcal{T}}$ in cases II and III but not in case I.

An alternative way to study the condensation (or absence of it) is to study the largest degree in the tree. This is discussed in detail in Section 19 (in the more general setting of random allocations). We give here a short summary of the main results, showing a similar picture: the maximum degree is typically rather small (logarithmic) in case I but larger (of order n) in cases II and III, which can be interpreted as a condensation; however, there are exceptions in the latter cases, and we do not have general theorems covering all possible weight sequences.

The relation between the two ways of looking at condensation is discussed in Section 20.

We denote, as in Section 19, the maximum outdegree in the tree $\mathcal{T}_n$ by $Y_{(1)}$; we use further the notation in Theorem 7.1 and Section 8.

Case Ia: $\nu > 1$. In this case $0 < \tau < \rho \leqslant \infty$, and we have a logarithmic bound due to Meir and Moon [86] (Theorem 19.3):

$$Y_{(1)} \leqslant \frac{1}{\log(\rho/\tau)} \log n + o_{\mathrm{p}}(\log n); \tag{9.1}$$

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

$$\frac{Y_{(1)}}{\log n} \xrightarrow{\mathrm{p}} \frac{1}{\log(\rho/\tau)}. \tag{9.2}$$

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 $O(1)$ (Theorem 19.16). The distribution of $Y_{(1)}$ is asymptotically given by a discretised Gumbel distribution (Theorem 19.19), 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 (Theorems 19.16 and 19.23).

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

$$d_{\mathrm{TV}}\bigl(Y_{(1)}, \xi_{(1)}\bigr) \to 0 \tag{9.3}$$

(Theorem 19.7 and Corollary 19.11). Since $\mathbb{E}\,\xi^2 < \infty$, this implies in particular

$$Y_{(1)} = o_{\mathrm{p}}(n^{1/2}). \tag{9.4}$$

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

$$Y_{(1)} = o_{\mathrm{p}}(n) \tag{9.5}$$

(Theorem 19.2), and this is (more or less) best possible (Example 19.27). However, (9.3) does not always hold in this case (Example 19.27).

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 [67] showed that

$$Y_{(1)} = (1-\nu)n + o_{\mathrm{p}}(n), \tag{9.6}$$

while the second largest node degree $Y_{(2)} = o_{\mathrm{p}}(n)$ (Theorem 19.34 and Remark 19.35). 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 (Example 19.37).

Case III: $\nu = \rho = 0$. This is similar to case II. In some regular cases we have (9.6), which now says $Y_{(1)} = n + o_{\mathrm{p}}(n)$, and then necessarily $Y_{(2)} = o_{\mathrm{p}}(n)$ (Example 19.36), but there are exceptions in other cases with an irregular weight sequence (Examples 19.38 and 19.39).

10. Examples of simply generated random trees

One of the reasons for the interest in simply generated trees is that many kinds of random trees occuring in various applications can be seen as simply generated random trees and conditioned Galton–Watson trees. We give some important examples here, see further Aldous [3, 4], Devroye [32] and Drmota [33].

We see from Theorem 7.1 and Section 8 that any simply generated random tree defined by a weight sequence with $\rho > 0$ can be defined by an equivalent probability weight sequence, and then the tree is the corresponding conditioned Galton–Watson tree. Moreover, the probability weight sequence (π_k) defined in (7.1) is the canonical choice of offspring distribution. Recall that (π_k) is characterised by having mean 1, whenever this is possible (i.e., in case I), i.e., we prefer to have critical Galton–Watson trees.

Example 10.1 (ordered trees)**.** The simplest example is to take $w_k = 1$ for every $k \geqslant 0$. Thus every tree has weight 1, and $\mathcal{T}_n$ is a uniformly random ordered rooted tree with n nodes. Further, Z_n is the number of such trees; thus Z_n is the Catalan number C_{n-1}, see Remark 2.2 and (2.1). (For this reason, these random trees are sometimes called *Catalan trees*.)

We have

$$\Phi(t) = \sum_{k=0}^{\infty} t^k = \frac{1}{1-t} \tag{10.1}$$

and

$$\Psi(t) = \frac{t\Phi'(t)}{\Phi(t)} = \frac{t}{1-t}. \tag{10.2}$$

Thus $\rho = 1$ and $\nu = \infty$ (cf. Lemma 3.1(iv)), and $\Psi(\tau) = 1$ yields $\tau = 1/2$. Hence (7.1) yields the canonical probability weight sequence

$$\pi_k = 2^{-k-1}, \qquad k \geqslant 0. \tag{10.3}$$

In other words, the uniformly random ordered rooted tree is the conditioned Galton–Watson tree with geometric offspring distribution $\xi \sim \mathrm{Ge}(1/2)$. (This is the geometric distribution with mean 1. Any other geometric distribution yields an equivalent weight sequence, and thus the same conditioned Galton–Watson tree.)

The size-biased random variable $\widehat{\xi}$ in (5.2) has the distribution

$$\mathbb{P}(\widehat{\xi}=k)=k\pi_k=k2^{-k-1}, \qquad k\geqslant 1; \tag{10.4}$$

thus $\widehat{\xi}-1$ has a negative binomial distribution $\mathrm{NBin}(2,1/2)$. It follows that in the infinite tree $\widehat{\mathcal{T}}$, if v is a node on the spine (for example the root) and $d^{\mathrm{L}}(v), d^{\mathrm{R}}(v)$ are the numbers of children of it to the left and right of the spine, respectively, then

$$\begin{aligned}\mathbb{P}\bigl(d^{\mathrm{L}}(v)=j \text{ and } d^{\mathrm{R}}(v)=k\bigr)&=\frac{1}{j+k+1}\,\mathbb{P}(\widehat{\xi}=j+k+1)=2^{-j-k-2}\\&=2^{-j-1}\cdot 2^{-k-1}, \qquad j,k\geqslant 0;\end{aligned} \tag{10.5}$$

thus $d^{\mathrm{L}}(v)$ and $d^{\mathrm{R}}(v)$ are independent and both have the same distribution $\mathrm{Ge}(1/2)$ as ξ.

We have $\sigma^2 := \operatorname{Var}\xi=\tau\Psi'(\tau)=2$, see Theorem 7.1 and (8.1), and $\mathbb{E}\,\widehat{\xi}=\sigma^2+1=3$, see (5.5).

Example 10.2 (unordered trees)**.** We have assumed that our trees are ordered, but it is possible to consider unordered labelled rooted trees too by imposing a random order on the set of children of each node. Note first that for ordered trees, the ordering of the children implicitly yields a labelling of all nodes as in Section 6. Hence, any ordered tree with n nodes can be explicitly labeled by $1,\dots,n$ in exactly $n!$ ways, and a uniformly random labelled ordered rooted tree is the same as a uniformly random unlabelled ordered rooted tree with a random labelling. (For unordered trees, a uniformly random labelled tree is different from a uniformly random unlabelled tree. We consider only labelled unordered trees here. In fact, unlabelled unordered trees are not simply generated trees; more formally, there is no weight sequence such that the corresponding simply generated random tree, with the orderings of the children of each node ignored, is a uniformly random unlabelled unordered tree.)

An unordered labelled rooted tree with outdegrees d_i corresponds to $\prod_i d_i!$ different ordered labelled rooted trees. If we take $w_k=1/k!$, we give each of these ordered trees weight $\prod_i d_i!^{-1}$, so their total weight is 1. Hence, the simply generated random tree with the weight sequence $(1/k!)$ yields, by ignoring the orderings of the children of each node, a uniformly random unordered labelled rooted tree.

In this sense, a uniformly random unordered labelled rooted tree is equivalent to a simply generated random tree with $w_k=1/k!$, and with a minor abuse of notation, we may say that a uniformly random unordered labelled rooted tree is simply generated (with $w_k=1/k!$).

The number of unordered labelled *unrooted* trees with n nodes is n^{n-2}, see e.g. [103, Section 5.3], a result given by Cayley [22] and known as Cayley's formula. (Although attributed by Cayley to Borchardt [17] and even earlier found by Sylvester [104], see e.g. [103, p. 66].) Equivalently, the number of unordered labelled rooted trees with n nodes is n^{n-1}. Hence random such trees are sometimes called *Cayley trees*. However, this name is also used for regular infinite trees.

We have, with $w_k = 1/k!$,

$$\Phi(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} = e^t \tag{10.6}$$

and

$$\Psi(t) = \frac{t\Phi'(t)}{\Phi(t)} = t. \tag{10.7}$$

Thus $\nu = \infty$ and $\Psi(\tau) = 1$ yields $\tau = 1$. Hence (7.1) yields the canonical probability weight sequence

$$\pi_k = \frac{e^{-1}}{k!}, \qquad k \geqslant 0. \tag{10.8}$$

In other words, the uniformly random labelled unordered rooted tree is (equivalent to) the conditioned Galton–Watson tree with Poisson offspring distribution $\xi \sim \mathrm{Po}(1)$. (Any other Poisson distribution yields an equivalent weight sequence, and thus the same conditioned Galton–Watson tree.)

The size-biased random variable $\widehat{\xi}$ in (5.2) has the distribution

$$\mathbb{P}(\widehat{\xi} = k) = k\pi_k = \frac{e^{-1}}{(k-1)!}, \qquad k \geqslant 1; \tag{10.9}$$

thus $\widehat{\xi} - 1$ has also the Poisson distribution $\mathrm{Po}(1)$, i.e., $\widehat{\xi} - 1 \overset{\mathrm{d}}{=} \xi$. (It is only for a Poisson distribution that $\widehat{\xi} - 1 \overset{\mathrm{d}}{=} \xi$.)

We have $\sigma^2 := \operatorname{Var} \xi = \tau\Psi'(\tau) = 1$ and $\mathbb{E}\,\widehat{\xi} = \sigma^2 + 1 = 2$, cf. (8.1) and (5.5).

The partition function is given by

$$Z_n(\boldsymbol{\pi}) = \mathbb{P}(|\mathcal{T}| = n) = \frac{n^{n-1}e^{-n}}{n!}. \tag{10.10}$$

This is a special case of the *Borel distribution* in (12.29) below; Borel [18] proved a result equivalent to (10.10) for a queueing problem, see also Otter [93], Tanner [107], Dwass [36], Takács [106], Pitman [99], Example 12.6 and Theorem 15.5 below. Equivalently, using (4.3),

$$Z_n(\mathbf{w}) = e^n Z_n(\boldsymbol{\pi}) = \frac{n^{n-1}}{n!}. \tag{10.11}$$

Recall that Z_n is defined by the sum (2.5) over unlabelled ordered rooted trees; if we sum over *labelled* ordered rooted trees, we obtain $n!\,Z_n$, which by the argument above corresponds to weight 1 on each labelled unordered rooted tree; i.e., the number of labelled unordered rooted trees is $n!\,Z_n(\mathbf{w}) = n^{n-1}$. Thus (10.11) is equivalent to Cayley's formula for the number of unordered trees given above.

By (10.11), the generating function $\mathcal{Z}(z)$ is $\sum_{n=1}^{\infty} n^{n-1}z^n/n!$, known as the the *tree function*; see (12.22)–(12.25) in Example 12.6.

Example 10.3 (binary trees I)**.** The name *binary tree* is used in (at least) two different, but related, meanings. The first version (Drmota [33, Section 1.2.1]), sometimes called *full binary tree* or *strict binary tree*, is an ordered rooted tree where every node has outdegree 0 or 2. We obtain a uniformly random full binary tree by taking the weight sequence with $w_0 = w_2 = 1$, and $w_k = 0$ for $k \neq 0, 2$. Note that this weight sequence has span 2; this is the standard example of a weight sequence with span > 1. As a consequence, a full binary tree of size n exists only if n is odd. (This is easily seen directly; see Corollary 15.6 for a general result.)

We have

$$\Phi(t) = 1 + t^2 \tag{10.12}$$

and

$$\Psi(t) = \frac{t\Phi'(t)}{\Phi(t)} = \frac{2t^2}{1+t^2}. \tag{10.13}$$

Thus $\rho = \infty$, $\nu = 2$ (cf. Lemma 3.1(v)), and $\Psi(\tau) = 1$ yields $\tau = 1$. Hence (7.1) yields the canonical probability weight sequence

$$\pi_k = \tfrac{1}{2}, \qquad k = 0, 2. \tag{10.14}$$

In other words, the random full binary tree is the conditioned Galton–Watson tree with offspring distribution $\xi = 2X$ where $X \sim \mathrm{Be}(1/2)$. (In the Galton–Watson tree $\mathcal{T}$, thus each node gets either twins or no children, each outcome with probability 1/2.)

The size-biased random variable $\widehat{\xi}$ has $\mathbb{P}(\widehat{\xi} = 2) = 1$ by (10.14) and (5.2), so $\widehat{\xi} = 2$ and $\widehat{\xi} - 1 = 1$ a.s.

We have $\sigma^2 := \operatorname{Var}\xi = 1$ and $\mathbb{E}\,\widehat{\xi} = \sigma^2 + 1 = 2$, cf. (8.1) and (5.5).

Example 10.4 (binary trees II)**.** The second version of a *binary tree* (Drmota [33, Example 1.3]) is a rooted tree where every node has at most one *left child* and at most one *right child*. Thus, each outdegree is 0, 1 or 2; if there are two children they are ordered, and, moreover, if there is only one child, it is marked as either left or right. (There is a one-to-one correspondence between binary trees of this type with n nodes and the full binary trees in Example 10.3 with $2n+1$ nodes, mapping a binary tree T to a full binary tree T', where T' is obtained from T by adding $2-d$ *external nodes* at every node with outdegree d; conversely, we obtain T by deleting all leaves in T' and keeping only the nodes that have outdegree 2 in T' (the *internal nodes*).)

There are thus two types of nodes with outdegree 1, but only one type each of nodes with outdegrees 0 or 2. If we ignore the type of child at each node with only one child, we obtain an ordered tree with all outdegrees $\leqslant 2$, and each such tree with n_1 nodes of degree 1 corresponds to 2^{n_1} binary trees. This number equals the weight given by the weight sequence $w_0 = 1$, $w_1 = 2$, $w_2 = 1$, and $w_k = 0$ for $k \geqslant 3$, i.e., $w_k = \binom{2}{k}$. Hence, a simply generated random tree with this weight sequence has the same distribution as the ordered tree obtained from a uniformly distributed random binary tree; conversely, we may obtain a uniformly distributed random binary tree by taking a simply generated random tree with $w_k = \binom{2}{k}$ and randomly labelling each single child as left or right.

In this sense, we may say that a uniformly distributed random binary tree is (equivalent to) a simply generated random tree with $w_k = \binom{2}{k}$.

The choice $w_k = \binom{2}{k}$ yields

$$\Phi(t) = 1 + 2t + t^2 = (1+t)^2 \tag{10.15}$$

and

$$\Psi(t) = \frac{t\Phi'(t)}{\Phi(t)} = \frac{2t}{1+t}. \tag{10.16}$$

Thus $\rho = \infty$, $\nu = 2$, and $\Psi(\tau) = 1$ yields $\tau = 1$. Hence (7.1) yields the canonical probability weight sequence

$$\pi_k = \frac{1}{4}\binom{2}{k}, \qquad k \geqslant 0. \tag{10.17}$$

In other words, a uniformly random binary tree of this type is (equivalent to) the conditioned Galton–Watson tree with binomial offspring distribution $\xi \sim \mathrm{Bi}(2, 1/2)$. (Any other distribution $\mathrm{Bi}(2, p)$, $0 < p < 1$, is equivalent and yields the same conditioned Galton–Watson tree.)

The size-biased random variable $\widehat{\xi}$ has by (5.2) $\mathbb{P}(\widehat{\xi} = 1) = \mathbb{P}(\widehat{\xi} = 2) = \frac{1}{2}$; thus $\widehat{\xi} - 1 \sim \mathrm{Bi}(1, 1/2)$.

We have $\sigma^2 := \operatorname{Var} \xi = 1/2$ and $\mathbb{E}\,\widehat{\xi} = \sigma^2 + 1 = 3/2$, cf. (8.1) and (5.5).

Example 10.5 (Motzkin trees)**.** A *Motzkin tree* is a ordered rooted tree with each outdegree $\leqslant 2$. The difference from Example 10.4 is that there is only one type of a single child. Thus we count such trees and obtain uniformly random Motzkin trees by taking $w_0 = w_1 = w_2 = 1$ and $w_k = 0$, $k \geqslant 3$. (We thus have the same set of trees as in Example 10.4, but different probability distributions on it.)

We have

$$\Phi(t) = 1 + t + t^2 \tag{10.18}$$

and

$$\Psi(t) = \frac{1+2t}{1+t+t^2}. \tag{10.19}$$

Thus $\rho = \infty$, $\nu = 2$, and $\Psi(\tau) = 1$ yields $\tau = 1$. Hence (7.1) yields the canonical probability weight sequence

$$\pi_k = \tfrac{1}{3}, \qquad k = 0, 1, 2. \tag{10.20}$$

In other words, a uniformly random Motzkin tree is the conditioned Galton–Watson tree with offspring distribution ξ uniform on $\{0, 1, 2\}$.

The size-biased random variable $\widehat{\xi}$ has, by (5.2) and (10.20), the distribution $\mathbb{P}(\widehat{\xi}=1)=\frac{1}{3}$, $\mathbb{P}(\widehat{\xi}=2)=\frac{2}{3}$; thus $\widehat{\xi}-1\sim \operatorname{Bi}(1,2/3)$.

We have $\sigma^2 := \operatorname{Var}\xi = 2/3$ and $\mathbb{E}\,\widehat{\xi}=\sigma^2+1=5/3$, cf. (8.1) and (5.5).

Example 10.6 (d-ary trees)**.** In a *d-ary* tree, each node has d positions where a child may be attached, and there is at most one child per position. (Trees with children attached at different positions are regarded as different trees.) This generalises the binary trees in Example 10.4, which is the special case $d=2$.

Since k children may be attached in $\binom{d}{k}$ ways (with a given order), the argument in Example 10.4 shows that a uniformly random d-ary tree is equivalent to a simply generated random tree with $w_k=\binom{d}{k}$. We have

$$\Phi(t)=(1+t)^d \tag{10.21}$$

and

$$\Psi(t)=\frac{t\Phi'(t)}{\Phi(t)}=\frac{dt}{1+t}. \tag{10.22}$$

Thus $\rho=\infty$, $\nu=\omega=d$, and $\Psi(\tau)=1$ yields $\tau=1/(d-1)$. Hence (7.1) yields the canonical probability weight sequence

$$\pi_k=\binom{d}{k}(d-1)^{d-k}d^{-d}=\binom{d}{k}\Bigl(\frac{1}{d}\Bigr)^k\Bigl(\frac{d-1}{d}\Bigr)^{d-k},\qquad k\geqslant 0. \tag{10.23}$$

In other words, a uniformly random d-ary tree is (equivalent to) the conditioned Galton–Watson tree with binomial offspring distribution $\xi\sim\operatorname{Bi}(d,1/d)$. (Any other distribution $\operatorname{Bi}(d,p)$, $0<p<1$, is equivalent and yields the same conditioned Galton–Watson tree.)

The size-biased random variable $\widehat{\xi}$ has the distribution

$$\mathbb{P}(\widehat{\xi}=k)=k\pi_k=\binom{d-1}{k-1}\Bigl(\frac{1}{d}\Bigr)^{k-1}\Bigl(\frac{d-1}{d}\Bigr)^{d-k},\qquad k\geqslant 1; \tag{10.24}$$

thus $\widehat{\xi}-1$ has the Binomial distribution $\operatorname{Bi}(d-1,1/d)$.

We have $\sigma^2 := \operatorname{Var}\xi = 1-1/d$ and $\mathbb{E}\,\widehat{\xi}=\sigma^2+1=2-1/d$, cf. (8.1) and (5.5).

Example 10.7. Let β be a real constant and let $w_k=(k+1)^{-\beta}$. (The case $\beta=0$ is Example 10.1.) Then $\rho=1$.

If $-\infty<\beta\leqslant 1$, then $\Phi(\rho)=\infty$, so $\nu=\infty$ by (3.10) and Lemma 3.1(iv).

If $\beta>1$, then $\Phi(\rho)=\zeta(\beta)<\infty$ and

$$\nu=\Psi(1)=\frac{\sum_k kw_k}{\Phi(1)}=\frac{\zeta(\beta-1)-\zeta(\beta)}{\zeta(\beta)},\qquad \beta>2, \tag{10.25}$$

while $\nu=\Psi(1)=\infty$ if $\beta\leqslant 2$. Hence, see also Bialas and Burda [13],

$$\nu=1\iff\zeta(\beta-1)=2\zeta(\beta)\iff\beta=\beta_0=2.47875\dots \tag{10.26}$$

and $\nu>1\iff-\infty<\beta<\beta_0$. (It can be shown that ν is a decreasing function of β for $\beta>2$.) In the case $\beta=\beta_0$, when thus $\nu=1$, we further have $\sigma^2 = \infty$ by (8.1), since $\Phi''(1) = \infty$ when $\beta \leqslant 3$. This is thus case Iβ, in the notation of Section 8.

In the case $\beta > \beta_0$ we thus have $0 < \nu < 1$, and $\mathcal{T}_n$ converges to a random tree $\widehat{\mathcal{T}}$ with one node of infinite degree, see Theorem 7.1 and Section 5. If $\beta \leqslant \beta_0$, then $\nu \geqslant 1$ and the limit tree $\widehat{\mathcal{T}}$ is locally finite. We thus see a phase transition at $\beta = \beta_0$ when we vary β in this example.

Note, however, that there is nothing special with the rate of decrease $k^{-\beta_0}$; the value of β_0 depends on the exact form of our choice of the weights w_k in this example, and reflects the values for small k rather than the asymptotic behaviour. For example, as remarked by Bialas and Burda [13], just changing w_0 would change β_0 to any desired value in $(2, \infty)$. With a different w_0, $\Phi(1) = \zeta(\beta) - 1 + w_0$, and a modification of (10.25) shows that the critical value β_0 yielding $\nu = 1$ is given by, see [13],

$$2\zeta(\beta_0) - \zeta(\beta_0 - 1) = 1 - w_0. \tag{10.27}$$

In particular, $\beta_0 > 3$ for $w_0 < 1 + \zeta(2) - 2\zeta(3) = 0.24082\dots$; in this case, for the critical $\beta = \beta_0$, we then have $\nu = 1$ and $\sigma^2 < \infty$, see (8.1).

See [13] for some further analytic properties. For example, if $\beta_0 < 3$ (for example when $w_0 = 1$), then, as $\beta \nearrow \beta_0$, we have $1 - \tau \sim c(\beta_0 - \beta)^{1/(\beta_0 - 2)}$, where $c > 0$ and the exponent can take any value > 1.

Example 10.8. Take $w_k = k!$. The generating function $\Phi(t) = \sum_{k=0}^{\infty} k!\, t^k$ has radius of convergence $\rho = 0$ so we are in case III, and there exists no equivalent conditioned Galton–Watson tree.

Theorem 7.1 shows that $\mathcal{T}_n$ converges to an infinite star, see Remark 7.6 and Example 5.1. This means that the root degree converges in probability to ∞, and that the outdegree of any fixed child converges to 0 in probability, i.e., equals 0 w.h.p. Note, however, that we cannot draw the conclusion that the outdegrees of *all* children of the root are 0 w.h.p.; Theorem 7.1 and symmetry imply that the proportion of children of the root with outdegree > 0 tends to 0, but the number of such children may still be large. (Theorem 7.11(ii) yields the same conclusion.)

In fact, for this particular example $w_k = k!$, it is shown by Janson, Jonsson and Stefánsson [64], using direct calculations, that w.h.p. all subtrees attached to the root have size 1 or 2, and that the number of such subtrees of size 2 has an asymptotic Poisson distribution Po(1). (This number thus w.h.p. equals N_1, and $l_2(\mathcal{T}_n)$, and also the number of children of the root with at least one child.)

Example 10.9. If we instead take $w_k = k!^{\alpha}$ with $0 < \alpha < 1$, then as in Example 10.8, $\rho = 0$ and $\mathcal{T}_n$ converges to the infinite star in Example 5.1. In this case, if (for simplicity) $1/\alpha \notin \mathbb{N}_1$, then $N_i(\mathcal{T}_n)/n^{1-i\alpha} \xrightarrow{\text{p}} i!^{\alpha}$ for $1 \leqslant i \leqslant \lfloor 1/\alpha \rfloor$, while $N_i = 0$ w.h.p. for each fixed $i > \lfloor 1/\alpha \rfloor$; furthermore, among the subtrees attached to the root, w.h.p. there are subtrees of all sizes $\leqslant \lfloor 1/\alpha \rfloor + 1$, and all possible shapes of these trees, with the number of each type tending to ∞ in probability, but no larger subtrees. See Janson, Jonsson and Stefánsson [64] for details.

If we take $w_k = k!^\alpha$ with $\alpha > 1$, then w.h.p. $\mathcal{T}_n$ is a star with $n-1$ leaves, so $N_d = 0$ for $1 \leqslant d < n-1$.

See also the examples in Section 12.

11. Balls-in-boxes

The *balls-in-boxes* model is a model for random allocation of m (unlabelled) balls in n (labelled) boxes; here $m \geqslant 0$ and $n \geqslant 1$ are given integers. The set of possible allocations is thus

$$\mathcal{B}_{m,n} := \Big\{(y_1, \dots, y_n) \in \mathbb{N}_0^n : \sum_{i=1}^n y_i = m\Big\}, \tag{11.1}$$

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

$$w(\mathbf{y}) := \prod_{i=1}^n w_{y_i}. \tag{11.2}$$

Given m and n, we choose a random allocation $B_{m,n}$ with probability proportional to its weight, i.e.,

$$\mathbb{P}(B_{m,n} = \mathbf{y}) = \frac{w(\mathbf{y})}{Z(m,n)}, \qquad \mathbf{y} \in \mathcal{B}_{m,n}, \tag{11.3}$$

where the normalizing factor $Z(m,n)$, again called the *partition function*, is given by

$$Z(m,n) = Z(m,n;\mathbf{w}) := \sum_{\mathbf{y} \in \mathcal{B}_{m,n}} w(\mathbf{y}). \tag{11.4}$$

We consider only m and n such that $Z(m,n) > 0$; otherwise $B_{m,n}$ is undefined. See further Lemma 13.3. We write $B_{m,n} = (Y_1^{(m,n)}, \dots, Y_n^{(m,n)})$, which we usually simplify to $(Y_1, \dots, Y_n)$, omitting the superscripts.

Remark 11.1. The names balls-in-boxes and balls-in-bins are used in the literature for several different allocation models. We use balls-in-boxes for the model defined here, following e.g. Bialas, Burda and Johnston [14].

Example 11.2 (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 $w(\mathbf{y}) = \mathbb{P}\big((\xi_1, \dots, \xi_n) = \mathbf{y}\big)$ for any $\mathbf{y} = (y_1, \dots, y_n)$. Hence

$$Z(m,n) = \mathbb{P}\big((\xi_1, \dots, \xi_n) \in \mathcal{B}_{m,n}\big) = \mathbb{P}(S_n = m), \tag{11.5}$$

where we define

$$S_n := \sum_{i=1}^n \xi_i. \tag{11.6}$$

Moreover, $B_{m,n}$ has the same distribution as $(\xi_1, \dots, \xi_n)$ conditioned on $S_n = m$:

$$(Y_1^{(m,n)}, \dots, Y_n^{(m,n)}) \overset{\mathrm{d}}{=} \big((\xi_1, \dots, \xi_n) \mid S_n = m\big). \tag{11.7}$$

We will use this setting (and notation) several times below. (This construction of a random allocation $B_{m,n}$ is used by Kolchin [76] and there called the *general scheme of allocation*.)

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 in Section 4.

Lemma 11.3. *Suppose that we replace the weights (w_k) by equivalent weights $(\widetilde{w}_k)$ where $\widetilde{w}_k := ab^k w_k$ with $a, b > 0$ as in (4.1). Then the weight of an allocation $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{B}_{m,n}$ is changed to*

$$\widetilde{w}(\mathbf{y}) = a^n b^m w(\mathbf{y}), \tag{11.8}$$

and the partition function $Z(m, n) = Z(m, n; \mathbf{w})$ is changed to

$$\widetilde{Z}(m, n) := Z(m, n; \widetilde{\mathbf{w}}) = a^n b^m Z(m, n), \tag{11.9}$$

while the distribution of $B_{m,n}$ is invariant. Thus $B_{m,n}$ depends only on the equivalence class of the weight sequence.

Proof. We have, by the definition (11.2),

$$\widetilde{w}(\mathbf{y}) = \prod_{i=1}^n \widetilde{w}_{y_i} = \prod_{i=1}^n ab^{y_i} w_{y_i} = a^n b^{\sum_{i=1}^n y_i} \prod_{i=1}^n w_{y_i} = a^n b^m w(\mathbf{y}), \tag{11.10}$$

which shows (11.8), and (11.9) follows by (11.4). Consequently, for every $\mathbf{y} \in \mathcal{B}_{m,n}$, we have $\widetilde{w}(\mathbf{y})/\widetilde{Z}(m, n) = w(\mathbf{y})/Z(m, n)$ so the probability $\mathbb{P}(B_{m,n} = \mathbf{y})$ in (11.3) is unchanged, which completes the proof. □

Our aim is to describe the asymptotic distribution of the random allocation $B_{m,n}$ as $m, n \to \infty$; we consider the case when $m/n \to \lambda$ for some real λ, and assume for simplicity that $0 \leqslant \lambda < \omega = \omega(\mathbf{w})$. (Cases with $m/n \to \infty$ are interesting too in some applications, for example in Section 19.7, but will not be considered here. See e.g. Kolchin, Sevast'yanov and Chistyakov [77], Kolchin [76] and Pavlov [96] for such results in special cases.) The first step is to note that the distribution of $B_{m,n} = (Y_1, \dots, Y_n)$ is *exchangeable*, i.e., invariant under any permutation of $Y_1, \dots, Y_n$. Hence, the distribution is completely described by the (joint) distribution of the numbers of boxes with a certain number of balls, so it suffices to study these numbers.

For any allocation of balls $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{N}_0^n$, and $k \geqslant 0$, let

$$N_k(\mathbf{y}) := |\{i : y_i = k\}|, \tag{11.11}$$

the number of boxes with exactly k balls. Thus, if $\mathbf{y} \in \mathcal{B}_{m,n}$, then

$$\sum_{k=0}^\infty N_k(\mathbf{y}) = n \qquad \text{and} \qquad \sum_{k=0}^\infty k N_k(\mathbf{y}) = m. \tag{11.12}$$

We thus want to find the asymptotic distribution of the random variables $N_k(B_{m,n})$, $k = 0, 1, \dots$. Our main result is the following, which will be proved in Section 14 together with the other theorems in this section.

Theorem 11.4. *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

$$\pi_k := \frac{w_k \tau^k}{\Phi(\tau)}, \qquad k \geqslant 0. \tag{11.13}$$

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

$$\mu = \Psi(\tau) = \min(\lambda, \nu) \tag{11.14}$$

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

$$N_k(B_{m,n})/n \xrightarrow{\mathrm{p}} \pi_k. \tag{11.15}$$

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 7.1 are the same as in Theorem 11.4 with $\lambda = 1$. Indeed, we will later see that the random trees correspond to $m = n - 1$ and thus $\lambda = 1$.

Remark 11.5. The argument in Remark 7.4 extends and shows that τ is the (unique) minimum point in $[0, \rho]$ of $\Phi(t)/t^\lambda$; i.e.,

$$\frac{\Phi(\tau)}{\tau^\lambda} = \inf_{0\leqslant t\leqslant\rho} \frac{\Phi(t)}{t^\lambda} = \inf_{0\leqslant t<\infty} \frac{\Phi(t)}{t^\lambda}. \tag{11.16}$$

By (11.15), there are roughly $n\pi_k$ boxes with k balls. 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 $\lambda > \nu$, (11.14) 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 (11.15) for fixed k. In physical terminology, this can be regarded as condensation of part of the mass (= balls). We study this further in Section 19.6. The simplest case is that there is a single giant box with $\approx (\lambda - \nu)n$ balls. We shall see that this happens in an important case (Theorem 19.34; see also Bialas, Burda and Johnston [14, Fig. 1] for some numerical examples), but that there are also other possibilities (Examples 19.37–19.39).

Recall that for simply generated random trees, which as said above correspond to balls-in-boxes with $\lambda = 1$, Theorem 7.1 too shows that there is a condensation when $\nu < \lambda = 1$ (since then $\mu < 1$ by (7.2)); in this case the condensation appears as a node of infinite degree in the random limit tree $\widehat{\mathcal{T}}$ of type (T2), see Section 5. We shall in Section 20 study the relation between the forms of the condensation shown in Theorems 7.1 and 11.4.

We further have the following, essentially equivalent, version of Theorem 11.4, where we assume only that m/n is bounded, but not necessarily convergent.

Theorem 11.6. *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* $m/n \leqslant C$ *for some* $C < \omega$.

Define the function $\tau : [0, \infty) \to [0, \infty]$ *by* $\tau(x) := \sup\{t \leqslant \rho : \Psi(t) \leqslant x\}$. *Then* $\tau(x)$ *is the unique number in* $[0, \rho]$ *such that* $\Psi(\tau(x)) = x$ *when* $x \leqslant \nu$, *and* $\tau(x) = \rho$ *when* $x \geqslant \nu$; *furthermore, the function* $x \mapsto \tau(x)$ *is continuous. We have* $0 \leqslant \tau(m/n) < \infty$ *and* $0 < \Phi(\tau(m/n)) < \infty$, *and for every* $k \geqslant 0$,

$$\frac{N_k(B_{m,n})}{n} - \frac{w_k(\tau(m/n))^k}{\Phi(\tau(m/n))} \xrightarrow{\mathrm{p}} 0. \tag{11.17}$$

Furthermore, for any $C < \omega$, *this holds uniformly as* $n \to \infty$ *for all* $m = m(n)$ *with* $m/n \leqslant C$.

Returning to the random variables $Y_1, \dots, Y_n$, we have the following result, which is shown by a physicists' proof by Bialas, Burda and Johnston [14].

Theorem 11.7. *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* $m/n \to \lambda$ *where* $0 \leqslant \lambda < \omega$, *and let* $(\pi_k)_{k\geqslant 0}$ *be as in Theorem 11.4. Then, for every* $\ell \geqslant 1$ *and* $y_1, \dots, y_\ell \geqslant 0$,

$$\mathbb{P}(Y_1^{(m,n)} = y_1, \dots, Y_\ell^{(m,n)} = y_\ell) \to \prod_{i=1}^{\ell} \pi_{y_i}. \tag{11.18}$$

In other words, for every fixed ℓ, *the random variables* $Y_1, \dots, Y_\ell$ *converge jointly to independent random variables with the distribution* $(\pi_k)_{k\geqslant 0}$.

A more fancy way of describing the same result is that the sequence $Y_1, \dots, Y_n$, arbitrarily extended to infinite length, converges in distribution, as an element of $\mathbb{N}_0^\infty$, to a sequence of i.i.d. random variables with the distribution $(\pi_k)_{k\geqslant 0}$. (See e.g. [15, Problem 3.7].)

Remark 11.8. We have assumed $w_0 > 0$ in the results above for convenience, and because this condition is necessary when discussing simply generated trees, which is our main topic. The balls-in-boxes model makes sense also when $w_0 = 0$, but this case is easily reduced to the case $w_0 > 0$: Let $\alpha := \min\{k : w_k > 0\}$. If $\alpha > 0$, then this means that each box has to have at least α balls. (In particular, we need $m \geqslant \alpha n$.) There is an obvious correspondence between such allocations in $\mathcal{B}_{m,n}$ and allocations in $\mathcal{B}_{m-\alpha n,n}$ obtained by removing α balls from each box. Formally, if $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{B}_{m,n}$ let $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \dots, \widetilde{y}_n)$ with $\widetilde{y}_i := y_i - \alpha$, and note that if we shift the weight sequence to $\widetilde{w}_k := w_{k+\alpha}$, then $\widetilde{w}(\widetilde{\mathbf{y}}) = w(\mathbf{y})$; thus $B_{m,n}$ has the same distribution as $B_{m-\alpha n,n}$ for $\widetilde{\mathbf{w}}$, with α extra balls added in each box. It follows easily that the results above hold also in the case $w_0 = 0$. (We interpret $w_k \tau^k / \Phi(\tau)$ for $\tau = 0$ as the appropriate limit value. Note also that it is essential to use (3.2) and not (3.3) when $w_0 = 0$.)

Remark 11.9. Similarly, we can always reduce to the case $\operatorname{span}(\mathbf{w}) = 1$: If $\operatorname{span}(\mathbf{w}) = d$, then the number of balls in each box has to be a multiple of d, so we may instead consider an allocation of m/d "superballs", each consisting of d balls. This means replacing each Y_i by Y_i/d and using the weight sequence (w_{dk}). We prefer, however, to allow a general span in our theorems, for ease of our applications to simply generated trees where the corresponding reduction is more complicated. (For trees, we may replace each branch by a d-fold branch. In the probability weight sequence case with Galton–Watson trees, this replaces the random variable ξ by $(\xi_1 + \dots + \xi_d)/d$, with $\xi_i \overset{\mathrm{d}}{=} \xi$ i.i.d., but the roots gets a different offspring distribution ξ/d; more generally, for a general weight sequence $\mathbf{w}$, we replace $\Phi(t)$ by $\Phi(t^{1/d})^d$, except at the root where we use different weights with the generating function $\Phi(t^{1/d})$. We will not use this and leave the details to the reader.)

Remark 11.10. We have assumed $m/n \to \lambda < \omega$ in Theorems 11.4 and 11.7, and similarly $m/n \leqslant C < \omega$ in Theorem 11.6; hence, for n large at least, $m/n < \omega$. In fact, $m/n \leqslant \omega$ is trivially necessary, see Lemma 13.3. When $\omega < \infty$, the only remaining case (assuming m/n converges) is thus $m/n \to \omega$ with $m/n \leqslant \omega$; in this case, it is easy to see that (11.15) and (11.18) hold with $\pi_\omega = 1$ and $\pi_k = 0$, $k \neq \omega$. (This can be seen as a limiting case of (11.13) with $\tau = \infty$.)

In fact, if $\omega < \infty$, so the boxes have a finite maximum capacity ω, then the complementation $y_i \mapsto \omega - y_i$ yields a bijection of $\mathcal{B}_{m,n}$ onto $\mathcal{B}_{\omega n - m,n}$, which preserves weights if (w_k) simultaneously is reflected to $\widetilde{\mathbf{w}} := (w_{\omega - k})$. Hence, $B_{m,n}$ corresponds to $B_{\omega n - m,n}$ (for $\widetilde{\mathbf{w}}$), and results for $m/n \to \omega < \infty$ follow from results for $m/n \to 0$.

As said above, we do not consider the case $\omega = \infty$ and $m/n \to \infty$, when the average occupancy tends to infinity.

12. Examples of balls-in-boxes

Apart from the connection with simply generated trees, see Section 15, the balls-in-boxes model is interesting in its own right.

We begin with three classic examples of balls-in-boxes, see e.g. Feller [38, II.5] and Kolchin [76], followed by further examples from probability theory, combinatorics and statistical physics, including several examples of random forests. (We return to these examples of random forests in Section 19.7, where we study the size of the largest tree in them.)

Example 12.1 (Maxwell–Boltzmann statistics; multinomial distribution)**.** Consider a uniform random allocation of m *labelled* balls in n boxes. This is the same as throwing m balls into n boxes at random, independently and with each ball uniformly distributed. (In statistical mechanics, this is known as the *Maxwell–Boltzmann statistics.*) It is elementary that the resulting random allocation $(Y_1, \dots, Y_n)$ has a multinomial distribution

$$\mathbb{P}\bigl((Y_1, \dots, Y_n) = (y_1, \dots, y_n)\bigr) = n^{-m}\binom{m}{y_1, \dots, y_n} = m!\, n^{-m}\prod_{i=1}^{n}\frac{1}{y_i!}. \tag{12.1}$$

If we take $w_k = 1/k!$, we see that the probabilities in (12.1) and (11.3) are proportional, and thus must be identical, so the weight sequence $(1/k!)$ yields the uniform random allocation of labelled balls. We see also that then

$$Z(m, n) = n^m/m!. \tag{12.2}$$

Alternatively, we may take a Poisson distribution $\mathrm{Po}(a)$: $w_k = a^k e^{-a}/k!$; this is an equivalent weight sequence for any $a > 0$. We see directly that then $S_n \sim \mathrm{Po}(na)$ so (11.5) yields

$$Z(m, n) = (na)^m e^{-na}/m!; \tag{12.3}$$

hence we see again that (11.3) and (12.1) agree.

Comparing with Example 10.2, and using Lemma 17.1 below, we see that the multiset of degrees in a random unordered labelled tree of size n has exactly the distribution obtained when throwing $n-1$ balls into n boxes at random.

With $w_k = 1/k!$ we have, as in Example 10.2, (10.6)–(10.7) and $\rho = \omega = \nu = \infty$. Hence, if $m/n \to \lambda$, we have $\tau = \lambda$ and thus $\pi_k = \lambda^k e^{-\lambda}/k!$, so (π_k) is the $\mathrm{Po}(\lambda)$ distribution, which thus is the canonical choice of weights. (In the asymptotic case; for given m and n one might choose $\mathrm{Po}(m/n)$, cf. (11.17).)

Theorem 11.7 (or (11.15)) shows that if $m/n \to \lambda < \infty$, then the asymptotic distribution of the numbers of balls in a given urn is $\mathrm{Po}(\lambda)$.

The idea to study the multinomial distribution as a vector of i.i.d. Poisson variables conditioned on the sum is an old one that has been used repeatedly, see e.g. Kolchin, Sevast'yanov and Chistyakov [77], Holst [52, 53], Kolchin [76], Janson [55].

Example 12.2 (Bose–Einstein statistics)**.** The weight sequence $w_k = 1$ yields a uniform distribution over all allocations of m identical and indistinguishable balls in n boxes; thus each allocation $(Y_1, \dots, Y_n) \in \mathcal{B}_{m,n}$ has the same probability $1/|\mathcal{B}_{m,n}| = 1/\binom{n+m-1}{m}$.

This is known as *Bose–Einstein statistics* in statistical quantum mechanics; it is the distribution followed by *bosons.* (In the simple case with no forces acting on them.)

Comparing with Example 10.1, and using Lemma 17.1 below, we see that the multiset of degrees in a random ordered tree of size n has exactly the distribution obtained by a uniform random allocation of $n-1$ balls into n boxes.

As in Example 10.1 we have (10.1)–(10.2) and $\rho = 1$, $\nu = \infty$. If $m/n \to \lambda < \infty$, then the equation $\Psi(\tau) = \lambda$ is, by (10.2), $\tau/(1-\tau) = \lambda$, and thus

$$\tau = \frac{\lambda}{1+\lambda}. \tag{12.4}$$

Any geometric distribution $\mathrm{Ge}(p)$ with $0 < p < 1$ is a weight sequence equivalent to (w_k), and (12.4) shows that the canonical choice (7.1) is, using (10.1),

$$\pi_k = (1-\tau)\tau^k = \frac{\lambda^k}{(\lambda+1)^{k+1}}, \tag{12.5}$$

which is the distribution $\mathrm{Ge}(1-\tau) = \mathrm{Ge}(1/(\lambda+1))$. By Theorem 11.7, this is also the asymptotic distribution of balls in a given urn.

See also Holst [52, 53] and Kolchin [76].

Example 12.3 (Fermi–Dirac statistics)**.** The other type of particles in statistical quantum mechanics is *fermions*; they exclude each other (the *Pauli exclusion principle*) so all allocations of them have to satisfy $Y_i \leqslant 1$, i.e., $Y_i \in \{0, 1\}$. A random allocation uniform among all such possibilities is known as *Fermi–Dirac statistics*; this is thus equivalent to a uniform random choice of one of the $\binom{n}{m}$ subsets of m boxes.

We obtain this distribution by the choice $w_0 = w_1 = 1$ and $w_k = 0$ for $k \geqslant 2$; thus

$$\Phi(t) = 1 + t \tag{12.6}$$

and

$$\Psi(t) = \frac{t}{1+t}. \tag{12.7}$$

We have $\rho = \infty$ and $\nu = \omega = 1$. (Formally, (12.6) is the case $d = 1$ of (10.21), but note that we assume $d \geqslant 2$ in Example 10.6.)

If $m/n \to \lambda < 1$, we thus have a rather trivial example of the general theory with $\tau/(1+\tau) = \lambda$ and thus

$$\tau = \frac{\lambda}{1-\lambda}, \tag{12.8}$$

and $(\pi_k) = (1-\lambda, \lambda, 0, 0, \dots)$, i.e., the Bernoulli distribution $\mathrm{Be}(\lambda)$. (Any Bernoulli distribution $\mathrm{Be}(p)$ with $0 < p < 1$ is equivalent.)

Since $\omega = 1$, the corresponding conditioned Galton–Watson tree is trivially the deterministic path P_n, a case which we have excluded above.

Example 12.4 (Pólya urn [53])**.** Consider a multicolour Pólya urn containing balls of n different colours, see Eggenberger and Pólya [37]. Initially, the urn contains $a > 0$ balls of each colour. Balls are drawn at random, one at a time. After each drawing, the drawn ball is replaced together with $b > 0$ additional balls of the same colour. (It is natural to take a and b to be integers, but the model is easily interpreted also for arbitrary real $a, b > 0$, see e.g. [58].)

Make m draws, and let Y_i be the number of times that a ball of colour i is drawn; then $(Y_1, \dots, Y_n)$ is a random allocation in $\mathcal{B}_{m,n}$.

A straightforward calculation, see [37], [66], [53], shows that

$$\begin{aligned}\mathbb{P}\big((Y_1, \dots, Y_n) = (y_1, \dots, y_n)\big)\\ &= \binom{m}{y_1, \dots, y_n} \frac{\prod_{i=1}^n a(a+b)\cdots(a+(y_i-1)b)}{na(na+b)\cdots(na+(m-1)b)}\\ &= \frac{\prod_{i=1}^n \binom{a/b+y_i-1}{y_i}}{\binom{na/b+m-1}{m}}.\end{aligned} \tag{12.9}$$

Hence, as noted by Holst [53], this equals the random allocation given by the weights

$$w_k = \binom{a/b+k-1}{k} = (-1)^k \binom{-a/b}{k}, \qquad k = 0, 1, \dots. \tag{12.10}$$

Note that the case $a = b$ yields $w_k = 1$ and the uniform random allocation in Example 12.2 (Bose–Einstein statistics). We have

$$\Phi(t) = \sum_{k=0}^{\infty} \binom{a/b+k-1}{k} t^k = (1-t)^{-a/b}, \tag{12.11}$$

with radius of convergence $\rho = 1$, and thus

$$\Psi(t) = \frac{a}{b} \cdot \frac{t}{1-t}. \tag{12.12}$$

Hence, $\nu = \Psi(1) = \infty$, and for any $\lambda \in [0, \infty)$,

$$\tau = \frac{b\lambda}{a + b\lambda}. \tag{12.13}$$

The equivalent probability weight sequences are, by Lemma 4.1, given by

$$\frac{t^k w_k}{\Phi(t)} = \binom{a/b+k-1}{k} t^k (1-t)^{a/b}, \qquad 0 < t < 1, \tag{12.14}$$

which is the negative binomial distribution $\mathrm{NBin}(a/b, 1-t)$ (where the parameter a/b is not necessarily an integer). The canonical choice, which by Theorems 11.4 and 11.7 is the asymptotic distribution of the number of balls of a given colour, is $\mathrm{NBin}(a/b, 1-\tau) = \mathrm{NBin}(a/b, a/(a+b\lambda))$. See also Holst [53] and Kolchin [76].

Note that the case $b = 0$ (excluded above) means drawing with replacement; this is Example 12.1, which thus can be seen as a limit case. (This corresponds to the Poisson limit $\mathrm{NBin}(a/b, a/(a+b\lambda)) \xrightarrow{\mathrm{d}} \mathrm{Po}(\lambda)$ as $b \to 0$.)

Example 12.5 (drawing without replacement)**.** Consider again an urn with balls of n colours, with initially a balls of each colour. (This time, $a \geqslant 1$ is an integer.) Draw m balls without replacement, and let as above Y_i be the number of drawn balls of colour i. (The case $a = 1$ yields the Fermi–Dirac statistics in Example 12.3.)

Formally, this is the case $b = -1$ of Example 12.4, and a similar calculation shows that

$$\mathbb{P}\big((Y_1, \dots, Y_n) = (y_1, \dots, y_n)\big) = \frac{\prod_{i=1}^n \binom{a}{y_i}}{\binom{na}{m}}; \tag{12.15}$$

hence this is the random allocation given by the weights

$$w_k = \binom{a}{k}, \qquad k = 0, 1, \dots \tag{12.16}$$

We have thus $\Phi(t) = (1 + t)^a$, exactly as in Example 10.6, with $d = a$.

The equivalent probability weight sequences are the binomial distributions $\mathrm{Bi}(a, p)$, $0 < p < 1$, and the canonical choice is, for $0 < \lambda < a$, $(\pi_k) = \mathrm{Bi}(a, \lambda/a)$, i.e.

$$\pi_k = \binom{a}{k}\left(\frac{\lambda}{a}\right)^k\left(\frac{a-\lambda}{a}\right)^{a-k} = \binom{a}{k}\frac{\lambda^k(a-\lambda)^{a-k}}{a^a}. \tag{12.17}$$

See also Holst [53] and Kolchin [76].

Note that taking the limit as $a \to \infty$, we obtain drawing with replacement, which is Example 12.1; this corresponds to the Poisson limit $\mathrm{Bi}(a, \lambda/a) \overset{\mathrm{d}}{\longrightarrow} \mathrm{Po}(\lambda)$ as $a \to \infty$.

Example 12.6 (random rooted forests [76])**.** Consider *labelled rooted forests* consisting of n unordered rooted trees with together m nodes, all of which are labelled. (Thus $m \geqslant n$.) We may assume that the n roots are labelled $1, \dots, n$; let T_i be the tree with root i and let $t_i := |T_i|$. Then the node sets $V(T_i)$ form a partition of $\{1, \dots, m\}$, so $\sum_{i=1}^n t_i = m$ and $(t_1, \dots, t_n)$ is an allocation in $\mathcal{B}_{m,n}$, with each $t_i \geqslant 1$. Furthermore, given $(t_1, \dots, t_n) \in \mathcal{B}_{m,n}$ with all $t_i \geqslant 1$, the node sets $V(T_i)$ can be chosen in $\binom{m-n}{t_1-1,\dots,t_n-1}$ ways, and given $V(T_i)$, the tree T_i can by Cayley's formula be chosen in $t_i^{t_i-2}$ ways. (The trees are rooted but the roots are given.) Hence, the number of forests with the allocation $(t_1, \dots, t_n)$ is

$$\binom{m-n}{t_1-1,\dots,t_n-1}\prod_{i=1}^n t_i^{t_i-2} = (m-n)!\prod_{i=1}^n \frac{t_i^{t_i-2}}{(t_i-1)!} = (m-n)!\prod_{i=1}^n \frac{t_i^{t_i-1}}{t_i!}. \tag{12.18}$$

Hence, a uniformly random labelled rooted forest corresponds to a random allocation $B_{m,n}$ with the weight sequence

$$w_k = \frac{k^{k-1}}{k!}, \quad k \geqslant 1, \qquad \text{and} \qquad w_0 = 0. \tag{12.19}$$

Note that here $w_0 = 0$ unlike almost everywhere else in the present paper; in the notation of Remark 11.8, we have $\alpha = 1$. (As discussed in Remark 11.8, we can reduce to the case $w_0 > 0$ by considering $(t_1 - 1, \dots, t_n - 1)$, which is an allocation in $\mathcal{B}_{m-n,n}$; this means that we count only non-root nodes.

We prefer, however, to keep the setting above with $w_0 = 0$, noting that the results above still hold by Remark 11.8.)

If $F^{\mathrm{r}}_{m,n}$ denotes the number of labelled rooted forests with m labelled nodes of which n are given as roots, then (12.18) implies

$$F^{\mathrm{r}}_{m,n} = (m-n)!\, Z(m,n). \tag{12.20}$$

It is well-known that $F^{\mathrm{r}}_{m,n} = nm^{m-n-1}$, a formula also given by Cayley [22], see e.g. [103, Proposition 5.3.2] or [99]; thus

$$Z(m,n) = \frac{nm^{m-n-1}}{(m-n)!}. \tag{12.21}$$

We have

$$\Phi(t) := \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} t^k = T(t), \tag{12.22}$$

the well-known *tree function* (known by this name since it is the exponential generating function for rooted unordered labelled trees, cf. Example 10.2). Note that $T(z)$ satisfies the functional equation

$$T(z) = ze^{T(z)}; \tag{12.23}$$

see e.g. [40, Section II.5]. Equivalently,

$$z = T(z)e^{-T(z)}, \tag{12.24}$$

which by differentiation leads to

$$T'(z) = \frac{T(z)}{z(1-T(z))}\ . \tag{12.25}$$

Hence,

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} = \frac{1}{1-T(t)}\ . \tag{12.26}$$

By (12.22) and Stirling's formula, $\Phi(t)$ has radius of convergence $\rho = e^{-1}$. Furthermore, (12.24) implies that $\Phi(\rho) = T(e^{-1}) = 1$. Hence, (12.26) yields $\nu = \Psi(\rho) = \infty$, and if $1 \leqslant \lambda < \infty$, then $\lambda = \Psi(\tau)$ is solved by

$$T(\tau) = 1 - \frac{1}{\lambda} = \frac{\lambda-1}{\lambda} \tag{12.27}$$

and thus, using (12.24),

$$\tau = \frac{\lambda-1}{\lambda} e^{-(\lambda-1)/\lambda}. \tag{12.28}$$

The probability weight sequences equivalent to (w_k) are by Lemma 4.1 given by, substituting $x = T(t)$, and thus $t = xe^{-x}$ by (12.24),

$$p_k = \frac{t^k}{T(t)} w_k = \frac{k^{k-1}t^k}{T(t)k!} = \frac{(kx)^{k-1}e^{-kx}}{k!}, \qquad k \geqslant 1, \tag{12.29}$$

where $0 \leqslant t \leqslant e^{-1}$ and thus $0 \leqslant x \leqslant 1$. This is known as a *Borel distribution*; it appears for example as the distribution of the size $|\mathcal{T}|$ of the Galton–Watson tree with offspring distribution $\mathrm{Po}(x)$. (This was first proved by Borel [18]. It follows by Theorem 15.5 below, with the probability weight sequence $\mathrm{Po}(x)$; see also Otter [93], Tanner [107], Dwass [36], Takács [106], Pitman [99].) It follows that the random rooted forest considered here has the same distribution as the forest defined by a Galton–Watson process with starting with n individuals (the roots) and $\mathrm{Po}(x)$ offspring distribution, conditioned to have total size m; cf. Example 12.8 below. See further Kolchin [76] and Pavlov [96].

In particular, the choice $x = 1$ ($t = e^{-1}$) in (12.29) yields the equivalent probability weight sequence

$$\widetilde{w}_k = e^{-k} w_k = \frac{k^{k-1} e^{-k}}{k!}, \qquad k \geqslant 1, \tag{12.30}$$

which by Stirling's formula satisfies the asymptotic power-law

$$\widetilde{w}_k \sim \frac{1}{\sqrt{2\pi}} k^{-3/2}, \qquad \text{as } k \to \infty. \tag{12.31}$$

Moreover, the canonical distribution for a given $\lambda \geqslant 1$ is, using (12.28),

$$\pi_k = \frac{k^{k-1}\tau^k}{T(\tau)k!} = \frac{k^{k-1}}{k!} \left(\frac{\lambda-1}{\lambda}\right)^{k-1} e^{-k(\lambda-1)/\lambda}. \tag{12.32}$$

By Theorems 11.4 and 11.7, and Remark 11.8, this is the asymptotic distribution of the size of a given (or random) tree in the forest, say T_1. The asymptotic distribution of $|T_1|$ is thus the distribution of the size $|\mathcal{T}|$ of a Galton–Watson tree with offspring distribution $\mathrm{Po}(1 - 1/\lambda)$. Moreover, T_1 is, given its size $|T_1|$, uniformly distributed over all trees on $|T_1|$ nodes, and the same is true for the Poisson Galton–Watson tree $\mathcal{T}$ by Example 10.2. Consequently, $T_1 \xrightarrow{\mathrm{d}} \mathcal{T}$ as $n \to \infty$ with $m/n \to \lambda$. (We may regard T_1 as an ordered tree, ordering the children of a node e.g. by their labels.)

The same random allocation $B_{m,n}$ also describes the block lengths in hashing with linear probing; see Janson [56]. Indeed, there is a one-to-one correspondence between hash tables and rooted forests, see e.g. Knuth [75, Exercise 6.4-31] and Chassaing and Louchard [24].

Example 12.7 (random unrooted forests)**.** Consider *labelled unrooted forests* consisting of n trees with together m nodes, all of which are labelled. (Thus $m \geqslant n$.) We may assume that the n trees are labelled $T_1, \dots, T_n$; let $t_i := |T_i|$. As in Example 12.6, the node sets $V(T_i)$ form a partition of $\{1, \dots, m\}$, so $\sum_{i=1}^n t_i = m$ and $(t_1, \dots, t_n)$ is an allocation in $\mathcal{B}_{m,n}$, with each $t_i \geqslant 1$. In the unrooted case, given $(t_1, \dots, t_n) \in \mathcal{B}_{m,n}$ with all $t_i \geqslant 1$, the node sets $V(T_i)$ can be chosen in $\binom{m}{t_1,\dots,t_n}$ ways, and given $V(T_i)$, the tree T_i can by Cayley's formula be chosen in $t_i^{t_i-2}$ ways. Hence, the number of unrooted forests with the allocation $(t_1, \dots, t_n)$ is

$$\binom{m}{t_1, \dots, t_n} \prod_{i=1}^{n} t_i^{t_i-2} = m! \prod_{i=1}^{n} \frac{t_i^{t_i-2}}{t_i!}. \tag{12.33}$$

Hence, a uniformly random labelled unrooted forest corresponds to a random allocation $B_{m,n}$ with the weight sequence

$$w_k = \frac{k^{k-2}}{k!}, \quad k \geqslant 1, \qquad \text{and} \qquad w_0 = 0. \tag{12.34}$$

As in Example 12.6, we have $w_0 = 0$, but this is no problem by Remark 11.8.

If $F^{\mathrm{u}}_{m,n}$ denotes the number of labelled unrooted forests with m labelled nodes and n labelled trees, then (12.33) implies

$$F^{\mathrm{u}}_{m,n} = m!\, Z(m,n). \tag{12.35}$$

There is no simple general formula for $F^{\mathrm{u}}_{m,n}$, as there is for the rooted forests in Example 12.6, and hence no simple formula for $Z(m,n)$. Asymptotics are given by Britikov [20]. (See Example 18.16 for one case. The asymptotic formula when $m/n \to \lambda > 2$ follows similary from Theorem 19.34(ii), and when $m/n \to \lambda < 2$ with $m = \lambda n + o(\sqrt{n})$ from Theorem 18.12.)

We have

$$\Phi(t) := \sum_{k=1}^{\infty} \frac{k^{k-2}}{k!} t^k = T(t) - \tfrac{1}{2} T(t)^2, \tag{12.36}$$

where $T(t)$ is the tree function in (12.22). (The latter equality is well-known, see e.g. [40, II.5.3]; it can be shown e.g. by showing that both sides have the same derivative $T(t)/t$; there are also combinatorial proofs.) Hence, using (12.25),

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} = \frac{T(t)}{\Phi(t)} = \frac{1}{1 - T(t)/2}; \tag{12.37}$$

cf. the similar (12.26) in the rooted case.

As for (12.22), Φ has the radius of convergence $\rho = e^{-1}$, but now, by (12.37), $\nu = \Psi(\rho) = 2$ is finite, so there is a phase transition at $\lambda = 2$. The parameter τ is by the definition in Theorem 7.1 and (12.37) given by $T(\tau) = 2 - 2/\lambda = 2(\lambda - 1)/\lambda$ for $\lambda \leqslant 2$; thus, using (12.24),

$$\tau = \begin{cases} 2\frac{\lambda-1}{\lambda} e^{-2(\lambda-1)/\lambda}, & \lambda \leqslant 2. \\ e^{-1}, & \lambda \geqslant 2. \end{cases} \tag{12.38}$$

The probability weight sequences equivalent to (w_k) are by Lemma 4.1 given by, again substituting $t = xe^{-x}$ or $x = T(t)$,

$$p_k = \frac{k^{k-2} t^k}{T(t)(1 - T(t)/2)k!} = \frac{x(kx)^{k-2} e^{-kx}}{(1 - x/2)k!}, \qquad k \geqslant 1, \tag{12.39}$$

where $0 \leqslant t \leqslant e^{-1}$ and thus $0 \leqslant x \leqslant 1$. In particular, taking $x = 1$ ($t = e^{-1}$), we obtain the equivalent probability weight sequence

$$\widetilde{w}_k = 2w_k e^{-k} = \frac{2k^{k-2}e^{-k}}{k!}, \qquad k \geqslant 1, \tag{12.40}$$

which by Stirling's formula satisfies the asymptotic power-law

$$\widetilde{w}_k \sim \frac{2}{\sqrt{2\pi}} k^{-5/2}, \qquad \text{as } k \to \infty. \tag{12.41}$$

Moreover, the canonical distribution for a given $\lambda \geqslant 1$ is, by (12.38) and (12.39), for $k \geqslant 1$,

$$\pi_k = \frac{k^{k-2}\tau^k}{T(\tau)(1 - T(\tau)/2)k!} = \begin{cases} \frac{k^{k-2}}{k!}\lambda\Bigl(2\frac{\lambda-1}{\lambda}\Bigr)^{k-1} e^{-2k(\lambda-1)/\lambda}, & \lambda \leqslant 2, \\ \frac{2k^{k-2}e^{-k}}{k!}, & \lambda \geqslant 2. \end{cases} \tag{12.42}$$

By Theorems 11.4 and 11.7, and Remark 11.8, this is the asymptotic distribution of the size of a given (or random) tree in the forest, say T_1.

We shall see in Theorem 19.49 that the phase transition at $\lambda = 2$ is seen clearly in the size of the largest tree in the forest: if $m/n \to \lambda < 2$, then the largest tree is of size $O_{\mathrm{p}}(\log n)$, while if $m/n \to \lambda > 2$, then there is a unique giant tree of size $(\lambda - 2)n + o_{\mathrm{p}}(n)$; for details see Theorems 19.34 and 19.49, and, more generally, Łuczak and Pittel [83]. This is thus an example of the condensation discussed after Theorem 11.4 (and similar to the condensation in Theorem 7.1 when $\nu < 1$).

Example 12.8 (simply generated forests and Galton–Watson forests)**.** A *simply generated forest* is a sequence $(T_1, \dots, T_n)$ of rooted trees, with weight

$$w(T_1, \dots, T_n) := \prod_{i=1}^{n} w(T_i), \tag{12.43}$$

where $w(T_i)$ is given by (2.3), for some fixed weight sequence $\mathbf{w}$. A *simply generated random forest* with n trees and m nodes, where n and m are given with $m \geqslant n$, is such a forest chosen at random, with probability proportional to its weight. Note that in the special case $n = 1$, this is the same as a simply generated random tree defined in Section 2. More generally, for any n, a simply generated random forest $(T_1, \dots, T_n)$ is, conditioned on the sizes $|T_1|, \dots, |T_n|$, a sequence of independent simply generated random trees with the given sizes (all defined by the same weight sequence $\mathbf{w}$). Moreover, the sizes $(|T_1|, \dots, |T_n|)$ form an allocation in $\mathcal{B}_{m,n}$, and it is easily seen that this is a random allocation $B_{m,n}$ defined by the weight sequence $(Z_k)_{k=0}^{\infty}$, where Z_k is the partition function (2.5) for simply generated trees with weight sequence $\mathbf{w}$ (and $Z_0 = 0$).

A simply generated random forest can thus be obtained by a two-stage process, combining the constructions in Sections 2 and 11. Note that equivalent weight sequences $\mathbf{w}$ yield equivalent weight sequences (Z_k) by (4.3), and thus the same simply generated random forest.

In the special case when $\mathbf{w}$ is a probability weight sequence, we also define a *Galton–Watson forest* with n trees, for a given n, as a sequence $(\mathcal{T}_1, \dots, \mathcal{T}_n)$ of n i.i.d. Galton–Watson trees; it describes the evolution of a Galton–Watson process started with n particles. (It can also be seen as a single Galton–Watson tree $\mathcal{T}$ with the root chopped off, conditioned on the root degree being n, provided that this root degree is possible.) Note that the probability distribution of the forest is given by the weights in (12.43). Hence, in the probability weight sequence case, the simply generated random forest equals the *conditioned Galton–Watson forest* with n trees and m nodes, defined as a Galton–Watson forest with n trees conditioned on the total size being m; in other words, it describes a Galton–Watson process started with n particles conditioned on the total size being m.

Random forests of this type are studied by Pavlov [96], see also Flajolet and Sedgewick [40, Example III.21].

For example, taking $w_k = 1/k!$, we have by (10.11) $Z_k = k^{k-1}/k!$, $k \geqslant 1$; this is the weight sequence used in Example 12.6, so we obtain the same random allocation of tree sizes as there; moreover, given the tree sizes, the trees are uniformly random labelled unordered rooted trees by Example 10.2. Consequently, for this weight sequence, the simply generated random forest is the random labelled forest with unordered rooted trees in Example 12.6. The same random forest is obtained by the equivalent probability weight sequence $w_k = x^k e^{-x}/k!$, with $0 < x \leqslant 1$, so it equals also the conditioned Galton–Watson forest with offspring distribution $\mathrm{Po}(x)$, cf. Example 12.6.

Another example is obtained by taking $w_k = 1$ for all $k \geqslant 0$. Then every forest has weight 1, so the this simply generated random forest is a uniformly random forest of ordered rooted trees. (An *ordered rooted forest*.) By Example 10.1, the weight sequence (Z_k) is then given by the Catalan numbers in (2.1): $Z_k = C_{k-1} = (2k-2)!/(k!\,(k-1)!)$, $k \geqslant 1$.

Further examples are given by starting with the other examples of random trees in Section 10.

We shall see in Theorem 18.11 that if the weight sequence $\mathbf{w}$ is as in Theorem 7.1, and further $\mathrm{span}(\mathbf{w}) = 1$, $\nu \geqslant 1$ and $\sigma^2 < \infty$, then

$$Z_k \sim \frac{\tau}{\sqrt{2\pi\sigma^2}} \left(\frac{\Phi(\tau)}{\tau}\right)^k k^{-3/2}. \tag{12.44}$$

Recalling $\mathcal{Z}(\tau/\Phi(\tau)) = \tau$ by (7.6), we may replace Z_k by the equivalent probability weight sequence

$$\widetilde{Z}_k := \frac{Z_k}{\mathcal{Z}(\tau/\Phi(\tau))} \left(\frac{\tau}{\Phi(\tau)}\right)^k = \frac{Z_k}{\tau} \left(\frac{\tau}{\Phi(\tau)}\right)^k \sim \frac{1}{\sqrt{2\pi\sigma^2}} k^{-3/2}, \tag{12.45}$$

so we have the asymptotic behaviour $\widetilde{Z}_k \sim c k^{-3/2}$ for every such weight sequence $\mathbf{w}$, where only the constant $c = 1/\sqrt{2\pi\sigma^2}$ depends on $\mathbf{w}$. This explains why random forests of this type have similar asymptotic behaviour, in contrast to the unrooted forests in Example 12.7 which are given by random allocations defined by a weight sequence $\sim ck^{-5/2}$, see (12.41); see further Example 12.10.

Example 12.9. Let, as in Example 10.7, $w_k = (k+1)^{-\beta}$ for some real constant β. Then $\rho = 1$. As shown in Example 10.7, $\nu = \infty$ if $\beta \leqslant 2$, and $\nu < \infty$ if $\beta > 2$; in the latter case, ν is given by (10.25). This example is studied further in e.g. Bialas, Burda and Johnston [14].

Example 12.10 (power-law)**.** More generally, suppose that $w_k \sim ck^{-\beta}$ as $k \to \infty$, for some real constant β and $c > 0$, i.e., that w_k asymptotically satisfies a power-law. Qualitatively, we have the same behaviour as in Examples 10.7 and 12.9, but numerical values such as the critical β in (10.26) will in general be different.

We repeat some easy facts: first, $\rho = 1$, $\omega = \infty$ and $\operatorname{span}(\mathbf{w}) = 1$.

If $-\infty < \beta \leqslant 1$, then $\Phi(\rho) = \Phi(1) = \infty$; hence $\nu = \infty$ by Lemma 3.1(iv).

If $1 < \beta \leqslant 2$, then $\Phi(\rho) < \infty$ but $\Phi'(\rho) = \sum_{k=0}^{\infty} kw_k = \infty$; hence again $\nu = \Psi(\rho) = \infty$ by (3.11).

On the other hand, if $\beta > 2$, then $\Phi(1) < \infty$ and $\Phi'(1) < \infty$, and thus $\nu < \infty$ by (3.11). Summarising:

$$\nu < \infty \iff \beta > 2. \tag{12.46}$$

In the case $\beta > 2$, there is thus a phase transition when we vary λ.

Suppose $\beta > 2$, so $\nu < \infty$. If $\lambda \geqslant \nu$, then $\tau = \rho = 1$, and the canonical distribution (π_k) is by (11.13) given simply by $\pi_k = w_k/\Phi(1)$. This distribution then has mean $\mu = \nu < \infty$ by (11.14); since $\pi_k \asymp k^{-\beta}$ as $k \to \infty$, the variance $\sigma^2 = \infty$ if $2 < \beta \leqslant 3$, while $\sigma^2 < \infty$ when $\beta > 3$.

Note that Examples 12.6 and 12.7 with random forests are of this type, provided we replace w_k by the equivalent $\widetilde{w}_k := e^{-k} w_k$; Stirling's formula shows that $\widetilde{w}_k \sim ck^{-\beta}$ where $\beta = 3/2$ for rooted forests and $\beta = 5/2$ for unrooted forests, see (12.31) and (12.41) (with a different choice of constant factor in the latter). The different values of β explains the different asymptotical behaviours of these two types of random forests: by the results above, the tail behaviour of w_k implies that $\nu = \infty$ for rooted forests but $\nu < \infty$ for unrooted forests, as we have shown by explicit calculations in Examples 12.6 and 12.7. Recall that this means that there is a phase transition and condensation for high m/n in the unrooted case but not in the rooted case.

More generally, (12.45) shows that simply generated random forests under weak assumptions have the same power-law behaviour of the weight sequence with $\beta = 3/2$ as the special case of (unordered) rooted forests in Example 12.6. Thus $\nu = \infty$ and there is no phase transition. (At least not in the range $m = O(n)$ that we consider. Pavlov [96] show a phase transition at $m = \Theta(n^2)$.)

Example 12.11 (unlabelled forests)**.** Consider, as Pavlov [97], rooted forests consisting of n rooted *unlabelled* unordered trees, assuming that the trees, or equivalently the roots, are labelled $1, \dots, n$, but otherwise the nodes are unlabelled. A uniformly random forest of this type with m nodes can be seen as balls-in-boxes with the weight sequence (t_k), where t_k is the number of unlabelled unordered rooted trees with k nodes. In this case there is no simple formula for the generating function $\Phi(z)$, but there is a functional equation, from which it can be shown that $t_k \sim c_1 k^{-3/2}\rho^{-k}$, where $\rho \approx 0.3382$ as usual is the radius of convergence of $\Phi(z)$ and $c_1 \approx 0.4399$, see Otter [92] or, e.g., Drmota [33, Section 3.1.5]. Furthermore, $\Phi(\rho) = 1$; thus $(t_k\rho^k)$ gives an equivalent probability weight sequence with $t_k\rho^k \sim c_1 k^{-3/2}$ as $k \to \infty$. The asymptotic behaviour of the weight sequence is thus the same as for labelled rooted forests in Example 12.6, and more generally for Galton–Watson forests (under weak conditions) in Example 12.8, and we expect the same type of asymptotic behaviour in spite of the fact that the unlabelled forest is not simply generated; this is seen in detail in Pavlov [96] for the size of the largest tree. In particular, we have $\nu = \infty$ by Example 12.10 and (12.46), and thus there is no phase transition at finite λ.

Similarly, Bernikovich and Pavlov [12] considered *unrooted* forests consisting of n unordered trees labelled $1, \dots, n$ with a total of m unlabelled nodes. These are described by the weight sequence $(\check{t}_k)$ where $\check{t}_k$ is the number of unrooted unlabelled unordered trees with k nodes. Again, there is no no simple formula for the generating function $\check{\Phi}(z) := \sum_k \check{t}_k z^k$, but there is the relation $\check{\Phi}(z) = \Phi(z) - \frac{1}{2}\Phi(z)^2 + \frac{1}{2}\Phi(z^2)$ found by Otter [92], which leads to the asymptotic formula $\check{t}_k \sim c_2 k^{-5/2}\rho^{-k}$, where ρ is as above and $c_2 \approx 0.5347$, see also Drmota [33, Section 3.1.5]. In this case, $(\check{t}_k\rho^k/\check{\Phi}(\rho))$ gives an equivalent probability weight sequence which is $\sim (c_2/\check{\Phi}(\rho))k^{-5/2}$ as $k \to \infty$, which is the same type of asymptotic behaviour as for the weight sequence for labelled unrooted forests in Example 12.7; we thus expect the same type of asymptotic behaviour as for those forests. In particular, $\nu < \infty$ by Example 12.10; a numerical calculation gives $\nu := \rho\check{\Phi}'(\rho)/\check{\Phi}(\rho) \approx 2.0513$, see Bernikovich and Pavlov [12].

Note that both types of "unlabelled" forests considered here have the n trees labelled $1, \dots, n$ (but the individual nodes are not labelled). Completely unlabelled forests cannot be described by balls-in-boxes (as far as we know), since the number of (non-isomorphic) ways to label the trees depends on the forest.

Example 12.12 (the backgammon model)**.** The model with $w_k = 1/k!$ for $k \geqslant 1$ as in Example 12.1, but $w_0 > 0$ arbitrary, was considered by Ritort [100] and Franz and Ritort [41, 42], who called it the *backgammon model.* We have

$$\Phi(t) = w_0 + \sum_{k=1}^{\infty} \frac{t^k}{k!} = e^t + w_0 - 1 \tag{12.47}$$

and

$$\Psi(t) = \frac{te^t}{\Phi(t)} = \frac{t}{1 + (w_0 - 1)e^{-t}}. \tag{12.48}$$

Thus $\rho = \nu = \infty$. The equation $\Psi(\tau) = \lambda$ can be written

$$(\tau - \lambda)e^{\tau} = (w_0 - 1)\lambda, \tag{12.49}$$

and the solution can be written

$$\tau = \lambda + W\bigl((w_0 - 1)\lambda e^{-\lambda}\bigr) = \lambda - T\bigl((1 - w_0)\lambda e^{-\lambda}\bigr), \tag{12.50}$$

where $W(z)$ is the Lambert W function [26] defined by $W(z)e^{W(z)} = z$ and $T(z)$ is the tree function in (12.22) (analytically extended to all real $z < e^{-1}$); note that $W(z) = -T(-z)$ by (12.24), see [26].

The canonical probability weight sequence (11.13) is, using (12.48) and $\Psi(\tau) = \lambda$,

$$\pi_k = \frac{\tau^k}{\Phi(\tau)k!} = \frac{\lambda\tau^{k-1}e^{-\tau}}{k!} = \frac{\lambda}{\tau} \cdot \frac{\tau^k e^{-\tau}}{k!}, \qquad k \geqslant 1, \tag{12.51}$$

and $\pi_0 = \lambda\tau^{-1}e^{-\tau}w_0$.

Example 12.13 (random permutations and recursive forests)**.** Consider *permutations* of $\{1, \dots, m\}$ with exactly n cycles. We want to list the cycle lengths in some order; for convenience, we consider all possible orders and define a *cycle-labelled permutation* to be a permutation with the cycles labelled $1, \dots, n$, in arbitrary order. Given a cycle-labelled permutation with exactly n cycles, let y_i be the length of the i:th cycle. Then $(y_1, \dots, y_n)$ is an allocation in $\mathcal{B}_{m,n}$ with each $y_i \geqslant 1$, and for each such $(y_1, \dots, y_n) \in \mathcal{B}_{m,n}$, the number of cycle-labelled permutations with y_i elements in cycle i is

$$\binom{m}{y_1, \dots, y_n} \prod_{i=1}^{n} (y_i - 1)! = m! \prod_{i=1}^{n} \frac{1}{y_i}, \tag{12.52}$$

since there are $(y-1)!$ cycles with y given elements. Consequently, a uniformly random permutation of $\{1, \dots, m\}$ with exactly n cycles (listed in random order, say) corresponds to a random allocation $B_{m,n}$ defined by the weight sequence

$$w_k = \frac{1}{k}, \quad k \geqslant 1, \qquad \text{and} \qquad w_0 = 0. \tag{12.53}$$

Note that here, as in Example 12.6, $w_0 = 0$, and Remark 11.8 applies with $\alpha = 1$.

The number of permutations with n (unlabelled) cycles is by (12.52)

$$m!\, Z(m, n)/n!, \tag{12.54}$$

where we divide by $n!$ in order to ignore the labelling above.

The same balls-in-boxes model with $w_k = 1/k$, $k \geqslant 1$, also describes *random recursive forests*, see Pavlov and Loseva [98].

We have

$$\Phi(t) = \sum_{k=1}^{\infty} \frac{t^k}{k} = -\log(1 - t) \tag{12.55}$$

with radius of convergence $\rho = 1$ and

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} = \frac{t}{-(1-t)\log(1-t)}, \tag{12.56}$$

so $\nu = \Psi(1) = \infty$, cf. Example 12.10 ($\beta = 1$).

The equivalent probability weight sequences are by Lemma 4.1 given by

$$p_k = \frac{x^k}{k|\ln(1-x)|}, \qquad 0 < x < 1, \tag{12.57}$$

with probability generating function $\Phi(xz)/\Phi(x) = \log(1-xz)/\log(1-x)$. This distribution is called the *logarithmic distribution.* See further Kolchin, Sevast'yanov and Chistyakov [77] and Kolchin [76].

By Remark 11.8, we obtain results on random permutations with m cycles as $m/n \to \lambda \in [1, \infty)$, see for example Kazimirov [71]. However, it is of greater interest to consider random permutations without constraining the number of cycles. This can be done using methods similar to the ones used here, but is outside the scope of the present paper; see e.g. Kolchin, Sevast'yanov and Chistyakov [77], Kolchin [76] and Arratia, Barbour and Tavaré [7]. Note that even if we condition on the number of cycles, a typical random permutation of $\{1, \dots, m\}$ has about $\log m$ cycles, so we are interested in the case $n \approx \log m$ and thus $m/n \to \infty$, which we do not considered here.

Other random objects that can be decomposed into components can be studied similarly, for example random mappings [76]; our results apply only to random objects with a given number of components (in some cases), but similar methods are useful for the general case; see Kolchin [76] and Arratia, Barbour and Tavaré [7].

13. Preliminaries

Proof of Lemma 3.1. (i): Since $\Phi'(t) = \sum_{k=0}^{\infty} k w_k t^{k-1}$ has the same radius of convergence ρ as Ψ, and $\Phi(t) \geqslant w_0 > 0$ for $t \geqslant 0$, it is immediate that Ψ is well-defined, finite and continuous for $t \in [0, \rho)$. Furthermore, if $0 < t < \rho$, then $t\Psi'(t)$ is by (4.10) the variance of a non-degenerate random variable, and thus $t\Psi'(t) > 0$. Hence $\Psi(t)$ is increasing, completing the proof of (i).

(ii): If $\Phi(\rho) = \infty$, the claim is just the definition of $\Psi(\rho)$ in Section 2. (Note that the existence of the limit follows from (i).) We may thus assume $\Phi(\rho) < \infty$; then $t \nearrow \rho$ implies $\Phi(t) \to \Phi(\rho) < \infty$ and $\Phi'(t) \to \Phi'(\rho) \leqslant \infty$ by monotone convergence, and thus

$$\Psi(t) := \frac{t\Phi'(t)}{\Phi(t)} \to \frac{\rho\Phi'(\rho)}{\Phi(\rho)} = \Psi(\rho).$$

(iii): The case $\rho = 0$ is trivial, and the case $\rho > 0$ follows from (i) and (ii).

(iv): For any $\ell \geqslant 0$,

$$\Psi(t)-\ell=\frac{\sum_{k=0}^{\infty}(k-\ell)w_kt^k}{\sum_{k=0}^{\infty}w_kt^k}\geqslant\frac{\sum_{k=0}^{\ell-1}(k-\ell)w_kt^k}{\sum_{k=0}^{\infty}w_kt^k}. \tag{13.1}$$

If $\rho<\infty$ and $\Phi(\rho)=\infty$, we thus have,

$$\Psi(t)-\ell\geqslant\frac{O(1)}{\Phi(t)}\to 0 \qquad \text{as } t\nearrow\rho,$$

so $\Psi(\rho)-\ell\geqslant 0$. Since ℓ is arbitrary, this shows $\Psi(\rho)=\infty$, proving (iv).

(v): If $\rho=\infty$, choose ℓ with $w_\ell>0$. Then (13.1) implies

$$\Psi(t)-\ell\geqslant\frac{-\ell\sum_{k=0}^{\ell-1}w_kt^k}{w_\ell t^\ell}\to 0 \qquad \text{as } t\to\infty,$$

so $\Psi(\infty)-\ell\geqslant 0$. Hence, $\Psi(\infty)\geqslant\sup\{\ell : w_\ell>0\}=\omega$.

Conversely,

$$\Psi(t)=\frac{\sum_{k=0}^{\omega}kw_kt^k}{\sum_{k=0}^{\omega}w_kt^k}\leqslant\omega \qquad \text{for all } t\in[0,\rho),$$

so $\Psi(\rho)\leqslant\omega$, completing the proof of (v).

Finally, (3.9) follows from (i) and (ii). ☐

Remark 13.1. Alternatively, the fact that $\Psi(t)$ is increasing can also be seen as follows: Let $0<a<b<\rho$ and let Y be a random variable with distribution $\mathbb{P}(Y=k)=w_ka^k/\Phi(a)$ (cf. Lemma 4.2). Then $\Psi(a)=\mathbb{E}\,Y$ and $\Psi(b)=\mathbb{E}\big(Y(b/a)^Y\big)/\mathbb{E}(b/a)^Y$, so $\Psi(a)\leqslant\Psi(b)$ is equivalent to the correlation inequality $\mathbb{E}\big(Y(b/a)^Y\big)\geqslant\mathbb{E}\,Y\,\mathbb{E}(b/a)^Y$, which says that the two random variables $f(Y):=Y$ and $g(Y):=(b/a)^Y$ are positively correlated; it is well-known that this holds (as long as the expectations are finite) for any two increasing functions f and g and any Y, see [50, Theorem 236] where the result is attributed to Chebyshev, and it is easy to see that, in fact, strict inequality holds in the present case. (The latter inequality is an analogue of Harris' correlation inequality [51] for variables Y with values in a discrete cube $\{0,1\}^N$; in fact, the inequalities have a common extension to variables with values in $\mathbb{R}^N$. Cf. also the related FKG inequality, which extends Harris' inequality; see for example [48] where also its history is described.)

For a third proof that $\Psi(t)$ is increasing, note that (3.7) shows that Ψ is (strictly) increasing if and only if $\log\Phi(e^x)$ is (strictly) convex, which is an easy consequence of Hölder's inequality, (See e.g. [31, Lemma 2.2.5(a)] and note that $\Phi(e^x)=\sum_{k=0}^{\infty}e^{kx}w_k$ is the moment generating function of (w_k) in the case that (w_k) is a probability weight sequence.)

Lemma 3.1 shows that Ψ is a bijection $[0,\rho]\to[0,\Psi(\rho)]=[0,\nu]$, so it has a well-defined inverse $\Psi^{-1}:[0,\nu]\to[0,\rho]$. We extend this inverse to $[0,\infty)$ as follows.

Lemma 13.2. *For $x \geqslant 0$ define $\tau = \tau(x) \in [0,\infty]$ by*

$$\tau(x) := \sup\{t \leqslant \rho : \Psi(t) \leqslant x\}. \tag{13.2}$$

Then $\tau(x)$ is the unique number in $[0,\rho]$ such that $\Psi(\tau(x)) = x$ when $x \leqslant \nu$, and $\tau(x) = \rho$ when $x \geqslant \nu$. Furthermore, the function $x \mapsto \tau(x)$ is continuous, and, for any $x \geqslant 0$,

$$\Psi(\tau(x)) = \min(x,\nu). \tag{13.3}$$

If $x < \omega$, then $0 \leqslant \tau(x) < \infty$ and $0 < \Phi(\tau(x)) < \infty$. On the other hand, if $x \geqslant \omega$, then $\tau(x) = \Phi(\tau(x)) = \infty$.

Proof. By Lemma 3.1 and the definition (3.10), Ψ is an increasing continuous bijection $[0,\rho] \to [0,\Psi(\rho)] = [0,\nu]$; thus if $0 \leqslant x \leqslant \nu$, there exists a unique $\Psi^{-1}(x) \in [0,\rho]$ with $\Psi(\Psi^{-1}(x)) = x$, and (13.2) yields $\tau(x) = \Psi^{-1}(x)$. Since Ψ is a continuous bijection of one compact space onto another, its inverse $\Psi^{-1} : [0,\nu] \to [0,\rho]$ is continuous too; thus $x \mapsto \tau(x) = \Psi^{-1}(x)$ is continuous on $[0,\nu]$. Furthermore, (13.3) holds for $x \leqslant \nu$.

If $x \geqslant \nu = \Psi(\rho)$, then (13.2) yields $\tau(x) = \rho$, and thus $\Psi(\tau(x)) = \Psi(\rho) = \nu$, so (13.3) holds in this case too.

Combining the two cases we see that $x \mapsto \tau(x)$ is continuous on $[0,\infty)$, and that (13.3) holds.

Now suppose that $x < \omega$ and $\tau(x) = \infty$. Since $\tau(x) \leqslant \rho$ we then have $\rho = \infty$, and Lemma 3.1(v) yields $\Psi(\tau(x)) = \Psi(\rho) = \omega > x$, contradicting (13.3). Thus $\tau(x) < \infty$ when $x < \omega$. Furthermore, if $\Phi(\tau(x)) = \infty$, then $\tau(x) = \rho$, since $\Phi(t) < \infty$ for $t < \rho$, and thus $\Phi(\rho) = \infty$. If further $x < \omega$, and thus $\rho = \tau(x) < \infty$ as just shown, then Lemma 3.1(iv) would give $\Psi(\tau(x)) = \Psi(\rho) = \infty$, again contradicting (13.3) since $x < \infty$. Thus $\Phi(\tau(x)) < \infty$ when $x < \omega$.

Conversely, if $x \geqslant \omega$, then $\omega < \infty$, so $\Phi(t)$ is a polynomial and $\rho = \infty$. Lemma 3.1(v) shows that $\Psi(\rho) = \omega \leqslant x$, so (13.2) yields $\tau(x) = \rho = \infty$, whence also $\Phi(\tau(x)) = \Phi(\infty) = \infty$. ☐

Next, we investigate when $Z(m,n) > 0$. We say than an allocation $(y_1,\dots,y_n)$ of m balls in n boxes is *good* if it has positive weight, i.e., if $y_i \in \operatorname{supp}(\mathbf{w})$ for every i. Thus, $Z(m,n) > 0$ if and only if there is a good allocation in $\mathcal{B}_{m,n}$; in this case, the random allocation $B_{m,n}$ is defined and is always good.

Provided m is not too small or too large, the m and n for which good allocations exist are easily characterised; the following lemma shows that a simple necessary condition also is sufficient. (The exact behaviour for very small m is complicated. The largest m such that $Z(m,n) = 0$ for all n is called the *Frobenius number* of the set $\operatorname{supp}(\mathbf{w})$; it is a well-known, and in general difficult, problem to compute this, see e.g. [10]. The case when m is close to ωn (with finite ω) is essentially the same by the symmetry in Remark 11.10.)

Lemma 13.3. *Suppose that $w_0 > 0$.*

(i) *If* $Z(m,n) > 0$*, then* $\operatorname{span}(\mathbf{w}) \mid m$ *and* $0 \leqslant m \leqslant \omega n$*.*
(ii) *If* $\omega < \infty$*, then there exists a constant* C *(depending on* $\mathbf{w}$*) such that if* $\operatorname{span}(\mathbf{w}) \mid m$ *and* $C \leqslant m \leqslant \omega n - C$*, then* $Z(m,n) > 0$*.*
(iii) *If* $\omega = \infty$*, then for each* $C' < \infty$*, there exists a constant* C *(depending on* $\mathbf{w}$ *and* C'*) such that if* $\operatorname{span}(\mathbf{w}) \mid m$ *and* $C \leqslant m \leqslant C'n$*, then* $Z(m,n) > 0$*.*

Proof. (i): $Z(m,n) > 0$ if and only if $m = \sum_{i=1}^n y_i$ for some y_i with $w_{y_i} > 0$, i.e., $y_i \in \operatorname{supp}(\mathbf{w})$. This implies $0 \leqslant y_i \leqslant \omega$ and $\operatorname{span}(\mathbf{w}) \mid y_i$ for each i, and the necessary conditions in (i) follow immediately.

(ii): We may for convenience assume that $\operatorname{span}(\mathbf{w}) = 1$, see Remark 11.9; then, by (3.3), $\operatorname{supp}(\mathbf{w}) \setminus \{0\}$ is a finite set of integers with greatest common divisor 1. Thus, by a well-known theorem by Schur, see e.g. [109, 3.15.2] or [40, Proposition IV.2], there is a constant C_1 such that every integer $m \geqslant C_1$ can be written as a finite sum $m = \sum_i y_i$ with $y_i \in \operatorname{supp}(\mathbf{w})$ (repetitions are allowed); i.e. we have a good allocation of m balls in some number $\ell(m)$ boxes. Choose one such allocation for each $m \in [C_1, C_1 + \omega)$, and let C_2 be the maximum number of boxes in any of them.

If $C_1 \leqslant m \leqslant \omega n - C_2\omega$, let $a := \lfloor (m - C_1)/\omega \rfloor$. Then $m - a\omega \in [C_1, C_1 + \omega)$, and has thus a good allocation in at most C_2 boxes. We add a boxes with ω balls each, and have obtained a good allocation of m balls using at most

$$C_2 + a = C_2 + \lfloor (m - C_1)/\omega \rfloor \leqslant C_2 + \lfloor (\omega n - C_2\omega - C_1)/\omega \rfloor \leqslant n$$

boxes. Hence we may add empty boxes and obtain a good allocation in $\mathcal{B}_{m,n}$. (Recall that $0 \in \operatorname{supp}(\mathbf{w})$.) Thus $Z(m,n) > 0$ when $C_1 \leqslant m \leqslant \omega n - C_2\omega$.

(iii): We may again assume $\operatorname{span}(\mathbf{w}) = 1$. Let K be a large integer and consider the truncated weight sequence $\mathbf{w}^{(K)} = (w_k^{(K)})$ defined by

$$w_k^{(K)} := \begin{cases} w_k, & k \leqslant K, \\ 0, & k > K; \end{cases} \tag{13.4}$$

we assume that $K \in \operatorname{supp}(\mathbf{w})$ and that K is so large that $K \geqslant C' + 1$ and $\operatorname{span}(\mathbf{w}^{(K)}) = \operatorname{span}(\mathbf{w}) = 1$. Then $\omega(\mathbf{w}^{(K)}) = K$, and (ii) shows that for some C_3, if $C_3 \leqslant m \leqslant Kn - C_3$, then $Z(m,n;\mathbf{w}) \geqslant Z(m,n;\mathbf{w}^{(K)}) > 0$. Hence, if $C_3 \leqslant m \leqslant C'n$ and $Z(m,n) = 0$, then $Kn - C_3 < m \leqslant C'n \leqslant (K-1)n$, and thus $n < C_3$, whence $m < C'C_3$. Consequently, if $C'C_3 \leqslant m \leqslant C'n$, then $Z(m,n) > 0$. ∎

Remark 13.4. In the case $\omega = \infty$, it is not always true that there is a constant C such that $Z(m,n) > 0$ whenever $m \geqslant C$. For example, suppose that $w_k = 1$ when $k = 0$ or $k = j!$ for some $j \geqslant 0$, and $w_k = 0$ otherwise. Then $Z(m,n) = 0$ when $m = (n+1)! - 1$ and $n \geqslant 2$.

Remark 13.5. Lemma 13.3 is easily modified for the case $w_0 = 0$; if $\alpha := \min\{k : w_k > 0\}$ as in Remark 11.8, then the necessary condition (i) is $\alpha n \leqslant m \leqslant \omega n$ and $\operatorname{span}(\mathbf{w}) \mid (m - \alpha n)$, and again this is sufficient if m stays away from the boundaries.

14. Proofs of Theorems 11.4–11.7

We now prove the theorems in Section 11, which we for the reader's convenience repeat first.

Theorem 11.4. *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

$$\pi_k := \frac{w_k \tau^k}{\Phi(\tau)}, \qquad k \geqslant 0. \tag{11.13}$$

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

$$\mu = \Psi(\tau) = \min(\lambda, \nu) \tag{11.14}$$

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

$$N_k(B_{m,n})/n \xrightarrow{\mathrm{p}} \pi_k. \tag{11.15}$$

Theorem 11.6. *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 $m/n \leqslant C$ for some $C < \omega$.*

Define the function $\tau : [0, \infty) \to [0, \infty]$ by $\tau(x) := \sup\{t \leqslant \rho : \Psi(t) \leqslant x\}$. Then $\tau(x)$ is the unique number in $[0, \rho]$ such that $\Psi(\tau(x)) = x$ when $x \leqslant \nu$, and $\tau(x) = \rho$ when $x \geqslant \nu$; furthermore, the function $x \mapsto \tau(x)$ is continuous. We have $0 \leqslant \tau(m/n) < \infty$ and $0 < \Phi(\tau(m/n)) < \infty$, and for every $k \geqslant 0$,

$$\frac{N_k(B_{m,n})}{n} - \frac{w_k(\tau(m/n))^k}{\Phi(\tau(m/n))} \xrightarrow{\mathrm{p}} 0. \tag{11.17}$$

Furthermore, for any $C < \omega$, this holds uniformly as $n \to \infty$ for all $m = m(n)$ with $m/n \leqslant C$.

Theorem 11.7. *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 $m/n \to \lambda$ where $0 \leqslant \lambda < \omega$, and let $(\pi_k)_{k\geqslant 0}$ be as in Theorem 11.4. Then, for every $\ell \geqslant 1$ and $y_1, \dots, y_\ell \geqslant 0$,*

$$\mathbb{P}(Y_1^{(m,n)} = y_1, \dots, Y_\ell^{(m,n)} = y_\ell) \to \prod_{i=1}^{\ell} \pi_{y_i}. \tag{11.18}$$

In other words, for every fixed ℓ, the random variables $Y_1, \dots, Y_\ell$ converge jointly to independent random variables with the distribution $(\pi_k)_{k\geqslant 0}$.

We begin with some lemmas. First we state and prove a version of the local central limit theorem (for integer-valued variables) that is convenient for our application below. We will need it for a triangular array, where the variables we sum depend on n.

We define the span of an integer-valued random variable to be the span of its distribution, defined as in (3.2).

Lemma 14.1. *Let ξ and $\xi^{(1)}, \xi^{(2)}, \dots$ be integer-valued random variables with $\xi^{(n)} \xrightarrow{\mathrm{d}} \xi$ as $n \to \infty$, and let $S_n^{(n)} := \sum_{i=1}^n \xi_i^{(n)}$, where $\xi_i^{(n)}$ are independent copies of $\xi^{(n)}$. Suppose further that ξ is non-degenerate, with span d and finite variance $\sigma^2 > 0$, and that $\sup_n \mathbb{E}|\xi^{(n)}|^3 < \infty$. If $d > 1$, we assume for simplicity that $d \mid \xi$ and $d \mid \xi^{(n)}$ for each n.*

Let $m = m(n)$ be a sequence of integers that are multiples of d, and assume that $\mathbb{E}\,\xi^{(n)} = m(n)/n$. Then, as $n \to \infty$,

$$\mathbb{P}(S_n^{(n)} = m) = \frac{d + o(1)}{\sqrt{2\pi\sigma^2 n}}. \tag{14.1}$$

Proof. The proof uses standard arguments, see e.g. Kolchin [76, Theorem 1.4.2]; we only have to check uniformity in $\xi^{(n)}$ of our estimates.

If the span $d > 1$, we may divide ξ, $\xi^{(n)}$ and m by d, and reduce to the case $d = 1$. Hence we assume in the proof that $\operatorname{span}(\xi) = 1$.

Let $\varphi(t) := \mathbb{E}\,e^{\mathrm{i}t\xi}$ and $\varphi_n(t) := \mathbb{E}\,e^{\mathrm{i}t\xi^{(n)}}$ be the characteristic functions of ξ and $\xi^{(n)}$. Further, let $\widetilde{\varphi}_n(t) := e^{-\mathrm{i}tm/n}\varphi_n(t)$ be the characteristic function of the centred random variable $\xi^{(n)} - \mathbb{E}\,\xi^{(n)} = \xi^{(n)} - m/n$.

Then $S_n^{(n)}$ has characteristic function $\varphi_n(t)^n$, and thus, by the inversion formula and a change of variables,

$$\begin{aligned}
\mathbb{P}(S_n^{(n)} = m) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathrm{i}mt}\varphi_n(t)^n \,\mathrm{d}t = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widetilde{\varphi}_n(t)^n \,\mathrm{d}t \\
&= \frac{1}{2\pi\sqrt{n}}\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \widetilde{\varphi}_n(x/\sqrt{n})^n \,\mathrm{d}x \\
&= \frac{1}{2\pi\sqrt{n}}\int_{-\infty}^{\infty} \widetilde{\varphi}_n(x/\sqrt{n})^n \mathbf{1}\{|x| < \pi\sqrt{n}\} \,\mathrm{d}x.
\end{aligned} \tag{14.2}$$

Let σ_n^2 be the variance of $\xi^{(n)}$. Since $\mathbb{E}|\xi^{(n)}|^3$ are uniformly bounded, $\sigma_n^2 < \infty$; moreover, the random variables $\xi^{(n)}$ are uniformly square integrable and it follows from $\xi^{(n)} \xrightarrow{\mathrm{d}} \xi$ that $\sigma_n^2 \to \sigma^2$. (See e.g. Gut [49, Theorems 5.4.2 and 5.4.9] for this standard argument.) In particular, $\sigma^2/2 \leqslant \sigma_n^2 \leqslant 2\sigma^2$ for all sufficiently large n; we consider in the remainder of the proof only such n.

Since $\widetilde{\varphi}_n(t)$ is the characteristic function of $\xi^{(n)} - \mathbb{E}\,\xi^{(n)}$ which has mean 0 and, by assumption, an absolute third moment that is uniformly bounded, we have by a standard expansion (see e.g. [49, Theorems 4.4.1])

$$\widetilde{\varphi}_n(t) = 1 - \tfrac{1}{2}\sigma_n^2 t^2 + O(\mathbb{E}|\xi^{(n)}|^3 |t|^3) = 1 - \tfrac{1}{2}\sigma_n^2 t^2 + O(|t|^3), \tag{14.3}$$

uniformly in all n and t. In particular, for any fixed real x,

$$\widetilde{\varphi}_n(x/\sqrt{n}) = 1 - \frac{\sigma_n^2 x^2}{2n} + O(n^{-3/2}) = 1 - \frac{\sigma^2 x^2 + o(1)}{2n}, \tag{14.4}$$

and thus
$$\widetilde{\varphi}_n(x/\sqrt{n})^n \to e^{-\sigma^2x^2/2}. \tag{14.5}$$
We are aiming at estimating the integral in (14.2) by dominated convergence, so we also need a suitable bound that is uniform in n.

We write (14.3) as
$$\Bigl|\widetilde{\varphi}_n(t) - (1-\tfrac12\sigma_n^2t^2)\Bigr| \leqslant C_1|t|^3. \tag{14.6}$$
Let $\delta := \min\{\sigma^{-1}, \sigma^2/8C_1\} > 0$. Then, if $|t| \leqslant \delta$, recalling our assumption $\sigma^2/2 \leqslant \sigma_n^2 \leqslant 2\sigma^2$, we have $1-\frac12\sigma_n^2t^2 \geqslant 1-\sigma^2t^2 \geqslant 0$ and, by (14.6),
$$|\widetilde{\varphi}_n(t)| \leqslant 1-\tfrac12\sigma_n^2t^2 + C_1|t|^3 \leqslant 1-\tfrac14\sigma^2t^2 + C_1\delta t^2 \leqslant 1-\tfrac18\sigma^2t^2. \tag{14.7}$$
For $\delta \leqslant |t| \leqslant \pi$ we claim that there exists n_0 and $\eta > 0$ such that if $n \geqslant n_0$ and $\delta \leqslant |t| \leqslant \pi$, then
$$|\widetilde{\varphi}_n(t)| \leqslant 1-\eta. \tag{14.8}$$
In fact, if this were not true, then there would exist sequences $n_k \geqslant k$ and $t_k \in [\delta,\pi]$ (by symmetry, it suffices to consider $t>0$) such that $|\varphi_{n_k}(t_k)| = |\widetilde{\varphi}_{n_k}(t_k)| > 1-1/k$. By considering a subsequence, we may assume that $t_k \to t_\infty$ as $k\to\infty$ for some $t_\infty \in [\delta,\pi]$. Since $\xi_n \xrightarrow{\mathrm{d}} \xi$, $\varphi_{n_k}(t) \to \varphi(t)$ uniformly for $|t| \leqslant \pi$, and thus $\varphi_{n_k}(t_k) \to \varphi(t_\infty)$. It follows that $|\varphi(t_\infty)| = 1$ for some $t_\infty \in [\delta,\pi]$, but this is impossible when $\operatorname{span}(\xi) = 1$, as is well-known (and easily seen from $\mathbb{E}\, e^{\mathrm{i}t_\infty(\xi-\xi')} = |\varphi(t_\infty)|^2 = 1$, where ξ' is an independent copy of ξ). This contradiction shows that (14.8) holds.

We can combine (14.7) and (14.8); we let $c_1 := \min\{\sigma^2/8, \eta/\pi^2\}$ and obtain, for $n \geqslant n_0$,
$$|\widetilde{\varphi}_n(t)| \leqslant 1-c_1t^2 \leqslant \exp(-c_1t^2), \qquad |t| \leqslant \pi,$$
and thus
$$|\widetilde{\varphi}_n(x/\sqrt{n})|^n \leqslant \exp(-c_1x^2), \qquad |x| \leqslant \pi\sqrt{n}.$$
This justifies the use of dominated convergence in (14.2), and we obtain by (14.5)
$$\begin{aligned} 2\pi\sqrt{n}\,\mathbb{P}(S_n^{(n)} = m) &= \int_{-\infty}^{\infty} \widetilde{\varphi}_n(x/\sqrt{n})^n \mathbf{1}\bigl\{|x| < \pi\sqrt{n}\bigr\}\,\mathrm{d}x \\ &\to \int_{-\infty}^{\infty} e^{-\sigma^2x^2/2}\,\mathrm{d}x = \sqrt{2\pi/\sigma^2}, \end{aligned}$$
which yields (14.1). (Recall that we have assumed $d=1$.) $\square$

Remark 14.2. A simple modification of the proof shows that the result still holds if the condition $\mathbb{E}\,\xi^{(n)} = m(n)/n$ is relaxed to $m(n) = n\,\mathbb{E}\,\xi^{(n)} + o(\sqrt{n})$. Furthermore, for any $m = m(n)$, $\mathbb{P}(S_n^{(n)} = m) \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} |\widetilde{\varphi}_n(t)|^n\,\mathrm{d}t$, and it follows by the proof above that
$$\mathbb{P}(S_n^{(n)} = m) \leqslant \frac{d+o(1)}{\sqrt{2\pi\sigma^2 n}}, \tag{14.9}$$

uniformly in all $m \in \mathbb{Z}$.

Moreover, both Lemma 14.1 and the remarks above hold, with only minor modifications in the proof, also if the condition $\sup_n \mathbb{E}\,|\xi^{(n)}|^3 < \infty$ is relaxed to uniform square integrability of $\xi^{(n)}$. In particular, if $\xi^{(n)} = \xi$, this assumption is not needed at all; then the assumption $\sigma^2 < \infty$ is the only moment condition that we need. (This is the classical local central limit theorem for discrete distributions, see e.g. Gnedenko and Kolmogorov [46, § 49] or Kolchin [76, Theorem 1.4.2].)

We use Lemma 14.1 to obtain lower bounds of the (rather weak) type $\exp(o(n))$ for $\mathbb{P}(S_n = m)$ in the case of a probability weight sequence, for suitable m. We treat the cases $\rho > 1$ and $\rho = 1$ separately.

Lemma 14.3. *Let* $\mathbf{w}$ *be a probability weight sequence with* $0 < w_0 < 1$ *and* $\rho > 1$. *Let* $\xi_1, \xi_2, \dots$ *be i.i.d. random variables with distribution* $\mathbf{w}$ *and let* $S_n := \sum_{i=1}^n \xi_i$.

Assume that $m = m(n)$ *are integers that are multiples of* $d := \operatorname{span}(\mathbf{w})$, *and that* $m(n)/n \to \mathbb{E}\,\xi_1$. *Then*

$$\mathbb{P}(S_n = m) = Z(m,n) = e^{o(n)}.$$

Proof. Let $\xi := \xi_1$ and $\lambda := \mathbb{E}\,\xi = \Phi'(1) = \Psi(1)$. Since $\rho > 1$, we have $\nu > \Psi(1) = \lambda$. Thus, by assumption, $m/n \to \lambda < \nu$, so $m/n < \nu$ for all large n; we consider in the sequel only such n. By Lemma 3.1 we may then define $\tau_n \in [0, \rho)$ by $\Psi(\tau_n) = m/n$. Since Ψ^{-1} is continuous on $[0, \nu)$, and $\Psi(1) = \mathbb{E}\,\xi = \lambda$, we have

$$\tau_n = \Psi^{-1}(m/n) \to \Psi^{-1}(\lambda) = 1 \qquad \text{as } n \to \infty. \tag{14.10}$$

Let $\xi^{(n)}$ have the conjugate distribution

$$\mathbb{P}(\xi^{(n)} = k) = \frac{\tau_n^k}{\Phi(\tau_n)} w_k, \qquad k \geqslant 0; \tag{14.11}$$

by Lemma 4.2 this is a probability distribution with expectation

$$\mathbb{E}\,\xi^{(n)} = \Psi(\tau_n) = m/n. \tag{14.12}$$

The conditions of Lemma 14.1 are easily verified: Since $\tau_n \to 1$ by (14.10), we have $\mathbb{P}(\xi^{(n)} = k) \to w_k = \mathbb{P}(\xi = k)$ and thus $\xi^{(n)} \xrightarrow{\mathrm{d}} \xi$. Furthermore, taking any $\tau_* \in (1, \rho)$ and considering only n that are so large that $\tau_n < \tau_*$,

$$\mathbb{E}\,|\xi^{(n)}|^3 = \sum_{k=0}^{\infty} k^3 \frac{\tau_n^k}{\Phi(\tau_n)} w_k \leqslant \frac{1}{\Phi(0)} \sum_{k=0}^{\infty} k^3 \tau_*^k w_k < \infty.$$

Furthermore, if $d = \operatorname{span}(\xi)$, then $w_k > 0 \implies d \mid k$ by (3.3); thus $d \mid \xi$ and $d \mid \xi^{(n)}$ (a.s.). Lemma 14.1 thus applies, and if $\mathbf{w}^{(n)}$ denotes the distribution of $\xi^{(n)}$ in (14.11), then by (11.5) and (14.1),

$$Z(m,n;\mathbf{w}^{(n)}) = \mathbb{P}\biggl(\sum_{i=1}^n \xi_i^{(n)} = m\biggr) \sim \frac{d}{\sqrt{2\pi\sigma^2 n}}, \tag{14.13}$$

where $\sigma^2 := \operatorname{Var}\xi$. By (11.9), we have $Z(m,n;\mathbf{w}^{(n)}) = \Phi(\tau_n)^{-n}\tau_n^m Z(m,n)$, and thus, recalling that $\tau_n \to 1 < \rho$ and hence $\Phi(\tau_n) \to \Phi(1) = 1$,

$$\begin{aligned}\mathbb{P}(S_n = m) = Z(m,n) &= \tau_n^{-m}\Phi(\tau_n)^n Z(m,n;\mathbf{w}^{(n)}) \\ &= \exp\bigl(-m\log\tau_n + n\log\Phi(\tau_n) + \log Z(m,n;\mathbf{w}^{(n)})\bigr) \\ &= \exp\bigl(o(n)\bigr). \qquad \square\end{aligned}$$

Lemma 14.4. *Let* $\mathbf{w}$ *be a probability weight sequence with* $0 < w_0 < 1$ *and* $\rho = 1$. *Let* $\xi_1, \xi_2, \dots$ *be i.i.d. random variables with distribution* $\mathbf{w}$ *and let* $S_n := \sum_{i=1}^n \xi_i$.

Assume that $m = m(n)$ *are integers that are multiples of* $d := \operatorname{span}(\mathbf{w})$, *and that* $m(n)/n \to \lambda < \infty$ *with* $\lambda \geqslant \mathbb{E}\,\xi_1$. *Then*

$$\mathbb{P}(S_n = m) = e^{o(n)}.$$

Proof. Let K be a large integer and consider the truncated weight sequence $\mathbf{w}^{(K)} = (w_k^{(K)})$ defined by, as in (13.4),

$$w_k^{(K)} := \begin{cases} w_k, & k \leqslant K, \\ 0, & k > K, \end{cases} \tag{14.14}$$

having generating function $\Phi_K(t) = \sum_{k=0}^K w_k t^k$, and the corresponding $\Psi_K(t) := t\Phi_K'(t)/\Phi_K(t)$. We assume that K is so large that $\operatorname{span}(\mathbf{w}^{(K)}) = \operatorname{span}(\mathbf{w})$, and that $K > k$ for some $k > \lambda$ with $w_k > 0$. (Such $k > \lambda$ exists since $\rho < \infty$.) Thus the weight sequence $\mathbf{w}^{(K)}$ has, by Lemma 3.1(v), $\nu(\mathbf{w}^{(K)}) = \Psi_K(\infty) = \omega(\mathbf{w}^{(K)}) > \lambda$. Hence, by Lemma 3.1 again, there exists $\tau_K \in [0,\infty)$ such that $\Psi_K(\tau_K) = \lambda$. Thus the probability distribution $\boldsymbol{\pi}^{(K)} = (\pi_k^{(K)})$ defined by

$$\pi_k^{(K)} := \frac{\tau_K^k}{\Phi_K(\tau_K)} w_k^{(K)} \tag{14.15}$$

has expectation λ. Since this distribution has finite support it has radius of convergence $\rho_K = \infty$; furthermore, $m/n \to \lambda$ by assumption. Hence Lemma 14.3 applies to $\boldsymbol{\pi}^{(K)}$ and yields

$$Z(m,n;\boldsymbol{\pi}^{(K)}) = e^{o(n)}. \tag{14.16}$$

By (11.9) and (14.15),

$$Z(m,n;\boldsymbol{\pi}^{(K)}) = \Phi_K(\tau_K)^{-n}\tau_K^m Z(m,n;\mathbf{w}^{(K)}). \tag{14.17}$$

Moreover, $Z(m,n;\mathbf{w}) \geqslant Z(m,n;\mathbf{w}^{(K)})$ since $w_k \geqslant w_k^{(K)}$ for each k. Hence, by (14.16) and (14.17),

$$\begin{aligned}Z(m,n;\mathbf{w}) \geqslant Z(m,n;\mathbf{w}^{(K)}) &= \tau_K^{-m}\Phi_K(\tau_K)^n Z(m,n;\boldsymbol{\pi}^{(K)}) \\ &= \tau_K^{-m}\Phi_K(\tau_K)^n e^{o(n)}. \end{aligned}\tag{14.18}$$

This holds for every large fixed K.

If $0 < t < \rho = 1$, then $\Phi_K(t) \to \Phi(t)$ and $\Phi'_K(t) \to \Phi'(t)$ as $K \to \infty$, so $\Psi_K(t) \to \Psi(t) < \Psi(1) = \mathbb{E}\,\xi_1 \leqslant \lambda$. Hence, for large K, $\Psi_K(t) < \lambda = \Psi_K(\tau_K)$, so $\tau_K > t$. Consequently, $\liminf_{K\to\infty} \tau_K \geqslant 1$.

On the other hand, if $t > \rho = 1$, let $\ell := \lceil\lambda\rceil + 1 > \lambda$, and assume $K \geqslant \ell$. Then

$$\Psi_K(t) = \frac{\sum_{k=0}^K kw_k t^k}{\sum_{k=0}^K w_k t^k} \geqslant \frac{\sum_{k=\ell}^K \ell w_k t^k}{\sum_{k=0}^K w_k t^k} = \ell - \frac{\sum_{k=0}^{\ell-1} \ell w_k t^k}{\Phi_K(t)} \to \ell > \lambda, \tag{14.19}$$

as $K \to \infty$, since $\Phi_K(t) \to \Phi(t) = \infty$. Hence, for large K, $\Psi_K(t) > \lambda = \Psi_K(\tau_K)$, and thus $\tau_K < t$. Consequently, $\limsup_{K\to\infty} \tau_K \leqslant 1$.

Combining these upper and lower bounds, we have

$$\tau_K \to 1, \qquad \text{as } K \to \infty.$$

If we take $t < 1$, we thus have for large K, $\tau_K > t$ and hence $\Phi_K(\tau_K) > \Phi_K(t)$. Thus, $\liminf_{K\to\infty} \Phi_K(\tau_K) \geqslant \lim_{K\to\infty} \Phi_K(t) = \Phi(t)$ for every $t < 1$, so

$$\liminf_{K\to\infty} \Phi_K(\tau_K) \geqslant \Phi(1) = 1.$$

Given any $\varepsilon > 0$, we may thus take K so large that $\tau_K < e^{\varepsilon}$ and $\Phi_K(\tau_K) > e^{-\varepsilon}$. Then (14.18) yields

$$Z(m, n; \mathbf{w}) \geqslant e^{-\varepsilon m - \varepsilon n + o(n)} \geqslant e^{-\varepsilon m - 2\varepsilon n}$$

for large n. Since ε is arbitrary and $m = O(n)$, this shows $Z(m, n; \mathbf{w}) \geqslant e^{o(n)}$, and the result follows since $Z(m, n) \leqslant 1$ for any probability weight sequence by (11.5). □

We next prove Theorems 11.4 and 11.6. Theorem 11.6 follows easily from Theorem 11.4, so it may seem natural to prove Theorem 11.4 first. However, our proof of Theorem 11.4 uses in one case Theorem 11.6 (for another case). We will therefore first show that Theorem 11.6 follows from Theorem 11.4, and then show Theorem 11.4.

Proof of Theorem 11.6 from Theorem 11.4. We prove that Theorem 11.4 for some weight sequence (w_k) implies Theorem 11.6 for the same weights. The assertions about τ follow from Lemma 13.2, so we turn to (11.17).

Consider a subsequence of $(m(n), n)$. It suffices to show that every such subsequence has a subsubsequence such that (11.17) holds. (See e.g. [49, Section 5.7], [65, p. 12] or [15, Theorem 2.3] for this standard argument.)

Since $m/n \leqslant C$ by assumption, we can select a subsubsequence such that $m/n \to \lambda$ for some $\lambda \leqslant C < \omega$. Then Theorem 11.4 applies and thus (along the subsubsequence),

$$\frac{N_k(B_{m,n})}{n} - \frac{w_k(\tau(\lambda))^k}{\Phi(\tau(\lambda))} = \frac{N_k(B_{m,n})}{n} - \pi_k \xrightarrow{\text{p}} 0. \tag{14.20}$$

Furthermore, since $m/n \to \lambda$ and $x \mapsto \tau(x)$ is continuous, $\tau(m/n) \to \tau(\lambda)$ (along the subsubsequence); hence

$$\frac{w_k(\tau(m/n))^k}{\Phi(\tau(m/n))} - \frac{w_k(\tau(\lambda))^k}{\Phi(\tau(\lambda))} \to 0. \tag{14.21}$$

Combining (14.20) and (14.21), we see that (11.17) holds along the subsubsequence, which as said above completes the proof of (11.17).

That (11.17) holds uniformly is, in fact, automatic since we have shown it for an arbitrary $m(n)$ (although we stated it for emphasis): Let $X_{m,n}$ denote the left-hand side of (11.17), and let $\varepsilon > 0$. Choose $m(n)$ as the integer $m \in [0, Cn]$ that maximises $\mathbb{P}(|X_{m,n}| > \varepsilon)$. Since (11.17) says that $\mathbb{P}(|X_{m(n),n}| > \varepsilon) \to 0$, we have $\sup_{m \leqslant Cn} \mathbb{P}(|X_{m,n}| > \varepsilon) \to 0$. □

Proof of Theorem 11.4. First, Lemma 13.2 shows that τ defined by (i) and (ii) is well-defined and equals $\tau(\lambda)$ defined in Lemma 13.2; since $\lambda < \omega$ we have $\tau < \infty$ and $\Phi(\tau) < \infty$. Further, (13.3) yields

$$\Psi(\tau) = \min(\lambda, \nu). \tag{14.22}$$

Since $\tau < \infty$ and $\Phi(\tau) < \infty$, π_k is well-defined by (11.13); furthermore, by Lemma 4.2 and (14.22), (π_k) is a probability distribution with mean and variance as asserted.

We now turn to proving (11.15), the main assertion. We study three cases separately.

Case (a): $\tau > 0$. Then $\boldsymbol{\pi} = (\pi_k)$ is a probability weight sequence equivalent to $\mathbf{w} = (w_k)$, so we may replace (w_k) by (π_k) without changing $B_{m,n}$. Note that this changes ρ and τ to $\rho(\boldsymbol{\pi}) = \rho(\mathbf{w})/\tau$ and $\tau(\boldsymbol{\pi}) = \tau(\mathbf{w})/\tau = 1$ by (4.4) and (4.5). We may thus assume that (w_k) equals the probability weight sequence (π_k), and that $\rho \geqslant \tau = 1$. By (14.22), then $\Psi(1) = \min(\lambda, \nu)$.

We employ the notation of Example 11.2. Note that by (11.14),

$$\mathbb{E}\,\xi_1 = \Psi(1) = \min(\lambda, \nu) \leqslant \lambda. \tag{14.23}$$

Moreover, if $\rho > 1$, then $\nu = \Psi(\rho) > \Psi(1)$ by Lemma 3.1, so (14.22) shows that in this case,

$$\mathbb{E}\,\xi_1 = \Psi(1) = \lambda. \tag{14.24}$$

The allocation $(\xi_1, \dots, \xi_n)$ (with a random sum S_n) consists of n i.i.d. components, so

$$N_k(\xi_1, \dots, \xi_n) = \sum_{i=1}^{n} \mathbf{1}\{\xi_i = k\} \sim \mathrm{Bi}(n, \pi_k) \tag{14.25}$$

has a binomial distribution. For every k and $\varepsilon > 0$, we have by Chernoff's inequality, see e.g. [65, Theorem 2.1 or Remark 2.5],

$$\mathbb{P}\big(|N_k(\xi_1, \dots, \xi_n) - n\pi_k| > \varepsilon n\big) \leqslant \exp(-c_\varepsilon n), \tag{14.26}$$

for some constant $c_\varepsilon > 0$ depending on ε.

We condition on $S_n = m$, recalling that

$$B_{m,n} \overset{\mathrm{d}}{=} \bigl((\xi_1, \dots, \xi_n) \mid S_n = m\bigr). \tag{14.27}$$

When $\rho > 1$ we apply Lemma 14.3, using $m/n \to \lambda$ and (14.24), and when $\rho = 1$ we apply Lemma 14.4, using (14.23). In both cases we obtain $\mathbb{P}(S_n = m) = \exp(o(n))$ and thus by (14.27),

$$\begin{aligned}\mathbb{P}\bigl(|N_k(B_{m,n}) - n\pi_k| > \varepsilon n\bigr) &= \mathbb{P}\bigl(|N_k(\xi_1, \dots, \xi_n) - n\pi_k| > \varepsilon n \mid S_n = m\bigr) \\ &\leqslant \frac{\mathbb{P}\bigl(|N_k(\xi_1, \dots, \xi_n) - n\pi_k| > \varepsilon n\bigr)}{\mathbb{P}(S_n = m)} \leqslant \exp\Bigl(-c_\varepsilon n + o(n)\Bigr) \to 0.\end{aligned}$$

Since ε is arbitrary, this shows that

$$\frac{N_k(B_{m,n})}{n} - \pi_k \xrightarrow{\mathrm{p}} 0$$

as asserted, which completes the proof when $\tau > 0$.

Case (b): $\tau = 0$ and $\rho > 0$. We write N_k for $N_k(B_{m,n})$. By (11.13) we have $\pi_0 = 1$ and $\pi_k = 0$ for $k > 0$; hence, (11.15) says that $N_0/n \xrightarrow{\mathrm{p}} 1$ and $N_k/n \xrightarrow{\mathrm{p}} 0$ for $k > 0$.

Since $\tau < \rho$, we are in case (i), so $\lambda = \Psi(\tau) = \Psi(0) = 0$. In other words, $m/n \to 0$. The result is trivial (and deterministic) in this case. We have

$$\frac{1}{n}\sum_{k=1}^{\infty} N_k \leqslant \frac{1}{n}\sum_{k=1}^{\infty} kN_k = \frac{m}{n} \to \lambda = 0. \tag{14.28}$$

Hence $N_k/n \to 0 = \pi_k$ for every $k \geqslant 1$. Moreover, (14.28) also implies

$$\frac{N_0}{n} = \frac{n - \sum_{k=1}^{\infty} N_k}{n} \to 1 = \pi_0, \tag{14.29}$$

which completes the proof when $\tau = 0 < \rho$.

Case (c): $\rho = 0$. We write again N_k for $N_k(B_{m,n})$, recalling that this is a random variable. In this case $\nu = 0$ and $\tau = \rho = 0$ for every $\lambda \geqslant 0$. By (11.13) we thus have $\pi_0 = 1$ and $\pi_k = 0$ for $k > 0$; hence, as in case (b), we have to show that $N_0/n \xrightarrow{\mathrm{p}} 1$ and $N_k/n \xrightarrow{\mathrm{p}} 0$ for $k > 0$. By assumption, m/n converges, so the sequence m/n is bounded; let C be a large constant such that $m/n \leqslant C$. Further, let K be a large integer; we assume $K > 2C$ and (for simplicity) $w_K > 0$. (Note that such K exist since $\omega = \infty$ when $\rho = 0$.)

We say that a box is *small* if it contains at most K balls, and *large* otherwise. Let $N' := \sum_0^K N_k$ be the number of small boxes and $M' := \sum_0^K kN_k$ the number of balls in them. Note first that by our assumptions, $m/n \leqslant C < K/2$. Hence,

$$m \geqslant m - M' = \sum_{K+1}^{\infty} kN_k \geqslant K\sum_{K+1}^{\infty} N_k = K(n - N') \geqslant \frac{2m}{n}(n - N'). \tag{14.30}$$

Thus, $n - N' \leqslant n/2$ and $N' \geqslant n/2$; in particular $N' \to \infty$. Moreover,

$$0 \leqslant \frac{M'}{N'} \leqslant \frac{m}{n/2} \leqslant 2C < K. \tag{14.31}$$

The weight $w(\mathbf{y})$ in (11.2) factorizes as the product over the small boxes times the product over the large boxes. Thus, if we condition on M' and N', and moreover on the set of the N' boxes that are small, then the allocations of the small boxes and the large boxes are independent; moreover, the allocations to the small boxes form a random allocation of the type $B_{M',N'}$ for the truncated weight sequence $\mathbf{w}^{(K)}$ given by (13.4) above. By assumption, $w_K > 0$, and thus the truncated sequence has $\omega^{(K)} := \omega(\mathbf{w}^{(K)}) = K$.

The truncated weight sequence $\mathbf{w}^{(K)}$ has a polynomial generating function $\Phi^{(K)}(t) = \sum_0^K w_k t^k$ with an infinite radius of convergence $\rho^{(K)} = \infty$. We have already proved Theorem 11.4 in this case, and thus Theorem 11.6 also holds in this case, by the proof above. Applying Theorem 11.6 to the truncated weight sequence and the allocations of small boxes we see that there exists a continuous function $\tau_K : [0, K) \to [0, \infty)$ such that, conditioned on (M', N'),

$$\frac{N_k}{N'} - \frac{w_k(\tau_K(M'/N'))^k}{\Phi^{(K)}\bigl(\tau_K(M'/N')\bigr)} \xrightarrow{\mathrm{p}} 0, \qquad k \leqslant K. \tag{14.32}$$

Moreover, (14.32) holds uniformly in all (M', N') by Theorem 11.6 and (14.31). Hence, denoting the left-hand side of (14.32) by X, we have for every $\varepsilon > 0$ $\mathbb{P}(|X| > \varepsilon \mid M', N') \leqslant \delta(n)$, for some function $\delta(n) \to 0$. Taking the expectation, it follows that also $\mathbb{P}(|X| > \varepsilon) \leqslant \delta(n) \to 0$, and thus (14.32) holds also unconditionally. Thus,

$$\frac{N_k}{N'} = \frac{w_k(\tau_K(M'/N'))^k}{\Phi^{(K)}\bigl(\tau_K(M'/N')\bigr)} + o_{\mathrm{p}}(1), \qquad k \leqslant K. \tag{14.33}$$

By (14.31), $M'/N' \leqslant 2C$, and thus, using Lemma 13.2 and $2C < K = \omega(\mathbf{w}^{(K)})$, $\tau_K(M'/N') \leqslant \tau_K(2C) < \infty$. Hence, with $C_1 := \tau_K(2C)$,

$$w_0 \leqslant \Phi^{(K)}(\tau_k(M'/N')) \leqslant \Phi^{(K)}(C_1) = C_2,$$

say. Taking $k = 0$ in (14.33) we now find

$$\frac{N_0}{N'} = \frac{w_0}{\Phi^{(K)}\bigl(\tau_K(M'/N')\bigr)} + o_{\mathrm{p}}(1) \geqslant \frac{w_0}{C_2} + o_{\mathrm{p}}(1). \tag{14.34}$$

Since $N' \geqslant n/2$ this shows that there exists $c_1 > 0$ (for example $c_1 := w_0/(3C_2)$) such that w.h.p.

$$\frac{N_0}{n} \geqslant c_1. \tag{14.35}$$

It follows further from (14.34) that we can invert (14.33) for $k = 0$ (since $x \mapsto x^{-1}$ is continuous for $x > 0$); thus

$$\frac{N'}{N_0} = \frac{\Phi^{(K)}\bigl(\tau_K(M'/N')\bigr)}{w_0} + o_{\mathrm{p}}(1). \tag{14.36}$$

Multiplying (14.33) and (14.36) we find the simpler relation

$$\frac{N_k}{N_0} = \frac{w_k}{w_0}(\tau_K(M'/N'))^k + o_{\mathrm{p}}(1), \qquad k \leqslant K. \tag{14.37}$$

Let $\ell := \min\{k > 0 : w_k > 0\}$ be the smallest non-zero index with positive weight, and define a random variable by

$$\tau_* := \left(\frac{w_0 N_\ell}{w_\ell N_0}\right)^{1/\ell}. \tag{14.38}$$

It follows from (14.37), with $k = \ell$, that $\tau_* = \tau_K(M'/N') + o_{\mathrm{p}}(1)$. Consequently, (14.37) yields

$$\frac{N_k}{N_0} = \frac{w_k}{w_0}\tau_*^k + o_{\mathrm{p}}(1), \qquad k \leqslant K. \tag{14.39}$$

We have so far worked with a fixed, large K. However, the definition (14.38) does not depend on the choice of K, and since K may be chosen arbitrarily large, we see that, in fact, (14.39) holds for every $k \geqslant 0$, with the same (random) τ_*.

Fix again $K > 0$, and sum (14.39) for $k \leqslant K$. This yields

$$\frac{n}{N_0} \geqslant \sum_0^K \frac{N_k}{N_0} = \sum_0^K \frac{w_k}{w_0}\tau_*^k + o_{\mathrm{p}}(1) = \frac{\Phi^{(K)}(\tau_*)}{w_0} + o_{\mathrm{p}}(1). \tag{14.40}$$

Recall that $N_0/n \geqslant c_1$ w.h.p. by (14.35). We thus have from (14.40)

$$\Phi^{(K)}(\tau_*) \leqslant w_0\frac{n}{N_0} + o_{\mathrm{p}}(1) \leqslant w_0/c_1 + 1 \tag{14.41}$$

w.h.p. By assumption, $\rho = 0$, so $\Phi(t) = \infty$ for every $t > 0$. Hence, for every $\varepsilon > 0$ we have $\Phi^{(K)}(\varepsilon) \to \Phi(\varepsilon) = \infty$ as $K \to \infty$, so we may choose K with $\Phi^{(K)}(\varepsilon) > w_0/c_1 + 1$. Then (14.41) shows that $\tau_* < \varepsilon$ whp; since $\varepsilon > 0$ is arbitrary, this says that

$$\tau_* \xrightarrow{\mathrm{p}} 0.$$

We substitute this in (14.39), and obtain $N_k/N_0 \xrightarrow{\mathrm{p}} 0$ for every $k \geqslant 1$; hence also

$$N_k/n \xrightarrow{\mathrm{p}} 0, \qquad k \geqslant 1. \tag{14.42}$$

Finally, we return to (14.30), and see that

$$K(n - N') \leqslant m \leqslant Cn. \tag{14.43}$$

Let $\varepsilon > 0$ and choose $K > C/\varepsilon$; then (14.43) yields $n - N' < \varepsilon n$ and thus $N' > (1-\varepsilon)n$. Further, by (14.42),

$$N_0 = N' - \sum_1^K N_k = N' + o_{\mathrm{p}}(n) > (1-\varepsilon)n + o_{\mathrm{p}}(n),$$

so w.h.p. $N_0 > (1-2\varepsilon)n$. This shows that $N_0/n \xrightarrow{\mathrm{p}} 1$, which together with (14.42) completes the proof in the case $\rho = 0$. ☐

This completes the proof of Theorem 11.4, and thus also of Theorem 11.6.

Proof of Theorem 11.7. Conditioned on the numbers $N_k = N_k(B_{m,n})$, $k = 0, 1, \dots$, the numbers $Y_1, \dots, Y_n$ are obtain by placing N_0 0's, N_1 1's, $\dots$, in (uniformly) random order; thus the conditional probability is

$$\mathbb{P}(Y_1 = y_1, \dots, Y_\ell = y_\ell \mid N_0, N_1, \dots) = \prod_{i=1}^{\ell} \frac{N_{y_i} - c_i}{n - i + 1} = \prod_{i=1}^{\ell} \frac{N_{y_i} + O(1)}{n + O(1)}, \tag{14.44}$$

where $c_i := |\{j < i : y_j = y_i\}|$. By Theorem 11.4, this product converges in probability to $\prod_{i=1}^{\ell} \pi_{y_i}$ as $n \to \infty$, and the result follows by taking the expectation (using dominated convergence). □

15. Trees and balls-in-boxes

The proofs of the results for random trees are based on a connection with the balls-in-boxes model. This connection is well-known, see e.g. Otter [93], Dwass [36], Kolchin [76], Pitman [99], but for completeness we give full proofs.

We consider a fixed weight sequence $\mathbf{w} = (w_k)$ and the corresponding random trees $\mathcal{T}_n$ and random allocations $B_{m,n}$; we write as above $B_{m,n} = (Y_1, \dots, Y_n) = (Y_1^{(m,n)}, \dots, Y_n^{(m,n)})$.

We begin with some deterministic considerations. The idea is to regard the outdegrees of the nodes of a tree T as an allocation; we regard the nodes as both balls and boxes, and if v is a node, we put the children of v as balls in box v. There are two complications, which will be dealt with in detail below: we have to specify an ordering of the nodes and we will not obtain all allocations.

Let T be a finite tree, with $|T| = n$. Take the nodes in some prescribed order $v_1, \dots, v_n$, for definiteness we use the depth-first order (this is the lexicographic order on V_∞), and list the outdegrees as $d_1 = d^+(v_1), \dots, d_n = d^+(v_n)$. We call this the *degree sequence* of T and denote it by $\Lambda(T) := (d_1, \dots, d_n)$. Note that the tree T can be reconstructed from $(d_1, \dots, d_n)$, so T is determined by $\Lambda(T) = (d_1, \dots, d_n)$.

By (2.2), $d_1 + \dots + d_n = n - 1$, so $(d_1, \dots, d_n)$ can be seen as an allocation of $n - 1$ balls in n boxes: $\Lambda(T) = (d_1, \dots, d_n) \in \mathcal{B}_{n-1,n}$. Consequently, Λ is an injective map $\mathfrak{T}_n \to \mathcal{B}_{n-1,n}$. Note also that Λ preserves the weight:

$$w(T) = w(\Lambda(T)) \tag{15.1}$$

by the definitions (2.3) and (11.2). However, not every allocation corresponds to a tree, so Λ is not onto. We begin by characterizing the image $\Lambda(\mathfrak{T}_n)$. We use a simple and well-known extension of (2.2).

Lemma 15.1. *Let T be a tree and T' a subtree with the same root. Let $\partial T' := \{v \in V(T) \setminus V(T') : v \sim w \text{ for some } w \in T'\}$ be the set of nodes outside T' with a parent inside it. Then,*

$$\sum_{v \in T'} d_T^+(v) = |T'| + |\partial T'| - 1. \tag{15.2}$$

Proof. The set of children of the nodes in T' consists of $\bigl(V(T') \setminus \{o\}\bigr) \cup \partial T'$. □

Lemma 15.2. *A sequence $(d_1, \dots, d_n) \in \mathbb{N}_0^n$ is the degree sequence of a tree $T \in \mathfrak{T}_n$ if and only if*

$$\sum_{i=1}^{k} d_i \geqslant k, \qquad 1 \leqslant k < n, \tag{15.3}$$

$$\sum_{i=1}^{n} d_i = n - 1. \tag{15.4}$$

Of course, (15.4) is just the requirement that $(d_1, \dots, d_n) \in \mathcal{B}_{n-1,n}$.

Proof. For any $k \leqslant n$, the nodes $v_1, \dots, v_k$ form a subtree T_k of T, and Lemma 15.1 yields

$$\sum_{i=1}^{k} d_T^+(v_i) = |\partial T_k| + k - 1, \tag{15.5}$$

which yields (15.3) since $|\partial T_k| \geqslant 1$ when $k < n$.

Conversely, if $(d_1, \dots, d_n)$ satisfies (15.3)–(15.4), a tree with degree sequence $(d_1, \dots, d_n)$ is easily constructed. (We construct the tree by assigning degrees in depth-first order. First, v_1 is the root and gets d_1 children. Next, v_2 is the first child of the root and gets d_2 children. If $d_2 > 0$, then v_3 is the first child of v_2, but if $d_2 = 0$, we backtrack and let v_3 be the second child of v_1; in any case, v_3 gets d_3 children, and so on. The point is that (15.3) assures that the construction will not stop before we have n nodes.) □

The amazing fact is that for any allocation in $\mathcal{B}_{n-1,n}$, exactly one of its cyclic shifts satisfies (15.3). (In particular, exactly $1/n$ of all allocations satisfy (15.3).) To see this, it is simplest to consider the sequence $(d_i - 1)_{i=1}^n$; we state a more general result that we will use later, see e.g. Takács [105], Wendel [108], Pitman [99].

Lemma 15.3. *Let $x_1, \dots, x_n \in \{-1, 0, 1, \dots\}$ with $x_1 + \dots + x_n = -r \leqslant 0$. For $j \in \mathbb{Z}$, let $x_1^{(j)}, \dots, x_n^{(j)}$ be the cyclic shift defined by $x_i^{(j)} := x_{i+j}$ with the index taken modulo n, and consider the corresponding partial sums $S_k^{(j)} := \sum_{i=1}^k x_i^{(j)}$, $k = 0, \dots, n$. Then there are exactly r values of $j \in \{1, \dots, n\}$ such that*

$$S_k^{(j)} > -r, \qquad 0 \leqslant k < n. \tag{15.6}$$

Note that $S_0^{(j)} = 0$ and $S_n^{(j)} = -r$ for every i. The condition (15.6) thus says that the walk $S_0^{(j)}, \dots, S_n^{(j)}$ first reaches $-r$ at time n. The case $r = 0$ is trivial: since $S_0^{(j)} = 0$, (15.6) then is never satisfied for $k = 0$.

Proof. We extend the definition of x_j for all $j \in \mathbb{Z}$ by taking the index modulo n; thus $x_{j+n} = x_j$. We further define S_k for all $k \in \mathbb{Z}$ by $S_0 = 0$ and $S_k - S_{k-1} = x_k$, $k \in \mathbb{Z}$; thus $S_k = \sum_{i=1}^{k} x_i$ when $k \geqslant 0$ and $S_k = -\sum_{i=k+1}^{0} x_i$ when $k < 0$. Then $S_{k+n} = S_k - r$ for all $k \in \mathbb{Z}$, and $S_k^{(j)} = S_{k+j} - S_j$.

Let further

$$M_k := \min_{-\infty < i \leqslant k} S_i = \min_{k-n < i \leqslant k} S_i;$$

note that M_k is finite and $M_{k+n} = M_k - r$. Moreover, $M_{k+1} \leqslant M_k$ and $M_{k+1} - M_k$ is 0 or -1, since $S_{k+1} = S_k + x_{k+1} \geqslant S_k - 1$. We have

$$\begin{aligned} S_k^{(j)} > -r, \text{ for } 0 \leqslant k < n &\iff S_{k+j} - S_j > -r, \text{ for } 0 \leqslant k < n \\ &\iff S_{k+j} + r > S_j, \text{ for } 0 \leqslant k < n \\ &\iff S_{k+j-n} > S_j, \text{ for } 0 \leqslant k < n \\ &\iff S_i > S_j, \text{ for } j - n \leqslant i < j \\ &\iff M_{j-1} > S_j \\ &\iff M_{j-1} > M_j. \end{aligned}$$

In each interval of n integers, M decreases by r in steps of 1, so there are exactly r steps down, which completes the proof. □

Corollary 15.4. *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 $T \in \mathfrak{T}_n$.*

Proof. Let $x_i := d_i - 1$. Then $\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} d_i - k$, so (15.3) is equivalent to $\sum_{i=1}^{k} x_i \geqslant 0$ for $k < n$, which for the shifted sequence is (15.6) with $r = 1$; further, $\sum_{i=1}^{n} x_i = n - 1 - n = -1$. Hence the result follows by Lemma 15.3 with $r = 1$. □

We now use our fixed weight sequence (w_k). We begin with the partition function for simply generated trees. This was proved (in the probability weight sequence case, which is no real loss of generality) by Otter [93], see also Dwass [36]; an algebraic proof uses the Lagrange inversion formula [79], see e.g. Boyd [19] and Drmota [33, Theorem 2.11]; Kolchin [76] gives a different proof by induction. See also Pitman [99] where the relation between different approaches is discussed.

Theorem 15.5.

$$Z_n = \frac{1}{n} Z(n-1, n).$$

Proof. By Corollary 15.4, the mapping $(T, j) \mapsto \Lambda(T)^{(j)}$, where ${}^{(j)}$ denotes a cyclic shift as in Lemma 15.3, is a bijection of $\mathfrak{T}_n \times \{1, \dots, n\} \to \mathcal{B}_{n-1,n}$. Consequently, by (11.4), (15.1) and (2.5), since the weight $w(\mathbf{y})$ is not changed by cyclic shifts,

$$Z(n-1, n) = \sum_{T \in \mathfrak{T}_n} \sum_{j=1}^{n} w\bigl(\Lambda(T)^{(j)}\bigr) = \sum_{T \in \mathfrak{T}_n} n w(\Lambda(T)) = \sum_{T \in \mathfrak{T}_n} n w(T) = n Z_n.$$

□

Corollary 15.6. *Suppose that $w_0 > 0$ and $\omega(\mathbf{w}) \geqslant 2$, with $d := \operatorname{span}(\mathbf{w}) \geqslant 1$. If $Z_n > 0$, then $n \equiv 1 \pmod d$. Conversely, for some n_0 (depending on $\mathbf{w}$), if $n \equiv 1 \pmod d$ and $n \geqslant n_0$, then $Z_n > 0$.*

Proof. By Theorem 15.5, $Z_n > 0 \iff Z(n-1,n) > 0$. The result follows from Lemma 13.3. □

In the same way we can compute various probabilities for the random tree $\mathcal{T}_n$. We begin with the root degree $d^+(o)$; note that for any tree T, v_1 is the root o, so $d^+(o) = d^+(v_1) = d_1$.

Lemma 15.7. *For any $d \geqslant 0$ and $n \geqslant 2$,*

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = d) = \frac{n}{n-1} d\,\mathbb{P}(Y_1^{(n-1,n)} = d). \tag{15.7}$$

Thus, the distribution of the root degree $d^+_{\mathcal{T}_n}(o)$ of $\mathcal{T}_n$ is the size-biased distribution of $Y_1^{(n-1,n)}$.

Lemma 15.7 is a special case of Lemma 15.9 below, but we prefer to study this simpler case first because it shows the main ideas in the proof without the complications (notational and others) in the more general version.

Proof. Consider an allocation $(d_1,\dots,d_n) \in \mathcal{B}_{n-1,n}$; if $d_1 = d$, then $d_2,\dots,d_n$ is an allocation in $\mathcal{B}_{n-1-d,n-1}$. Furthermore, by Lemma 15.2, an allocation $(d_2,\dots,d_n) \in \mathcal{B}_{n-1-d,n-1}$ is obtained by dropping the first term from the degree sequence of a tree $T \in \mathfrak{T}_n$ with $d_1 = d$ if and only if

$$d + \sum_{i=2}^{k} d_i \geqslant k, \qquad 1 \leqslant k < n, \tag{15.8}$$

or, equivalently,

$$\sum_{i=1}^{k} d_{i+1} \geqslant k + 1 - d, \qquad 0 \leqslant k < n-1.$$

We use Lemma 15.3 again, now with $x_i = d_{i+1} - 1$ and $r = d$ and see that for any $(d_2,\dots,d_n) \in \mathcal{B}_{n-1-d,n-1}$, exactly $r = d$ of the $n-1$ cyclic shifts of $d_2,\dots,d_n$ satisfy (15.8). Thus, by considering all trees T with $d_1 = d$ and the $n-1$ cyclic shifts of $d_2,\dots,d_n$, we obtain each allocation $(d_1,\dots,d_n) \in \mathcal{B}_{n-1,n}$ with $d_1 = d$ exactly $r = d$ times. (It is possible that some shifts of $(d_2,\dots,d_n)$ coincide, but this does not matter.) Consequently, using (2.4) and (11.3), and recalling $d^+(o) = d_1$,

$$\begin{aligned}
(n-1)Z_n\,\mathbb{P}(d^+_{\mathcal{T}_n}(o) = d) &= (n-1) \sum_{T \in \mathfrak{T}_n:\, d_1(T) = d} w(T) \\
&= d \sum_{(d_1,\dots,d_n) \in \mathcal{B}_{n-1,n}:\, d_1 = d} w\big((d_1,\dots,d_n)\big) \\
&= d\, Z(n-1,n)\,\mathbb{P}(Y_1 = d).
\end{aligned}$$

This yields the result by Theorem 15.5. □

Remark 15.8. More explicitly we have

$$\begin{aligned}Z(n-1,n)\,\mathbb{P}(Y_1^{(n-1,n)}=d) &= \sum_{(d_1,\dots,d_n)\in\mathcal{B}_{n-1,n}:\,d_1=d} w\big((d_1,\dots,d_n)\big)\\ &= \sum_{(d_2,\dots,d_n)\in\mathcal{B}_{n-1-d,n-1}} w_d w\big((d_2,\dots,d_n)\big)\\ &= w_d Z(n-1-d,n-1),\end{aligned}$$

and thus

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o)=d) = dw_d\frac{n}{n-1}\cdot\frac{Z(n-1-d,n-1)}{Z(n-1,n)}. \tag{15.9}$$

Proof of Theorem 7.10. We have $\mathbb{P}(Y_1^{(n-1,n)}=d)\to\pi_d$ by Theorem 11.7 (with $m=n-1$ and $\lambda=1$), and (7.9) follows from Lemma 15.7.

The space $\overline{\mathbb{N}}_0$ is compact, so every sequence of random variables in it is tight, and therefore has a subsequence converging in distribution, see [15, Section 6]. It follows from (7.9) that if $d^+_{\mathcal{T}_n}(o)\xrightarrow{\mathrm{d}}X$ along a subsequence, then $\mathbb{P}(X=k)=k\pi_k$ for every $k\in\mathbb{N}_0$, and thus $\mathbb{P}(X=\infty)=1-\sum_{k=0}^\infty k\pi_k=1-\mu$. Consequently, $X\overset{\mathrm{d}}{=}\widehat{\xi}$ so $d^+_{\mathcal{T}_n}(o)\xrightarrow{\mathrm{d}}\widehat{\xi}$ for every convergent subsequence, which means that the entire sequence converges to $\widehat{\xi}$, see [15, Theorem 2.3]. ☐

This proves the part of Theorem 7.1 that describes the root degree. It remains to consider all other nodes. This will be done by similar arguments. We begin with a generalization of Lemma 15.7.

Lemma 15.9. *Let $T'\in\mathfrak{T}_{\mathrm{f}}$ be a fixed finite subtree of the Ulam–Harris tree U_∞, let $\ell:=|T'|$ be its size and let $v_1,\dots,v_\ell$ be its nodes in depth-first order, and let $d'_1,\dots,d'_\ell$ be its degree sequence. (I.e., $d'_i=d^+_{T'}(v_i)$.) Suppose that $d_1,\dots,d_\ell\in\mathbb{N}_0$ and that $d_i\geqslant d'_i$ for every i. Then, for every $n>\ell$,*

$$\begin{aligned}&\mathbb{P}\big(d^+_{\mathcal{T}_n}(v_i)=d_i \text{ for } i=1,\dots,\ell\big)\\ &\qquad=\Big(\sum_{i=1}^{\ell}d_i-\ell+1\Big)\frac{n}{n-\ell}\,\mathbb{P}\Big(Y_i^{(n-1,n)}=d_i \text{ for } i=1,\dots\ell\Big).\end{aligned} \tag{15.10}$$

Note that $d^+_T(v_i)\geqslant d'_i$ for $i=1,\dots,\ell$ implies that $T\supset T'$.

Proof. We have earlier used the depth-first order of the nodes to define the degree sequence, but many other orders could be used. In this proof, we consider only trees T that contain the given T' as a subtree, and then we choose the order which first takes the nodes of T' in depth-first order (this is $v_1,\dots,v_\ell$), and then the remaining nodes of T in depth-first order; let $\Lambda'(T)$ be the degree sequence in this order.

Let $\mathcal{A}_n$ be the set of trees $T\in\mathfrak{T}_n$ with $d^+_T(v_i)=d_i$ for all i (which implies $T\supset T'$). If $T\in\mathcal{A}_n$, then the degree sequence $\Lambda'(T)$ thus begins with the given $d_1,\dots,d_\ell$; furthermore, it satifies (15.3)–(15.4). Conversely, every sequence beginning with the given $d_1, \dots, d_\ell$ that satifies (15.3)–(15.4) is the degree sequence $\Lambda'(T)$ of a unique tree in $\mathcal{A}_n$. Note also that (15.3) is automatically satisfied for $k < \ell$, since then $d_i \geqslant d'_i$ for $i \leqslant k$ and $\sum_{i=1}^k d'_i \geqslant k$ by Lemma 15.2 applied to T'.

Let $D := d_1 + \dots + d_\ell$. Consider a sequence $(d_1, \dots, d_n) \in \mathcal{B}_{n-1,n}$ beginning with the given $d_1, \dots, d_\ell$, and let $x_i := d_{\ell+i} - 1$, for $i = 1, \dots, n - \ell$. Then $(d_1, \dots, d_n)$ satisfies (15.3) if and only if

$$D + \sum_{i=1}^{k} (x_i + 1) \geqslant \ell + k$$

for $k = 0, \dots, n - \ell - 1$, which is equivalent to

$$\sum_{i=1}^{k} x_i \geqslant -(D - \ell), \qquad 0 \leqslant k < n - \ell.$$

Furthermore,

$$\sum_{i=1}^{n-\ell} x_i = \sum_{\ell+1}^{n} d_i - (n - \ell) = (n - 1 - D) - (n - \ell) = -(D - \ell + 1).$$

Lemma 15.3 with $r = D - \ell + 1$ thus shows that of the $n - \ell$ cyclic permutations of $d_{\ell+1}, \dots, d_n$, exactly $D - \ell + 1$ yield a degree sequence $\Lambda'(T)$ of a tree $T \in \mathcal{A}_n$. In other words, if we take the degree sequences $\Lambda'(T)$ for all trees $T \in \mathcal{A}_n$ and make these $n - \ell$ permutations of each of them, then we obtain every allocation $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{B}_{n-1,n}$ with $y_i = d_i$, $i = 1, \dots, \ell$, exactly $D - \ell + 1$ times each. Consequently,

$$\begin{aligned}
(n-\ell) Z_n \,\mathbb{P}(\mathcal{T}_n \in \mathcal{A}_n) &= (n-\ell) \sum_{T \in \mathcal{A}_n} w(T) = \sum_{T \in \mathcal{A}_n} (n-\ell) w(\Lambda'(T)) \\
&= \sum_{\mathbf{y} \in \mathcal{B}_{n-1,n}\colon\, y_i = d_i \text{ for } i \leqslant \ell} (D - \ell + 1) w(\mathbf{y}) \\
&= (D - \ell + 1) Z(n-1, n) \,\mathbb{P}(Y_i = d_i \text{ for } i \leqslant \ell).
\end{aligned}$$

The result follows by Theorem 15.5. $\square$

Remark 15.10. Arguing as in Remark 15.8, we obtain from Lemma 15.9 the explicit formula, generalizing (15.9), with $D := \sum_{i=1}^{\ell} d_i$ and other notations as above,

$$\begin{aligned}
&\mathbb{P}\bigl(d^+_{\mathcal{T}_n}(v_i) = d_i \text{ for } i = 1, \dots, \ell\bigr) \\
&\qquad = \frac{n}{n-\ell} (D - \ell + 1) \frac{w_{d_1} \cdots w_{d_\ell} Z(n - D - 1, n - \ell)}{Z(n-1, n)}. \qquad (15.11)
\end{aligned}$$

Remark 15.11. Note that Lemma 15.9 (or (15.11)) shows that the probability remains exactly the same if we permute $d_1, \dots, d_\ell$, provided that the permuted sequence $(d_{\sigma(i)})$ still is allowed, i.e., $d_{\sigma(i)} \geqslant d'_i$ for all $i \leqslant \ell$. However, if the latter condition fails for some i, then the probability typically becomes 0. (This is an interesting case of a symmetry that is not complete.)

For example, considering only the root o and its first child 1, we have

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = d \text{ and } d^+_{\mathcal{T}_n}(1) = d') = \mathbb{P}(d^+_{\mathcal{T}_n}(o) = d' \text{ and } d^+_{\mathcal{T}_n}(1) = d)$$

whenever $d, d' \geqslant 1$; however, if, say, $d \geqslant 1$ and $d' = 0$, then the right-hand side is 0 while the left-hand side in general is not.

Remark 15.12. Lemma 15.9 extends with minor modifications (mainly notational) to arbitrary finite rooted subtrees T' of U_∞ (not necessarily satisfying (6.1)). We omit the details.

16. Proof of Theorem 7.1

For convenience, we first repeat the theorem.

Theorem 7.1. *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*

$$\pi_k := \frac{\tau^k w_k}{\Phi(\tau)}, \qquad k \geqslant 0; \tag{7.1}$$

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

$$\mu = \Psi(\tau) = \min(\nu, 1) \leqslant 1 \tag{7.2}$$

and variance $\sigma^2 = \tau\Psi'(\tau) \leqslant \infty$. *Let* $\widehat{\mathcal{T}}$ *be the infinite modified Galton–Watson tree constructed in Section 5 for the distribution* $(\pi_k)_{k\geqslant 0}$. *Then* $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ *as* $n \to \infty$, *in the topology defined in Section 6.*

Furthermore, in case (i), $\mu = 1$ *(the critical case) and* $\widehat{\mathcal{T}}$ *is locally finite with an infinite spine; in case* (ii) $\mu = \nu < 1$ *(the subcritical case) and* $\widehat{\mathcal{T}}$ *has a finite spine ending with an explosion.*

Proof. First, as in the proof of Theorem 11.4, Lemma 13.2 shows that τ defined by (i) and (ii) is well-defined and equals $\tau(1)$ defined in Lemma 13.2; since $1 < 2 \leqslant \omega$ we have $\tau < \infty$ and $\Phi(\tau) < \infty$. Further, (13.3) yields $\Psi(\tau) = \min(1, \nu)$. Hence, by Lemma 4.2, (π_k) is a probability distribution with mean and variance as asserted. (This is a special case of the corresponding claims in Theorem 11.4, with $\lambda = 1$. We have $\lambda = 1$ here since we relate the random trees to allocations with $m = n - 1$, and thus $m/n \to 1$.)

The claims in the final paragraph are obvious from (7.2) and the construction in Section 5.

We turn to the main assertion, $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$. Since $\mathfrak{T}$ is a compact metric space, any sequence of random trees in $\mathfrak{T}$ is tight, and has thus a convergent subsequence. (See e.g. [15, Section 6].) In particular, this holds for $\mathcal{T}_n$.

Consider a limiting random tree $\overline{\mathcal{T}}$ in $\mathfrak{T}$ such that $\mathcal{T}_n \xrightarrow{\mathrm{d}} \overline{\mathcal{T}}$ along some subsequence. We will show that then $\overline{\mathcal{T}} \overset{\mathrm{d}}{=} \widehat{\mathcal{T}}$, regardless of the subsequence; this implies $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ for the full sequence, which then completes the proof.

We have defined $\mathfrak{T}$ in Section 6 such that $\mathfrak{T} \subset \overline{\mathbb{N}}_0^{V_\infty}$ using the embedding $T \mapsto (d_T^+(v))_{v \in V_\infty}$. In order to show $\overline{\mathcal{T}} \overset{\mathrm{d}}{=} \widehat{\mathcal{T}}$, it thus suffices to show that the distributions agree on cylinder sets, i.e., that $\bigl(d^+(v_1), \dots, d^+(v_\ell)\bigr) \in \overline{\mathbb{N}}_0^\ell$ has the same distribution for $\overline{\mathcal{T}}$ and $\widehat{\mathcal{T}}$, for any finite set $V = \{v_1, \dots, v_\ell\} \subset V_\infty$. Since $\overline{\mathbb{N}}_0^\ell$ is a countable set, this is equivalent to

$$\mathbb{P}\bigl(d_{\overline{\mathcal{T}}}^+(v_1) = d_1, \dots, d_{\overline{\mathcal{T}}}^+(v_\ell) = d_\ell\bigr) = \mathbb{P}\bigl(d_{\widehat{\mathcal{T}}}^+(v_1) = d_1, \dots, d_{\widehat{\mathcal{T}}}^+(v_\ell) = d_\ell\bigr), \tag{16.1}$$

for any finite set $V = \{v_1, \dots, v_\ell\} \subset V_\infty$ and any $d_1, \dots, d_\ell \in \overline{\mathbb{N}}_0$.

It thus suffices to show (16.1). Furthermore, given any finite set $V \subset V_\infty$, we may enlarge it to a finite set V satisfying (6.2)–(6.4), i.e., a set that is the node set of some finite tree in $\mathfrak{T}_{\mathrm{f}}$. It thus suffices to show (16.1) for $V = V(T')$ with $T' \in \mathfrak{T}_{\mathrm{f}}$.

We make one more reduction. Suppose that $V = V(T')$ with $T' \in \mathfrak{T}_{\mathrm{f}}$ and that (16.1) contains a condition $d^+(v_i) = d_i$ with $d_i < d_{T'}^+(v_i)$. Let $v := v_i$ and let u be the last child of v in T'; thus (recalling the notation in Section 6) $u = vj$ for some integer $j = d_{T'}^+(v) > d_i$. By (6.5), any tree $T \in \mathfrak{T}$ with $d_T^+(v) = d_i$ has $d_T^+(u) = 0$, and further (e.g. by (6.5) and induction) $d_T^+(s) = 0$ for every descendant s of u. Thus, letting T'_u denote the subtree of T' rooted at u, for any $s \in T'_u$, the event $\{d_{\widehat{\mathcal{T}}}^+(v) = d_i \text{ and } d_{\widehat{\mathcal{T}}}^+(s) > 0\}$ is impossible and has probability 0; furthermore, the same holds for $\overline{\mathcal{T}}$, i.e., $\mathbb{P}\bigl(d_{\overline{\mathcal{T}}}^+(v) = d_i \text{ and } d_{\overline{\mathcal{T}}}^+(s) > 0\bigr) = 0$. Consequently, if (16.1) contains a condition $d^+(v_j) = d_j$ with $v_j \in T'_u$ and $d_j > 0$, then both sides are trivially 0. On the other hand, if $d_j = 0$ for all $v_j \in T'_u$, then the conditions $d^+(v_j) = d_j$ are redundant in (16.1) and may be deleted, so we may replace T' by the smaller tree with T'_u removed. Repeating this pruning, if necessary, we see that it suffices to show (16.1) for $V = V(T')$ when $T' \in \mathfrak{T}_{\mathrm{f}}$ is a finite tree and $d_i \geqslant d_{T'}^+(v_i)$ for every i.

Recall that d_i in (16.1) may be infinite. We study three different cases separately.

Case (a): Every $d_i < \infty$. This is the case treated in Lemma 15.9; we take the limit as $n \to \infty$ in (15.10) and obtain by Theorem 11.7 (with $m = n - 1$ and $\lambda = 1 < \omega(\mathbf{w})$), letting again $D := \sum_{i=1}^{\ell} d_i$,

$$\mathbb{P}\bigl(d_{\mathcal{T}_n}^+(v_i) = d_i \text{ for } i = 1, \dots, \ell\bigr) \to (D - \ell + 1) \prod_{i=1}^{\ell} \pi_{d_i}.$$

Since we have assumed $\mathcal{T}_n \xrightarrow{\mathrm{d}} \overline{\mathcal{T}}$ along a subsequence, this yields

$$\mathbb{P}\big(d^+_{\overline{\mathcal{T}}}(v_i) = d_i \text{ for } i = 1, \dots, \ell\big) = (D - \ell + 1) \prod_{i=1}^{\ell} \pi_{d_i}. \tag{16.2}$$

Now consider the modified Galton–Watson tree $\widehat{\mathcal{T}}$. (Recall its construction in Section 5.) If the tree $\widehat{\mathcal{T}}$ has $d^+_{\widehat{\mathcal{T}}}(v_i) = d_i < \infty$ for all $v_i \in T'$, then the spine has to extend outside T'. The first point on the spine outside T' is a node in $\partial T'$ (regarding T' as a subtree of $\widehat{\mathcal{T}}$). The condition $d^+_{\widehat{\mathcal{T}}}(v_i) = d_i$ for $v_i \in T'$ determines the boundary $\partial T'$ of T' in $\widehat{\mathcal{T}}$, which thus not depend on $\widehat{\mathcal{T}}$, and Lemma 15.1 shows that $|\partial T'| = D - \ell + 1$.

Fix a node $u \in \partial T'$, and consider the event $\mathcal{E}_u$ that the spine of $\widehat{\mathcal{T}}$ passes through u and that $d^+_{\widehat{\mathcal{T}}}(v_i) = d_i$ for $i = 1, \dots, \ell$. The event $\mathcal{E}_u$ thus specifies the nodes in T' that are special in the construction of $\widehat{\mathcal{T}}$ (viz. the nodes on the path from o to u), and for each special node it specifies which of its children will be special; furthermore it specifies the number of children for each node in T', special or not. Recall that the probability that a special node has $d < \infty$ children, with a given one of them being special, is π_d, just as the probability that a normal node has d children. Thus, by independence, for every $u \in \partial T'$, $\mathbb{P}(\mathcal{E}_u) = \prod_{i=1}^{\ell} \pi_{d_i}$. This probability thus does not depend on u, so summing over the $D - \ell + 1$ nodes $u \in \partial T'$ we obtain

$$\mathbb{P}\big(d^+_{\widehat{\mathcal{T}}}(v_i) = d_i \text{ for } i = 1, \dots, \ell\big) = \sum_{u \in \partial T'} \mathbb{P}(\mathcal{E}_u) = (D - \ell + 1) \prod_{i=1}^{\ell} \pi_{d_i},$$

which together with (16.2) shows (16.1) in this case. (Cf. Remark 5.7 for a similar argument.)

Case (b): Exactly one $d_i = \infty$. Suppose that $d_j = \infty$ and $d_i < \infty$ for $i \neq j$. Define, for $0 \leqslant k \leqslant \infty$,

$$\mathcal{A}_k := \{T \in \mathfrak{T} : d^+_T(v_i) = d_i \text{ for } i \neq j \text{ and } d^+_T(v_j) = k\}.$$

We thus want to show $\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{A}_\infty) = \mathbb{P}(\widehat{\mathcal{T}} \in \mathcal{A}_\infty)$. We define further

$$\mathcal{A}_{\geqslant K} := \bigcup_{K \leqslant k \leqslant \infty} \mathcal{A}_k,$$

and note that since $\mathcal{T}_n \xrightarrow{\mathrm{d}} \overline{\mathcal{T}}$ (along a subsequence), we have (along the subsequence), for any finite K,

$$\mathbb{P}(\mathcal{T}_n \in \mathcal{A}_{\geqslant K}) \to \mathbb{P}(\overline{\mathcal{T}} \in \mathcal{A}_{\geqslant K}). \tag{16.3}$$

We define also (for finite k) the analogous

$$\mathcal{B}_k := \{(y_1, \dots, y_n) \in \mathcal{B}_{n-1,n} : y_j = k \text{ and } y_i = d_i \text{ for } i \leqslant \ell \text{ with } i \neq j\}.$$

Then Lemma 15.9 can be written, with $D' := \sum_{i \neq j} d_i$, for $k < \infty$,

$$\mathbb{P}(\mathcal{T}_n \in \mathcal{A}_k) = (k + D' - \ell + 1) \frac{n}{n - \ell} \mathbb{P}(B_{n-1,n} \in \mathcal{B}_k). \tag{16.4}$$

Consider, for simplicity, $k > k_0 := \max_{i\neq j} d_i$. Then (14.44) shows that, with $N_i = N_i(B_{n-1,n})$,

$$\mathbb{P}(B_{n-1,n} \in \mathcal{B}_k) = \mathbb{E}\,\mathbb{P}(B_{n-1,n} \in \mathcal{B}_k \mid N_0, N_1, \dots) = \mathbb{E}\Bigl(\frac{N_k}{n}\prod_{i\neq j}\frac{N_{d_i}+O(1)}{n+O(1)}\Bigr)$$
$$= \mathbb{E}\Bigl(\frac{N_k}{n}\prod_{i\neq j}\frac{N_{d_i}}{n} + O\Bigl(\frac{N_k}{n^2}\Bigr)\Bigr).$$

(The implicit constants in the O's in this proof may depend on ℓ and $d_1, \dots, d_\ell$, but not on n or k.) Consequently, by (16.4),

$$\mathbb{P}(\mathcal{T}_n \in \mathcal{A}_k) = \bigl(k+O(1)\bigr)\bigl((1+O(n^{-1})\bigr)\,\mathbb{E}\Bigl(\frac{N_k}{n}\prod_{i\neq j}\frac{N_{d_i}}{n} + O\Bigl(\frac{N_k}{n^2}\Bigr)\Bigr)$$
$$= \bigl(1+O(k^{-1})\bigr)\,\mathbb{E}\Bigl(\frac{kN_k}{n}\prod_{i\neq j}\frac{N_{d_i}}{n}\Bigr) + O\Bigl(\mathbb{E}\Bigl(\frac{kN_k}{n^2}\Bigr)\Bigr).$$

Summing over $k \geqslant K$, we obtain for any $K > k_0$, using $\sum_{k=0}^\infty kN_k = n-1$ for any allocation $B_{n-1,n}$,

$$\begin{aligned}\mathbb{P}(\mathcal{T}_n \in \mathcal{A}_{\geqslant K}) &= \sum_{k=K}^\infty \mathbb{P}(\mathcal{T}_n \in \mathcal{A}_k)\\ &= \bigl(1+O(K^{-1})\bigr)\,\mathbb{E}\Bigl(\frac{\sum_{k\geqslant K} kN_k}{n}\prod_{i\neq j}\frac{N_{d_i}}{n}\Bigr) + O\Bigl(\mathbb{E}\Bigl(\frac{\sum_{k\geqslant K} kN_k}{n^2}\Bigr)\Bigr)\\ &= \bigl(1+O(K^{-1})\bigr)\,\mathbb{E}\Bigl(\frac{n-1-\sum_{k<K} kN_k}{n}\prod_{i\neq j}\frac{N_{d_i}}{n}\Bigr) + O(n^{-1}).\end{aligned} \tag{16.5}$$

By Theorem 11.4, for any fixed K, as $n\to\infty$,

$$\frac{n-1-\sum_{k<K} kN_k}{n}\prod_{i\neq j}\frac{N_{d_i}}{n} \overset{\mathrm{p}}{\longrightarrow} \bigl(1-\sum_{k<K} k\pi_k\bigr)\prod_{i\neq j}\pi_{d_i}.$$

By dominated convergence, the expectation converges to the same limit, and thus (16.3) and (16.5) yield, for $K > k_0$,

$$\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{A}_{\geqslant K}) = \bigl(1+O(K^{-1})\bigr)\Bigl(1-\sum_{k<K} k\pi_k\Bigr)\prod_{i\neq j}\pi_{d_i}. \tag{16.6}$$

Finally, let $K\to\infty$ to obtain

$$\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{A}_\infty) = \Bigl(1-\sum_{k<\infty} k\pi_k\Bigr)\prod_{i\neq j}\pi_{d_i} = (1-\mu)\prod_{i\neq j}\pi_{d_i}. \tag{16.7}$$

Now consider $\widehat{\mathcal{T}}$. If $d^+_{\widehat{\mathcal{T}}}(v_j) = d_j = \infty$, then the spine ends with an explosion at v_j. This fixes the spine, and the event that $d^+_{\widehat{\mathcal{T}}}(v_i) = d_i$ for $i \neq j$ then means, just as in case (a) when we considered a specific $\mathcal{E}_u$, that we have specified the number of children to be d_i for these nodes, and for the special nodes (except v_j) we have also specified which child is special. The probability of this is π_{d_i} for each $i \neq j$, and the probability that the special node v_j has an infinite number of children is, by (5.2), $1-\mu$. Hence, by independence,

$$\mathbb{P}(\widehat{\mathcal{T}} \in \mathcal{A}_\infty) = (1-\mu)\prod_{i\neq j} \pi_{d_i}, \tag{16.8}$$

which together with (16.7) shows $\mathbb{P}(\overline{\mathcal{T}} \in \mathcal{A}_\infty) = \mathbb{P}(\widehat{\mathcal{T}} \in \mathcal{A}_\infty)$, which is (16.1) in this case.

Case (c): More than one $d_i = \infty$. By the definition of the modified Galton–Watson tree $\widehat{\mathcal{T}}$, there is at most one node with infinite degree, so in this case,

$$\mathbb{P}\big(d^+_{\widehat{\mathcal{T}}}(v_i) = d_i \text{ for } i = 1,\dots,\ell\big) = 0.$$

This means that the sum of these probabilities for all sequences $(d_1,\dots,d_n)$ with at most one infinite value is 1. But we have shown that for such sequences, the probability is the same for $\overline{\mathcal{T}}$ as for $\widehat{\mathcal{T}}$, so the probabilities for $\overline{\mathcal{T}}$ for these sequences also sum up to 1. Consequently, if more than one $d_i = \infty$, then

$$\mathbb{P}\big(d^+_{\overline{\mathcal{T}}}(v_i) = d_i \text{ for } i = 1,\dots,\ell\big) = 0$$

too, which shows (16.1) in this case.

This shows that (16.1) holds for any $v_1,\dots,v_m$ such that $\{v_1,\dots,v_m\} = V(T')$ where $T' \in \mathfrak{T}_{\mathrm{f}}$ is a finite tree and $(d_1,\dots,d_n)$ is any sequence in $\overline{\mathbb{N}}_0^m$ with $d_i \geqslant d^+_{T'}(v_i)$ for every i. As discussed above, this implies (16.1) in full generality and thus $\overline{\mathcal{T}} \overset{\mathrm{d}}{=} \widehat{\mathcal{T}}$, which shows that $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$. □

17. Proofs of Theorems 7.11 and 7.12

We begin by stating another version of the correspondence between simply generated trees and the balls-in-boxes model.

Lemma 17.1. *We may couple $\mathcal{T}_n$ and $B_{n-1,n}$ such that the degree sequence $\Lambda(\mathcal{T}_n)$ is a cyclic shift of $B_{n-1,n}$, and, conversely, $B_{n-1,n}$ is a uniformly random cyclic shift of $\Lambda(\mathcal{T}_n)$.*

Proof. Let $B_{n-1,n} = (Y_1,\dots,Y_n)$ and let $(Y_{\sigma(1)},\dots,Y_{\sigma(n)})$ be the unique cyclic shift of $(Y_1,\dots,Y_n)$ that is the degree sequence of a tree in $\mathfrak{T}_n$, see Corollary 15.4. Then $(Y_{\sigma(1)},\dots,Y_{\sigma(n)}) \overset{\mathrm{d}}{=} \Lambda(\mathcal{T}_n)$, as a consequence of Corollary 15.4 and the invariance of the weight $w(Y_1,\dots,Y_n)$ under cyclic shifts. Consequently, we may couple $B_{n-1,n}$ and $\mathcal{T}_n$ such that $(Y_{\sigma(1)},\dots,Y_{\sigma(n)}) = \Lambda(\mathcal{T}_n)$, and the result follows. □

Proof of Theorem 7.11. We use the coupling in Lemma 17.1. Then N_d in Theorem 7.11 equals $N_d(B_{n-1,n})$ in Theorem 11.4, and thus (7.12) follows by (11.15).

We obtain (7.11) as a simple consequence of (7.12), using $\mathbb{P}(d^+_{\mathcal{T}_n}(v) = d \mid N_d) = N_d/n$ and thus $\mathbb{P}(d^+_{\mathcal{T}_n}(v) = d) = \mathbb{E}\, N_d/n$, cf. the proof of Theorem 11.7. Alternatively, we can arrange so that $d^+_{\mathcal{T}_n}(v) = Y_1$, and the result then follows by Theorem 11.7. □

Proof of Theorem 7.12. We use again the coupling in Lemma 17.1. Let T be a fixed tree of size ℓ and let its degree sequence be $(\bar{d}_1, \dots, \bar{d}_\ell)$. Recall that we have defined the degree sequence using depth-first search. It follows that if a tree has degree sequence $(d_1, \dots, d_n)$ and a node v is visited as node v_j in the depth-first search, then the subtree rooted at v has degree sequence $(d_j, \dots, d_k)$, where we stop when this is a degree sequence of a tree, i.e., when it satisfies the condition in Lemma 15.2. In particular, the subtree rooted at v equals T if and only if $(d_j, \dots, d_{j+\ell-1}) = (\bar{d}_1, \dots, \bar{d}_\ell)$. (Clearly, this is impossible if $j > n - \ell + 1$, since then a tree would be completed with less size than ℓ.)

Consequently, N_T equals the number of substrings $(\bar{d}_1, \dots, \bar{d}_\ell)$ in $(Y_1, \dots, Y_n)$, regarded as a cyclic sequence. In other words, if we let I_j be the indicator of the event $(Y_j, \dots, Y_{j+\ell-1}) = (\bar{d}_1, \dots, \bar{d}_\ell)$, where we define $Y_i := Y_{i-n}$ for $i > n$, then

$$N_T = \sum_{j=1}^{n} I_j. \tag{17.1}$$

In particular, taking the expectation and using the rotational symmetry,

$$\mathbb{P}(\mathcal{T}_{n;v} = T) = \frac{1}{n}\,\mathbb{E}\, N_T = \mathbb{E}\, I_1 = \mathbb{P}\big((Y_1, \dots, Y_\ell) = (\bar{d}_1, \dots, \bar{d}_\ell)\big),$$

and thus Theorem 11.7 yields

$$\mathbb{P}(\mathcal{T}_{n;v} = T) \to \prod_{i=1}^{\ell} \pi_{\bar{d}_i} = \mathbb{P}(\mathcal{T} = T),$$

which proves (7.13).

In order to show the stronger result (7.14), we condition as in the proof of Theorem 11.7 on $N_0, N_1 \dots$ and obtain, see (14.44),

$$\begin{aligned}\mathbb{E}(I_j \mid N_0, N_1, \dots) &= \mathbb{P}\big((Y_1, \dots, Y_\ell) = (\bar{d}_1, \dots, \bar{d}_\ell) \mid N_0, N_1, \dots\big) \\ &= \prod_{i=1}^{\ell} \frac{N_{\bar{d}_i} - c_i}{n - i + 1} = \prod_{i=1}^{\ell} \frac{N_{\bar{d}_i}}{n} + O\left(\frac{1}{n}\right),\end{aligned} \tag{17.2}$$

where $c_i := |\{j < i : \bar{d}_j = \bar{d}_i\}|$. If $|j - k| \geqslant \ell$ and $|j - k \pm n| \geqslant \ell$ (i.e., j and k have distance at least ℓ, regarded as point on a circle of length n), then similarly, with $c'_i := |\{j \leqslant \ell : \bar{d}_j = \bar{d}_i\}|$,

$$\mathbb{E}(I_j I_k \mid N_0, N_1, \dots) = \prod_{i=1}^{\ell} \frac{N_{\bar{d}_i} - c_i}{n - i + 1} \prod_{i=1}^{\ell} \frac{N_{\bar{d}_i} - c_i - c'_i}{n - \ell - i + 1},$$

and it follows that

$$\operatorname{Cov}(I_j, I_k \mid N_0, N_1, \dots) = O(1/n). \tag{17.3}$$

For j and k of distance less than ℓ, we use the trivial

$$|\operatorname{Cov}(I_j, I_k \mid N_0, N_1, \dots)| \leqslant 1. \tag{17.4}$$

There are less than n^2 pairs (j,k) of the first type and $O(n)$ pairs of the second type, and thus by (17.1) and (17.3)–(17.4),

$$\operatorname{Var}(N_T \mid N_0, N_1, \dots) = \sum_{j=1}^{n}\sum_{k=1}^{n} \operatorname{Cov}(I_j, I_k \mid N_0, N_1, \dots) = O(n).$$

Consequently, $N_T/n - \mathbb{E}(N_T/n \mid N_0, N_1, \dots) \xrightarrow{\mathrm{p}} 0$, and thus by (17.1), (17.2) and Theorem 11.4,

$$\begin{aligned} \frac{N_T}{n} &= \mathbb{E}\Big(\frac{N_T}{n} \Bigm| N_0, N_1, \dots\Big) + o_{\mathrm{p}}(1) = \prod_{i=1}^{\ell} \frac{N_{\bar d_i}}{n} + o_{\mathrm{p}}(1) \\ &\xrightarrow{\mathrm{p}} \prod_{i=1}^{\ell} \pi_{\bar d_i} = \mathbb{P}(\mathcal{T} = T). \end{aligned}$$

□

18. Asymptotics of the partition functions

We have a simple asymptotic result for the partition function $Z(m,n)$ (to the first order in the exponent, at least if $\rho > 0$):

Theorem 18.1. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be a 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* $\operatorname{span}(\mathbf{w}) \mid m$*, $m \to \infty$ and $m/n \to \lambda$ where $0 \leqslant \lambda < \omega$, and let τ be as in Theorem 11.4.*

(i) *If $\rho > 0$, then*

$$\frac{1}{n} \log Z(m,n) \to \log \Phi(\tau) - \lambda \log \tau \in (-\infty, \infty). \tag{18.1}$$

(ii) *If $\rho = 0$ and $\lambda > 0$, then*

$$\frac{1}{n} \log Z(m,n) \to \infty. \tag{18.2}$$

In both cases, the result can be written

$$\frac{1}{n} \log Z(m,n) \to \log \inf_{0 \leqslant t \leqslant \rho} \frac{\Phi(t)}{t^{\lambda}} = \log \inf_{0 \leqslant t < \infty} \frac{\Phi(t)}{t^{\lambda}} \leqslant \infty. \tag{18.3}$$

If $0 \leqslant \lambda \leqslant \nu$ and $\rho > 0$, the limit can also be written $\log \Phi(\tau) - \Psi(\tau) \log \tau$.

The formula (18.1) is shown by a physicists' proof by Bialas, Burda and Johnston [14].

Remark 18.2. If $\lambda = 0$, then $\tau = 0$, and we interpret the right-hand side of (18.1) as $\log \Phi(0) = \log w_0$; this is in accordance with (18.3).

It is easily seen that the result holds, with this limit, also in the rather trivial case when m is bounded, provided $Z(m,n) > 0$.

Remark 18.3. If $\omega < \infty$, then the result holds also when $\lambda = \omega$, provided $Z(m,n) > 0$, if we let $\tau = \infty$ as in Remark 11.10 and interpret the right-hand side of (18.1) as the limit value $\log w_\omega$, which again is in accordance with (18.3). This follows from Remark 18.2 by the symmetry argument in Remark 11.10.

Remark 18.4. Using the function $\tau(x)$ defined in Theorem 11.6, the result (18.1) can also be written, using the continuity of $\tau(x)$ and an extra argument (which we omit) when $\lambda = 0$,

$$\log Z(m,n) = n \log \Phi(\tau(x)) - m \log \tau(x) + o(n) \tag{18.4}$$

or, equivalently,

$$Z(m,n) = \Phi(\tau(x))^n \tau(x)^{-m} e^{o(n)}. \tag{18.5}$$

As in Theorem 11.6, it suffices here that $m/n \leqslant C < \omega$ (and $m \to \infty$).

Proof of Theorem 18.1. Note that the assumptions imply that $Z(m,n) > 0$ (at least for n, and thus m, large) by Lemma 13.3. The equivalence between (18.1)–(18.2) and (18.3) follows from (11.16).

(i): Assume first $\lambda > 0$. Since $\rho > 0$ and $\lambda > 0$, we then have $\tau > 0$. Thus $\mathbf{w} = (w_k)$ is equivalent to $\boldsymbol{\pi} = (\pi_k)$, and Lemma 11.3 yields

$$Z(m,n) = Z(m,n;\mathbf{w}) = \Phi(\tau)^n \tau^{-m} Z(m,n;\boldsymbol{\pi}).$$

We saw in the proof of Theorem 11.4, case (a), that Lemmas 14.3 and 14.4 yield $Z(m,n;\boldsymbol{\pi}) = \exp(o(n))$, and thus

$$Z(m,n) = \exp\bigl(n \log \Phi(\tau) - m \log \tau + o(n)\bigr),$$

which yields (18.1).

It remains to consider the case $\lambda = 0$. Then $m/n \to 0$, and we may assume $m < n/2$. In any allocation of m balls, there are at most m non-empty boxes. Let us mark $2m$ boxes, including all non-empty boxes. For each choice of the marked boxes, we have in them an allocation in $\mathcal{B}_{m,2m}$, and only empty boxes outside; since there are $\binom{n}{2m}$ choices of marked boxes,

$$Z(m,n) \leqslant \binom{n}{2m} w_0^{n-2m} Z(m,2m). \tag{18.6}$$

On the other hand, any allocation of m balls in $2m$ boxes can be extended to an allocation in $\mathcal{B}_{m,n}$ with the last $n-2m$ boxes empty; thus

$$Z(m,n) \geqslant w_0^{n-2m} Z(m,2m). \tag{18.7}$$

We have, by Stirling's formula, using $m/n \to \lambda = 0$,

$$\frac{1}{n} \log \binom{n}{2m} \leqslant \frac{1}{n} \log \Bigl(\frac{en}{2m}\Bigr)^{2m} = \frac{2m}{n} \log \frac{e}{2} - \frac{2m}{n} \log \frac{m}{n} \to 0. \tag{18.8}$$

Moreover, by the case $\lambda > 0$ just proved, we have from (18.1) $\log Z(m,2m) = O(m) = o(n)$. Consequently, (18.6)–(18.8) yield

$$\log Z(m,n) = (n-2m) \log w_0 + o(n) = n \log w_0 + o(n),$$

showing (18.1) in the case $\lambda = 0$.

(ii): As in the proof of Lemma 14.4, we use the truncated weight sequence $\mathbf{w}^{(K)}$ defined in (14.14), where K is so large that $\operatorname{span}(\mathbf{w}^{(K)}) = \operatorname{span}(\mathbf{w})$ and $\omega(\mathbf{w}^{(K)}) > \lambda$, and we let again Φ_K and Ψ_K be the corresponding functions for $\mathbf{w}^{(K)}$ and define τ_K by $\Psi_K(\tau_K) = \lambda$.

For any $t > 0$, $\Phi_K(t) \to \Phi(t) = \infty$ as $K \to \infty$, and thus (14.19) holds, showing that for large K, $\Psi_K(t) > \lambda$ and thus $\tau_K < t$. Since t is arbitrary, this shows that $\tau_K \to 0$ as $K \to \infty$. Applying (i) to $\mathbf{w}^{(K)}$ and its partition function Z_K we obtain, for every large K,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log Z(m,n) \geqslant \lim_{n\to\infty} \frac{1}{n} \log Z_K(m,n) &= \log \Phi_K(\tau_K) - \lambda \log \tau_K \\ &\geqslant \log w_0 - \lambda \log \tau_K.\end{aligned}$$

As $K \to \infty$, $\tau_K \to 0$ so the right-hand side tends to ∞, which completes the proof. ∎

Remark 18.5. The case $\rho = 0$ and $\lambda = 0$ is excluded from Theorem 18.1; in this case, almost anything can happen. To see this, note first that by (18.6)–(18.8), if $m/n \to \lambda = 0$, then

$$\frac{1}{n} \log Z(m,n) = \log w_0 + \frac{1}{n} \log Z(m,2m) + o(1). \tag{18.9}$$

Furthermore, by Theorem 18.1(ii), $\frac{1}{m} \log Z(m,2m) \to \infty$ as $m \to \infty$, and hence $m/\log Z(m,2m) \to 0$. We can choose $m = m(n) \to \infty$ with $m/n \to 0$ so rapidly that $m/n \ll m/\log Z(m,2m)$; then $\frac{1}{n} \log Z(m,2m) \to 0$ and (18.9) yields $\frac{1}{n} \log Z(m,n) \to \log w_0 = \log \Phi(0)$.

We can also choose m with $m/n \to 0$ so slowly that $m/n \gg m/\log Z(m,2m)$; then $\frac{1}{n} \log Z(m,2m) \to \infty$ and (18.9) yields $\frac{1}{n} \log Z(m,n) \to \infty$.

Moreover, we can choose $m(n)$ oscillating between these two cases, and then $\liminf \frac{1}{n} \log Z(m,n) = \log \Phi(0)$ and $\limsup \frac{1}{n} \log Z(m,n) = \infty$, and we can arrange so that every number in $[\log \Phi(0), \infty)$ is a limit point of some subsequence.

For many weight sequences with $\rho = 0$, one can choose $m(n)$ such that $\frac{1}{n} \log Z(m,n) \to a$ for any given $a \in [\log \Phi(0), \infty]$. For example for $w_k = k!$ as in Example 10.8, we have by [64] and Theorem 15.5 $Z(n-1,n) \sim en!$ and it follows, arguing similarly to (18.6) and (18.7), that $\frac{1}{m} \log Z(m,2m) = \log m + O(1)$, so taking $m \sim an/\log n$, we obtain $\frac{1}{n} \log Z(m,n) \to a$ by (18.9).

However, if w_k increases very rapidly, it may be impossible to obtain convergence of the full sequence to a limit different from $\log \Phi(0)$ or ∞, so we can only achieve convergence of subsequences. For example, if $w_0 = 1$ and $w_{k+1} \geqslant Z(k,2k)^2$, then $Z(k+1, 2(k+1)) \geqslant w_{k+1} \geqslant Z(k,2k)^2$, and it follows easily from (18.9) that $\limsup \frac{1}{n} \log Z(m,n) \geqslant 2 \liminf \frac{1}{n} \log Z(m,n)$.

We apply Theorem 18.1 to simply generated trees.

Theorem 18.6. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be any weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 2$. Suppose that $n \to \infty$ with $n \equiv 1 \pmod{\operatorname{span}(\mathbf{w})}$, and let τ be as in Theorem 7.1. Then*

$$\frac{1}{n}\log Z_n \to \log\Phi(\tau) - \log\tau = \log\inf_{0\leqslant t<\infty}\frac{\Phi(t)}{t} \in (-\infty,\infty].$$

The limit is finite if $\rho > 0$, and $+\infty$ if $\rho = 0$.

Proof. An immediate consequence of Theorems 15.5 and 18.1. □

For probability weight sequences, Theorem 18.6 can be expressed as follows, cf. Remark 7.9.

Theorem 18.7. *Let $\mathcal{T}$ be a Galton–Watson tree with offspring distribution ξ, and assume that $\mathbb{P}(\xi = 0) > 0$ and $\mathbb{P}(\xi > 1) > 0$. Suppose that $n \to \infty$ with $n \equiv 1 \pmod{\operatorname{span}(\xi)}$, and let τ be as in Theorem 7.1. Then*

$$\frac{1}{n}\log\mathbb{P}(|\mathcal{T}| = n) \to \log\Phi(\tau) - \log\tau = \log\inf_{0\leqslant t<\infty}\frac{\Phi(t)}{t} \in (-\infty,0].$$

If $\mathbb{E}\,\xi = 1$, or if $\mathbb{E}\,\xi < 1$ and $\rho = 1$, then the limit is 0; otherwise it is strictly negative. In other words, $\mathbb{P}(|\mathcal{T}| = n)$ decays exponentially fast in the supercritial case (then $\tau < 1$) and in the subcritical case with $\rho > 1$ (then $\tau > 1$), but only subexponentially in the critical case and in the subcritical case with $\rho = 1$ (then $\tau = 1$).

Proof. We have $\mathbb{P}(|\mathcal{T}| = n) = Z_n$, see Section 2, and we apply Theorem 18.6. Since now (w_k) is a probability weight sequence, we have $\rho \geqslant 1$ and $\inf_{0\leqslant t<\infty}\Phi(t)/t \leqslant \Phi(1)/1 = 1$, with equality if and only if $\tau = 1$, see Remark 7.4. The final claims follow using the definition of τ in Theorem 7.1. □

When $\rho > 0$ and $\lambda > 0$ (which are equivalent to $\tau > 0$), we can also prove stronger "local" versions of Theorems 18.1 and 18.6, showing that the partition function behaves smoothly for small changes in m or n.

Theorem 18.8. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be a 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 $m/n \to \lambda$ where $0 < \lambda < \omega$, and let τ be as in Theorem 11.4. If $\rho > 0$, then, for every fixed $k \in \mathbb{Z}$ such that $\operatorname{span}(\mathbf{w}) \mid k$,*

$$\frac{Z(m+k,n)}{Z(m,n)} \to \tau^{-k}. \tag{18.10}$$

Proof. For any $k \geqslant 0$, by (11.2)–(11.4),

$$\mathbb{P}(Y_1 = k) = \frac{w_k Z(m-k,n-1)}{Z(m,n)}, \tag{18.11}$$

and thus

$$\frac{\mathbb{P}(Y_1 = k)}{\mathbb{P}(Y_1 = 0)} = \frac{w_k Z(m-k,n-1)}{w_0 Z(m,n-1)}.$$

Since Theorem 11.7 yields

$$\frac{\mathbb{P}(Y_1 = k)}{\mathbb{P}(Y_1 = 0)} \to \frac{\pi_k}{\pi_0} = \tau^k \frac{w_k}{w_0},$$

we see (replacing n by $n+1$) that (18.10) holds when $-k \in \operatorname{supp}(\mathbf{w})$. Furthermore, the set of $k \in \mathbb{Z}$ such that (18.10) holds for any allowed sequence $m(n)$ is easily seen to be a subgroup of $\mathbb{Z}$ (since we may replace m by $m \pm k'$ for any fixed k'). Consequently, by (3.3), this set contains every multiple of $\operatorname{span}(\mathbf{w})$. ∎

Theorem 18.9. *Let $\mathbf{w} = (w_k)_{k \geqslant 0}$ be a 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 $m/n \to \lambda$ where $0 \leqslant \lambda < \omega$, and let τ be as in Theorem 11.4. Then,*

$$\frac{Z(m, n+1)}{Z(m,n)} \to \Phi(\tau). \tag{18.12}$$

Proof. By (18.11) with $k = 0$ and Theorem 11.7,

$$\frac{w_0 Z(m, n-1)}{Z(m,n)} = \mathbb{P}(Y_1 = 0) \to \pi_0 = \frac{w_0}{\Phi(\tau)},$$

and the result follows since $w_0 \neq 0$. ∎

For trees we have a corresponding result:

Theorem 18.10. *Let $\mathbf{w} = (w_k)_{k \geqslant 0}$ be a weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 2$. If $\rho > 0$ and $\operatorname{span}(\mathbf{w}) = 1$, then*

$$\frac{Z_{n+1}}{Z_n} \to \frac{\Phi(\tau)}{\tau}.$$

Proof. By Theorems 15.5 and 18.8–18.9,

$$\frac{Z_{n+1}}{Z_n} = \frac{nZ(n, n+1)}{(n-1)Z(n-1,n)} = \frac{n}{n-1} \cdot \frac{Z(n, n+1)}{Z(n,n)} \cdot \frac{Z(n,n)}{Z(n-1,n)} \to \Phi(\tau)\tau^{-1}.$$

∎

We assumed here span 1 for convenience only; if $\operatorname{span}(\mathbf{w}) = d$, we instead obtain, by a similar argument, $Z_{n+d}/Z_n \to (\Phi(\tau)/\tau)^d$.

In the case $\nu \geqslant 1$ and $\sigma^2 = \tau \Psi'(\tau) < \infty$ (which is automatic if $\nu > 1$), i.e. our case Iα, Theorem 18.6 can be sharpened substantially as follows, see Otter [93], Meir and Moon [85], Kolchin [76], Drmota [33].

Theorem 18.11. *Let $\mathbf{w} = (w_k)$, τ and σ^2 be as in Theorem 7.1, and let $d := \operatorname{span}(\mathbf{w})$. If $\nu \geqslant 1$ and $\sigma^2 < \infty$, then, for $n \equiv 1 \pmod d$,*

$$Z_n \sim \frac{d}{\sqrt{2\pi\sigma^2}} \cdot \frac{\Phi(\tau)^n \tau^{1-n}}{n^{3/2}} = d\sqrt{\frac{\Phi(\tau)}{2\pi\Phi''(\tau)}} \left(\frac{\Phi(\tau)}{\tau}\right)^n n^{-3/2}. \tag{18.13}$$

Proof. Replacing (w_k) by (π_k) and using (4.3), we see that it suffices to consider the case of a probability weight sequence with $\tau = \Phi(\tau) = 1$. By Theorem 15.5, (11.5) and (8.1), in this case the result is equivalent to

$$\mathbb{P}(S_n = n-1) \sim \frac{d}{\sqrt{2\pi\sigma^2 n}},$$

which is the local central limit theorem in this case, see e.g. Kolchin [76, Theorem 1.4.2] or use Lemma 14.1 and Remark 14.2. ☐

There is a corresponding improvement of Theorem 18.1.

Theorem 18.12. *Let* $\mathbf{w} = (w_k)$, $m = m(n)$, τ *amd* σ^2 *be as in Theorem 11.4, and let* $d := \operatorname{span}(\mathbf{w})$. *If* $0 < \lambda < \nu$, *or* $\lambda = \nu$ *and* $\sigma^2 < \infty$, *then, for* $m = \lambda n + o(\sqrt{n})$ *with* $m \equiv 0 \pmod d$,

$$Z(m,n) \sim \frac{d}{\sqrt{2\pi\sigma^2 n}} \Phi(\tau)^n \tau^{-m}. \tag{18.14}$$

Proof. Again it suffices to consider the case of a probability weight sequence with $\tau = \Phi(\tau) = 1$; this time using (11.9). In this case the result is by (11.5) equivalent to

$$\mathbb{P}(S_n = m) \sim \frac{d}{\sqrt{2\pi\sigma^2 n}},$$

which again is the local central limit theorem and follows e.g. by Lemma 14.1 and Remark 14.2. ☐

Remark 18.13. The asymptotic formula (18.14) holds for arbitrary $m = m(n)$ with $0 < c \leqslant m/n \leqslant C < \omega$ and $m \equiv 0 \pmod d$, and either $C < \nu$ or $C = \nu$ and $\Phi''(\rho) < \infty$ (which means that $\Psi'(\rho) < \infty$ and thus the distribution (11.13) has finite variance for $\tau = \nu$), provided τ is replaced by $\tau(m/n)$ given by $\Psi(\tau(m/n)) = m/n$. (Cf. Theorem 11.6.) The proof is essentially the same (as in the proof of Theorem 11.6, it suffices to consider subsequences where $m(n)/n$ converges); we omit the details.

In the case $\nu = \lambda$ ($\nu = 1$ in the tree case) and $\sigma^2 = \infty$, we have no general results but we can obtain similar more precise versions of Theorems 18.6 and 18.1 in the important case of a power-law weight sequence, Example 12.10. (We need $1 < \alpha \leqslant 2$ here; if $\alpha \leqslant 1$, then $\nu = \infty > \lambda$, and if $\alpha > 2$, then $\sigma^2 < \infty$ so Theorems 18.11 and 18.12 apply, see Example 12.10 with $\beta = \alpha + 1$. Note also that $\operatorname{span}(\mathbf{w}) = 1$.) The case $\lambda > \nu$ is treated in Theorem 19.34 and Remark 19.35.

Theorem 18.14. *Suppose for some* $c > 0$ *and* α *with* $1 < \alpha \leqslant 2$,

$$w_k \sim c k^{-\alpha-1} \qquad \text{as } k \to \infty. \tag{18.15}$$

(i) *If* $\nu = 1$, *then,*

$$Z_n \sim \frac{\Phi(1)^{1/\alpha}}{c^{1/\alpha}\Gamma(-\alpha)^{1/\alpha}|\Gamma(-1/\alpha)|} \Phi(1)^n n^{-1-1/\alpha}, \qquad \text{when } 1 < \alpha < 2, \tag{18.16}$$

and

$$Z_n \sim \Big(\frac{\Phi(1)}{\pi c}\Big)^{1/2}\Phi(1)^n n^{-3/2}(\log n)^{-1/2}, \qquad \textit{when } \alpha=2. \tag{18.17}$$

(ii) *If* $m=\nu n+o(n^{1/\alpha})$, *then*

$$Z(m,n) \sim \frac{\Phi(1)^{1/\alpha}}{c^{1/\alpha}\Gamma(-\alpha)^{1/\alpha}|\Gamma(-1/\alpha)|}\Phi(1)^n n^{-1/\alpha}, \qquad \textit{when } 1<\alpha<2, \tag{18.18}$$

and

$$Z(m,n) \sim \Big(\frac{\Phi(1)}{\pi c}\Big)^{1/2}\frac{\Phi(1)^n}{\sqrt{n\log n}}, \qquad \textit{when } \alpha=2. \tag{18.19}$$

Proof. This time, we did not assume $w_0>0$, but we may do so without loss of generality in the proof. In fact, if $w_0=0$, then $\nu>1$, so in (i) we always have $w_0>0$, and in (ii) we can reduce to the case $w_0>0$ by the method in Remark 11.8.

(i) follows from Theorem 15.5 and (ii), taking $m=n-1$; hence it suffices to prove (18.18)–(18.19).

We have $\rho=1$, and in the usual notation $\lambda=\nu$ and thus $\tau=\rho=1$. We reduce to the probability weight sequence case by dividing each w_k by $\Phi(1)$ (which changes c to $c/\Phi(1)$). Let ξ be a random variable with the distribution $(\pi_k)=(w_k)$. Then $\mathbb{E}\,\xi=\nu$. Furthermore, (18.15) yields

$$\mathbb{P}(\xi\geqslant k)=\sum_{l=k}^{\infty} w_l \sim c\alpha^{-1}k^{-\alpha}. \tag{18.20}$$

Hence ξ is in the domain of attraction of an α-stable distribution, see Feller [39, Section XVII.5]. More precisely, if we first consider the case $1<\alpha<2$, then there exists an α-stable random variable X_α such that

$$\frac{S_n-n\nu}{n^{1/\alpha}} \xrightarrow{\mathrm{d}} X_\alpha. \tag{18.21}$$

(The distribution of X_α is given by (19.93) and (19.113) below.) Moreover, a local limit law holds, see e.g. Gnedenko and Kolmogorov [46, § 50], Ibragimov and Linnik [54, Theorem 4.2.1] or Bingham, Goldie and Teugels [16, Corollary 8.4.3], which says

$$\mathbb{P}(S_n=\ell)=n^{-1/\alpha}\Big(g\Big(\frac{\ell-n\nu}{n^{1/\alpha}}\Big)+o(1)\Big), \tag{18.22}$$

uniformly for all integers ℓ, where g is the density function of X_α. In particular,

$$Z(m,n)=\mathbb{P}(S_n=m)\sim n^{-1/\alpha}g(0). \tag{18.23}$$

The results in [39, Sections XVII.5–6] show, if we keep track of the constants (see e.g. [63] for calculations), that

$$g(0)=(c\,\Gamma(-\alpha))^{-1/\alpha}|\Gamma(-1/\alpha)|^{-1}, \tag{18.24}$$

and (18.18) follows.

In the case $\alpha = 2$, [39, Section XVII.5] similarly yields

$$\frac{S_n}{\sqrt{n \log n}} \xrightarrow{\mathrm{d}} N(0, c/2); \tag{18.25}$$

again a local limit theorem holds by [54, Theorem 4.2.1] or [16, Corollary 8.4.3], and thus

$$\mathbb{P}(S_n = \ell) = \frac{1}{\sqrt{n \log n}} \Big(g\Big(\frac{\ell - n\nu}{\sqrt{n \log n}} \Big) + o(1) \Big), \tag{18.26}$$

uniformly in $\ell \in \mathbb{Z}$, where now $g(x)$ is the density function $(\pi c)^{-1/2} e^{-x^2/c}$ of $N(0, c/2)$. In particular,

$$Z(m,n) = \mathbb{P}(S_n = m) \sim \frac{1}{\sqrt{n \log n}} g(0) = \frac{1}{\sqrt{n \log n}} \cdot \frac{1}{\sqrt{\pi c}}, \tag{18.27}$$

which proves (18.19). $\square$

Remark 18.15. The proof shows that (18.15) can be relaxed to (18.20) together with $\operatorname{span}(\mathbf{w}) = 1$.

Example 18.16. Let $F^{\mathrm{u}}_{m,n}$ be the number of labelled unrooted forests with m labelled nodes and n labelled trees, see Example 12.7. Using the weights $w_k = k^{k-2}/k!$ and $\widetilde{w}_k = e^{-k} w_k \sim (2\pi)^{-1/2} k^{-5/2}$, we have by (12.35) and (11.9)

$$F^{\mathrm{u}}_{m,n} = m!\, Z(m,n;\mathbf{w}) = m!\, e^m Z(m,n;\widetilde{\mathbf{w}}). \tag{18.28}$$

At the phase transition $m = 2n$, Theorem 18.14 applies to $\widetilde{\mathbf{w}}$ with $\alpha = 3/2$. We have $c = (2\pi)^{-1/2}$ and, by (12.36), $\Phi(1) = \Phi(\rho) = 1/2$. Hence (18.18) yields, after simplifications,

$$\frac{F^{\mathrm{u}}_{2n,n}}{2n!} = Z(2n,n;\mathbf{w}) = e^{2n} Z(2n,n;\widetilde{\mathbf{w}}) \sim \frac{2^{-2/3} 3^{-1/3}}{\Gamma(1/3)} e^{2n} 2^{-n} n^{-2/3}. \tag{18.29}$$

(The constant can also be written $2^{-5/3} 3^{1/6} \pi^{-1} \Gamma(2/3)$.) A more general result is proved by the same method by Britikov [20]. Flajolet and Sedgewick [40, Proposition VIII.11], show (18.29) by a different method (although there is a computational error in the constant given in the result there).

We end this section by considering the behaviour of the generating function $\mathcal{Z}(z) := \sum_{n=1}^{\infty} Z_n z^n$. The following immediate corollary of Theorem 18.6 was shown by Otter [93], see Minami [89] and, for $\nu > 1$, Flajolet and Sedgewick [40, Proposition IV.5]. See also also Remark 7.5.

Corollary 18.17. *Let $(w_k)_{k\geqslant 0}$ and τ be as in Theorem 7.1, and let $\rho_{\mathcal{Z}}$ be the radius of convergence of the generating function $\mathcal{Z}(z) := \sum_{n=1}^{\infty} Z_n z^n$. Then $\rho_{\mathcal{Z}} = \tau/\Phi(\tau)$.* $\square$

Moreover, by (7.6), $\mathcal{Z}(\rho_{\mathcal{Z}}) = \tau < \infty$. Since the generating function $\mathcal{Z}(z)$ has non-negative coefficients, it follows that $\mathcal{Z}(z)$ is continuous on the closed disc $|z| \leqslant \rho_{\mathcal{Z}}$, and $|\mathcal{Z}(z)| \leqslant \tau$ there. If we, for simplicity, assume that span($\mathbf{w}$) $= 1$, then $|\mathcal{Z}(z)| < |\mathcal{Z}(\rho_{\mathcal{Z}})| = \tau$ for $|z| \leqslant \rho_{\mathcal{Z}}$, $z \neq \rho_{\mathcal{Z}}$. Since $|Z| < \tau$ implies

$$|\Phi(Z) - Z\Phi'(Z)| = \left| w_0 - \sum_{k=1}^{\infty}(k-1)w_k Z^k \right| \geqslant w_0 - \sum_{k=1}^{\infty}(k-1)w_k |Z|^k$$
$$> w_0 - \sum_{k=1}^{\infty}(k-1)w_k \tau^k = \Phi(\tau) - \tau\Phi'(\tau) = 0,$$

it follows that $\Phi(Z) - Z\Phi'(Z) \neq 0$ if $Z = \mathcal{Z}(z)$ with $|z| = \rho_{\mathcal{Z}}$, $z \neq \rho_{\mathcal{Z}}$; hence the implicit function theorem and (3.13) show that $\mathcal{Z}(z)$ has an analytic continuation to some neighbourhood of z. Consequently, $\mathcal{Z}$ then can be extended across $|z| = \rho_{\mathcal{Z}}$ everywhere except at $z = \rho_{\mathcal{Z}}$. (If span($\mathbf{w}$) $= d$, the same holds except at $z = \rho_{\mathcal{Z}} e^{2\pi \mathrm{i} j/d}$, $j \in \mathbb{Z}$.)

In our case Ia ($\nu > 1$, or equivalently $\tau < \rho$), much more is known: $\mathcal{Z}$ has a square root singularity at $\rho_{\mathcal{Z}}$ with a local expansion of $\mathcal{Z}(z)$ as an analytic function of $\sqrt{1 - z/\rho_{\mathcal{Z}}}$:

$$\mathcal{Z}(z) = \tau - b\sqrt{1 - z/\rho_{\mathcal{Z}}} + \dots, \tag{18.30}$$

where, with $\sigma^2 := \operatorname{Var} \xi$ given by (8.1),

$$b := \sqrt{\frac{2\Phi(\tau)}{\Phi''(\tau)}} = \sqrt{2}\,\frac{\tau}{\sigma}, \tag{18.31}$$

see Meir and Moon [85], Flajolet and Sedgewick [40, Theorem VI.6] and Drmota [33, Section 3.1.4 and Theorem 2.19]; in particular, $\mathcal{Z}$ then extends analytically to a neighbourhood of ρ cut at the ray $[\rho, \infty)$. In fact, this extends (in a weaker form) to the case $\nu \geqslant 1$ and $\sigma^2 < \infty$ (case Iα): (18.30) holds in a suitable region, with an error term $o(\sqrt{1 - z/\rho_{\mathcal{Z}}})$, see Janson [59].

Remark 18.18. In the case $\nu > 1$, (18.30) and (18.31) yield another proof of (18.13) by standard singularity analysis, see e.g. Drmota [33, Theorem 3.6] and Flajolet and Sedgewick [40, Theorem VI.6 and VII.2]; this argument can be extended to the case $\nu \geqslant 1$ and $\sigma^2 < \infty$, see Drmota [33, Remark 3.7] and Janson [59, Appendix]. When $\nu > 1$, an expansion with further terms can also be obtained, see Minami [89] and Flajolet and Sedgewick [40, Theorem VI.6].

In the other cases ($\sigma^2 = \infty$ or $\nu < 1$), the asymptotic behaviour of $\mathcal{Z}$ at the singularity $\rho_{\mathcal{Z}}$ depends on the behaviour of $\Phi(z)$ at its singularity ρ. It seems difficult to say anything detailed in general, so we study only a few examples. We assume $\nu \leqslant 1$ and $\omega > 1$; thus Lemma 3.1 implies that $\rho < \infty$, $\Phi(\rho) < \infty$ and $\Phi'(\rho) < \infty$. We assume also $\rho > 0$ and span($\mathbf{w}$) $= 1$.

Example 18.19. Suppose that $0 < \rho < \infty$ and that $\Phi(z)$ has an analytic extension to a sector $D_{\rho,\delta} := \{z : |\arg(\rho - z)| < \pi/2 + \delta \text{ and } |z - \rho| < \delta\}$ for some $\delta > 0$, and that in this sector $D_{\rho,\delta}$, for some $a \neq 0$ and non-integer

$\alpha > 1$, and some $f(z)$ analytic at ρ (which can be taken as a polynomial of degree $< \alpha$),

$$\Phi(z) = f(z) + a(\rho - z)^\alpha + o\bigl(|\rho - z|^\alpha\bigr), \qquad \text{as } z \to \rho. \tag{18.32}$$

(We have to have $\alpha > 1$ since $\Phi'(\rho) < \infty$. For $\alpha \geqslant 2$ integer, see instead Example 18.20.) If we assume that Φ has no further singularities on $|z| = \rho$, this implies by singularity analysis, see Flajolet and Sedgewick [40, Section VI.3],

$$w_k \sim \frac{a}{\Gamma(-\alpha)} k^{-\alpha-1} \rho^{\alpha-k}, \qquad \text{as } k \to \infty. \tag{18.33}$$

The converse does not hold in general, but can be expected if the weight sequence is very regular. For example, (18.32) holds (in the plane cut at $[\rho, \infty)$) if $w_k = (k+1)^{-\beta}$, $k \geqslant 1$, as in Example 10.7, with $\beta = \alpha + 1 > 2$, $\rho = 1$ and $a = \Gamma(-\alpha)$, see e.g. [40, Section VI.8].

Let $F(Z) := Z/\Phi(Z)$, so (3.13) can be written

$$F(\mathcal{Z}(z)) = z. \tag{18.34}$$

Since $\nu \leqslant 1$, we have $\tau = \rho$, and thus by Corollary 18.17 and (7.6) $\rho_{\mathcal{Z}} = F(\tau) = F(\rho)$ and $\mathcal{Z}(\rho_{\mathcal{Z}}) = \rho$. Note that

$$F'(\rho) = \frac{\Phi(\rho) - \rho\Phi'(\rho)}{\Phi(\rho)^2} = \frac{1 - \Psi(\rho)}{\Phi(\rho)} = \frac{1-\nu}{\Phi(\rho)}. \tag{18.35}$$

If $\nu < 1$, then (18.35) yields $F'(\rho) > 0$ and (18.34) shows that $\rho - \mathcal{Z}(z) \sim F'(\rho)^{-1}(\rho_{\mathcal{Z}} - z)$ as $z \to \rho_{\mathcal{Z}}$. Moreover, F is defined in a sector $D_{\rho,\delta}$, and its image contains some similar sector $D_{\rho_{\mathcal{Z}},\delta'}$ (with $0 < \delta' < \delta$) such that $\mathcal{Z}(z)$ extends analytically to $D_{\rho_{\mathcal{Z}},\delta'}$ by (18.34), and it follows easily by (18.34) and (18.32) that in $D_{\rho_{\mathcal{Z}},\delta'}$, with some $f_1(z)$ analytic at $\rho_{\mathcal{Z}}$,

$$\mathcal{Z}(z) = f_1(z) + a_1(\rho_{\mathcal{Z}} - z)^\alpha + o\bigl(|\rho_{\mathcal{Z}} - z|^\alpha\bigr), \qquad \text{as } z \to \rho_{\mathcal{Z}}, \tag{18.36}$$

where

$$a_1 = a\frac{\rho\Phi(\rho)^{\alpha-1}}{(1-\nu)^{\alpha+1}}. \tag{18.37}$$

As noted above, $\mathcal{Z}(z)$ has no other singularities on $|z| = \rho$, and singularity analysis [40] applies and shows, using (18.33),

$$Z_n \sim \frac{a_1}{\Gamma(-\alpha)} n^{-\alpha-1} \rho_{\mathcal{Z}}^{\alpha-n} \sim \frac{\rho}{(1-\nu)^{\alpha+1}} \Phi(\rho)^{n-1} w_n. \tag{18.38}$$

However, we will show in greater generality in Theorem 19.34 and Remark 19.35 (by a straightforward reduction to the case $\rho = 1$ using (4.3)) that (18.33) always implies (18.38) when $\nu < 1$, without any assumption like (18.32) on $\Phi(z)$.

If $\nu = 1$, we assume $1 < \alpha < 2$, since (18.32) with $\alpha > 2$ implies $\Phi''(\rho) < \infty$ and thus $\sigma^2 < \infty$, so (18.30) and Theorem 18.11 would apply. We now have $F'(\rho) = 0$, and (18.32)–(18.34) yield, in some domain $D_{\rho_{\mathcal{Z}},\delta'}$,

$$\mathcal{Z}(z) = \rho - \left(\frac{\Phi(\rho)}{a}\right)^{1/\alpha} \left(1 - \frac{z}{\rho_{\mathcal{Z}}}\right)^{1/\alpha} + \dots. \tag{18.39}$$

Singularity analysis yields

$$Z_n \sim \frac{1}{|\Gamma(-1/\alpha)|}\left(\frac{\Phi(\rho)}{a}\right)^{1/\alpha} n^{-1-1/\alpha}\rho_{\mathcal{Z}}{}^{-n}. \tag{18.40}$$

However, we have already proved in Theorem 18.14(i) (assuming $\rho = 1$, without loss of generality) that (18.33) implies (18.40) in this case, without any assumption like (18.32) on $\Phi(z)$.

Example 18.20. If $\alpha \geqslant 2$ is an integer, (18.32) does not exhibit a singularity. Instead we consider $\mathbf{w}$ with, for some f analytic at ρ,

$$\Phi(z) = f(z) + a(\rho - z)^\alpha \log(\rho - z) + O\big(|\rho - z|^\alpha\big), \tag{18.41}$$

as $z \to \rho$ in some sector $D_{\rho,\delta}$. This includes the case $w_k = (k+1)^{-\alpha-1}$, see Flajolet and Sedgewick [40, Section VI.8].

In the case $\nu < 1$, we obtain as above

$$\mathcal{Z}(z) = f_1(z) + a_1(\rho_{\mathcal{Z}} - z)^\alpha \log(\rho_{\mathcal{Z}} - z) + O\big(|\rho_{\mathcal{Z}} - z|^\alpha\big), \tag{18.42}$$

as $z \to \rho_{\mathcal{Z}}$ in some sector, with $f_1(z)$ analytic at $\rho_{\mathcal{Z}}$ and a_1 given by (18.37). We again obtain by singularity analysis

$$Z_n \sim \frac{\rho}{(1-\nu)^{\alpha+1}}\Phi(\rho)^{n-1} w_n, \tag{18.43}$$

which is another instance of (19.118).

In the case $\nu = 1$, we consider only $\alpha = 2$, since $\sigma^2 < \infty$ if $\alpha > 2$. Then (18.41) yields (we have $a < 0$ in this case)

$$\mathcal{Z}(z) = \rho - \left(\frac{2\Phi(\rho)}{-a}\right)^{1/2}\big(1 - z/\rho_{\mathcal{Z}}\big)^{1/2}\big(-\log\left(1 - z/\rho_{\mathcal{Z}}\right)\big)^{-1/2} + \dots. \tag{18.44}$$

Singularity analysis [40, Theorems VI.2–3] gives another proof of (18.17) in the special case (18.41) (again assuming $\rho = 1$, as we may).

Example 18.21. Define $\mathbf{w}$ by $\Phi(z) = w_0 + \sum_{j=0}^\infty 2^{-2j} z^{2^j}$, for some $w_0 > 0$; thus $\operatorname{supp}(\mathbf{w})$ is the lacunary sequence $\{0\} \cup \{2^j\}$. Then $\rho = 1$, $\Phi(\rho) = w_0 + 4/3$ and $\Phi'(\rho) = 2$; hence $\nu = \Psi(\rho) = 2/(w_0 + 4/3)$. The function $\Phi(z)$ is analytic in the unit disc and has the unit circle as a natural boundary; it cannot be extended analytically at any point. (See e.g. Rudin [101, Remark 16.4 and Theorem 16.6].)

Taking $w_0 > 2/3$, we have $\nu < 1$; hence, $F'(\rho) > 0$ by (18.35). Thus F maps the unit circle onto a closed curve Γ that goes vertically through $F(1) = \rho_z$, and since F cannot be continued analytically across the unit circle, $\mathcal{Z}(z)$ cannot be continued analytically across the curve Γ. In particular, $\mathcal{Z}(z)$ is not analytic in any sector $D_{\rho_{\mathcal{Z}},\delta'}$.

19. Largest degrees and boxes

Consider a random allocation $B_{m,n} = (Y_1, \dots, Y_n)$ and arrange $Y_1, \dots, Y_n$ in decreasing order as $Y_{(1)} \geqslant Y_{(2)} \geqslant \dots$. Thus, $Y_{(1)}$ is the largest number of balls in any box, $Y_{(2)}$ is the second largest, and so on.

By Lemma 17.1, we may also consider the random tree $\mathcal{T}_n$ by taking $m = n-1$; then $Y_{(1)}$ is the largest outdegree in $\mathcal{T}_n$, $Y_{(2)}$ is the second largest outdegree, and so on.

As usual, we consider asymptotics as $n \to \infty$ and $m/n \to \lambda$. (Thus $\lambda = 1$ in the tree case.) We usually ignore the cases $m/n \to 0$ and $m/n \to \infty$; these are left to the reader as open problems. (See e.g. Kolchin, Sevast'yanov and Chistyakov [77], Kolchin [76], Pavlov [96] and Kazimirov [70] for examples of such results.)

The results in Sections 7 and 11 suggest that $Y_{(1)}$ is small when $\lambda < \nu$, but large (perhaps of order n) when $\lambda > \nu$, which is one aspect of the phase transition at $\lambda = \nu$. We will see that this roughly is correct, but that the full story is somewhat more complicated.

We study the cases $\lambda \leqslant \nu$ and $\lambda > \nu$ separately; we also consider separately several subcases of the first case where we can give more precise results.

We first note that the case $\omega < \infty$, when the box capacities (node degrees in the tree case) are bounded is trivial: w.h.p. the maximum is attained in many boxes.

Theorem 19.1. *Let $\mathbf{w} = (w_k)_{k \geqslant 0}$ be a weight sequence with $w_0 > 0$ and $\omega < \infty$. Suppose that $n \to \infty$ and $m = m(n)$ with $m/n \to \lambda > 0$. Then $Y_{(j)} = \omega$ w.h.p. for every fixed j.*

Proof. Clearly, each $Y_i \leqslant \omega$, so $Y_{(j)} \leqslant Y_{(1)} \leqslant \omega$.

We assume tacitly, as always, that $B_{m,n}$ exists, i.e. $Z(m,n) > 0$, and thus $m \leqslant \omega n$, so $\lambda \leqslant \omega$. By Theorem 11.4 if $\lambda < \omega$, and Remark 11.10 if $\lambda = \omega$, $N_\omega(B_{m,n})/n \xrightarrow{\mathrm{p}} \pi_\omega > 0$. In particular, $N_\omega(B_{m,n}) \xrightarrow{\mathrm{p}} \infty$, and thus $\mathbb{P}(Y_{(j)} = \omega) \to 1$. $\square$

19.1. The case $\lambda \leqslant \nu$.

In the case $\lambda \leqslant \nu$, we show that, indeed, all Y_i are small. Theorems 19.2–19.3 yield (w.h.p.) a bound $o(n)$ when $\lambda = \nu$, and a much stronger logarithmic bound $O(\log n)$ when $\lambda < \nu$. (In the tree case, we have $\lambda = 1$, so these are the cases $\nu = 1$ and $\nu > 1$.)

Example 19.27 shows that in general, the bound $o(n)$ when $\lambda = \nu$ is essentially best possible; at least, for any given $\varepsilon > 0$, we can have $Y_{(1)} > n^{1-\varepsilon}$ w.h.p. (We do not know precisely how fast $Y_{(1)}$ can grow; see Problem 19.31.)

Theorem 19.2. *Let $\mathbf{w} = (w_k)_{k \geqslant 0}$ be a 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 $m/n \to \lambda$ where $0 \leqslant \lambda < \infty$. If $\lambda \leqslant \nu$, then $Y_{(1)} = o_{\mathrm{p}}(n)$.*

Equivalently, $Y_{(1)}/n \xrightarrow{\mathrm{p}} 0$.

Proof. The case $\lambda = 0$ is trivial, since $Y_{(1)}/n \leqslant m/n \to \lambda$. The case $\lambda = \omega$ is also trivial, since then $\omega < \infty$ and $Y_{(1)} \leqslant \omega$. As above, $\lambda > \omega$ is impossible. Hence we may assume $0 < \lambda < \omega$ and $\nu \geqslant \lambda > 0$, which implies $\tau > 0$, where $\Psi(\tau) = \lambda$, cf. Theorem 11.4. We may then for convenience replace (w_k) by the equivalent weight sequence (π_k) in (11.13); we may thus assume that $\mathbf{w}$ is a probability weight sequence with $\tau = 1$, and thus $\rho \geqslant \tau = 1$, and then the corresponding random variable ξ has $\mathbb{E}\,\xi = \lambda$.

By (18.11) and symmetry, for any $k \geqslant 0$,

$$\mathbb{P}(Y_{(1)} = k) \leqslant n\,\mathbb{P}(Y_1 = k) = n\frac{w_k Z(m-k, n-1)}{Z(m,n)}. \tag{19.1}$$

Furthermore, $w_k = \pi_k = \mathbb{P}(\xi = k) \leqslant 1$ and, using Example 11.2, $Z(m,n) = \mathbb{P}(S_n = m) = e^{o(n)}$ by Lemma 14.3 ($\rho > 1$) or 14.4 ($\rho = 1$). We turn to estimating $Z(m-k, n-1)$.

Let $0 < \varepsilon < \lambda$, and define τ_ε by $\Psi(\tau_\varepsilon) = \lambda - \varepsilon$. Since $\Psi(\tau) = \lambda$, we have $0 < \tau_\varepsilon < \tau = 1$.

For each n, choose $k = k(n) \in [\varepsilon n, m]$ such that $Z(m-k, n-1)$ is maximal. We have $\varepsilon \leqslant k/n \leqslant m/n \to \lambda$; choose a subsequence such that k/n converges, say $k/n \to \gamma$ with $\varepsilon \leqslant \gamma \leqslant \lambda$. Then, along the subsequence, $(m-k)/(n-1) \to \lambda - \gamma$.

By Theorem 18.1 (and Remark 18.2, ignoring the trivial case $Z(m-k, n-1) = 0$), using $\tau_\varepsilon < 1$, $\gamma \geqslant \varepsilon$ and (11.16),

$$\begin{aligned}\frac{1}{n}\log Z(m-k, n-1) &\to \log \inf_{t \geqslant 0} \frac{\Phi(t)}{t^{\lambda-\gamma}} \leqslant \log \inf_{0 \leqslant t \leqslant \tau_\varepsilon} \frac{\Phi(t)}{t^{\lambda-\gamma}} \\ &\leqslant \log \inf_{0 \leqslant t \leqslant \tau_\varepsilon} \frac{\Phi(t)}{t^{\lambda-\varepsilon}} = \log \frac{\Phi(\tau_\varepsilon)}{\tau_\varepsilon^{\lambda-\varepsilon}} =: c_\varepsilon,\end{aligned}$$

say, where Remark 11.5 shows that, since $\tau_\varepsilon \neq 1$,

$$c_\varepsilon < \log\bigl(\Phi(1)/1^{\lambda-\varepsilon}\bigr) = 0. \tag{19.2}$$

We have shown that

$$\limsup_{n\to\infty} \frac{1}{n}\log Z(m-k, n-1) \leqslant c_\varepsilon \tag{19.3}$$

for $k = k(n)$ and any subsequence such that k/n converges; it follows that (19.3) holds for the full sequence. In other words,

$$\log Z(m-k, n-1) \leqslant c_\varepsilon n + o(n) \tag{19.4}$$

for our choice $k = k(n)$ that maximises the left-hand side, and thus uniformly for all $k \in [\varepsilon n, m]$. Using (19.4) and, as said above, Lemma 14.4 in (19.1) we obtain, recalling (19.2),

$$\mathbb{P}(Y_{(1)} \geqslant \varepsilon n) = \sum_{k=\varepsilon n}^{m} \mathbb{P}(Y_{(1)} = k) \leqslant mne^{c_\varepsilon n + o(n)} e^{o(n)} = e^{c_\varepsilon n + o(n)} \to 0.$$

In other words, for any $\varepsilon > 0$, $Y_{(1)} < \varepsilon n$ w.h.p., which is equivalent to $Y_{(1)} = o_{\mathrm{p}}(n)$. $\square$

The following logarithmic bound when $\lambda < \nu$ is essentially due to Meir and Moon [86] (who studied the tree case).

Theorem 19.3. *Let* $\mathbf{w} = (w_k)_{k\geqslant 0}$ *be a 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* $m/n \to \lambda$. *Assume* $0 < \lambda < \nu$, *and define* $\tau \in (0, \rho)$ *by* $\Psi(\tau) = \lambda$.

(i) *Then* $\tau < \rho$ *and*

$$Y_{(1)} \leqslant \frac{1}{\log(\rho/\tau)} \log n + o_{\mathrm{p}}(\log n). \tag{19.5}$$

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

(iii) *If further* $w_k^{1/k} \to 1/\rho$ *as* $k \to \infty$, *then, for every fixed* $j \geqslant 1$,

$$\frac{Y_{(j)}}{\log n} \xrightarrow{\mathrm{p}} \frac{1}{\log(\rho/\tau)}. \tag{19.6}$$

Recall that $1/\rho = \limsup_{k\to\infty} w_k^{1/k}$, see (3.5), so the extra assumption $w_k^{1/k} \to 1/\rho$ as $k \to \infty$ in (iii) holds unless the weight sequence is rather irregular. (The proof shows that the assumption can be weakened to $\mathbb{P}(\xi \geqslant k)^{1/k} \to \tau/\rho$.)

It is not difficult to show Theorem 19.3 directly, but we prefer to postpone the proof and use parts of the more refined Theorem 19.7 below, in order to avoid some repetitions of arguments. Further results, under additional assumptions, are given in Sections 19.3–19.4.

We conjecture that Theorem 19.3 holds also for $\lambda = 0$. Since then $\tau = 0$, this means the following. (This seems almost obvious given the result for positive λ in Theorem 19.3, where the constant $1/\log(\rho/\tau) \to 0$ as $\lambda \to 0$ and thus $\tau \to 0$, but there is no general monotonicity and we leave this as an open problem.)

Conjecture 19.4. *If* $\rho > 0$ *and* $m/n \to 0$, *then* $Y_{(1)} = o_{\mathrm{p}}(\log n)$.

19.2. **The subcase $\sigma^2 < \infty$.** In the case $\sigma^2 := \operatorname{Var} \xi < \infty$ (which includes the case $\lambda < \nu$), there is a much more precise result, which says that, simply, the largest numbers $Y_{(1)}, Y_{(2)} \dots$ asymptotically have the same distribution as the largest elements in the i.i.d. sequence $\xi_1, \dots, \xi_n$. (Provided we choose the distribution of ξ correctly, and possibly depending on n, see below for details.) In other words, the conditioning in Example 11.2 then has asymptotically no effect on the largest elements of the sequence. (When $\sigma^2 = \infty$ this is no longer necessarily true, however, as we shall see in Example 19.27.)

In order to state this precisely, we now assume that $\omega = \infty$ (see Theorem 19.1 otherwise) and $0 < \lambda \leqslant \nu$, and define as usual τ by $\Psi(\tau) = \lambda$, and let ξ be a random variable with the distribution in (11.13).

If $m/n \leqslant \nu$, we further define τ_n by $\Psi(\tau_n) = m/n$, and let $\xi^{(n)}$ be the random variable with the distribution in (14.11). We will only use τ_n and $\xi^{(n)}$ in the case $\lambda < \nu$, so $m/n \to \lambda < \nu$ and τ_n really is defined (at least for large n); furthermore $\tau_n \to \tau < \rho$ and $\xi^{(n)} \xrightarrow{\mathrm{d}} \xi$.

We further let $\xi_1, \dots, \xi_n$ and (when $\lambda < \nu$) $\xi_1^{(n)}, \dots, \xi_n^{(n)}$ be i.i.d. sequences of copies of ξ and $\xi^{(n)}$, respectively, and we arrange them in decreasing order as $\xi_{(1)} \geqslant \dots \geqslant \xi_{(n)}$ and $\xi_{(1)}^{(n)} \geqslant \dots \geqslant \xi_{(n)}^{(n)}$. Finally, we introduce the counting variables, for any subset $A \subseteq \mathbb{N}_0$,

$$N_A := |\{i \leqslant n : Y_i \in A\}|, \tag{19.7}$$

$$\overline{N}_A := |\{i \leqslant n : \xi_i \in A\}|, \tag{19.8}$$

$$\overline{N}_A^{(n)} := |\{i \leqslant n : \xi_i^{(n)} \in A\}|. \tag{19.9}$$

(N_A and $\overline{N}_A$ also depend on n, but as usual, we for simplicity do not show this in the notation.) Note that $\overline{N}_A$ and $\overline{N}_A^{(n)}$ simply have binomial distributions $\overline{N}_A \sim \mathrm{Bi}(n, \mathbb{P}(\xi \in A))$ and $\overline{N}_A^{(n)} \sim \mathrm{Bi}(n, \mathbb{P}(\xi^{(n)} \in A))$.

We have

$$Y_{(j)} \leqslant k \iff N_{[k+1,\infty)} < j, \tag{19.10}$$

and similarly for $\xi_{(j)}$ and $\xi_{(j)}^{(n)}$. Thus it is elementary to obtain asymptotic results for the maximum $\xi_{(1)}$ of i.i.d. variables, and more generally for $\xi_{(j)}$ and $\xi_{(j)}^{(n)}$, see e.g. Leadbetter, Lindgren and Rootzén [82].

We introduce three different probability metrics to state the results. For discrete random variables X and Y with values in $\mathbb{N}_0$ (the case we are interested in here), we define the *Kolmogorov distance*

$$d_{\mathrm{K}}(X,Y) := \sup_{x \in \mathbb{N}_0} |\,\mathbb{P}(X \leqslant x) - \mathbb{P}(Y \leqslant x)| \tag{19.11}$$

and the *total variation distance*

$$d_{\mathrm{TV}}(X,Y) := \sup_{A \subseteq \mathbb{N}_0} |\,\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|. \tag{19.12}$$

In order to treat also the case with variables tending to ∞, we further define the *modified Kolmogorov distance*

$$\widetilde{d}_{\mathrm{K}}(X,Y) := \sup_{x \in \mathbb{N}_0} \frac{|\,\mathbb{P}(X \leqslant x) - \mathbb{P}(Y \leqslant x)|}{1+x}. \tag{19.13}$$

For $\widetilde{d}_{\mathrm{K}}$, we also allow random variables in $\overline{\mathbb{N}}_0$, i.e., we allow the value ∞. (Furthermore, the definitions of d_{K} and d_{TV} and the results for them in the lemma below extend to random variables with values in $\mathbb{Z}$. The definitions extend further to random variables with values in $\mathbb{R}$ for d_{K}, and in any space for d_{TV}, but not all properties below hold in this generality.)

Note that these distances depend only on the distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$, so $d(\mathcal{L}(X), \mathcal{L}(Y))$ might be a better notation, but we find it convenient to allow both notations, as well as the mixed $d(X, \mathcal{L}(Y))$.

It is obvious that the three distances above are metrics on the space of probability measures on $\mathbb{N}_0$ (or on $\overline{\mathbb{N}}_0$).

We collect a few simple, and mostly well-known, facts for these three metrics in a lemma; the proofs are left to the reader.

Lemma 19.5. (i) *For any random variables X and Y with values in $\mathbb{N}_0$,*

$$\widetilde{d}_{\mathrm{K}}(X,Y) \leqslant d_{\mathrm{K}}(X,Y) \leqslant d_{\mathrm{TV}}(X,Y).$$

(ii) *For any X and $X_1, X_2, \dots$ with values in $\mathbb{N}_0$,*

$$X_n \xrightarrow{\mathrm{d}} X \iff d_{\mathrm{TV}}(X_n,X) \to 0 \iff d_{\mathrm{K}}(X_n,X) \to 0 \\ \iff \widetilde{d}_{\mathrm{K}}(X_n,X) \to 0.$$

(iii) *For any X and $X_1, X_2, \dots$ with values in $\overline{\mathbb{N}}_0$,*

$$X_n \xrightarrow{\mathrm{d}} X \iff \widetilde{d}_{\mathrm{K}}(X_n,X) \to 0.$$

In particular,

$$X_n \xrightarrow{\mathrm{p}} \infty \iff \widetilde{d}_{\mathrm{K}}(X_n,\infty) \to 0.$$

(iv) *For any X_n and X'_n with values in $\overline{\mathbb{N}}_0$, $\widetilde{d}_{\mathrm{K}}(X_n,X'_n) \to 0 \iff \bigl|\mathbb{P}(X_n \leqslant x) - \mathbb{P}(X'_n \leqslant x)\bigr| \to 0$ for every fixed $x \geqslant 0$.*

(v) *For any X_n and X'_n, $d_{\mathrm{TV}}(X_n,X'_n) \to 0 \iff$ there exists a coupling (X_n,X'_n) with $X_n = X'_n$ w.h.p. (We denote this also by $X_n \overset{\mathrm{d}}{\approx} X'_n$.)*

(vi) *The supremum in (19.12) is attained, and the absolute value sign is redundant. In fact, if $A := \{i : \mathbb{P}(X=i) > \mathbb{P}(Y=i)\}$, then $d_{\mathrm{TV}}(X,Y) = \mathbb{P}(X \in A) - \mathbb{P}(Y \in A)$.*

(vii) *For any X and Y with values in $\mathbb{N}_0$,*

$$d_{\mathrm{TV}}(X,Y) = \sum_{x\in\mathbb{N}_0} \bigl(\mathbb{P}(X=x) - \mathbb{P}(Y=x)\bigr)_+ = \tfrac{1}{2} \sum_{x\in\mathbb{N}_0} |\,\mathbb{P}(X=x) - \mathbb{P}(Y=x)|.$$

□

Remark 19.6. The three metrics are, by Lemma 19.5(ii), equivalent in the usual sense that they define the same topology, but they are not uniformly equivalent. For example, if $X_n \sim \mathrm{Po}(n)$, $X'_n := 2\lfloor X_n/2 \rfloor$ (i.e., X_n rounded down to an even integer) and $X''_n := X'_n + 1$, then $d_{\mathrm{K}}(X'_n, X''_n) \to 0$ as $n \to \infty$, but $d_{\mathrm{TV}}(X'_n, X''_n) = 1$.

We define $\mathrm{Po}(\infty)$ as the distribution of a random variable that equals ∞ identically.

After all these preliminaries, we state the result (together with some supplementary results). There are really two versions; it turns out that for general sequences $m(n)$, we have to use the random variables $\xi^{(n)}$, with $\mathbb{E}\,\xi^{(n)} = m(n)/n$ exactly tuned to $m(n)$, but under a weak assumption we can replace $\xi^{(n)}$ by ξ and obtain a somewhat simpler statement, which we choose as our main formulation. (This goes back to Meir and Moon [87], who proved (i) in the tree case, assuming $\lambda < \nu$; see also Kolchin, Sevast'yanov and Chistyakov [77, Theorem 1.6.1] and Kolchin [76, Theorem 1.5.2] for $Y_{(1)}$ in the special case in Example 12.1.)

Theorem 19.7. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be a weight sequence with $w_0 > 0$ and $\omega = \infty$. Suppose that $n \to \infty$ and $m = m(n)$ with $m = \lambda n + o(\sqrt{n})$ where $0 < \lambda \leqslant \nu$, and use the notation above. Suppose further that $\sigma^2 := \operatorname{Var}\xi < \infty$. (This is redundant when $\lambda < \nu$.)*

(i) *If (possibly for n in a subsequence) $h(n)$ are integers such that $nP(\xi \geqslant h(n)) \to \alpha$, for some $\alpha \in [0,\infty]$, then*

$$N_{[h(n),\infty)} := |\{i : Y_i \geqslant h(n)\}| \xrightarrow{\mathrm{d}} \operatorname{Po}(\alpha).$$

(ii) *If $h(n)$ are integers such that $nP(\xi \geqslant h(n)) \to 0$, then w.h.p. $Y_{(1)} < h(n)$.*

(iii) *If $h(n)$ are integers such that $nP(\xi \geqslant h(n)) \to \infty$, then, for every fixed j, w.h.p. $Y_{(j)} \geqslant h(n)$.*

(iv) *For any sequence $h(n)$, $\widetilde{d}_{\mathrm{K}}\bigl(N_{[h(n),\infty)}, \overline{N}_{[h(n),\infty)}\bigr) \to 0$.*

(v) *For every fixed j, $d_{\mathrm{K}}\bigl(Y_{(j)}, \xi_{(j)}\bigr) \to 0$.*

(vi) *$d_{\mathrm{TV}}\bigl(Y_{(1)}, \xi_{(1)}\bigr) \to 0$.*

If $\lambda < \nu$, the condition $m = \lambda n + o(\sqrt{n})$ can be weakened to $m/n = \lambda + o(1/\log n)$.

Moreover, if $\lambda < \nu$, then the results hold for any $m = m(n)$ with $m/n \to \lambda$, provided ξ is replaced by $\xi^{(n)}$, $\overline{N}$ by $\overline{N}^{(n)}$ and $\xi_{(j)}$ by $\xi^{(n)}_{(j)}$.

Remark 19.8. In the version with $\xi^{(n)}$, we do not need λ at all. By considering subsequences, it follows that it suffices that $0 < c \leqslant m/n \leqslant C < \nu$. (Cf. Theorem 11.6.) Furthermore, this version extends to the case $\lambda = \nu$ and $m/n \leqslant \nu$, but we have ignored this case for simplicity.

Problem 19.9. *Is Theorem 19.7 (in the $\xi^{(n)}$ version) true also for $\lambda = 0 < \nu$?*

The total variation approximation in (vi) is stronger than the Kolmogorov distance approximation in (v), and our proof is considerably longer, but for many purposes (v) is enough. We conjecture that total variation approximation holds for every $Y_{(j)}$, and not just for $Y_{(1)}$; presumably this can be shown by a modification of the proof for $Y_{(1)}$ below, but we have not checked the details and leave this as an open problem. Furthermore, we believe that the result extends to the joint distribution of finitely many $Y_{(j)}$. (The corresponding result in (v), using a multivariate version of the Kolmogorov distance, is easily verified by the methods below.)

Problem 19.10. *Does $d_{\mathrm{TV}}\bigl(Y_{(j)}, \xi_{(j)}\bigr) \to 0$ hold for every fixed j, under the assumptions of Theorem 19.7?*

Proof of Theorem 19.7. As in the proof of Theorem 19.2, we may replace (w_k) by the equivalent weight sequence (π_k) in (11.13). We may thus assume that $\mathbf{w}$ is a probability weight sequence with $\tau = 1$, and thus $\rho > \tau = 1$, and the corresponding random variable ξ has $\mathbb{E}\,\xi = \lambda$. We consider first the version with ξ, assuming $m = \lambda n + o(\sqrt{n})$, and discuss afterwards the modifications for $\xi^{(n)}$.

We begin by looking again at (18.11):

$$\mathbb{P}(Y_1 = k) = \frac{w_k Z(m-k, n-1)}{Z(m,n)}. \tag{19.14}$$

When $m = \lambda n + o(\sqrt{n})$, we may apply Lemma 14.1 and Remark 14.2 and thus, with $d := \operatorname{span}(\mathbf{w})$,

$$Z(m,n) = \mathbb{P}(S_n = m) = \frac{d + o(1)}{\sqrt{2\pi\sigma^2 n}}. \tag{19.15}$$

Moreover, by (14.9), for any k,

$$Z(m-k, n-1) = \mathbb{P}(S_{n-1} = m-k) \leqslant \frac{d + o(1)}{\sqrt{2\pi\sigma^2 n}}. \tag{19.16}$$

Consequently, (19.14) yields, uniformly for all k,

$$\mathbb{P}(Y_1 = k) \leqslant (1 + o(1)) w_k = (1 + o(1))\, \mathbb{P}(\xi = k). \tag{19.17}$$

In particular, we may sum over $k \geqslant K$ and obtain, for any $K = K(n)$,

$$\mathbb{P}(Y_1 \geqslant K) \leqslant (1 + o(1))\, \mathbb{P}(\xi \geqslant K). \tag{19.18}$$

Since, by assumption, $\mathbb{E}\,\xi^2 < \infty$, we have $\mathbb{P}(\xi \geqslant K) = o(K^{-2})$ as $K \to \infty$. Hence, for every fixed $\delta > 0$, $\mathbb{P}(\xi \geqslant \delta\sqrt{n}) = o(n^{-1})$. It follows that there exists a sequence $\delta_n \to 0$ such that $\mathbb{P}(\xi \geqslant \delta_n\sqrt{n}) = o(n^{-1})$. Consequently, defining $B(n) := \delta_n\sqrt{n}$, we have $B(n) = o(\sqrt{n})$ and

$$\mathbb{P}(\xi \geqslant B(n)) = o(n^{-1}), \tag{19.19}$$

and thus, by (19.18) and symmetry,

$$\mathbb{P}(Y_{(1)} \geqslant B(n)) \leqslant n\, \mathbb{P}(Y_1 \geqslant B(n)) = n\big(1 + o(1)\big)\, \mathbb{P}(\xi \geqslant B(n)) = o(1). \tag{19.20}$$

Hence, $Y_{(1)} < B(n)$ w.h.p.

Similarly, $\mathbb{P}(\xi_{(1)} \geqslant B(n)) \leqslant n\, \mathbb{P}(\xi_1 \geqslant B(n)) = o(1)$, so $\xi_{(1)} < B(n)$ w.h.p.

(i): Write, for convenience, $N := N_{[h(n),B(n)]}$, and note that w.h.p. $Y_{(1)} \leqslant B(n)$ and then $N = N_{[h(n),\infty)}$. (We assume for simplicity $h(n) \leqslant B(n)$; otherwise we let $N := 0$, leaving the trivial modifications in this case to the reader.)

Moreover, for $k \leqslant B(n) = o(\sqrt{n})$, we have $(m-k) - (n-1)\lambda = o(\sqrt{n})$, and thus Remark 14.2 shows that, for any $k = k(n) \leqslant B(n)$,

$$Z(m-k, n-1) = \mathbb{P}(S_{n-1} = m-k) = \frac{d + o(1)}{\sqrt{2\pi\sigma^2 n}}. \tag{19.21}$$

Since we here may take $k = k(n)$ that maximises or minimises this for $k \leqslant B(n)$, it follows that (19.21) holds uniformly for all $k \leqslant B(n)$. Consequently, by (19.14), (19.15) and (19.21),

$$\mathbb{P}(Y_1 = k) = (1 + o(1)) w_k = (1 + o(1))\, \mathbb{P}(\xi = k), \tag{19.22}$$

uniformly for all $k \leqslant B(n)$. By the assumption and (19.19), this yields

$$\begin{aligned} \mathbb{E}\, N &= n \sum_{k=h(n)}^{B(n)} \mathbb{P}(Y_1 = k) = n \sum_{k=h(n)}^{B(n)} \bigl(1+o(1)\bigr)\, \mathbb{P}(\xi = k) \\ &= \bigl(1+o(1)\bigr) n\, \mathbb{P}\bigl(h(n) \leqslant \xi \leqslant B(n)\bigr) \\ &= \bigl(1+o(1)\bigr) n \bigl(\mathbb{P}(\xi \geqslant h(n)) - \mathbb{P}(\xi > B(n))\bigr) \to \alpha. \end{aligned}$$

Similarly, again using the symmetry as well as Lemma 14.1 and Remark 14.2,

$$\begin{aligned} \mathbb{E}\, N(N-1) &= n(n-1)\, \mathbb{P}\bigl(Y_1, Y_2 \in [h(n), B(n)]\bigr) \\ &= n(n-1) \sum_{k_1,k_2=h(n)}^{B(n)} \mathbb{P}(Y_1 = k_1 \text{ and } Y_2 = k_2) \\ &= n(n-1) \sum_{k_1,k_2=h(n)}^{B(n)} \frac{w_{k_1} w_{k_2} Z(m-k_1-k_2, n-2)}{Z(m,n)} \\ &= n(n-1) \sum_{k_1,k_2=h(n)}^{B(n)} \mathbb{P}(\xi = k_1)\, \mathbb{P}(\xi = k_2) \bigl(1+o(1)\bigr) \\ &= \bigl(1+o(1)\bigr) n^2 \bigl(\mathbb{P}(\xi \geqslant h(n)) - \mathbb{P}(\xi > B(n))\bigr)^2 \\ &\to \alpha^2. \end{aligned}$$

Moreover, the same argument works for any factorial moment $\mathbb{E}(N)_\ell$ and yields $\mathbb{E}(N)_\ell \to \alpha^\ell$ for every $\ell \geqslant 1$. If $\alpha < \infty$, we thus obtain $N \xrightarrow{\mathrm{d}} \mathrm{Po}(\alpha)$ by the method of moments, and the result follows, since $N = N_{[h(n),\infty)}$ w.h.p.

If $\alpha = \infty$, this argument yields

$$\mathbb{E}(N)_\ell \sim \bigl(n\, \mathbb{P}(\xi \geqslant h(n))\bigr)^\ell \to \infty \tag{19.23}$$

for every $\ell \geqslant 1$, and we make a thinning: Let A be a constant and let $q := A/\bigl(n\, \mathbb{P}(\xi \geqslant h(n))\bigr)$; then $q \to A/\alpha = 0$. We consider only n that are so large that $q < 1$. We then randomly, and independently, mark each box with probability q. Let N' be the random number of marked boxes i such that $Y_i \in [h(n), B(n)]$. Then, for every $\ell \geqslant 1$, using (19.23),

$$\mathbb{E}(N')_\ell = (n)_\ell q^\ell\, \mathbb{P}\bigl(Y_1, \dots, Y_\ell \in [h(n), B(n)]\bigr) = q^\ell\, \mathbb{E}(N)_\ell \to A^\ell. \tag{19.24}$$

Consequently, by the method of moments, $N' \xrightarrow{\mathrm{d}} \mathrm{Po}(A)$. In particular, this shows, for every fixed x,

$$\mathbb{P}(N < x) \leqslant \mathbb{P}(N' < x) \to \mathbb{P}(\mathrm{Po}(A) < x),$$

which can be made arbitrarily small by taking A large. Hence, $\mathbb{P}(N < x) \to 0$ for every fixed x, i.e., $N \xrightarrow{\mathrm{p}} \infty$ and thus $N_{[h(n),\infty)} \xrightarrow{\mathrm{p}} \infty$, as we claim in this case.

(ii): Part (i) applies with $\alpha = 0$, and yields $N_{[h(n),\infty)} \xrightarrow{\mathrm{p}} 0$, which means $N_{[h(n),\infty)} = 0$ w.h.p. Thus $Y_{(j)} < h(n)$ w.h.p. by (19.10).

(iii): Part (i) applies with $\alpha = \infty$, and yields $N_{[h(n),\infty)} \xrightarrow{\mathrm{p}} \infty$. Thus, for every fixed j, by (19.10), $\mathbb{P}(Y_{(j)} < h(n)) = \mathbb{P}(N_{[h(n),\infty)} < j) \to 0$.

(iv): Suppose not. Then there exists a sequence $h(n)$ and an $\varepsilon > 0$ such that, for some subsequence,

$$\widetilde{d}_{\mathrm{K}}\bigl(N_{[h(n),\infty)}, \overline{N}_{[h(n),\infty)}\bigr) > \varepsilon. \tag{19.25}$$

We may select a subsubsequence such that $n\,\mathbb{P}(\xi \geqslant h(n)) \to \alpha$ for some $\alpha \in [0,\infty]$; then $\widetilde{d}_{\mathrm{K}}\bigl(N_{[h(n),\infty)}, \mathrm{Po}(\alpha)\bigr) \to 0$ by (i) and Lemma 19.5(iii). Moreover, along the same subsubsequence, $\overline{N}_{[h(n),\infty)} \sim \mathrm{Bi}\bigl(n, \mathbb{P}(\xi \geqslant h(n)\bigr) \xrightarrow{\mathrm{d}} \mathrm{Po}(\alpha)$, by the standard Poisson approximation for binomial distributions (and rather trivially if $\alpha = \infty$); hence $\widetilde{d}_{\mathrm{K}}\bigl(\overline{N}_{[h(n),\infty)}, \mathrm{Po}(\alpha)\bigr) \to 0$. The triangle inequality yields $\widetilde{d}_{\mathrm{K}}\bigl(N_{[h(n),\infty)}, \overline{N}_{[h(n),\infty)}\bigr) \to 0$ along the subsubsequence, which contradicts (19.25). This contradiction proves (iv).

(v): Suppose not. Then, by (19.11), there is an $\varepsilon > 0$ and a subsequence such that for some $h(n)$,

$$\bigl|\mathbb{P}(Y_{(j)} \leqslant h(n)) - \mathbb{P}(\xi_{(j)} \leqslant h(n))\bigr| \geqslant \varepsilon. \tag{19.26}$$

However, by (19.10), (19.13) and (iv),

$$\begin{aligned}\bigl|\mathbb{P}(Y_{(j)} \leqslant h(n)) - {}&\mathbb{P}(\xi_{(j)} \leqslant h(n))\bigr| \\ &= \bigl|\mathbb{P}(N_{[h(n)+1,\infty)} \leqslant j-1) - \mathbb{P}(\overline{N}_{[h(n)+1,\infty)} \leqslant j-1)\bigr| \\ &\leqslant j\widetilde{d}_{\mathrm{K}}\bigl(N_{[h(n)+1,\infty)}, \overline{N}_{[h(n)+1,\infty)}\bigr) \to 0,\end{aligned}$$

which contradicts (19.26). This contradiction proves (v).

(vi): Let $A = A(n) := \{i : \mathbb{P}(Y_{(j)} = i) > \mathbb{P}(\xi_{(j)} = i)\}$; thus, see Lemma 19.5(vi),

$$d_{\mathrm{TV}}(Y_{(j)}, \xi_{(j)}) = \mathbb{P}(Y_{(j)} \in A) - \mathbb{P}(\xi_{(j)} \in A). \tag{19.27}$$

Let $\delta > 0$. For each n, we partition $\mathbb{N}_0$ into a finite family $\mathcal{P} = \{J_l\}_{l=1}^{L}$ of intervals as follows. First, each $i \in \mathbb{N}_0$ with $\mathbb{P}(\xi_{(1)} = i) \geqslant \delta/2$ is a singleton $\{i\}$; note that there are at most $2/\delta$ such i. The complement of the set of these i consists of at most $2/\delta + 1$ intervals $\tilde{J}_k$ (of which one is infinite). We partition each such interval $\tilde{J}_k$ further into intervals J_l with $\mathbb{P}(\xi_{(1)} \in J_l) \leqslant \delta$ by repeatedly chopping off the largest such subinterval starting at the left endpoint. Since only points with $\mathbb{P}(\xi_{(1)} = i) < \delta/2$ remain, each such interval J_l except the last in each $\tilde{J}_k$ satisfies $\mathbb{P}(\xi_{(1)} \in J_l) > \delta/2$. Hence, our final partition $\{J_l\}$ contains at most $2/\delta + 1$ intervals J_l with $\mathbb{P}(\xi_{(1)} \in J_l) < \delta/2$, while the number of intervals J_l with $\mathbb{P}(\xi_{(1)} \in J_l) \geqslant \delta/2$ is clearly at most $2/\delta$. Consequently, L, the total number of intervals, is at most $4/\delta + 1$.

We write $J_l = [a_l, b_l]$. We say that an interval $J_l \in \mathcal{P}$ is *fat* if $\mathbb{P}(\xi_{(1)} \in J_l) > \delta$, and *thin* otherwise. Note that by our construction, a fat interval is a singleton $\{a_l\}$.

Next, fix a large number D. We say that an interval $J_l = [a_l, b_l] \in \mathcal{P}$ is *good* if $n\,\mathbb{P}(\xi \geqslant a_l) \leqslant D$, and *bad* otherwise.

For any interval J_l,

$$\bigl|\mathbb{P}(Y_{(1)} \in J_l) - \mathbb{P}(\xi_{(1)} \in J_l)\bigr| \leqslant 2d_{\mathrm{K}}(Y_{(1)}, \xi_{(1)}) = o(1) \tag{19.28}$$

by (v).

Let $A_l := A \cap J_l$. Thus A is the disjoint union $\bigcup_l A_l$. (A, J_l and A_l depend on n.)

We note that if J_l is fat, then J_l is a singleton, and either $A_l = J_l$ or $A_l = \emptyset$; in both cases we have, using (19.28),

$$\mathbb{P}(Y_{(1)} \in A_l) - \mathbb{P}(\xi_{(1)} \in A_l) \leqslant 2d_{\mathrm{K}}(Y_{(1)}, \xi_{(1)}) = o(1). \tag{19.29}$$

We next turn to the good intervals. We claim that, uniformly for all good intervals J_l, as $n \to \infty$,

$$\mathbb{P}(Y_{(1)} \in A_l) \leqslant e^{\delta e^D}\,\mathbb{P}(\xi_{(1)} \in A_l) + o(1). \tag{19.30}$$

As usual, we suppose that this is not true and derive a contradiction. Thus, assume that there is an $\varepsilon > 0$ and, for each n in some subsequence, a good interval $J_l = [a_l, b_l]$ (depending on n) such that

$$\mathbb{P}(Y_{(1)} \in A_l) > e^{\delta e^D}\,\mathbb{P}(\xi_{(1)} \in A_l) + \varepsilon. \tag{19.31}$$

If J_l is fat, then (19.31) contradicts (19.29) for large n, so we may assume that J_l is thin, i.e., $\mathbb{P}(\xi_{(1)} \in J_l) \leqslant \delta$.

Let $A_l^{\mathrm{c}} := J_l \setminus A_l$ and $B_l := [b_l + 1, \infty)$. Let $\alpha_n := n\,\mathbb{P}(\xi \in A_l)$, $\beta_n := n\,\mathbb{P}(\xi \in B_l)$ and $\gamma_n := n\,\mathbb{P}(\xi \in A_l^{\mathrm{c}})$. The assumption that J_l is good implies that $\alpha_n + \beta_n + \gamma_n = n\,\mathbb{P}(\xi \geqslant a_l) \leqslant D$. By selecting a subsubsequence we may assume that $\alpha_n \to \alpha$, $\beta_n \to \beta$ and $\gamma_n \to \gamma$ for some real α, β, γ with $\alpha + \beta + \gamma \leqslant D$. Then (i) shows that $N_{B_l} \xrightarrow{\mathrm{d}} \mathrm{Po}(\beta)$; moreover, the proof extends easily (using joint factorial moments) to show that $N_{A_l} \xrightarrow{\mathrm{d}} \mathrm{Po}(\alpha)$, $N_{B_l} \xrightarrow{\mathrm{d}} \mathrm{Po}(\beta)$ and $N_{A_l^{\mathrm{c}}} \xrightarrow{\mathrm{d}} \mathrm{Po}(\gamma)$, jointly and with independent limits.

Similarly, by the method of moments or otherwise (this is a standard Poisson approximation of a multinomial distribution), $\overline{N}_{A_l} \xrightarrow{\mathrm{d}} \mathrm{Po}(\alpha)$, $\overline{N}_{B_l} \xrightarrow{\mathrm{d}} \mathrm{Po}(\beta)$ and $\overline{N}_{A_l^{\mathrm{c}}} \xrightarrow{\mathrm{d}} \mathrm{Po}(\gamma)$, jointly and with independent limits.

Note that

$$Y_{(1)} \in A_l \implies N_{A_l} \geqslant 1 \text{ and } N_{B_l} = 0.$$

Conversely,

$$N_{A_l} \geqslant 1 \text{ and } N_{B_l} = N_{A_l^{\mathrm{c}}} = 0 \implies Y_{(1)} \in A_l.$$

The corresponding results hold for $\xi_{(1)}$. Thus,

$$\mathbb{P}(Y_{(1)} \in A_l) \leqslant \mathbb{P}(N_{A_l} \geqslant 1,\ N_{B_l} = 0) \to \mathbb{P}\bigl(\mathrm{Po}(\alpha) \geqslant 1\bigr)\,\mathbb{P}\bigl(\mathrm{Po}(\beta) = 0\bigr) \tag{19.32}$$

and

$$\mathbb{P}(\xi_{(1)} \in A_l) \geqslant \mathbb{P}(\overline{N}_{A_l} \geqslant 1,\ \overline{N}_{B_l} = \overline{N}_{A_l^{\mathrm{c}}} = 0)$$

$$\to \mathbb{P}\big(\mathrm{Po}(\alpha) \geqslant 1\big)\,\mathbb{P}\big(\mathrm{Po}(\beta) = 0\big)\,\mathbb{P}\big(\mathrm{Po}(\gamma) = 0\big). \quad (19.33)$$

Since $\mathbb{P}\big(\mathrm{Po}(\gamma) = 0\big) = e^{-\gamma}$, (19.32)–(19.33) yield

$$\mathbb{P}(Y_{(1)} \in A_l) - e^{\gamma}\,\mathbb{P}(\xi_{(1)} \in A_l) \leqslant o(1). \quad (19.34)$$

Moreover, $\overline{N}_{J_l} = \overline{N}_{A_l} + \overline{N}_{A_l^{\mathrm{c}}} \xrightarrow{\mathrm{d}} \mathrm{Po}(\alpha+\gamma)$, and thus

$$\begin{aligned}\mathbb{P}(\xi_{(1)} \in J_l) = \mathbb{P}(\overline{N}_{J_l} \geqslant 1,\ \overline{N}_{B_l} = 0) &\geqslant \mathbb{P}(\overline{N}_{J_l} = 1,\ \overline{N}_{B_l} = 0)\\ &\to (\alpha+\gamma)e^{-\alpha-\gamma}e^{-\beta}. \quad (19.35)\end{aligned}$$

We are assuming that J_l is thin, i.e., $\mathbb{P}(\xi_{(1)} \in J_l) \leqslant \delta$, and thus (19.35) yields $(\alpha+\gamma)e^{-\alpha-\gamma}e^{-\beta} \leqslant \delta$ and consequently

$$\gamma \leqslant \alpha + \gamma \leqslant \delta e^{\alpha+\beta+\gamma} \leqslant \delta e^{D}.$$

Hence, (19.34) implies

$$\mathbb{P}(Y_{(1)} \in A_l) \leqslant e^{\delta e^D}\,\mathbb{P}(\xi_{(1)} \in A_l) + o(1),$$

which contradicts (19.31). This contradiction shows that (19.30) holds uniformly for all good intervals.

It remains to consider the bad intervals.

Let $J_\ell = [a_\ell, b_\ell]$ be the rightmost bad interval. If J_ℓ is fat we use (19.29) and if J_ℓ is thin we use (19.28) which gives

$$\mathbb{P}(Y_{(1)} \in A_\ell) \leqslant \mathbb{P}(Y_{(1)} \in J_\ell) \leqslant \mathbb{P}(\xi_{(1)} \in J_\ell) + o(1) \leqslant \delta + o(1).$$

In both cases,

$$\mathbb{P}(Y_{(1)} \in A_\ell) \leqslant \mathbb{P}(\xi_{(1)} \in A_\ell) + \delta + o(1). \quad (19.36)$$

Finally, let A^* be the union of the remaining bad intervals. Then $A^* = [0, a_\ell - 1]$ and by (v),

$$\mathbb{P}(Y_{(1)} \in A^*) = \mathbb{P}(Y_{(1)} < a_\ell) \leqslant \mathbb{P}(\xi_{(1)} < a_\ell) + o(1). \quad (19.37)$$

Furthermore, recalling $n\,\mathbb{P}(\xi \geqslant a_\ell) > D$ since J_ℓ is bad,

$$\mathbb{P}(\xi_{(1)} < a_\ell) = \mathbb{P}(\overline{N}_{[a_\ell,\infty)} = 0) = \big(1 - \mathbb{P}(\xi \geqslant a_\ell)\big)^n \leqslant e^{-n\,\mathbb{P}(\xi \geqslant a_\ell)} \leqslant e^{-D}. \quad (19.38)$$

We obtain by summing (19.30) for all good intervals together with (19.36) and (19.37), recalling that the number of intervals is bounded (for a fixed δ) and using (19.38),

$$\begin{aligned}\mathbb{P}(Y_{(1)} \in A) &= \sum_l \mathbb{P}(Y_{(1)} \in A_l)\\ &\leqslant e^{\delta e^D} \sum_l \mathbb{P}(\xi_{(1)} \in A_l) + o(1) + \delta + \mathbb{P}(\xi_{(1)} < a_\ell)\\ &\leqslant e^{\delta e^D}\,\mathbb{P}(\xi_{(1)} \in A) + o(1) + \delta + e^{-D}.\end{aligned}$$

Consequently,

$$\begin{aligned} d_{\mathrm{TV}}(Y_{(1)}, \xi_{(1)}) &= \mathbb{P}(Y_{(1)} \in A) - \mathbb{P}(\xi_{(1)} \in A) \\ &\leqslant \bigl(e^{\delta e^D} - 1\bigr)\, \mathbb{P}(\xi_{(1)} \in A) + \delta + e^{-D} + o(1) \\ &\leqslant \bigl(e^{\delta e^D} - 1\bigr) + \delta + e^{-D} + o(1), \end{aligned}$$

and thus

$$\limsup_{n\to\infty} d_{\mathrm{TV}}(Y_{(1)}, \xi_{(1)}) \leqslant \bigl(e^{\delta e^D} - 1\bigr) + \delta + e^{-D}. \tag{19.39}$$

Letting first $\delta \to 0$ and then $D \to \infty$, we obtain $d_{\mathrm{TV}}(Y_{(1)}, \xi_{(1)}) \to 0$, which proves (vi).

This completes the proof of the version with ξ and the assumption $m = \lambda n + o(n^{1/2})$. Now remove this assumption, but assume $\lambda < \nu$ and thus $\tau < \rho$. We consider only n with $0 < m/n < \nu$ and thus $0 < \tau_n < \rho$. Denote the distribution (14.11) of $\xi^{(n)}$ by $\mathbf{w}^{(n)}$ (this is a probability weight sequence equivalent to $\mathbf{w}$) and let $S_n^{(n)} := \xi_1^{(n)} + \cdots + \xi_n^{(n)}$. Then, by Example 11.2 applied to $\mathbf{w}^{(n)}$, in analogy with (19.14) (and equivalent to it by (11.9)),

$$\mathbb{P}(Y_1 = k) = \frac{w_k^{(n)} Z(m-k, n-1; \mathbf{w}^{(n)})}{Z(m,n;\mathbf{w}^{(n)})} = \frac{Z(m-k, n-1; \mathbf{w}^{(n)})}{Z(m,n;\mathbf{w}^{(n)})}\, \mathbb{P}(\xi^{(n)} = k). \tag{19.40}$$

Furthermore, for any $y \geqslant 0$, using (14.13),

$$\begin{aligned} \mathbb{P}(Y_{(1)} \geqslant y) \leqslant n\, \mathbb{P}(Y_1 \geqslant y) &= n\, \mathbb{P}\bigl(\xi_1^{(n)} \geqslant y \bigm| S_n^{(n)} = m\bigr) \\ &\leqslant n\, \mathbb{P}\bigl(\xi_1^{(n)} \geqslant y\bigr)\, \mathbb{P}\bigl(S_n^{(n)} = m\bigr)^{-1} \\ &\leqslant n\, \mathbb{P}(\xi^{(n)} \geqslant y) \cdot O(n^{1/2}). \end{aligned} \tag{19.41}$$

Choose $\tau_* \in (\tau, \rho)$. Then, for $s > 0$ and n so large than $\tau_n < \tau_*$, by (4.11),

$$\mathbb{P}(\xi^{(n)} \geqslant y) \leqslant e^{-sy} \frac{\Phi(e^s \tau_n)}{\Phi(\tau_n)} \leqslant e^{-sy} \frac{\Phi(e^s \tau_*)}{\Phi(0)}. \tag{19.42}$$

Choosing $s > 0$ with $e^s < \rho/\tau_*$, we thus find $\mathbb{P}(\xi^{(n)} \geqslant y) = O(e^{-sy})$ and, by (19.41),

$$\mathbb{P}(Y_{(1)} \geqslant y) = O\bigl(n^{3/2} e^{-sy}\bigr).$$

We now define $B(n) := 2s^{-1} \log n$, and obtain

$$\mathbb{P}(Y_{(1)} \geqslant B(n)) = O\bigl(n^{3/2} e^{-sB(n)}\bigr) = O\bigl(n^{-1/2}\bigr) \to 0. \tag{19.43}$$

Hence, $Y_{(1)} < B(n)$ w.h.p. Similarly, using (19.42) again,

$$\mathbb{P}(\xi^{(n)} \geqslant B(n)) = o(n^{-1}) \tag{19.44}$$

and thus $\mathbb{P}(\xi_{(1)}^{(n)} \geqslant B(n)) \leqslant n\, \mathbb{P}(\xi^{(n)} \geqslant B(n)) \to 0$, so $\xi_{(1)} < B(n)$ w.h.p.

We have shown that (19.19) (with $\xi^{(n)}$) and (19.20) hold. Moreover, Lemma 14.1 yields, see (14.13) again, $Z(m,n;\mathbf{w}^{(n)}) \sim d/(2\pi\sigma^2 n)^{1/2}$, and for $k \leqslant B(n) = O(\log n)$, the same argument yields also, using Remark 14.2,

$Z(m-k, n-1; \mathbf{w}^{(n)}) \sim d/(2\pi\sigma^2 n)^{1/2}$, because $m - k - (n-1)\,\mathbb{E}\,\xi^{(n)} = m - k - (n-1)m/n = -k + m/n = o(n^{1/2})$. Consequently, (19.40) yields

$$\mathbb{P}(Y_1 = k) = \bigl(1 + o(1)\bigr)\,\mathbb{P}(\xi^{(n)} = k), \tag{19.45}$$

uniformly for $k \leqslant B(n)$.

We can now argue exactly as above, using $\xi^{(n)}$, $\xi^{(n)}_{(j)}$ and $\overline{N}^{(n)}_A$, which proves this version of the theorem.

Finally, if $\lambda < \nu$ and $m/n = \lambda + o(1/\log n)$, then $\tau_n := \Psi^{-1}(m/n) = \tau + o(1/\log n)$, because Ψ^{-1} is differentiable on $(0, \nu)$. Since we are assuming $\tau = 1$ and $\Phi(\tau) = 1$ in the proof, we thus have, uniformly for all $k \leqslant B(n) = O(\log n)$,

$$\mathbb{P}(\xi^{(n)} = k) = \frac{\tau_n^k}{\Phi(\tau_n)} w_k = \bigl(1 + o(1)\bigr) w_k = \bigl(1 + o(1)\bigr)\,\mathbb{P}(\xi = k). \tag{19.46}$$

Since also $\mathbb{P}(\xi^{(n)} \geqslant B(n)) = o(n^{-1})$ and $\mathbb{P}(\xi \geqslant B(n)) = o(n^{-1})$, it follows that $n\,\mathbb{P}(\xi^{(n)} \geqslant h(n)) \to \alpha \iff n\,\mathbb{P}(\xi \geqslant h(n)) \to \alpha$, and thus we may in (i)–(iii) replace $\xi^{(n)}$ by ξ again. Finally, (iv)–(vi) follow as above in this case too. $\square$

Proof of Theorem 19.3. Recall that $\lambda < \nu \iff \tau < \rho$ by Lemma 3.1. We have $\nu > \lambda > 0$, so $\rho > 0$, and $\tau > 0$. Thus $1 < \rho/\tau \leqslant \infty$.

(i): Fix $a > 1/\log(\rho/\tau)$. Choose b with $e^{1/a} < b < \rho/\tau$. Then $0 \leqslant 1/\rho < (b\tau)^{-1}$. Choose c with $1/\rho < c < (b\tau)^{-1}$.

Since $\limsup_{k\to\infty} w_k^{1/k} = 1/\rho < c$, we have $w_k^{1/k} < c$ for large k, and then, defining τ_n and $\xi^{(n)}$ by (14.10)–(14.11),

$$\mathbb{P}(\xi^{(n)} = k) = \frac{\tau_n^k}{\Phi(\tau_n)} w_k \leqslant \frac{(c\tau_n)^k}{\Phi(0)}. \tag{19.47}$$

As $n \to \infty$, $c\tau_n \to c\tau < b^{-1}$. Let $h := \lfloor a \log n \rfloor$. For large n, (19.47) applies for $k \geqslant h$, and $c\tau_n < b^{-1} < 1$, and then

$$\mathbb{P}(\xi^{(n)} \geqslant h) \leqslant \sum_{k=h}^{\infty} \frac{(c\tau_n)^k}{\Phi(0)} \leqslant \sum_{k=h}^{\infty} \frac{b^{-k}}{w_0} = O\bigl(b^{-h}\bigr) = O\bigl(n^{-a\log b}\bigr)$$

Since $a \log b > 1$, thus $n\,\mathbb{P}(\xi^{(n)} \geqslant h) \to 0$, and Theorem 19.7(ii) yields $Y_{(1)} \leqslant h \leqslant a \log n$ w.h.p.

(ii): If $\rho = \infty$, then (i) applies with $\rho/\tau = \infty$ and thus $1/\log(\rho/\tau) = 0$.

(iii): If $\rho = \infty$, the result follows by (ii), so we may assume $1 = \tau < \rho < \infty$. Let $a := 1/\log(\rho/\tau)$ and $0 < \varepsilon < 1$. The upper bound $Y_{(j)} \leqslant Y_{(1)} \leqslant (a + \varepsilon) \log n$ w.h.p. follows from (i), and it remains to find a matching lower bound.

Let $k := \lceil (1-\varepsilon)a\log n\rceil$. Then, since $\tau_n \to \tau$,

$$\begin{aligned}\log \mathbb{P}(\xi^{(n)} = k) &= \log w_k + k\log\tau_n - \log\Phi(\tau_n)\\ &= -k(\log\rho + o(1)) + k(\log\tau + o(1)) + O(1)\\ &= -k\log(\rho/\tau) + o(k) = -(1-\varepsilon+o(1))\log n\end{aligned}$$

and thus

$$n\,\mathbb{P}(\xi^{(n)} \geqslant k) \geqslant n\,\mathbb{P}(\xi^{(n)} = k) = n^{\varepsilon+o(1)} \to \infty.$$

By Theorem 19.7(iii) (and the last sentence in Theorem 19.7), this implies w.h.p.

$$Y_{(j)} \geqslant k \geqslant (1-\varepsilon)a\log n$$

This completes the proof, since we can take ε arbitrarily small. □

Specialising Theorem 19.7 to the tree case ($m = n-1$), we obtain the following. (Recall that $\sigma^2 < \infty$ is automatic when $\nu > 1$.)

Corollary 19.11. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be a weight sequence with $w_0 > 0$ and $w_k > 0$ for some $k \geqslant 2$, and let ξ have the distribution given by (π_k) in (7.1). Suppose that $\nu \geqslant 1$ and $\sigma^2 := \operatorname{Var}\xi < \infty$. Then, as $n\to\infty$, for the largest degrees $Y_{(1)} \geqslant Y_{(2)} \geqslant \dots$ in $\mathcal{T}_n$, $d_{\mathrm{TV}}(Y_{(1)}, \xi_{(1)}) \to 0$ and, for every fixed j, $d_{\mathrm{K}}(Y_{(j)}, \xi_{(j)}) \to 0$.*

Proof. The case $\omega = \infty$ is a special case of Theorem 19.7, with $\lambda = 1$.

The case $\omega < \infty$ is trivial: for every fixed j, $Y_{(j)} = \omega$ w.h.p. by Theorem 19.1, and, trivially, $\xi_{(j)} = \omega$ w.h.p. □

The comparison with $\xi_{(j)}$ in Theorem 19.7 and Corollary 19.11 is appealing since $\xi_{(j)}$ is the j:th largest of n i.i.d. random variables. For applications it is often convenient to modify this a little by taking a Poisson number of variables instead.

Consider an infinite i.i.d. sequence $\xi_1, \xi_2, \dots$, let as above $\xi_{(j)}$ be the j:th largest among the first n elements of the sequence and define $\tilde\xi_{(j)}$ as the j:th largest among the first $N(n)$ elements $\xi_1, \dots, \xi_{N(n)}$, where $N(n) \sim \mathrm{Po}(n)$ is a random Poisson variable independent of $\xi_1, \xi_2, \dots$.

Lemma 19.12. *W.h.p. $\tilde\xi_{(j)} = \xi_{(j)}$ and thus $d_{\mathrm{TV}}(\tilde\xi_{(j)}, \xi_{(j)}) \to 0$ as $n\to\infty$ for every fixed $j \geqslant 1$.*

Proof. Let $n_\pm := \lfloor n \pm n^{2/3}\rfloor$, and let $\xi^-_{(j)}$ be the j:th largest of $\xi_1, \dots, \xi_{n_-}$. By symmetry, the positions of the j largest among $\xi_1, \dots, \xi_n$ are uniformly random (we resolve any ties in the ordering at random); thus the probability that one of them has index $> n_-$ is at most $j(n-n_-)/n = o(1)$. Hence, w.h.p. all j are among $\xi_1, \dots, \xi_{n_-}$, and then $\xi_{(j)} = \xi^-_{(j)}$.

Furthermore, w.h.p. $n_- \leqslant N(n) \leqslant n_+$, and a similar argument (using conditioning on $N(n)$) shows that w.h.p. $\tilde\xi_{(j)} = \xi^-_{(j)}$. Hence, w.h.p. $\xi_{(j)} = \xi^-_{(j)} = \tilde\xi_{(j)}$. Now use Lemma 19.5(v). □

We can thus replace $\xi_{(j)}$ by $\tilde{\xi}_{(j)}$ in Theorem 19.7 and Corollary 19.11. (We can similarly replace $\xi^{(n)}_{(j)}$ by $\tilde{\xi}^{(n)}_{(j)}$ defined in the same way.) The advantage is that, by standard properties of the Poisson distribution, the corresponding counting variables

$$\widetilde{N}_k := |\{i \leqslant N(n) : \xi = k\}|$$

are *independent* Poisson variables with $\widetilde{N}_k \sim \mathrm{Po}(n\,\mathbb{P}(\xi = k))$. We similarly define $\widetilde{N}_{[k,\infty)} := \sum_{l=k}^{\infty} \widetilde{N}_l \sim \mathrm{Po}\bigl(n\,\mathbb{P}(\xi \geqslant k)\bigr)$.

Remark 19.13. An equivalent way to express this is that the multiset $\Xi_n := \{\xi_i : i \leqslant N(n)\}$ is a Poisson process on $\mathbb{N}_0$ with intensity measure Λ_n given by $\Lambda_n\{k\} = n\,\mathbb{P}(\xi = k)$.

We thus have (exactly), for any j and k,

$$\mathbb{P}(\tilde{\xi}_{(j)} \leqslant k) = \mathbb{P}\bigl(\widetilde{N}_{[k+1,\infty)} < j\bigr) = \mathbb{P}\bigl(\mathrm{Po}(n\,\mathbb{P}(\xi > k)) < j\bigr); \tag{19.48}$$

in particular

$$\mathbb{P}(\tilde{\xi}_{(1)} \leqslant k) = e^{-n\,\mathbb{P}(\xi > k)}. \tag{19.49}$$

This gives the following special case of Theorem 19.7. (There is a similar version with $\xi^{(n)}$.)

Corollary 19.14. *Suppose that $w_0 > 0$ and $\omega = \infty$. Suppose further that $n \to \infty$ and $m = m(n)$ with $m = \lambda n + o(\sqrt{n})$ where $0 < \lambda \leqslant \nu$, and that either $\lambda < \nu$ or $\sigma^2 := \mathrm{Var}\,\xi < \infty$. Then, uniformly in all $k \geqslant 0$,*

$$\mathbb{P}(Y_{(j)} \leqslant k) = \mathbb{P}\bigl(\mathrm{Po}(n\,\mathbb{P}(\xi > k)) < j\bigr) + o(1) \tag{19.50}$$

for each fixed $j \geqslant 1$; in particular

$$\mathbb{P}(Y_{(1)} \leqslant k) = e^{-n\,\mathbb{P}(\xi > k)} + o(1). \tag{19.51}$$

Proof. Immediate by Theorem 19.7(v), Lemma 19.12 and (19.48)–(19.49). $\square$

Remark 19.15. Since $\widetilde{N}_{[h(n),\infty)} \geqslant j \iff \tilde{\xi}_{(j)} \geqslant h(n)$, it follows easily from Lemmas 19.12 and 19.5(iv) that for any sequence $h(n)$,

$$\widetilde{d}_{\mathrm{K}}\bigl(\overline{N}_{[h(n),\infty)}, \widetilde{N}_{[h(n),\infty)}\bigr) \to 0. \tag{19.52}$$

Hence, Theorem 19.7(iv) is equivalent to $\widetilde{d}_{\mathrm{K}}\bigl(N_{[h(n),\infty)}, \widetilde{N}_{[h(n),\infty)}\bigr) \to 0$, and thus

$$\widetilde{d}_{\mathrm{K}}\Bigl(N_{[h(n),\infty)}, \mathrm{Po}\bigl(n\,\mathbb{P}(\xi \geqslant h(n))\bigr)\Bigr) \to 0. \tag{19.53}$$

This is another, essentially equivalent, way to express the results above.

19.3. **The subcase $\lambda < \nu$.** When $\lambda < \nu$, we have $\tau < \rho$ and the random variable ξ has some finite exponential moment, cf. Section 8; hence the probabilities π_k decrease rapidly. Theorem 19.7 and Corollary 19.14 show that $Y_{(1)}$ (and each $Y_{(j)}$) has its distribution concentrated on k such that $\mathbb{P}(\xi \geqslant k)$ is of the order $1/n$. If the decrease of π_k is not too irregular, this implies strong concentration of $Y_{(1)}$, with, rougly speaking, $Y_{(1)} \approx k$ when $\mathbb{P}(\xi \geqslant k) \approx 1/n$. To make this precise, we define three versions of a suitable such estimate $k = k(n)$. Let, as above, $\pi_k = \mathbb{P}(\xi = k) = \tau^k w_k/\Phi(\tau)$ and let

$$\Pi_k := \mathbb{P}(\xi \geqslant k) = \sum_{l=k}^{\infty} \pi_l. \tag{19.54}$$

Define

$$k_1(n) := \max\{k : \pi_k \geqslant 1/n\}, \tag{19.55}$$

$$k_2(n) := \max\{k : \Pi_k \geqslant 1/n\}, \tag{19.56}$$

$$k_3(n) := \max\{k : \sqrt{\Pi_k \Pi_{k+1}} \geqslant 1/n\}. \tag{19.57}$$

Note that $k_1(n) \leqslant k_2(n)$ and $k_2(n) - 1 \leqslant k_3(n) \leqslant k_2(n)$.

We consider the typical case when w_{k+1}/w_k converges as $k \to \infty$. We assume implicitly that w_{k+1}/w_k is defined for all large k; thus $w_k > 0$ and $\omega = \infty$. If $w_{k+1}/w_k \to a$ as $k \to \infty$, then (3.5) yields $\rho = 1/a$; hence $\rho = \infty$ if $a = 0$ and $0 < \rho < \infty$ if $a > 0$.

Theorem 19.16. *Suppose that $w_0 > 0$ and that $w_{k+1}/w_k \to a < \infty$ as $k \to \infty$. Suppose further that $n \to \infty$ and $m = m(n)$ with $m = \lambda n + o(\sqrt{n})$ where $0 < \lambda < \nu$.*

(i) *Then, for each $j \geqslant 1$,*

$$Y_{(j)} = k_1(n) + O_{\mathrm{p}}(1) = k_2(n) + O_{\mathrm{p}}(1) = k_3(n) + O_{\mathrm{p}}(1).$$

(ii) *If $a = 0$, then, moreover, w.h.p.,*

$$|Y_{(j)} - k_1(n)| \leqslant 1, \qquad |Y_{(j)} - k_2(n)| \leqslant 1, \qquad Y_{(j)} \in \{k_3(n), k_3(n)+1\}.$$

Proof. (i): We have, as said above, $\rho = 1/a > 0$. Furthermore, since $\lambda < \nu$, we have $\tau < \rho$ and thus, as $k \to \infty$,

$$\frac{\pi_{k+1}}{\pi_k} = \tau \frac{w_{k+1}}{w_k} \to \tau a = \frac{\tau}{\rho} < 1. \tag{19.58}$$

It follows from (19.58) and (19.54), using dominated convergence, that, as $k \to \infty$,

$$\frac{\Pi_k}{\pi_k} = \sum_{i=0}^{\infty} \frac{\pi_{k+i}}{\pi_k} \to \sum_{i=0}^{\infty} (\tau a)^i = \frac{1}{1-\tau a}. \tag{19.59}$$

If ℓ is chosen such that $(\tau a)^\ell < 1 - \tau a$, then (19.59) and (19.58) imply $\Pi_{k+\ell}/\pi_k \to (\tau a)^\ell/(1-\tau a) < 1$ as $k \to \infty$, and thus, for large k, $\Pi_{k+\ell} <$

$\pi_k < \Pi_k$; hence, for large n, $k_1(n) \leqslant k_2(n) \leqslant k_1(n) + \ell$. Thus, recalling that $|k_2(n) - k_3(n)| \leqslant 1$,

$$k_1(n) = k_2(n) + O(1) = k_3(n) + O(1). \tag{19.60}$$

Furthermore, (19.58) and (19.59) yield also

$$\frac{\Pi_{k+1}}{\Pi_k} \to \tau a < 1. \tag{19.61}$$

By (19.56), $n\Pi_{k_2(n)} \geqslant 1 > n\Pi_{k_2(n)+1}$. This and (19.61) imply that if $\Omega(n)$ is any sequence with $\Omega(n) \to \infty$, then $n\Pi_{k_2(n)-\Omega(n)} \to \infty$ and $n\Pi_{k_2(n)+\Omega(n)} \to 0$. Consequently, recalling the definition (19.54), by Theorem 19.7(ii)–(iii) (or by Corollary 19.14) w.h.p. $Y_{(j)} \geqslant k_2(n) - \Omega(n)$ and $Y_{(j)} < k_2(n) + \Omega(n)$. Since $\Omega(n) \to \infty$ is arbitrary, this yields $Y_{(j)} = k_2(n) + O_{\mathrm{p}}(1)$. (See e.g. [62].) The result follows by (19.60).

(ii): When $a = 0$, (19.59) yields $\Pi_k \sim \pi_k$, (19.58) yields $\pi_{k+1}/\pi_k \to 0$ and (19.61) yields $\Pi_{k+1}/\Pi_k \to 0$ as $k \to \infty$. It follows easily from (19.55)–(19.57) that $n\Pi_{k_1(n)-1} \to \infty$, $n\Pi_{k_1(n)+2} \to 0$, $n\Pi_{k_2(n)-1} \to \infty$, $n\Pi_{k_2(n)+2} \to 0$, $n\Pi_{k_3(n)} \to \infty$, $n\Pi_{k_3(n)+2} \to 0$, and the results follow by Theorem 19.7(ii)–(iii). $\square$

If $a = 0$, i.e. $w_{k+1}/w_k \to 0$ as $k \to \infty$, thus $Y_{(1)}$ is asymptotically concentrated at one or two values. (This was shown, in the tree case, by Meir and Moon [88], after showing concentration to at most three values in [87]; see also Kolchin, Sevast'yanov and Chistyakov [77], Kolchin [76] and Carr, Goh and Schmutz [21] for special cases.) If $a > 0$, we still have a strong concentration, but not to any finite number of values as is seen by Theorem 19.19 below.

We consider two important examples, where we apply this to random trees, so $m = n-1$ and $\lambda = 1$. (Recall that $Y_{(1)}$ then is the largest outdegree in $\mathcal{T}_n$. The largest degree is w.h.p. $Y_{(1)} + 1$, since w.h.p. it is not attained at the root, e.g. because the root degree is $O_{\mathrm{p}}(1)$ by Theorem 7.10; this should be kept in mind when comparing with results in other papers.)

Example 19.17. For uniform random labelled ordered rooted trees, we have by Example 10.1 $\xi \sim \mathrm{Ge}(1/2)$ with $\pi_k = 2^{-k-1}$ and thus $\mathbb{P}(\xi \geqslant k) = 2^{-k}$. Hence $Y_{(1)}$ has asymptotically the same distribution as the maximum of n i.i.d. geometrically distributed random variables, which is a simple and well-studied example, see e.g. Leadbetter, Lindgren and Rootzén [82]. Explicitly, Corollary 19.14 applies and (19.51) yields, uniformly in $k \geqslant 0$,

$$\mathbb{P}(Y_{(1)} \leqslant k) = e^{-n2^{-k-1}} + o(1). \tag{19.62}$$

(This was, essentially, shown by Meir and Moon [87].)

One way to express this is to introduce a random variable W with the Gumbel distribution

$$\mathbb{P}(W \leqslant x) = e^{-e^{-x}}, \qquad -\infty < x < \infty. \tag{19.63}$$

Then (19.62) yields, uniformly for $k \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{P}(Y_{(1)} \leqslant k) &= \mathbb{P}\bigl(W < (k+1)\log 2 - \log n\bigr) + o(1)\\ &= \mathbb{P}\Bigl(\frac{W+\log n}{\log 2} < k+1\Bigr) + o(1)\\ &= \mathbb{P}\Bigl(\Bigl\lfloor\frac{W+\log n}{\log 2}\Bigr\rfloor \leqslant k\Bigr) + o(1). \end{aligned} \tag{19.64}$$

In other words, extending d_{K} to $\mathbb{Z}$-valued random variables,

$$d_{\mathrm{K}}\bigl(Y_{(1)}, \lfloor (W+\log n)/\log 2\rfloor\bigr) \to 0. \tag{19.65}$$

Thus, the maximum degree $Y_{(1)}$ can be approximated (in distribution) by $\lfloor (W+\log n)/\log 2\rfloor = \lfloor W/\log 2 + \log_2 n\rfloor$. Hence $Y_{(1)} - \log_2 n$ is tight but no asymptotic distribution exists; $Y_{(1)} - \log_2 n$ can be approximated by $\lfloor W/\log 2 + \log_2 n\rfloor - \log_2 n = \lfloor W/\log 2 + \{\log_2 n\}\rfloor - \{\log_2 n\}$ (where we let $\{x\} := x - \lfloor x\rfloor$ denote the fractional part of x), which shows convergence in distribution for any subsequence such that $\{\log_2 n\}$ converges to some $\alpha \in [0,1]$, but the limit depends on α. See further Janson [60, in particular Lemma 4.1 and Example 4.3].

In the same way we see that $Y_{(j)}$ can be approximated in distribution by $\lfloor W_j/\log 2 + \log_2 n\rfloor$ where W_j has the distribution

$$\mathbb{P}(W_j \leqslant x) = \mathbb{P}(\mathrm{Po}(e^{-x}) < j) = \sum_{i=0}^{j-1} \frac{e^{-ix}}{i!} e^{-e^{-x}}, \qquad -\infty < x < \infty, \tag{19.66}$$

with density function $e^{-jx}e^{-e^{-x}}/(j-1)!$; further $W_j \overset{\mathrm{d}}{=} -\log V_j$, where V_j has the Gamma distribution $\mathrm{Gamma}(j,1)$. (Cf. Leadbetter, Lindgren and Rootzén [82, Section 2.2] for the relation between the distributions of $\xi_{(j)}$ and $\xi_{(1)}$ in the i.i.d. case.)

Example 19.18. For uniform random labelled unordered rooted trees, we have by Example 10.2 $\xi \sim \mathrm{Po}(1)$ with $\pi_k = e^{-1}/k!$. We have $w_{k+1}/w_k \to 0$, so Theorem 19.16(ii) applies and shows that $Y_{(1)}$ is concentrated on at most two values, as proved by Kolchin [76, Theorem 2.5.2]; see also Meir and Moon [88] and Carr, Goh and Schmutz [21].

Explicitly, (19.51) yields (treating the rather trivial case $n > k^{1/2}\cdot k!$ separately)

$$\mathbb{P}(Y_{(1)} < k) = e^{-ne^{-1}/k!(1+O(1/k))} + o(1) = e^{-ne^{-1}/k!} + o(1) \tag{19.67}$$

which by Stirling's formula yields

$$\mathbb{P}(Y_{(1)} < k) = \exp\bigl(-e^{\log n - (k+\frac12)\log k + k - \log(e\sqrt{2\pi})}\bigr) + o(1) \tag{19.68}$$

uniformly in $k \geqslant 1$, cf. Carr, Goh and Schmutz [21]. It follows easily from Stirling's formula, or from (19.68), that $k_1(n), k_2(n), k_3(n) \sim \log n/\log\log n$, and more precise asymptotics can be found too; cf. [90], [87], [21].

In fact, the simple Example 19.17 is typical for the case $w_{k+1}/w_k \to a > 0$ as $k \to \infty$; then $Y_{(1)}$ always has asymptotically the same distribution as the maximum of i.i.d. geometric random variables, provided we adjust the number of these variables according to $\mathbf{w}$. We state some versions of this in the next theorem. For simplicity we consider only the maximum $Y_{(1)}$, and leave the extensions to $Y_{(j)}$ for general fixed j to the reader.

Theorem 19.19. *Suppose that $w_0 > 0$ and that $w_{k+1}/w_k \to a$ as $k \to \infty$, with $0 < a < \infty$. Suppose further that $n \to \infty$ and $m = m(n)$ with $m = \lambda n + o(\sqrt{n})$ where $0 < \lambda < \nu$.*

Let $q := \tau a = \tau/\rho < 1$. Let $k(n)$ be any sequence such that $\pi_{k(n)} = \Theta(1/n)$; equivalently, $k(n) = k_1(n) + O(1)$, and let $N = N(n)$ be integers such that

$$N \sim \frac{n\pi_{k(n)}q^{-k(n)}}{1-q} = \frac{nw_{k(n)}a^{-k(n)}}{\Phi(\tau)(1-q)}. \tag{19.69}$$

(i) *Let $\eta_1, \dots, \eta_N$ be i.i.d. random variables with a geometric distribution $\mathrm{Ge}(1-q)$, i.e., $\mathbb{P}(\eta_i = k) = (1-q)q^{-k}$, $k \geqslant 0$. Then*

$$Y_{(1)} \overset{\mathrm{d}}{\approx} \max_{i \leqslant N} \eta_i. \tag{19.70}$$

(ii) *Let W have the Gumbel distribution* (19.63)*. Then*

$$Y_{(1)} \overset{\mathrm{d}}{\approx} \lfloor W/\log(1/q) + \log_{1/q} N \rfloor. \tag{19.71}$$

(iii) *Let $b_n := n\pi_{k(n)}$; thus $b_n = \Theta(1)$. Then*

$$Y_{(1)} - k(n) \overset{\mathrm{d}}{\approx} \left\lfloor \bigl(W + \log(b_n/(1-q))\bigr)/\log(1/q) \right\rfloor. \tag{19.72}$$

Thus $Y_{(1)} - k(n)$ is tight, and converges for every subsequence such that b_n converges.

Hence $Y_{(1)} - k(n)$ converges for every subsequence such that b_n converges, but the limit depends on the subsequence so $Y_{(1)} - k(n)$ does not have a limit distribution. (For the distributions that appear as subsequence limits, see Janson [60, Examples 4.3 and 2.7].) Note that necessarily $k(n) \to \infty$ and thus $N \to \infty$ as $n \to \infty$.

We show first a simple lemma, similar to Lemma 19.5.

Lemma 19.20. *Let X_n and X'_n be integer-valued random variables and suppose that there exists a sequence of integers $k(n)$ such that $X_n - k(n)$ is tight. (Equivalently: $X_n = k(n) + O_{\mathrm{p}}(1)$.) Then the following are equivalent:*

(i) *$\mathbb{P}(X_n \leqslant k(n) + \ell) - \mathbb{P}(X'_n \leqslant k(n) + \ell) \to 0$ for each fixed $\ell \in \mathbb{Z}$;*
(ii) *$d_{\mathrm{K}}(X_n, X'_n) \to 0$;*
(iii) *$X_n \overset{\mathrm{d}}{\approx} X'_n$, i.e., $d_{\mathrm{TV}}(X_n, X'_n) \to 0$.*

Proof. By considering $X_n - k(n)$ and $X'_n - k(n)$ we may assume that $k(n) = 0$. Let $\varepsilon > 0$. Since X_n is tight, there exists L such that $\mathbb{P}(|X_n| > L) < \varepsilon$ for every n. Suppose that (i) holds. Then

$$\begin{aligned} d_{\mathrm{TV}}(X_n, X'_n) &= \sum_{\ell=-\infty}^{\infty} \bigl(\mathbb{P}(X_n = \ell) - \mathbb{P}(X'_n = \ell)\bigr)_+ \\ &\leqslant \sum_{\ell=-L}^{L} \bigl(\mathbb{P}(X_n = \ell) - \mathbb{P}(X'_n = \ell)\bigr)_+ + \mathbb{P}(|X_n| > L) \leqslant o(1) + \varepsilon. \end{aligned}$$

This shows (iii). The implications (iii) $\Longrightarrow$ (ii) and (ii) $\Longrightarrow$ (i) are trivial. $\square$

Proof of Theorem 19.19. By (19.58), $\pi_{k+1}/\pi_k \to q$ as $k \to \infty$, and it follows from (19.55) that $n\pi_{k_1(n)} \in [1, q^{-1} + o(1)]$. It follows further that $\pi_{k(n)} = \Theta(1/n) \iff k(n) = k_1(n) + O(1)$, as asserted, and then $\pi_{k(n)} q^{-k(n)} \sim \pi_{k_1(n)} q^{-k_1(n)}$; thus we may replace $k(n)$ by $k_1(n)$ in (19.69).

(i): For each fixed $\ell \in \mathbb{Z}$, by (19.59), (19.58) and (19.69),

$$\begin{aligned} n\,\mathbb{P}(\xi \geqslant k(n) + \ell) &= n\Pi_{k(n)+\ell} \sim n\pi_{k(n)+\ell}/(1-q) \sim n\pi_{k(n)} q^{\ell}/(1-q) \\ &\sim N q^{k(n)+\ell} = N\,\mathbb{P}(\eta_1 \geqslant k(n) + \ell); \end{aligned} \tag{19.73}$$

furthermore, this is $\Theta(1)$. Hence, (19.51) yields

$$\begin{aligned} \mathbb{P}\bigl(Y_{(1)} < k(n) + \ell\bigr) &= e^{-n\,\mathbb{P}(\xi \geqslant k(n)+\ell)} + o(1) = e^{-N\,\mathbb{P}(\eta_1 \geqslant k(n)+\ell)} + o(1) \\ &= \bigl(1 - \mathbb{P}(\eta_1 \geqslant k(n) + \ell)\bigr)^N + o(1) \\ &= \mathbb{P}\bigl(\max_{i \leqslant N} \eta_i < k(n) + \ell\bigr) + o(1), \end{aligned}$$

and (19.70) follows by Lemma 19.20, since $Y_{(1)} - k_1(n)$ is tight by Theorem 19.16.

(ii): As in (19.64), uniformly in $k \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{P}\Bigl(\max_{i \leqslant N} \eta_i < k\Bigr) &= \bigl(1 - q^k\bigr)^N = e^{-Nq^k} + o(1) \\ &= \mathbb{P}\bigl(W < k \log(1/q) - \log N\bigr) + o(1) \\ &= \mathbb{P}\left(\left\lfloor \frac{W + \log N}{\log(1/q)} \right\rfloor < k\right) + o(1). \end{aligned} \tag{19.74}$$

Hence, $d_{\mathrm{TV}}\bigl(\max_{i \leqslant N} \eta_i, \lfloor W/\log(1/q) + \log_{1/q} N \rfloor\bigr) \to 0$, and (19.71) follows from (19.70) and Lemma 19.20.

(iii): By (19.69), $\log_{1/q} N = k(n) + \log_{1/q}(b_n/(1-q)) + o(1)$, and (19.72) follows easily from (19.71), using Lemma 19.20 and the fact that W is absolutely continuous. $\square$

Remark 19.21. For later use we note that Theorem 19.19, as other results, extends to the case $w_0 = 0$ by the argument in Remark 11.8; we now have to assume $\lambda > \alpha := \min\{k : w_k > 0\}$. The extension of Theorem 19.19(i) is perhaps more subtle that other applications of this argument since N will change by a factor $\sim q^\alpha$, but (ii) and (iii) are straightforward, and then (i) follows by (19.74) and Lemma 19.20.

If the weight sequence is very irregular, $Y_{(1)}$ can fail to be concentrated even in the case $\lambda < \nu$.

Example 19.22. Let $\ell_j := 2^{2^j}$ and $\mathsf{S} := \{\ell_j\}_{j\geqslant 1}$. Let $w_k = 2^{-k}$ if $k \in \mathsf{S}$, $w_k = 0$ if $k \geqslant 2$ and $k \notin \mathsf{S}$, and choose $w_0 := \sum_{k\in\mathsf{S}}(k-1)w_k$ and $w_1 := 1 - \sum_{k\in\mathsf{S}} kw_k$. Then (w_k) is a probability weight sequence with $\mu := \sum_{k=0}^\infty kw_k = 1$. Furthermore, $\rho = 2$, $\Phi(\rho) = \infty$ and, by Lemma 3.1(iv), $\nu := \Psi(\rho) = \infty$. Choose $m = n-1$ (the tree case); thus $\lambda = 1 < \nu$ and $\tau = 1$ so $(\pi_k) = (w_k)$.

Note that $\ell_{j+1} = \ell_j^2$. If $n = 2^{\ell_j}$, then $\mathbb{P}(\xi \geqslant \ell_j) \sim 2^{-\ell_j} = n^{-1}$, $\mathbb{P}(\xi \geqslant \ell_{j+1}) \sim 2^{-\ell_{j+1}} \ll n^{-1}$ and $\mathbb{P}(\xi \geqslant \ell_{j-1}) \sim 2^{-\ell_{j-1}} \gg n^{-1}$; hence it follows from (19.51) that for the subsequence $n = 2^{\ell_j}$ with $j \in \mathsf{S}$, $\mathbb{P}(Y_{(1)} < \ell_{j+1}) \to 1$, $\mathbb{P}(Y_{(1)} < \ell_j) \to e^{-1}$ and $\mathbb{P}(Y_{(1)} < \ell_{j-1}) \to 0$. Thus, along this subsequence, $\mathbb{P}(Y_{(1)} = \ell_j) \to 1 - e^{-1}$ and $\mathbb{P}(Y_{(1)} = \ell_{j-1}) \to e^{-1}$, i.e., $\mathbb{P}\bigl(Y_{(1)} = \log_2 n\bigr) \to 1 - e^{-1}$ and $\mathbb{P}\bigl(Y_{(1)} = \log_2^{1/2} n\bigr) \to e^{-1}$.

19.4. **The subcase $w_{k+1}/w_k \to 0$ as $k \to \infty$.** We have seen in Theorem 19.16 that when $w_{k+1}/w_k \to 0$ as $k \to \infty$, the maximum $Y_{(1)}$ is asymptotically concentrated at one or two values. We shall see that for "most" (in a sense specified below) values of n, $Y_{(1)}$ is concentrated at one value, but there are also rather large transition regions where $Y_{(1)}$ takes two values with rather large probabilities.

We have, as said before Theorem 19.16, $\omega = \infty$ and $\rho = \infty$. Furthermore, by Lemma 3.1(v), $\nu = \infty$.

We define

$$n_k := \lfloor 1/\pi_k \rfloor, \tag{19.75}$$

noting that $n_{k+1}/n_k \sim \pi_k/\pi_{k+1} \to \infty$ as $k \to \infty$; in particular, $n_{k+1} > n_k$ (for large k, at least). The results above then can be stated as follows.

Theorem 19.23. *Suppose that $w_0 > 0$ and that $w_{k+1}/w_k \to 0$ as $k \to \infty$. Suppose further that $n \to \infty$ and $m = m(n)$ with $m = \lambda n + o(\sqrt{n})$ where $0 < \lambda < \infty$.*

(i) *Consider n in a subsequence such that for some $k(n)$ and some $x \in (0,\infty)$, $n/n_{k(n)} \to x$. Then*
$$\mathbb{P}\bigl(Y_{(1)} = k(n) - 1\bigr) \to e^{-x},$$
$$\mathbb{P}\bigl(Y_{(1)} = k(n)\bigr) \to 1 - e^{-x}.$$

(ii) *Let $\Omega_k \to \infty$ as $k \to \infty$. If $n \to \infty$ with $n \notin \bigcup_{k=1}^\infty [\Omega_k^{-1} n_k, \Omega_k n_k]$, then, for $k(n)$ such that $n_{k(n)} < n < n_{k(n)+1}$,*
$$\mathbb{P}\bigl(Y_{(1)} = k(n)\bigr) \to 1.$$

Proof. (i): Along the subsequence, using (19.54), (19.59) and (19.75),

$$n\,\mathbb{P}(\xi \geqslant k(n)) = n\Pi_{k(n)} \sim n\pi_{k(n)} \sim \frac{n}{n_{k(n)}} \to x. \tag{19.76}$$

Hence, (19.51) yields $\mathbb{P}(Y_{(1)} \leqslant k(n)-1) \to e^{-x}$. Furthermore, by (19.76) and (19.61), $n\,\mathbb{P}(\xi > k(n)) \to 0$ and $n\,\mathbb{P}(\xi \geqslant k(n)-1) \to \infty$; hence (19.51) yields $\mathbb{P}(Y_{(1)} \leqslant k(n)) \to 1$ and $\mathbb{P}(Y_{(1)} \leqslant k(n)-2) \to 0$.

(ii): We may assume $\Omega_k > 1$. Then the assumptions imply $\Omega_{k(n)}n_{k(n)} < n < \Omega_{k(n)+1}^{-1}n_{k(n)+1}$, where $k(n) \to \infty$ and thus $\Omega_{k(n)} \to \infty$ as $n \to \infty$. Hence, similarly to (19.76),

$$n\,\mathbb{P}\bigl(\xi \geqslant k(n)\bigr) \sim \frac{n}{n_{k(n)}} > \Omega_{k(n)} \to \infty,$$

$$n\,\mathbb{P}\bigl(\xi \geqslant k(n)+1\bigr) \sim \frac{n}{n_{k(n)+1}} < \Omega_{k(n)+1}^{-1} \to 0,$$

and the result follows by (19.51). $\square$

Roughly speaking, the values of n such that $Y_{(1)}$ takes two values with significant probabilities thus form intervals around each n_k, of the same length on a logarithmic scale; between these intervals, $Y_{(1)}$ is concentrated at one value.

Example 19.24. Consider again uniform random labelled unordered rooted trees, as in Example 19.18. We have $n_k = \lfloor k!/e \rfloor$. In this case, it is simpler to redefine $n_k := k!$; Theorem 19.23(ii) is unaffected but (i) is modified to

$$\mathbb{P}\bigl(Y_{(1)} = k(n)-1\bigr) \to e^{-x/e}, \tag{19.77}$$

$$\mathbb{P}\bigl(Y_{(1)} = k(n)\bigr) \to 1 - e^{-x/e}. \tag{19.78}$$

Cf. Carr, Goh and Schmutz [21].

Remark 19.25. We have for simplicity considered only the maximum value $Y_{(1)}$ in Theorem 19.23. It is easily seen, by minor modifications in the proof, that for any fixed j, in (ii) also $Y_{(j)} = k(n)$ w.h.p., while in (i) $Y_{(j)} \in \{k(n)-1, k(n)\}$ w.h.p., but the two probabilities have limits depending on j; in fact, the number of j such that $Y_{(j)} = k(n)$ converges in distribution to $\mathrm{Po}(x)$. We omit the details.

To make the statement about "most" n precise, recall that the *upper* and *lower densities* of a set $A \subseteq \mathbb{N}$ are defined as $\limsup_{n\to\infty} a(n)/n$ and $\liminf_{n\to\infty} a(n)/n$, where $a(n) := |\{i \leqslant n : i \in A\}|$; if they coincide, i.e., if the limit $\lim_{n\to\infty} a(n)/n$ exists, it is called the *density*. Similarly, the the *logarithmic density* of A is $\lim_{n\to\infty} \frac{1}{\log n}\sum_{i\leqslant n,\, i\in A}\frac{1}{i}$, when this limit exists, with *upper* and *lower logarithmic densities* defined using lim sup and lim inf. It is easily seen that if a set has a density, then it also has a logarithmic density, and the two densities coincide. (The converse does not hold.) Furthermore, define

$$p_n^* := \max_k \mathbb{P}(Y_{(1)} = k).$$

It follows from Theorem 19.16 that the second largest probability $\mathbb{P}(Y_{(1)} = k)$ is $1 - p_n^* + o(1)$. Thus, for n in a subsequence, $Y_{(1)}$ is asymptotically concentrated at one value if and only if $p_n^* \to 1$; if p_n^* stays away from 1, $Y_{(1)}$ takes two values with large probabilities.

Theorem 19.26. *Suppose that $w_0 > 0$ and that $w_{k+1}/w_k \to 0$ as $k \to \infty$. Suppose further that $n \to \infty$ and $m = m(n)$ with $m = \lambda n + o(\sqrt{n})$ where $0 < \lambda < \infty$.*

(i) *If $\frac{1}{2} < a < 1$, then the set $\{n : p_n^* < a\}$ has upper density $\log \frac{a}{1-a} / \log \frac{1}{1-a} > 0$ and lower density 0.*

(ii) *There exists a subsequence of n with upper density 1 and logarithmic density 1 such that $p_n^* \to 1$.*

Note that the upper density in (i) can be made arbitrarily close to 1 by taking a close to 1. This was observed by Carr, Goh and Schmutz [21] for the case in Example 19.24. (However, they failed to remark that the lower density nevertheless is 0.)

Proof. (i): Let $b_1 := -\log a$ and $b_2 := -\log(1-a)$; thus $0 < b_1 < b_2 < \infty$. Then $\max(e^{-x}, 1 - e^{-x}) < a \iff x \in (b_1, b_2)$, and it follows from Theorem 19.23 (and a uniformity in x implicit in the proof) that for any $\varepsilon > 0$, if $n \in \bigcup_k [(b_1 + \varepsilon)n_k, (b_2 - \varepsilon)n_k]$, then $p_n^* < a$ for large n, while if $n \notin \bigcup_k [(b_1 - \varepsilon)n_k, (b_2 + \varepsilon)n_k]$, then $p_n^* > a$ for large n. Since $n_{k+1}/n_k \to 0$ as $k \to \infty$, it is easily seen that for any b_1', b_2' with $0 < b_1' < b_2' < \infty$, $\bigcup_k [b_1' n_k, b_2' n_k]$ has upper density $(b_2' - b_1')/b_2'$ and lower density 0; it follows by taking $b_j' := b_j \pm \varepsilon$ and letting $\varepsilon \to 0$ that the set $\{n : p_n^* < a\}$ has upper density $(b_1 - b_2)/b_2$ and lower density 0.

(ii): Let Ω_k be an increasing sequence with $\Omega_k \nearrow \infty$ so slowly that $\log \Omega_k = o(\log(n_k/n_{k-1}))$. Let $A := \bigcup_k [\Omega_k^{-1} n_k, \Omega_k n_k]$. By Theorem 19.23(ii), $p_n^* \to 1$ as $n \to \infty$ with $n \notin A$, so it suffices to prove that A has lower density 0 and logarithmic density 0.

It is easily seen that for the upper logarithmic density of A, it suffices to consider $n \in \{\lfloor \Omega_k n_k \rfloor\}$, which gives

$$\limsup_{k\to\infty} \frac{\sum_{j=1}^{k} \sum_{i=\Omega_j^{-1} n_j}^{\Omega_j n_j} 1/i}{\log(\Omega_k n_k)} \leqslant \limsup_{k\to\infty} \frac{\sum_{j=1}^{k} \bigl(2 \log \Omega_j + O(1)\bigr)}{\sum_{j=1}^{k} \log(n_j/n_{j-1})} \to 0.$$

Hence the logarithmic density exists and is 0.

The lower density is at most, considering the subsequence $\lfloor \Omega_k^{-1} n_k \rfloor$,

$$\liminf_{k\to\infty} \frac{a(\Omega_k^{-1} n_k)}{\Omega_k^{-1} n_k} \leqslant \liminf_{k\to\infty} \frac{\Omega_{k-1} n_{k-1}}{\Omega_k^{-1} n_k} = \lim_{k\to\infty} \frac{\Omega_{k-1}\Omega_k}{n_k/n_{k-1}} = 0,$$

since $\Omega_{k-1} \leqslant \Omega_k < (n_k/n_{k-1})^{1/3}$ for large k. (Alternatively, it is a general fact that the lower density is at most the (lower) logarithmic density, for any set $A \subseteq \mathbb{N}$.) $\square$

19.5. **The subcase $\lambda = \nu$ and $\sigma^2 = \infty$.** We give two examples of the case $\lambda = \nu$ and $\sigma^2 = \infty$. (In both examples, we may assume that $\nu = 1$ and $m = n - 1$, so the examples apply to simply generated random trees.) The first example shows that Theorem 19.7 does not always hold if $\sigma^2 = \infty$; the second shows that it sometimes does.

Example 19.27. Let $1 < \alpha < 2$ and let (w_k) be a probability weight sequence with $w_0 > 0$ and $w_k \sim ck^{-\alpha-1}$ as $k \to \infty$, for some $c > 0$. (This is as in Example 12.10 with $\beta = \alpha + 1 \in (2, 3)$. If (w_k) is not a probability weight sequence, we may replace c by $c' := c/\Phi(1)$.) We have $\rho = 1$, and thus $\nu = \Psi(1) = \sum kw_k < \infty$. (We may obtain any desired $\nu > 0$, for example $\nu = 1$, by adjusting the first few w_k.)

We consider the case $m = \nu n + O(1)$; thus $m/n \to \lambda = \nu$. (This includes the tree case $m = n - 1$ in the case $\nu = 1$. Actually, it suffices to assume $m = \nu n + o(n^{1/\alpha})$.) Then $\tau = 1 = \rho$, and $\pi_k = w_k$.

The random variable ξ thus satisfies $\mathbb{E}\,\xi = \lambda = \nu$. Note that $\sigma^2 := \operatorname{Var}\xi = \infty$. (This is the main reason for taking $1 < \alpha < 2$; if we take $\alpha > 2$, then $\sigma^2 < \infty$ and Theorem 19.7 applies.) Furthermore,

$$\mathbb{P}(\xi \geqslant k) = \sum_{l=k}^{\infty} w_l \sim c\alpha^{-1}k^{-\alpha}. \tag{19.79}$$

As in the proof of Theorem 18.14, there exists by [39, Section XVII.5] a stable random variable X_α (satisfying (19.93) and (19.113)) such that

$$\frac{S_n - n\nu}{n^{1/\alpha}} \xrightarrow{\mathrm{d}} X_\alpha; \tag{19.80}$$

moreover, by [46, § 50], the local limit law (18.22) holds uniformly for all integers ℓ. Note that the density function g is bounded and uniformly continuous on $\mathbb{R}$, and that $g(0) > 0$ by (18.24). (In fact, $g(x) > 0$ for all x. See also [39, Section XVII.6] for an explicit formula for g as a power series; X_α is, after rescaling, the extreme case $\gamma = 2 - \alpha$, in the notation there.)

By (19.14) and (18.22),

$$\begin{aligned}\mathbb{P}(Y_1 = k) &= \frac{w_k\,\mathbb{P}(S_{n-1} = m - k)}{\mathbb{P}(S_n = m)} = w_k \frac{g(-k/n^{1/\alpha}) + o(1)}{g(0) + o(1)} \\ &= w_k \frac{g(-k/n^{1/\alpha}) + o(1)}{g(0)},\end{aligned} \tag{19.81}$$

uniformly in $k \geqslant 0$.

For a non-negative function f on $[0, \infty)$, define

$$X_n^f := \sum_{i=1}^{n} f(Y_i/n^{1/\alpha}). \tag{19.82}$$

In particular, if f is the indicator $\mathbf{1}\{a \leqslant x \leqslant b\}$ of an interval $[a, b]$, we write $X_n^{a,b}$ and have in the notation of (19.7)

$$X_n^{a,b} := |\{i \leqslant n : an^{1/\alpha} \leqslant Y_i \leqslant bn^{1/\alpha}\}| = N_{[an^{1/\alpha}, bn^{1/\alpha}]}. \tag{19.83}$$

Suppose that f is either the indicator of a compact interval $[a,b] \subset (0,\infty)$, or a continuous function with compact support in $(0,\infty)$ (or, more generally, any Riemann integrable function with support in a compact interval in $(0,\infty)$). Then, using (19.81) and dominated convergence,

$$\begin{aligned}
\mathbb{E}\, X_n^f &= n\sum_{k=0}^{\infty} f(k/n^{1/\alpha})\,\mathbb{P}(Y_1=k) = n\sum_{k=0}^{\infty} f(k/n^{1/\alpha}) w_k \frac{g(-k/n^{1/\alpha})+o(1)}{g(0)} \\
&= n^{1+1/\alpha}\int_0^\infty f(\lfloor xn^{1/\alpha}\rfloor/n^{1/\alpha}) w_{\lfloor xn^{1/\alpha}\rfloor} \frac{g(-\lfloor xn^{1/\alpha}\rfloor/n^{1/\alpha})+o(1)}{g(0)}\,\mathrm{d}x \\
&\to \int_0^\infty f(x) c x^{-\alpha-1}\frac{g(-x)}{g(0)}\,\mathrm{d}x. \qquad (19.84)
\end{aligned}$$

In the special case when $f(x) = \mathbf{1}\{a \leqslant x \leqslant b\}$ with $0 < a < b < \infty$, we further similarly obtain,

$$\begin{aligned}
\mathbb{E}\, X_n^{a,b}(X_n^{a,b}-1) &= n(n-1)\sum_{k,j\geqslant 0} f(k/n^{1/\alpha}) f(j/n^{1/\alpha})\,\mathbb{P}(Y_1=k,\ Y_2=j) \\
&= n(n-1)\sum_{k,j\geqslant 0} f(k/n^{1/\alpha}) f(j/n^{1/\alpha}) w_k w_l \frac{\mathbb{P}(S_{n-2}=m-k-j)}{\mathbb{P}(S_n=m)} \\
&\to c^2\int_0^\infty\int_0^\infty f(x)f(y)x^{-\alpha-1}y^{-\alpha-1}\frac{g(-x-y)}{g(0)}\,\mathrm{d}x\,\mathrm{d}y \\
&= c^2\int_a^b\int_a^b x^{-\alpha-1}y^{-\alpha-1}\frac{g(-x-y)}{g(0)}\,\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

and, more generally, for any $\ell \geqslant 1$,

$$\mathbb{E}(X_n^{a,b})_\ell \to c^\ell \int_a^b\cdots\int_a^b \prod_{i=1}^{\ell} x_i^{-\alpha-1}\frac{g(-x_1-\dots-x_\ell)}{g(0)}\,\mathrm{d}x_1\cdots\mathrm{d}x_\ell. \qquad (19.85)$$

For each such interval $[a,b]$, this integral is bounded by CR^ℓ for all $\ell \geqslant 1$, for some C and R (depending on a and b), and it follows by the method of moments that $X_n^{a,b} \overset{\mathrm{d}}{\longrightarrow} X_\infty^{a,b}$, where $X_\infty^{a,b}$ is determined by its factorial moments

$$\mathbb{E}(X_\infty^{a,b})_\ell = c^\ell \int_a^b\cdots\int_a^b \prod_{i=1}^{\ell} x_i^{-\alpha-1}\frac{g(-x_1-\dots-x_\ell)}{g(0)}\,\mathrm{d}x_1\cdots\mathrm{d}x_\ell. \qquad (19.86)$$

(It follows that $X_\infty^{a,b}$ has a finite moment generating function, so the method of moment applies.) Furthermore, joint convergence for several intervals holds by the same argument. It follows also (by some modifications or by approximation with step functions; we omit the details) that $X_n^f \overset{\mathrm{d}}{\longrightarrow} X_\infty^f$ for every continuous $f \geqslant 0$ with compact support and some X_∞^f.

Let Ξ_n be the multiset $\{Y_i/n^{1/\alpha} : Y_i > 0\}$, regarded as a point process on $(0,\infty)$. (I.e., formally we let Ξ_n be the discrete measure $\sum_{i:Y_i>0}\delta_{Y_i/n^{1/\alpha}}$. See e.g. Kallenberg [68] or [69] for details on point processes, or Janson [57,

§ 4] for a brief summary.) The convergence $X_n^f \xrightarrow{\mathrm{d}} X_\infty^f$ for every continuous $f \geqslant 0$ with compact support in $(0,\infty)$ implies, see [68, Lemma 5.1] or [69, Lemma 16.15 and Theorem 16.16], that Ξ_n converges in distribution, as a point process on $(0,\infty)$, to some point process Ξ on $(0,\infty)$. The distribution of Ξ is determined by (19.86), where $X_\infty^{a,b}$ is the number of points of Ξ in $[a,b]$. By (19.86) or (19.84), the intensity measure is given by

$$\mathbb{E}\,\Xi = cg(0)^{-1}x^{-\alpha-1}g(-x)\,\mathrm{d}x. \tag{19.87}$$

We can also consider infinite intervals. Let $a>0$. Then, using again (19.14) and noting that $\sum_{k=-\infty}^{\infty}\mathbb{P}(S_{n-1}=m-k)=1$,

$$\begin{aligned}\mathbb{E}\,X_n^{a,\infty} &= n\sum_{k\geqslant an^{1/\alpha}}\mathbb{P}(Y_1=a) = n\sum_{k\geqslant an^{1/\alpha}} w_k\frac{\mathbb{P}(S_{n-1}=m-k)}{\mathbb{P}(S_n=m)}\\ &\leqslant nC_1(an^{1/\alpha})^{-\alpha-1}\frac{\sum_{k\geqslant an^{1/\alpha}}\mathbb{P}(S_{n-1}=m-k)}{\mathbb{P}(S_n=m)}\\ &\leqslant C_1a^{-\alpha-1}n^{-1/\alpha}\frac{1}{n^{-1/\alpha}(g(0)+o(1))}\\ &\leqslant C_2a^{-\alpha-1}. \end{aligned}\tag{19.88}$$

By Fatou's lemma, (19.88) implies $\mathbb{E}\,X_\infty^{a,\infty}\leqslant C_2a^{-\alpha-1}<\infty$. Hence, $X_\infty^{a,\infty}<\infty$ a.s. for every $a>0$, and we may order the points in Ξ in decreasing order as

$$\Xi = \{\eta_j\}_{j=1}^J \qquad \text{with} \quad \eta_1\geqslant\eta_2\geqslant\dots. \tag{19.89}$$

(Here $J=X_\infty^{0,\infty}\leqslant\infty$ is the random number of points in Ξ. We shall see that $J=\infty$ a.s.)

The bound (19.88) is uniform in n, and tends to 0 as $a\to\infty$. It follows, see [57, Lemma 4.1], that if we regard Ξ_n and Ξ as point processes on $[0,\infty]$, the convergence $\Xi_n\xrightarrow{\mathrm{d}}\Xi$ on $(0,\infty)$ implies the stronger result

$$\Xi_n\xrightarrow{\mathrm{d}}\Xi \quad \text{on } [0,\infty]. \tag{19.90}$$

The points in Ξ_n, ordered in decreasing order, are $Y_{(1)}/n^{1/\alpha}\geqslant Y_{(2)}/n^{1/\alpha}\geqslant\dots$. If we extend (19.89) by defining $\eta_j:=0$ when $j>J$, the convergence (19.90) of point processes on $[0,\infty]$ is by [57, Lemma 4.4] equivalent to joint convergence of the ranked points, i.e.

$$Y_{(j)}/n^{1/\alpha}\xrightarrow{\mathrm{d}}\eta_j, \qquad j\geqslant 1 \text{ (jointly)}. \tag{19.91}$$

We claim that each $\eta_j>0$ a.s., and thus $J=X_\infty^{0,\infty}=\infty$ a.s. Suppose the opposite: $\mathbb{P}(\eta_j=0)=\delta>0$ for some j. Then, for every $\varepsilon>0$, $\liminf\mathbb{P}(Y_{(j)}/n^{1/\alpha}<\varepsilon)\geqslant\mathbb{P}(\eta_j<\varepsilon)\geqslant\delta$, and it follows that there exists a sequence $\varepsilon_n\to 0$ such that $\mathbb{P}(Y_{(j)}/n^{1/\alpha}<\varepsilon_n)\geqslant\delta/2$ for all n. We may assume that $\varepsilon_nn^{1/\alpha}\to\infty$. Let $A>0$ and take (for large n) $a_n:=\varepsilon_n$ and $b_n:=\bigl(\varepsilon_n^{-\alpha}-\alpha c^{-1}A\bigr)^{-1/\alpha}$. Then $a_n,b_n\to 0$. For $k\leqslant b_nn^{1/\alpha}=o(n^{1/\alpha})$, (18.22) implies $\mathbb{P}(S_{n-1} = m - k)/\mathbb{P}(S_n = m) \to 1$, and the argument in (19.84)–(19.85) yields, for each $\ell \geqslant 1$,

$$\mathbb{E}\big(X_n^{a_n,b_n}\big)_\ell \sim \left(n \sum_{k=a_n n^{1/\alpha}}^{b_n n^{1/\alpha}} w_k\right)^\ell \sim \left(c\int_{a_n}^{b_n} x^{-\alpha-1}\,\mathrm{d}x\right)^\ell$$
$$= \big(c\alpha^{-1}\big(a_n^{-\alpha} - b_n^{-\alpha}\big)\big)^\ell = A^\ell. \tag{19.92}$$

Hence, $X_n^{a_n,b_n} \xrightarrow{\mathrm{d}} \mathrm{Po}(A)$; in particular,

$$\delta/2 \leqslant \mathbb{P}(Y_{(j)}/n^{1/\alpha} < \varepsilon_n) \leqslant \mathbb{P}(X_n^{a_n,b_n} < j) \to \mathbb{P}(\mathrm{Po}(A) < j).$$

Taking A large enough, we can make $\mathbb{P}(\mathrm{Po}(A) < j) < \delta/2$, a contradiction which proves our claim.

We have shown that (19.91) holds with $\eta_j > 0$. Furthermore, since the intensity (19.87) is absolutely continuous, each η_j has an absolutely continuous distribution. Hence $Y_{(1)}$, and every $Y_{(j)}$, is of the order $n^{-1/\alpha}$, with a continuous limit distribution η_j (and thus no strict concentration at some constant times $n^{-1/\alpha}$).

Note that if we consider i.i.d. variables $\xi_1, \dots, \xi_n$, then $\{\xi_i/n^{1/\alpha} : \xi_i > 0\}$ converges (as is easily verified) to a Poisson process on $[0, \infty]$ with intensity $cx^{-\alpha-1}\,\mathrm{d}x$. This intensity differs from the intensity of Ξ in (19.87), and, since $g(-x) \to 0$ as $x \to \infty$, it is easy to see that $\xi_{(1)}/n^{1/\alpha}$ and $Y_{(1)}/n^{1/\alpha}$ have different limit distributions. Thus, Theorem 19.7 does not hold in this case. (However, $Y_{(1)}$ and $\xi_{(1)}$ are of the same order $n^{-1/\alpha}$.) Note also that, as an easy consequence of (19.86), the limiting point process Ξ in this example is *not* a Poisson process.

Remark 19.28. The distribution of the limiting point process Ξ in Example 19.27 is in principle determined by (19.86) and its extension to joint convergence for several $X_\infty^{a_i,b_i}$. This can be made more explicit as follows. (See Łuczak and Pittel [83] for similar calculations.)

It follows from Feller [39, Section XVII.5], see e.g. [63] for detailed calculations, that X_α has the characteristic function

$$\varphi(t) = \exp\big(c\,\Gamma(-\alpha)(-\mathrm{i}t)^\alpha\big), \qquad t \in \mathbb{R}. \tag{19.93}$$

(Note that $\Gamma(-\alpha) > 0$ and $\mathrm{Re}(-\mathrm{i}t)^\alpha < 0$ for $t \neq 0$ since $1 < \alpha < 2$.) The inversion formula gives

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}xt}\varphi(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}xt + c\Gamma(-\alpha)(-\mathrm{i}t)^\alpha}\,\mathrm{d}t, \tag{19.94}$$

and (19.86) yields

$$\begin{aligned}\mathbb{E}(X_\infty^{a,b})_\ell &= \frac{1}{2\pi g(0)} c^\ell \int_a^b \cdots \int_a^b \prod_{j=1}^{\ell} x_j^{-\alpha-1} \int_{-\infty}^{\infty} \prod_{j=1}^{\ell} e^{\mathrm{i}x_j t} \varphi(t)\,\mathrm{d}t\,\mathrm{d}x_1 \cdots \mathrm{d}x_\ell \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \Bigl(c \int_a^b x^{-\alpha-1} e^{\mathrm{i}tx}\,\mathrm{d}x\Bigr)^\ell \varphi(t)\,\mathrm{d}t. \qquad (19.95)\end{aligned}$$

In particular, $\mathbb{E}(X_\infty^{a,b})_\ell = O(C^\ell)$ for some $C < \infty$ (with C depending on a but not on b). Hence, $X_\infty^{a,b}$ has probability generating function, convergent for all complex z,

$$\begin{aligned}\mathbb{E}\, z^{X_\infty^{a,b}} &= \mathbb{E} \sum_{\ell=0}^{\infty} \binom{X_\infty^{a,b}}{\ell} (z-1)^\ell \\ &= \sum_{\ell=0}^{\infty} \frac{(z-1)^\ell}{\ell!} \cdot \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \Bigl(c \int_a^b x^{-\alpha-1} e^{\mathrm{i}tx}\,\mathrm{d}x\Bigr)^\ell \varphi(t)\,\mathrm{d}t \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl((z-1)c \int_a^b x^{-\alpha-1} e^{\mathrm{i}tx}\,\mathrm{d}x\Bigr) \varphi(t)\,\mathrm{d}t. \qquad (19.96)\end{aligned}$$

We can here let $b \to \infty$, so (19.96) holds for $b = \infty$ too. In particular, taking $z = 0$, we obtain, using (19.91), the limit distribution of $Y_{(1)}/n^{1/\alpha}$ as

$$\begin{aligned}&\mathbb{P}(\eta_1 \leqslant x) = \mathbb{P}(X_\infty^{x,\infty} = 0) \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl(-c \int_a^{\infty} x^{-\alpha-1} e^{\mathrm{i}tx}\,\mathrm{d}x\Bigr) \varphi(t)\,\mathrm{d}t \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl(-c \int_a^{\infty} x^{-\alpha-1} e^{\mathrm{i}tx}\,\mathrm{d}x + c\,\Gamma(-\alpha)(-\mathrm{i}t)^\alpha\Bigr)\,\mathrm{d}t \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl(c\Bigl(\int_0^a x^{-\alpha-1}\bigl(e^{\mathrm{i}tx} - 1 - \mathrm{i}tx\bigr)\,\mathrm{d}x - \frac{a^{-\alpha}}{\alpha} - \mathrm{i}t\frac{a^{1-\alpha}}{\alpha-1}\Bigr)\Bigr)\,\mathrm{d}t, \\ &\qquad (19.97)\end{aligned}$$

where the last equality holds because

$$\Gamma(-\alpha)u^\alpha = \int_0^\infty x^{-\alpha-1}\bigl(e^{-ux} - 1 + ux\bigr)\,\mathrm{d}u \qquad (19.98)$$

when $\operatorname{Re} u \geqslant 0$ and $1 < \alpha < 2$.

Furthermore, by extending (19.86) to joint factorial moments for several (disjoint) intervals, it follows similarly, for step functions f, that the random variable $X_\infty^f = \sum_{j=1}^\infty f(\eta_j)$ satisfies

$$\begin{aligned}\mathbb{E}\, e^{X_\infty^f} &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl(c\int_0^\infty \bigl(e^{f(x)}-1\bigr)x^{-\alpha-1}e^{\mathrm{i}tx}\,\mathrm{d}x\Bigr)\varphi(t)\,\mathrm{d}t \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl(c\int_0^\infty \bigl(e^{f(x)}-1\bigr)x^{-\alpha-1}e^{\mathrm{i}tx}\,\mathrm{d}x + c\,\Gamma(-\alpha)(-\mathrm{i}t)^\alpha\Bigr)\,\mathrm{d}t \\ &= \frac{1}{2\pi g(0)} \int_{-\infty}^{\infty} \exp\Bigl(c\Bigl(\int_0^\infty x^{-\alpha-1}\bigl(e^{f(x)+\mathrm{i}tx}-1-\mathrm{i}tx\bigr)\,\mathrm{d}x\Bigr)\Bigr)\,\mathrm{d}t.\end{aligned} \tag{19.99}$$

By taking limits, (19.99) extends to, e.g., any bounded measurable f with compact support in $(0,\infty]$. Since $\mathbb{E}\, e^{sX_\infty^f} = \mathbb{E}\, e^{X_\infty^{sf}}$ for $s \in \mathbb{R}$, this formula determines (in principle) the distribution of each X_∞^f and thus of Ξ.

Example 19.29. Let (w_k) be as in Example 19.27 but with $\alpha = 2$, i.e., $w_k \sim ck^{-3}$ as $k \to \infty$, for some $c > 0$. (Example 12.10 with $\beta = 3$.) We still have (19.79); further, $\rho = 1$, and thus $\nu = \Psi(1) = \sum kw_k < \infty$. (We may again obtain any desired $\nu > 0$, for example $\nu = 1$, by adjusting the first few w_k.)

As in Example 19.27. we consider the case $m = \nu n + O(1)$, including the tree case $m = n-1$ when $\nu = 1$. Thus, again, $m/n \to \lambda = \nu$, $\tau = 1 = \rho$, $\pi_k = w_k$, and the random variable ξ satisfies $\mathbb{E}\,\xi = \lambda = \nu$, while $\sigma^2 := \operatorname{Var}\xi = \infty$.

As in the proof of Theorem 18.14, we have the central limit theorem (18.25), and the local limit law (18.26) holds uniformly for all integers ℓ.

Choose $B(n) := n^{1/2}\log\log n = o(\sqrt{n\log n})$. Then, by (18.26),

$$Z(m,n) = \mathbb{P}(S_n = m) = \frac{g(0)+o(1)}{\sqrt{n\log n}} \tag{19.100}$$

and, uniformly for all $k \leqslant B(n)$,

$$Z(m-k,n-1) = \mathbb{P}(S_{n-1} = m-k) = \frac{g(0)+o(1)}{\sqrt{n\log n}}. \tag{19.101}$$

Hence, by (19.14), (19.22) holds. Furthermore, (18.26) yields also, since $g(0) = \max_{x\in\mathbb{R}} g(x)$,

$$Z(m-k,n-1) = \mathbb{P}(S_{n-1} = m-k) \leqslant \frac{g(0)+o(1)}{\sqrt{n\log n}}, \tag{19.102}$$

uniformly for all $k \geqslant 0$; hence (19.14) implies that (19.17)–(19.18) hold.

For our $B(n)$ we have by (19.79)

$$\mathbb{P}(\xi \geqslant B(n)) = O(B(n)^{-2}) = o(n^{-1}), \tag{19.103}$$

so (19.19) holds, and thus (19.20) holds.

The proof of Theorem 19.7 now holds without further modifications; hence the conclusions of Theorem 19.7 holds for this example, although $\sigma^2 = \infty$.

Note that in Example 19.27, although the asymptotic distributions of $Y_{(1)}$ and $\xi_{(1)}$ are different, they are still of the same order of magnitude. We do not know whether this is true in general. This question can be formulated more precisely as follows.

Problem 19.30. *In the case $\lambda = \nu$, do Theorem 19.7*(ii)–(iii) *hold also when $\sigma^2 = \infty$?*

In any case, we can ask about the possible rates of growth of $Y_{(1)}$, for example as follows, where we for definiteness consider the tree case $m = n-1$ (and thus $\lambda = 1$).

Problem 19.31. *For which sequences $\omega(n)$ does there exist a weight sequence with $\nu = 1$ such that, with $m = n-1$, $Y_{(1)} \geqslant \omega(n)$ w.h.p.?*

As remarked earlier, the answer is positive for $\omega(n) = n^{1-\varepsilon}$, for any $\varepsilon > 0$, as shown by Example 19.27 with $\alpha < 1/(1-\varepsilon)$.

19.6. **The case $\lambda > \nu$.** We turn to the case $\lambda > \nu$. Then, as briefly discussed in Section 11, the asymptotic formula for the numbers N_k in Theorem 11.4 accounts only for $\sum_{k=0}^{\infty} k\pi_k n = \mu n = \nu n$ balls, so there are $m - \nu n \approx (\lambda - \nu)n$ balls missing. A more careful treatment of the limits show that the explanation is that Theorem 11.4 really implies that the "small" boxes (i.e., those with rather few balls) have a total of about $\sum_{k=0}^{\infty} k\pi_k n = \mu n = \nu n$ balls, while the remaining $\approx (\lambda - \nu)n$ balls are in a few "large" boxes. One way to express this precisely is the following simple result.

Lemma 19.32. *Let $\mathbf{w} = (w_k)_{k\geqslant 0}$ be a weight sequence with $w_0 > 0$ and $\omega = \infty$. Suppose that $n \to \infty$ and $m = m(n)$ with $m/n \to \lambda$ where $\nu < \lambda < \infty$.*

(i) *For any sequence $K_n \to \infty$,*
$$\sum_{k\leqslant K_n} kN_k \geqslant \nu n + o_{\mathrm{p}}(n) \qquad \textit{and} \qquad \sum_{k>K_n} kN_k \leqslant (\lambda-\nu)n + o_{\mathrm{p}}(n).$$

(ii) *There exists a sequence $\Omega_n \to \infty$ such that for any sequence $K_n \to \infty$ with $K_n \leqslant \Omega_n$ we have*
$$\sum_{k\leqslant K_n} kN_k = \nu n + o_{\mathrm{p}}(n) \qquad \textit{and} \qquad \sum_{k>K_n} kN_k = (\lambda-\nu)n + o_{\mathrm{p}}(n).$$

Proof. The two statements in each part are equivalent, since
$$\sum_{k=0}^{\infty} kN_k = m = \lambda n + o(n). \tag{19.104}$$

(i): For every fixed ℓ, Theorem 11.4 implies
$$\frac{1}{n}\sum_{k\leqslant\ell} kN_k \xrightarrow{\mathrm{p}} \sum_{k\leqslant\ell} k\pi_k. \tag{19.105}$$

Let $\varepsilon > 0$. Since $\sum_{k=0}^{\infty} k\pi_k = \nu < \infty$, there exists ℓ such that $\sum_{k\leqslant\ell} k\pi_k > \nu - \varepsilon$, and (19.105) implies that w.h.p.

$$\frac{1}{n}\sum_{k\leqslant\ell} kN_k > \nu - \varepsilon.$$

Since ε is arbitrary, this implies $\sum_{k\leqslant K_n} kN_k \geqslant \nu n + o_{\mathrm{p}}(n)$.

(ii): For each fixed ℓ, $\sum_{k\leqslant\ell} k\pi_k < \sum_{k=0}^{\infty} k\pi_k = \nu$, and thus (19.105) implies $\mathbb{P}(\sum_{k\leqslant\ell} kN_k > \nu n) \to 0$. Hence, there exists an increasing sequence of integers n_ℓ such that if $n \geqslant n_\ell$, then $\mathbb{P}\bigl(\sum_{k\leqslant\ell} kN_k > \nu n\bigr) < 1/\ell$. Now define $\Omega_n = \ell$ for $n_\ell \leqslant n < n_{\ell+1}$. Then $\sum_{k\leqslant\Omega_n} kN_k \leqslant \nu n$ w.h.p., which together with (i) yields (ii). ☐

Consider the "large" boxes. One obvious possibility is that there is a single "giant" box with $\approx (\lambda-\nu)n$ balls; more formally, $(\lambda-\nu)n + o_{\mathrm{p}}(n)$ balls (a "monopoly"). Applying Lemma 19.32(i) with $K_n = o(n)$, we see that for every $\varepsilon > 0$, w.h.p. there are then less than εn balls in all other boxes with more than K_n balls each; thus, either $Y_{(2)} \leqslant K_n$ or $Y_{(2)} < \varepsilon n$. Consequently, this case is defined by

$$Y_{(1)} = (\lambda-\nu)n + o_{\mathrm{p}}(n), \tag{19.106}$$

$$Y_{(2)} = o_{\mathrm{p}}(n). \tag{19.107}$$

Equivalently, $Y_{(1)}/n \xrightarrow{\mathrm{p}} \lambda - \nu$ and $Y_{(2)}/n \xrightarrow{\mathrm{p}} 0$. This thus describes condensation of the missing balls to a single box.

We will see in Theorem 19.34 that, indeed, this is the case for the important example of weights with a power-law. Another, more extreme example is Example 10.8, $w_k = k!$, where $\nu = 0$, see Example 19.36.

However, if (w_k) is very irregular, (19.106)–(19.107) do not always hold. Examples 19.37 and 19.38 give examples where, at least for a subsequence, either $Y_{(2)}/n \xrightarrow{\mathrm{p}} a > 0$, so there are at least two giant boxes with order n balls each (an "oligopoly"), or $Y_{(1)}/n \xrightarrow{\mathrm{p}} 0$, so there is no giant box with order n balls, and the missing $(\lambda-\nu)n$ balls are distributed over a large number (necessarily $\to\infty$ as $n\to\infty$) of boxes, each with a large but $o(n)$ number of balls.

Example 19.33. We consider Example 12.10; $w_k \sim ck^{-\beta}$ as $k\to\infty$. If $\beta \leqslant 2$, then $\nu = \infty$, see (12.46), and thus $\lambda < \nu$ and Theorems 19.3 and 19.7 apply. We are interested in the case $\lambda > \nu$, so we assume $\beta > 2$. In this case, Jonsson and Stefánsson [67] showed (for the case of random trees) that when $\lambda > \nu$ we have the simple situation with condensation to a single giant box. We state this in the next theorem, which also includes further, more precise, results. (Note that the case $\lambda < \nu$ is covered by Theorems 19.3 and 19.7, with $Y_{(1)}$ of order $\log n$; the case $\lambda = \nu$ is studied in Examples 19.27 and 19.29 for $2 < \beta \leqslant 3$, and is covered by Theorem 19.7 when $\beta > 3$; in both cases $Y_{(1)}$ is of order $n^{-1/(\beta-1)} = o(n)$.)

Theorem 19.34. *Suppose that $w_k \sim ck^{-\beta}$ as $k \to \infty$ for some $c > 0$ and $\beta > 2$. Then $\nu < \infty$. Suppose further $m/n \to \lambda > \nu$. Let $\alpha := \beta - 1 > 1$ and $c' := c/\Phi(1)$.*

(i) *The random allocation $B_{m,n} = (Y_1, \dots, Y_n)$ has largest components*

$$Y_{(1)} = (\lambda - \nu)n + o_{\mathrm{p}}(n), \tag{19.108}$$

$$Y_{(2)} = o_{\mathrm{p}}(n). \tag{19.109}$$

(ii) *The partition function is asymptotically given by*

$$Z(m,n) \sim c(\lambda-\nu)^{-\beta}\Phi(1)^{n-1}n^{1-\beta}. \tag{19.110}$$

(iii) *Furthermore,*

$$\big(Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}\big) \overset{\mathrm{d}}{\approx} \Big(m - \sum_{i=1}^{n-1}\xi_i, \xi'_{(1)}, \dots, \xi'_{(n-1)}\Big), \tag{19.111}$$

where $\xi'_{(1)}, \dots, \xi'_{(n-1)}$ are the $n-1$ i.i.d. random variables $\xi_1, \dots, \xi_{n-1}$, with distribution (π_k), ordered in decreasing order.

(iv) *$Y_{(1)} = m - \nu n + O_{\mathrm{p}}(n^{1/\alpha})$ and*

$$n^{-1/\alpha}\big(m - \nu n - Y_{(1)}\big) \overset{\mathrm{d}}{\longrightarrow} X_\alpha, \tag{19.112}$$

where X_α is an α-stable random variable with Laplace transform

$$\mathbb{E}\, e^{-tX_\alpha} = \exp\big(c'\Gamma(-\alpha)t^\alpha\big), \qquad \operatorname{Re} t \geqslant 0. \tag{19.113}$$

(v) *$Y_{(2)} = O_{\mathrm{p}}(n^{1/\alpha})$ and*

$$n^{-1/\alpha}Y_{(2)} \overset{\mathrm{d}}{\longrightarrow} W, \tag{19.114}$$

where W has the Fréchet distribution

$$\mathbb{P}(W \leqslant x) = \exp\Big(-\frac{c'}{\alpha}x^{-\alpha}\Big), \qquad x \geqslant 0. \tag{19.115}$$

(vi) *More generally, for each $j \geqslant 2$, $Y_{(j)} = O_{\mathrm{p}}(n^{1/\alpha})$ and*

$$n^{-1/\alpha}Y_{(j)} \overset{\mathrm{d}}{\longrightarrow} W_j, \tag{19.116}$$

where W_j has the density function

$$c'x^{-\alpha-1}\frac{\big(c'\alpha^{-1}x^{-\alpha}\big)^{j-2}}{(j-2)!}\exp\big(-c'\alpha^{-1}x^{-\alpha}\big), \qquad x \geqslant 0, \tag{19.117}$$

and $c'\alpha^{-1}W_j^{-\alpha} \sim \Gamma(j-1,1)$.

Note that $\pi_k = w_k/\Phi(1)$ and that $\Gamma(-\alpha) > 0$ in (19.113).

Part (iii) shows that $Y_{(2)}, \dots, Y_{(n)}$ asymptotically are as order statistics of $n-1$ i.i.d. random variables ξ_i; thus the giant box absorbs the dependency between the variables $Y_1, \dots, Y_n$ introduced by the conditioning in (11.7).

Remark 19.35. Jonsson and Stefánsson [67] considered only trees, and thus $m = n-1$ and $\lambda = 1$, and then showed the tree versions of (i) and (ii). (They further showed Theorem 7.1 when $w_k \sim ck^{-\beta}$.) In the tree case (i) says that the random tree $\mathcal{T}_n$ has w.h.p. a node of largest degree $(1-\nu)n+o(n)$, while all other nodes have degrees $o(n)$; further, by Theorem 15.5, (ii) becomes

$$Z_n \sim c(1-\nu)^{-\beta}\Phi(1)^{n-1}n^{-\beta} \sim (1-\nu)^{-\beta}\Phi(1)^{n-1}w_n. \tag{19.118}$$

Proof of Theorem 19.34. We may assume that $w_0 > 0$ by the argument in Remark 11.8. Furthermore, using (11.9) for (ii), by dividing w_k (and c) by $\Phi(1)$, we may assume that (w_k) is a probability weight sequence, and thus $\Phi(1) = 1$. For $\lambda > \nu$ we have $\tau = \rho = 1$, and thus then $\pi_k = w_k$.

(i): $\Phi(t)$ has radius of convergence $\rho = 1$, and since $\beta > 2$, $\Phi(1) = \sum_k w_k < \infty$ and $\nu = \Phi'(1)/\Phi(1) < \infty$.

Consider as in Example 11.2 i.i.d. random variables $\xi_1, \dots, \xi_n$ with distribution $(\pi_k) = (w_k)$ and mean $\mu = \nu$.

Fix a small $\varepsilon > 0$. We assume that $\varepsilon < \lambda - \nu$.

By the law of large numbers, $S_{n-1}/n \xrightarrow{\text{p}} \mu = \nu$. We may thus find a sequence $\delta_n \to 0$ such that $|S_{n-1} - n\nu| \leqslant n\delta_n$ w.h.p.

Since $m/n - \nu - \delta_n \to \lambda - \nu > \varepsilon$, we have $m - \nu n - \delta_n n > \varepsilon n$ for large n; we consider only such n.

We separate the event $S_n = m$ into four disjoint cases (subevents):

$\mathcal{E}_1$: Exactly one $\xi_i > \varepsilon n$, and that ξ_i satisfies $|\xi_i - (m - \nu n)| \leqslant \delta_n n$.
$\mathcal{E}_2$: Exactly one $\xi_i > \varepsilon n$, and that ξ_i satisfies $|\xi_i - (m - \nu n)| > \delta_n n$.
$\mathcal{E}_3$: $\xi_i > \varepsilon n$ for at least two $i \in \{1, \dots, n\}$.
$\mathcal{E}_4$: All $\xi_i \leqslant \varepsilon n$.

We shall show that $\mathcal{E}_1$ is the dominating event. We define also the events

$\mathcal{E}_{1i}$: $S_n = m$, $|\xi_i - (m - \nu n)| \leqslant \delta_n n$ and $\xi_j \leqslant \varepsilon n$ for $j \neq i$.
$\mathcal{E}^*_{1i}$: $S_n = m$, $|\xi_i - (m - \nu n)| \leqslant \delta_n n$.
$\mathcal{E}^*_{2i}$: $S_n = m$, $|\xi_i - (m - \nu n)| > \delta_n n$, $\xi_i > \varepsilon n$.
$\mathcal{D}_{ij}$: $S_n = m$, $\xi_i > \varepsilon n$, $\xi_j > \varepsilon n$.

Then $\mathcal{E}_1$ is the disjoint union $\bigcup_{i=1}^n \mathcal{E}_{1i}$, so by symmetry

$$\mathbb{P}(\mathcal{E}_1) = n\,\mathbb{P}(\mathcal{E}_{11}). \tag{19.119}$$

Furthermore, for any i,

$$\mathcal{E}_{1i} \subseteq \mathcal{E}^*_{1i} \subseteq \mathcal{E}_{1i} \cup \bigcup_{j \neq i} \mathcal{D}_{ij}$$

and thus, again using symmetry,

$$\mathbb{P}(\mathcal{E}^*_{11}) \geqslant \mathbb{P}(\mathcal{E}_{11}) \geqslant \mathbb{P}(\mathcal{E}^*_{11}) - n\,\mathbb{P}(\mathcal{D}_{12}). \tag{19.120}$$

Using the fact that $|k-(m-\nu n)| \leqslant \delta_n n$ implies $w_k \sim ck^{-\beta} \sim c(\lambda n - \nu n)^{-\beta}$, together with $|S_{n-1} - n\nu| \leqslant \delta_n n$ w.h.p., we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{11}^*) &= \sum_{|k-(m-\nu n)|\leqslant \delta_n n} \mathbb{P}(\xi_1 = k,\, S_n = m) \\
&= \sum_{|k-(m-\nu n)|\leqslant \delta_n n} \mathbb{P}(\xi_1 = k)\, \mathbb{P}(S_{n-1} = m-k) \\
&= \sum_{|k-(m-\nu n)|\leqslant \delta_n n} c(\lambda-\nu)^{-\beta} n^{-\beta} \bigl(1+o(1)\bigr)\, \mathbb{P}(S_{n-1} = m-k) \\
&= c(\lambda-\nu)^{-\beta} n^{-\beta}\, \mathbb{P}\bigl(|S_{n-1} - n\nu| \leqslant \delta_n n\bigr) \bigl(1+o(1)\bigr) \\
&= c(\lambda-\nu)^{-\beta} n^{-\beta} \bigl(1+o(1)\bigr). \qquad (19.121)
\end{aligned}$$

Similarly, allowing the constants C_i here and below to depend on ε,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{2i}^*) &= \sum_{|k-(m-\nu n)| > \delta_n n,\, k>\varepsilon n} \mathbb{P}(\xi_i = k,\, S_n = m) \\
&\leqslant C_1 (\varepsilon n)^{-\beta} \sum_{|k-(m-\nu n)| > \delta_n n,\, k>\varepsilon n} \mathbb{P}(S_{n-1} = m-k) \\
&\leqslant C_2 n^{-\beta}\, \mathbb{P}\bigl(|S_{n-1} - \nu n| > \delta_n n\bigr) = o\bigl(n^{-\beta}\bigr). \qquad (19.122)
\end{aligned}$$

For any i and j, by symmetry,

$$\begin{aligned}
\mathbb{P}(\mathcal{D}_{ij}) &= \mathbb{P}(\xi_n > \varepsilon n,\, \xi_{n-1} > \varepsilon n,\, S_n = m) \\
&= \sum_{k>\varepsilon n} \mathbb{P}(\xi_n = k)\, \mathbb{P}(S_{n-1} = m-k,\, \xi_{n-1} > \varepsilon n) \\
&\leqslant C_3 (\varepsilon n)^{-\beta} \sum_{k>\varepsilon n} \mathbb{P}(S_{n-1} = m-k,\, \xi_{n-1} > \varepsilon n) \\
&\leqslant C_3 (\varepsilon n)^{-\beta}\, \mathbb{P}(\xi_{n-1} > \varepsilon n) \leqslant C_4 (\varepsilon n)^{1-2\beta}. \qquad (19.123)
\end{aligned}$$

Hence, (19.120) and (19.121) yield

$$\mathbb{P}(\mathcal{E}_{11}) = c(\lambda-\nu)^{-\beta} n^{-\beta} + o\bigl(n^{-\beta}\bigr) + O\bigl(n^{2-2\beta}\bigr) = c(\lambda-\nu)^{-\beta} n^{-\beta} + o\bigl(n^{-\beta}\bigr)$$

and hence, by (19.119),

$$\mathbb{P}(\mathcal{E}_1) = c(\lambda-\nu)^{-\beta} n^{1-\beta} + o\bigl(n^{1-\beta}\bigr). \qquad (19.124)$$

Furthermore, (19.122) yields

$$\mathbb{P}(\mathcal{E}_2) \leqslant \sum_{i=1}^{n} \mathbb{P}(\mathcal{E}_{2i}^*) = n\, \mathbb{P}(\mathcal{E}_{21}^*) = o\bigl(n^{1-\beta}\bigr), \qquad (19.125)$$

and (19.123) also yields

$$\mathbb{P}(\mathcal{E}_3) \leqslant \sum_{i<j} \mathbb{P}(\mathcal{D}_{ij}) \leqslant n^2\, \mathbb{P}(\mathcal{D}_{12}) = O\bigl(n^{3-2\beta}\bigr) = o\bigl(n^{1-\beta}\bigr). \qquad (19.126)$$

It remains to estimate $\mathbb{P}(\mathcal{E}_4)$. We define the truncated variables $\bar{\xi}_i := \xi_i \mathbf{1}\{\xi_i \leqslant \varepsilon n\}$ and $\bar{S}_n := \sum_{i=1}^n \bar{\xi}_i$. Thus $\mathcal{E}_4 \subseteq \{\bar{S}_n = m\}$ and hence, for every real s,

$$\mathbb{P}(\mathcal{E}_4) \leqslant e^{-sm}\, \mathbb{E}\, e^{s\bar{S}} = e^{-sm} \Bigl(\mathbb{E}\, e^{s\bar{\xi}_1}\Bigr)^n. \tag{19.127}$$

Let $s := a \log n/n$, for a constant $a > 0$ chosen later. Then,

$$\begin{aligned} \mathbb{E}\, e^{s\bar{\xi}_1} &= 1 + s\, \mathbb{E}\, \bar{\xi}_1 + \sum_{k=1}^{\varepsilon n} \pi_k \bigl(e^{sk} - 1 - sk\bigr) \\ &\leqslant 1 + s\nu + C_5 \sum_{k=1}^{2\beta/s} k^{-\beta} s^2 k^2 + C_5 \sum_{k=2\beta/s}^{\varepsilon n} k^{-\beta} e^{sk}. \end{aligned} \tag{19.128}$$

We have, treating the cases $2 < \beta < 3$, $\beta = 3$ and $\beta > 3$ separately, using $s \to 0$,

$$\sum_{k=1}^{2\beta/s} s^2 k^{2-\beta} \leqslant C_6 s^2 \max\Bigl(1, (2\beta/s)^{3-\beta}, \log(2\beta/s)\Bigr) = o(s).$$

Furthermore, for $k > 2\beta/s$,

$$\frac{k^{-\beta} e^{sk}}{(k+1)^{-\beta} e^{s(k+1)}} = \Bigl(1 + \frac{1}{k}\Bigr)^{\beta} e^{-s} \leqslant e^{\beta/k - s} \leqslant e^{s/2 - s} = e^{-s/2}.$$

Hence, the final sum in (19.128) is dominated by a geometric series

$$\sum_{k \leqslant \lfloor \varepsilon n \rfloor} (\lfloor \varepsilon n \rfloor)^{-\beta} e^{s\lfloor \varepsilon n \rfloor} e^{-s(\lfloor \varepsilon n \rfloor - k)/2} \leqslant C_7 s^{-1} n^{-\beta} e^{s\varepsilon n} = C_7 s^{-1} n^{-\beta} e^{a\varepsilon \log n}.$$

If we assume $a\varepsilon \leqslant \beta - 2$, the sum is thus $\leqslant C_8 n^{1-\beta+a\varepsilon} \leqslant C_8 n^{-1} = o(s)$. Consequently, (19.128) yields

$$\mathbb{E}\, e^{s\bar{\xi}_1} \leqslant 1 + s\nu + o(s) \leqslant \exp\bigl(s\nu + o(s)\bigr)$$

and thus (19.127) yields

$$\mathbb{P}(\mathcal{E}_4) \leqslant \exp\bigl(-sm + ns\nu + o(ns)\bigr) = \exp\bigl(-ns(\lambda - \nu + o(1))\bigr) = n^{-a(\lambda-\nu)+o(1)}. \tag{19.129}$$

We choose first $a := \beta/(\lambda - \nu)$ and then $\varepsilon < (\beta - 2)/a$, and see by (19.129) that then $\mathbb{P}(\mathcal{E}_4) = n^{-\beta+o(1)} = o\bigl(n^{1-\beta}\bigr)$. Combining (19.124), (19.125), (19.126) and (19.129), we find

$$\mathbb{P}(S_n = m) = \mathbb{P}(\mathcal{E}_1) + o\bigl(n^{1-\beta}\bigr) = c(\lambda - \nu)^{-\beta} n^{1-\beta} + o\bigl(n^{1-\beta}\bigr), \tag{19.130}$$

and, in particular, $\mathbb{P}(\mathcal{E}_1 \mid S_n = m) \to 1$. Consequently, by conditioning on $S_n = m$ we see that w.h.p. $|Y_{(1)} - (m - \nu n)| \leqslant \delta_n n$ and $Y_{(2)} \leqslant \varepsilon n$. Since ε can be chosen arbitrarily small, this completes the proof of (19.108)–(19.109).

(ii): $Z(m, n) = \mathbb{P}(S_n = m)$, so (19.110) follows from (19.130), since we assume $\Phi(1) = 1$.

(iii): Since $\mathcal{E}_1 \subseteq \{S_n = m\}$ and $\mathbb{P}(\mathcal{E}_1 \mid S_n = m) \to 1$,

$$(Y_1, \dots, Y_n) \overset{\mathrm{d}}{=} \bigl((\xi_1, \dots, \xi_n) \mid S_n = m\bigr) \overset{\mathrm{d}}{\approx} \bigl((\xi_1, \dots, \xi_n) \mid \mathcal{E}_1\bigr). \tag{19.131}$$

When we consider the ordered variables $Y_{(1)}, \dots, Y_{(n)}$, we may by symmetry condition on $\mathcal{E}_{1n}$ instead of $\mathcal{E}_1$. Note that $\mathcal{E}_{1n}$ is the event $(\xi_1, \dots, \xi_n) \in A$, where A is the set

$$\Bigl\{(x_1, \dots, x_n) : x_j \leqslant \varepsilon n \text{ for } j \leqslant n-1,\, x_n = m - \sum_{i=1}^{n-1} x_i,\, \Bigl|\sum_{i=1}^{n-1} x_i - \nu n\Bigr| \leqslant \delta_n n\Bigr\}.$$

Since $(x_1, \dots, x_n) \in A$ implies $|x_n - (m - \nu n)| \leqslant \delta_n n$, we then have, similarly to (19.121),

$$\mathbb{P}(\xi_n = x_n) \sim c x_n^{-\beta} \sim c(m - \nu n)^{-\beta} \sim c(\lambda - \nu)^{-\beta} n^{-\beta}.$$

Furthermore, $x_1, \dots, x_{n-1}$ determine x_n by $\sum_1^n x_i = m$. It follows that, uniformly for all $(x_1, \dots, x_n) \in A$,

$$\begin{aligned}
&\mathbb{P}\bigl((\xi_1, \dots, \xi_n) = (x_1, \dots, x_n)\bigr) \\
&\quad = \bigl(1 + o(1)\bigr) c(\lambda - \nu)^{-\beta} n^{-\beta}\, \mathbb{P}\bigl((\xi_1, \dots, \xi_{n-1}) = (x_1, \dots, x_{n-1})\bigr) \\
&\quad = \bigl(1 + o(1)\bigr) c(\lambda - \nu)^{-\beta} n^{-\beta}\, \mathbb{P}\bigl((\xi_1, \dots, \xi_{n-1}, m - S_{n-1}) = (x_1, \dots, x_n)\bigr).
\end{aligned}$$

Hence, since the factor $c(\lambda - \nu)^{-\beta} n^{-\beta}$ is a constant for each n,

$$\bigl((\xi_1, \dots, \xi_n) \mid \mathcal{E}_{1n}\bigr) \overset{\mathrm{d}}{\approx} \bigl((\xi_1, \dots, \xi_{n-1}, m - S_{n-1}) \mid \widetilde{\mathcal{E}}_n\bigr), \tag{19.132}$$

where $\widetilde{\mathcal{E}}_n$ is the event

$$\bigl\{(\xi_1, \dots, \xi_{n-1}, m - S_{n-1}) \in A\bigr\} = \bigl\{\xi_j \leqslant \varepsilon n \text{ for } j \leqslant n-1,\, \bigl|S_{n-1} - \nu n\bigr| \leqslant \delta_n n\bigr\}. \tag{19.133}$$

If $\widetilde{\mathcal{E}}_n$ holds, then $m - S_{n-1} \geqslant m - \nu n - \delta_n n > \varepsilon n$ (for large n), so the largest variable among $\xi_1, \dots, \xi_{n-1}, m - S_{n-1}$ is $m - S_{n-1}$. Hence, ordering the variables, we obtain using (19.131)–(19.132)

$$\bigl(Y_{(1)}, \dots, Y_{(n)}\bigr) \overset{\mathrm{d}}{\approx} \bigl((m - S_{n-1}, \xi'_{(1)}, \dots, \xi'_{(n-1)}) \mid \widetilde{\mathcal{E}}_n\bigr). \tag{19.134}$$

Finally, observe that $|S_{n-1} - \nu n| \leqslant \delta_n n$ w.h.p. and

$$\mathbb{P}\bigl(\xi_j > \varepsilon n \text{ for some } j \leqslant n-1\bigr) \leqslant n\, \mathbb{P}(\xi_1 > \varepsilon n) = O\bigl(n^{2-\beta}\bigr) \to 0.$$

Hence, $\mathbb{P}(\widetilde{\mathcal{E}}_n) \to 1$, and thus

$$\bigl((m - S_{n-1}, \xi'_{(1)}, \dots, \xi'_{(n-1)}) \mid \widetilde{\mathcal{E}}_n\bigr) \overset{\mathrm{d}}{\approx} \bigl(m - S_{n-1}, \xi'_{(1)}, \dots, \xi'_{(n-1)}\bigr). \tag{19.135}$$

The result (19.111) follows from (19.134) and (19.135).

(iv): By (iii), $m - n\nu - Y_{(1)} \overset{\mathrm{d}}{\approx} \sum_{i=1}^{n-1} \xi_i - n\nu$, and (19.112) follows by standard results on domains of attraction for stable distributions, see e.g. Feller [39, Section XVII.5].

(v): By (iii), $Y_{(2)} \overset{\mathrm{d}}{\approx} \xi'_{(1)}$, and (19.114) follows by standard results on the maximum of i.i.d. random variables, as in e.g. Leadbetter, Lindgren and Rootzén [82]: using $\mathbb{P}(\xi > x) \sim c\alpha^{-1}x^{-\alpha}$ as $x \to \infty$, we have

$$\begin{aligned}\mathbb{P}(Y_{(1)} \leqslant xn^{1/\alpha}) &= \mathbb{P}(\xi'_{(1)} \leqslant xn^{1/\alpha}) + o(1) = \mathbb{P}(\xi \leqslant xn^{1/\alpha})^{n-1} + o(1)\\ &= \bigl(1 - (c\alpha^{-1} + o(1))(xn^{1/\alpha})^{-\alpha}\bigr)^{n-1} + o(1)\\ &\to \exp\bigl(-c\alpha^{-1}x^{-\alpha}\bigr).\end{aligned}$$

(vi): Similar, cf. Leadbetter, Lindgren and Rootzén [82, Section 2.2]. □

Example 19.36. If we take $w_k = k!$, then $\nu = \rho = 0$. Consider the tree case $m = n - 1$. By Example 10.8, translating to balls-in-boxes, w.h.p. there are N_1 boxes with 1 ball each and a single box with the remaining $n-1-N_1$ balls, while all other boxes are empty; furthermore, $N_1 \xrightarrow{\mathrm{d}} \mathrm{Po}(1)$ so $N_1 = O_{\mathrm{p}}(1)$. Hence, $Y_{(1)} = n + O_{\mathrm{p}}(1)$ and $Y_{(2)} \leqslant 1$ w.h.p.

If we take $w_k = k!^\alpha$ with $0 < \alpha < 1$, and still $m = n - 1$, then by Example 10.9 and [64], $Y_{(1)} = n + O_{\mathrm{p}}(n^{1-\alpha}) = n + o_{\mathrm{p}}(n)$ and $Y_{(2)} \leqslant \lfloor 1/\alpha \rfloor$ w.h.p.

If we take $w_k = k!^\alpha$ with $\alpha > 1$, and still $m = n-1$, then by Example 10.9, w.h.p. there is a single box containing all $n - 1$ balls; thus $Y_{(1)} = n - 1$ and $Y_{(2)} = 0$ w.h.p.

In particular, (19.106)–(19.107) hold, with $\lambda = 1$ and $\nu = 0$, for all three cases. We guess that the same is true for any $\lambda < \infty$, but we have not checked the details.

Example 19.37. We consider the tree case $m = n-1$. Let $\mathsf{S} := \{k_0, k_1, \dots\}$ be an infinite set with $k_0 = 0$, $k_1 = 1$, $k_2 = 2$, and k_j for $j \geqslant 3$ chosen recursively as specified below. Let $w_k = (k + 1)^{-4}$ for $k \in \mathsf{S}$, and $w_k = 0$ otherwise; thus, $\operatorname{supp}(\mathbf{w}) = \mathsf{S}$. ($\mathsf{S} = \mathbb{N}_0$ gives Example 10.7 with $\beta = 4$.) Then $\rho = 1$ and

$$\nu = \Psi(1) = \frac{\sum_{k=0}^\infty kw_k}{\sum_{k=0}^\infty w_k} \leqslant \frac{\sum_{k=0}^\infty k(k+1)^{-4}}{w_0} = \zeta(3) - \zeta(4) < 0.2 < 1; \tag{19.136}$$

thus $\tau = \rho = 1$.

To begin with, we require that $k_j \geqslant jk_{j-1}$ for $j \geqslant 3$. Take $n = k_j$. A good allocation of $n - 1$ balls in n boxes has at most k_{j-1} balls in any box, since $n - 1 < k_j$, so

$$Y_{(1)} \leqslant k_{j-1} \leqslant k_j/j = n/j. \tag{19.137}$$

Hence, for n in the subsequence $\{k_j\}$, the random allocation $B_{n-1,n}$ has $Y_{(1)} = o(n)$.

Next, suppose that $k_0, \dots, k_{j-1}$ are given, and let $\mathbf{w}^{(k_{j-1})}$ be $\mathbf{w}$ truncated at k_{j-1} as in (13.4); for ease of notation we denote the corresponding generating function by $\Phi_j(t) := \sum_{i=0}^{j-1} w_{k_i}t^{k_i}$ and write $\Psi_j(t) := t\Phi'_j(t)/\Phi_j(t)$ and

$Z_j(m,n) := Z(m,n;\mathbf{w}^{(k_{j-1})})$. Note that (19.136) applies to each Ψ_j too, and thus

$$\Psi_j(1) < 0.2. \tag{19.138}$$

Take $n = 3k_j$ (where k_j is not yet determined). A good allocation with $n-1$ balls has at most 2 boxes with k_j balls, and for the remaining boxes the weights $\mathbf{w}$ and $\mathbf{w}^{(k_{j-1})}$ coincide. We thus obtain

$$\begin{aligned} Z(3k_j-1,3k_j) = Z_j(3k_j-1,3k_j) &+ 3k_j w_{k_j} Z_j(2k_j-1,3k_j-1) \\ &+ \binom{3k_j}{2} w_{k_j}^2 Z_j(k_j-1,3k_j-2). \end{aligned} \tag{19.139}$$

Let the three terms on the right-hand side be A_0, A_1, A_2, where A_i corresponds to the case when i boxes have k_j balls. The generating function Φ_j is a polynomial, with radius of convergence $\rho_j = \infty$ and, by Lemma 3.1, $\nu_j := \Psi_j(\infty) = \omega(\mathbf{w}^{(k_{j-1})}) = k_{j-1} \geqslant 2$. Define τ, τ' and τ'' by $\Psi_j(\tau_j) = 1$, $\Psi_j(\tau'_j) = 2/3$, $\Psi_j(\tau''_j) = 1/3$. Since $\Psi_j(1) < 1/3$ by (19.138), we have $1 < \tau''_j < \tau'_j < \tau_j < \infty$.

Theorem 18.1 applies to each term A_i in (19.139), with $\lambda = 1, \frac{2}{3}, \frac{1}{3}$, respectively; hence, as $k_j \to \infty$,

$$\log A_0 = 3k_j \log \frac{\Phi_j(\tau_j)}{\tau_j} + o(k_j), \tag{19.140}$$

$$\log A_1 = 3k_j \log \frac{\Phi_j(\tau'_j)}{(\tau'_j)^{2/3}} + o(k_j), \tag{19.141}$$

$$\log A_2 = 3k_j \log \frac{\Phi_j(\tau''_j)}{(\tau''_j)^{1/3}} + o(k_j). \tag{19.142}$$

By (11.16) and $\tau''_j > 1$,

$$\frac{\Phi_j(\tau_j)}{\tau_j} \leqslant \frac{\Phi_j(\tau''_j)}{\tau''_j} < \frac{\Phi_j(\tau''_j)}{(\tau''_j)^{1/3}}$$

and

$$\frac{\Phi_j(\tau'_j)}{(\tau'_j)^{2/3}} \leqslant \frac{\Phi_j(\tau''_j)}{(\tau''_j)^{2/3}} < \frac{\Phi_j(\tau''_j)}{(\tau''_j)^{1/3}}.$$

Hence, the constant multiplying k_j is larger in (19.142) than in (19.140) and (19.141), so by choosing k_j large enough, we obtain $A_2 > jA_1$ and $A_2 > jA_0$, and thus

$$\mathbb{P}(B_{3k_j-1,3k_j} \text{ has 2 boxes with } k_j \text{ balls}) = \frac{A_2}{A_0+A_1+A_2} > 1 - \frac{2}{j}. \tag{19.143}$$

This constructs recursively the sequence (k_j) and thus S and $\mathbf{w}$, and (19.143) shows that for n in the subsequence $(3k_j)_j$, $B_{n-1,n}$ w.h.p. has 2 boxes with $n/3$ balls each.

By Lemma 17.1, it follows that, for this subsequence, $\mathcal{T}_n$ w.h.p. has 2 nodes with outdegrees $n/3$.

To summarise, we have found a weight sequence with $0 < \nu < 1$ such that, with $m = n - 1$, for one subsequence

$$Y_{(1)}/n \to 0 \tag{19.144}$$

and for another subsequence w.h.p.

$$Y_{(1)} = Y_{(2)} = n/3. \tag{19.145}$$

Hence, neither (19.106) nor (19.107) holds. (It is easy to modify the construction such that for every $\ell \geqslant 1$, there is a subsequence with $Y_{(1)} = \cdots = Y_{(\ell)} = n/(\ell + 1)$.)

Example 19.38. Let $\mathsf{S} := \{0\} \cup \{2^i : i \geqslant 0\}$. We will construct a weight sequence $\mathbf{w}$ recursively with support $\operatorname{supp}(\mathbf{w}) = \mathsf{S}$ and $\rho = 0$. Let $w_0 = 1$.

Let $i \geqslant 0$. If $w_0, \dots, w_{2^{i-1}}$ are fixed and we let $w_{2^i} \to \infty$, then for every m with $2^i \leqslant m < 2^{i+1}$ and every n,

$$\mathbb{P}(B_{m,n} \text{ contains a box with } 2^i \text{ balls}) \to 1. \tag{19.146}$$

Hence, we can recursively choose w_{2^i} so large that, for every $i \geqslant 0$, if $2^i \leqslant m < 2^{i+1}$ and $2^i \leqslant n \leqslant 2^{2i}$, then, by (11.3),

$$\mathbb{P}(B_{m,n} \text{ contains a box with } 2^i \text{ balls}) > 1 - i^{-1}. \tag{19.147}$$

We further take $w_{2^i} \geqslant (2^i)!$; thus $\rho = 0$ and $\nu = 0$.

Consider the tree case, $m = n - 1$. Thus $\lambda = 1$. If $2^i < n \leqslant 2^{i+1}$, then (19.147) applies and shows that $B_{n-1,n}$ w.h.p. contains a box with 2^i balls, so w.h.p.

$$Y_{(1)} = 2^{\lfloor \log_2(n-1) \rfloor} = 2^{\lceil \log_2 n \rceil - 1}. \tag{19.148}$$

Hence, $Y_{(1)}/n$ w.h.p. is a (non-random) value that oscillates between $\frac{1}{2}$ and 1, depending on the fractional part $\{\log_2 n\}$ of $\log_2 n$. Consequently, (19.106) holds for subsequences such that $0 \neq \{\log_2 n\} \to 0$, but not in general.

Moreover, conditioned on the existence of a box with 2^i balls, the remainder of the allocation is a random allocation $B_{m-2^i,n-1}$ of the remaining $m - 2^i$ balls in $n - 1$ boxes. For example, if $n = 2^{i+1}$, so $m = 2^{i+1} - 1$, we have $m - 2^i = 2^i - 1$, and we can apply (19.147) again (with $i - 1$) to see that w.h.p. $Y_{(2)} = 2^{i-1} = n/4$. Continuing in the same way we see that for n in the subsequence (2^i), we have, for each fixed j, w.h.p.

$$Y_{(j)} = 2^{-j} n. \tag{19.149}$$

Hence neither (19.106) nor (19.107) holds in this case.

Similar results follow easily for other subsequences. For example, for n in the subsequence $(\lfloor r 2^i \rfloor)_{i \geqslant 1}$, where $\frac{1}{2} < r < 1$ and r has the infinite binary expansion $r = 2^{-\ell_1} + 2^{-\ell_2} + \dots$, with $1 = \ell_1 < \ell_2 < \dots$, we have w.h.p. $Y_{(j)} = 2^{-\ell_j} \lceil n/r \rceil$ for each fixed j.

Example 19.39. Let again $m = n - 1$, so $\lambda = 1$. Taking $w_k = k!$ for $k \in \operatorname{supp}(\mathbf{w}) = \{0\} \cup \{i! : i \geqslant 0\}$, we obtain an example with $\rho = 0$ and thus

$\nu = 0$ such that $Y_{(1)}/n \to 0$ for some subsequences, for example for $n = i!$ (since then $Y_{(1)} \leqslant (i-1)!$).

Problem 19.40. *Is $Y_{(1)}/n \xrightarrow{\mathrm{p}} 0$ possible when $0 < \nu < \lambda$? Example 19.37 shows that this is possible for a subsequence, but we conjecture that it is not possible for the full sequence, and, a little stronger, that there always is some $\varepsilon > 0$ and some subsequence along which $Y_{(1)} \geqslant \varepsilon n$ w.h.p.*

Problem 19.41. *Is $Y_{(1)}/n \xrightarrow{\mathrm{p}} 0$ possible when $\lambda > \nu = 0$? (Example 19.39 shows that this is possible for a subsequence.)*

We expect that bad behaviour as in the examples above only can occur for quite irregular weight sequences, but we have no general result beyond Theorem 19.34. We formulate two natural problems.

Problem 19.42. *Suppose that $w_k \geqslant w_{k+1}$ for all (large) k. Does this imply that* (19.106)–(19.107) *hold when $\lambda > \nu$?*

Problem 19.43. *Suppose that $w_{k+1}/w_k \to \infty$ as $k \to \infty$. (Hence, $\rho = 0$ and $\nu = 0$.) Does this imply that* (19.106)–(19.107) *hold when $\lambda > \nu$?*

19.7. **Applications to random forests.** We give some applications of the results above to the size of the largest tree(s) in different types of random forests witn n trees and $m \geqslant n$ nodes. We consider only the case $m/n \to \lambda$ with $1 < \lambda < \infty$; for simplicity we further assume that $m = \lambda n + O(1)$, although this can be relaxed and, moreover, the general case $m/n \to \lambda$ can be handled by using $\lambda_n := m/n$ and the corresponding $\tau_n := \tau(\lambda_n)$ as in Theorem 11.6; for details and for results in the cases $m = n + o(n)$ and $m/n \to \infty$, see Pavlov [94, 95, 96, 97], Kolchin [76], Łuczak and Pittel [83], Kazimirov and Pavlov [72] and Bernikovich and Pavlov [12].

The random forests considered here are described by balls-in-boxes with weight sequences with $w_0 = 0$ and $w_1 > 0$, see Section 12. As usual, we use (without further comments) the argument in Remark 11.8 to extend theorems above to the case $w_0 = 0$. (See Remark 19.21.)

We first consider random rooted forests as in Example 12.6. We have

$$w_k = \frac{k^{k-1}}{k!} \sim \frac{1}{\sqrt{2\pi}} k^{-3/2} e^k, \qquad \text{as } k \to \infty, \tag{19.150}$$

and thus $w_{k+1}/w_k \to e$ as $k \to \infty$. (Alternatively, we may use $\widetilde{w}_k := e^{-k} w_k \sim (2\pi)^{-1/2} k^{-3/2}$, see (12.30)–(12.31) and Example 12.10.) Since $\nu = \infty$, see Examples 12.6 and 12.10, $\lambda < \nu$ and Theorem 19.19 applies for any $\lambda \in (1, \infty)$.

We have $a = e$ and thus, by (12.28),

$$q := \tau e = \frac{\lambda - 1}{\lambda} e^{1/\lambda} \in (0, 1) \tag{19.151}$$

and, consequently,

$$\log(1/q) = -\log q = -\log\Bigl(1 - \frac{1}{\lambda}\Bigr) - \frac{1}{\lambda} > 0. \tag{19.152}$$

As $k \to \infty$, by (12.22), (12.27) and (19.150),

$$\pi_k = \frac{w_k \tau^k}{\Phi(\tau)} = \frac{\lambda}{\lambda - 1} w_k \tau^k \sim (2\pi)^{-1/2} \frac{\lambda}{\lambda - 1} k^{-3/2} q^k. \tag{19.153}$$

It follows that $\pi_{k(n)} = \Theta(1/n)$ for

$$k(n) = \frac{\log n - \frac{3}{2}\log\log n}{\log(1/q)} + O(1), \tag{19.154}$$

and then (19.69) yields

$$N \sim n \frac{\lambda}{\sqrt{2\pi}(\lambda - 1)(1 - q)} k(n)^{-3/2} \sim \frac{\lambda \log^{3/2}(1/q)}{\sqrt{2\pi}(\lambda - 1)(1 - q)} n \log^{-3/2} n. \tag{19.155}$$

Consequently, Theorem 19.19(ii) yields the following theorem for the maximal tree size $Y_{(1)}$; this is due to Pavlov [94, 96] (in a slightly different formulation), who also gives further results. We further use Theorem 19.16(i) to give a simple estimate for the size $Y_{(j)}$ of the j:th largest tree. (More precise limit results for $Y_{(j)}$ are also easily obtained from (19.50).)

Theorem 19.44. *For a random rooted forest, with $m = \lambda n + O(1)$ where $1 < \lambda < \infty$,*

$$Y_{(1)} \overset{\mathrm{d}}{\approx} \left\lfloor \frac{\log n - \frac{3}{2}\log\log n + \log b + W}{\log(1/q)} \right\rfloor, \tag{19.156}$$

where W has the Gumbel distribution (19.63) *and*

$$b := \frac{\lambda \log^{3/2}(1/q)}{\sqrt{2\pi}(\lambda - 1)(1 - q)} \tag{19.157}$$

with q given by (19.151)–(19.152)*.*

Furthermore, $Y_{(j)} = Y_{(1)} + O_{\mathrm{p}}(1)$ for each fixed j. □

Next, let us, more generally, consider a random simply generated forest as in Example 12.8, defined by a weight sequence $\mathbf{w}$. Then the tree sizes in the random forest are distributed as balls-in-boxes with the weight sequence $(Z_k)_{k=0}^{\infty}$, where Z_k is the partition function (2.5) for simply generated trees with weight sequence $\mathbf{w}$ (and $Z_0 = 0$).

We assume that $\nu(\mathbf{w}) \geqslant 1$; thus there exists $\tau_1 > 0$ such that $\Psi(\tau_1) = 1$, and then $\mathbf{w}' := (\tau_1^k w_k / \Phi(\tau_1))_k$ is an equivalent probability weight sequence with expectation 1, see Lemma 4.2. (τ_1 is the same as τ in Theorem 7.1, but here we need to consider several different τ's so we modify the notation.) This probability weight sequence $\mathbf{w}'$ defines the same random forest, which thus can be realized as a conditioned critical Galton–Watson forest. Recall from (4.10) and Theorem 7.1 that the probability distribution $\mathbf{w}'$ has variance $\sigma^2 = \tau_1 \Psi'(\tau_1)$; we assume that σ^2 is finite, which always holds if $\nu(\mathbf{w}) > 1$ and thus $\tau_1 < \rho(\mathbf{w})$. We further assume, for simplicity, that $\mathbf{w}$ has span 1. We then have the following generalization of Theorem 19.44, see Pavlov [95, 96], where also further results are given.

Theorem 19.45. *Consider a simply generated random forest defined by a weight sequence* $\mathbf{w}$*, and assume that* $m = \lambda n + O(1)$ *where* $1 < \lambda < \infty$*. Suppose that* $\nu(\mathbf{w}) \geqslant 1$ *and* $\operatorname{span}(\mathbf{w}) = 1$*. Define* $\tau_1 > 0$ *by* $\Psi(\tau_1) = 1$*, and assume that* $\sigma^2 := \tau_1 \Psi'(\tau_1) < \infty$ *(this is automatic if* $\nu(\mathbf{w}) > 1$*). Define further* $\tau_2 > 0$ *by*

$$\Psi(\tau_2) = 1 - 1/\lambda \tag{19.158}$$

and let

$$q := \frac{\tau_2}{\Phi(\tau_2)} \cdot \frac{\Phi(\tau_1)}{\tau_1}. \tag{19.159}$$

Then $0 < q < 1$ *and*

$$Y_{(1)} \overset{\mathrm{d}}{\approx} \left\lfloor \frac{\log n - \frac{3}{2}\log\log n + \log b + W}{\log(1/q)} \right\rfloor, \tag{19.160}$$

where W *has the Gumbel distribution* (19.63) *and*

$$b := \frac{\tau_1 \log^{3/2}(1/q)}{\tau_2 \sqrt{2\pi\sigma^2}(1-q)}. \tag{19.161}$$

Furthermore, $Y_{(j)} = Y_{(1)} + O_{\mathrm{p}}(1)$ *for each fixed* j*.*

Proof. Replace $\mathbf{w}$ by the equivalent probability weight sequence $\widetilde{\mathbf{w}} = (\widetilde{w}_k)$ with $\widetilde{w}_k := \tau_2^k w_k / \Phi(\tau_2)$. This probability weight sequence has expectation $\Psi(\tau_2) < 1$ by (4.9), and using it we realize the random forest as a conditioned subcritical Galton–Watson forest. The partition function $\widetilde{Z}_k$ for $\widetilde{\mathbf{w}}$ is by (4.3) and Theorem 18.11,

$$\widetilde{Z}_k = \frac{\tau_2^{k-1}}{\Phi(\tau_2)^k} Z_k \sim \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\tau_2^{k-1}}{\Phi(\tau_2)^k} \cdot \frac{\Phi(\tau_1)^k}{\tau_1^{k-1}} k^{-3/2}. \tag{19.162}$$

Moreover, by (2.6), $(\widetilde{Z}_k)$ is the distribution of the size of a Galton–Watson process with offspring distribution $\widetilde{\mathbf{w}}$. Since this offspring distribution is subcritical with expectation $\Psi(\tau_2) < 1$, the size distribution $(\widetilde{Z}_k)$ has finite mean

$$\sum_{k=0}^{\infty} k \widetilde{Z}_k = \frac{1}{1 - \Psi(\tau_2)} = \lambda, \tag{19.163}$$

by our choice of τ_2.

The sizes of the trees in the random forest are distributed as balls-in-boxes with the weight sequence $(\widetilde{Z}_k)$, see Example 12.8. We apply Theorem 19.19, translating w_k to $\widetilde{Z}_k$. By (19.162),

$$\widetilde{Z}_{k+1}/\widetilde{Z}_k \to a := \frac{\tau_2}{\Phi(\tau_2)} \cdot \frac{\Phi(\tau_1)}{\tau_1}, \qquad \text{as } k \to \infty. \tag{19.164}$$

Note further that (with this weight sequence $(\widetilde{Z}_k)$) τ in Theorem 19.19 is chosen such that the equivalent probability weight sequence $\bigl(\tau^k \widetilde{Z}_k / \widetilde{\mathcal{Z}}(\tau)\bigr)$ has expectation λ. We have already constructed $(\widetilde{Z}_k)$ such that it is a probability weight sequence with this expectation, see (19.163); hence we have $\tau = 1$ and $q = a$, which yields (19.159).

As in (19.154), $\pi_{k(n)} = \widetilde{Z}_{k(n)} = \Theta(1/n)$ for

$$k(n) = \frac{\log n - \frac{3}{2}\log\log n}{\log(1/q)} + O(1), \tag{19.165}$$

and then (19.69) yields, by (19.162),

$$N \sim n\frac{\tau_1}{\sqrt{2\pi\sigma^2}\tau_2(1-q)}k(n)^{-3/2} \sim \frac{\tau_1 \log^{3/2}(1/q)}{\tau_2\sqrt{2\pi\sigma^2}(1-q)} n \log^{-3/2} n. \tag{19.166}$$

The result (19.160) now follows from Theorem 19.19(ii). Finally, again, Theorem 19.16(i) gives the estimate for $Y_{(j)}$. □

Example 19.46. Consider a random ordered rooted forest. This is obtained by the weight sequence $w_k = 1$, see Example 12.8, and we have by (10.1)–(10.2) $\Phi(t) = 1/(1-t)$ and $\Psi(t) = t/(1-t)$. Hence, $\tau_1 = 1/2$ and $\sigma^2 = 2$ (see Example 10.1); furthermore, (19.158) is $\tau_2/(1-\tau_2) = 1 - 1/\lambda$, which has the solution

$$\tau_2 = \frac{\lambda - 1}{2\lambda - 1}. \tag{19.167}$$

Consequently, Theorem 19.45 says that (19.160) holds, with the parameters q and b given by, see (19.159) and (19.161),

$$q = \frac{\tau_2(1-\tau_2)}{\tau_1(1-\tau_1)} = 4\tau_2(1-\tau_2) = \frac{4\lambda(\lambda-1)}{(2\lambda-1)^2} = 1 - \frac{1}{(2\lambda-1)^2} \tag{19.168}$$

and

$$b = \frac{(2\lambda-1)^3}{4\sqrt{\pi}(\lambda-1)}\log^{3/2}(1/q). \tag{19.169}$$

Example 19.47. The random rooted unlabelled forest in Example 12.11 is described by a weight sequence that also satisfies $w_k \sim c_1 k^{-3/2}\rho^{-k}$ as $k \to \infty$, and we thus again obtain (19.160), although the parameters q and b now are implicitly defined using the generating function of the number of unlabelled rooted trees, see Pavlov [97].

Example 19.48. For the random recursive forest in Example 12.13, we have

$$w_k = k^{-1}, \qquad k \geqslant 1. \tag{19.170}$$

Thus Theorem 19.19 applies with $a = 1$ and $q = \tau \in (0,1)$ given by

$$\frac{q}{(1-q)|\log(1-q)|} = \lambda, \tag{19.171}$$

see (12.56). (Recall that $\nu = \infty$, so we can take any $\lambda > 1$ here.) In this case, see (12.55), $\pi_{k(n)} = k(n)^{-1}q^{k(n)}/|\log(1-q)| = \Theta(1/n)$ for

$$k(n) = \frac{\log n - \log\log n}{\log(1/q)} + O(1), \tag{19.172}$$

cf. (19.154), and then (19.69) yields

$$N \sim \frac{\log(1/q)}{(1-q)|\log(1-q)|} n \log^{-1} n. \tag{19.173}$$

Consequently, Theorem 19.19(ii) yields

$$Y_{(1)} \overset{\mathrm{d}}{\approx} \left\lfloor \frac{\log n - \log\log n + \log b + W}{\log(1/q)} \right\rfloor, \tag{19.174}$$

where W has the Gumbel distribution (19.63) and, using (19.171),

$$b := \frac{\log(1/q)}{(1-q)|\log(1-q)|} = \frac{\lambda \log(1/q)}{q}. \tag{19.175}$$

We thus obtain a result similar to the cases above, but with a different coefficient for $\log\log n$ in (19.174). See Pavlov and Loseva [98] for further results.

If we consider the random unrooted forest in Example 12.7, we find different results. In this case, the tree sizes are described by balls-in-boxes with the weight sequence $w_k = k^{k-2}/k!$, $k \geqslant 1$ (and $w_0 = 0$). Alternatively, we can use the probability weight sequence (12.40)

$$\widetilde{w}_k := 2w_k e^{-k} = \frac{2k^{k-2}e^{-k}}{k!}, \tag{19.176}$$

which by Stirling's formula satisfies (12.41)

$$\widetilde{w}_k \sim \frac{2}{\sqrt{2\pi}} k^{-5/2}, \qquad \text{as } k \to \infty. \tag{19.177}$$

Since we now have $\nu = 2 < \infty$, see Examples 12.7 and 12.10, there is a phase transition at $\lambda = 2$. We show in the theorem below that for $\lambda < 2$ we have a result similar to Theorems 19.44 and 19.45 with maximal tree size $Y_{(1)} = O_{\mathrm{p}}(\log n)$, but for $\lambda > 2$ there is a unique giant tree with size of order n. At the phase transition, with $m/n \to 2$, the result depends on the rate of convergence of m/n; if, for example, $m = 2n$ exactly, the maximal size is of order $n^{2/3}$; see further Łuczak and Pittel [83], where precise results for general $m = m(n)$ are given. (By the proof below, (iii) in the following theorem holds as soon as $m/n \to \lambda > 2$, but (i) and (ii) are more sensitive.)

Theorem 19.49. *Consider a random unrooted forest, and assume that $m = \lambda n + O(1)$ where $1 < \lambda < \infty$.*

(i) *If $1 < \lambda < 2$, let*

$$q := 2\frac{\lambda - 1}{\lambda} e^{2/\lambda - 1}. \tag{19.178}$$

Then $0 < q < 1$ and

$$Y_{(1)} \overset{\mathrm{d}}{\approx} \left\lfloor \frac{\log n - \frac{5}{2}\log\log n + \log b + W}{\log(1/q)} \right\rfloor, \tag{19.179}$$

where W has the Gumbel distribution (19.63) *and*

$$b := \frac{\lambda^2 \log^{5/2}(1/q)}{2\sqrt{2\pi}(\lambda-1)(1-q)} \tag{19.180}$$

Furthermore, $Y_{(j)} = Y_{(1)} + O_{\mathrm{p}}(1)$ for each fixed j.

(ii) *If $\lambda = 2$, then*

$$Y_{(j)}/n^{2/3} \xrightarrow{\mathrm{d}} \eta_j \tag{19.181}$$

for each j, where $\eta_j > 0$ are some random variables. The distribution of η_1 is given by (19.97) *with $\alpha = 3/2$ and $c = (2/\pi)^{1/2}$.*

(iii) *If $2 < \lambda < \infty$, then $Y_{(1)} = (\lambda-2)n + O_{\mathrm{p}}(n^{2/3})$. More precisely,*

$$n^{-2/3}\bigl(m - 2n - Y_{(1)}\bigr) \xrightarrow{\mathrm{d}} X, \tag{19.182}$$

where X is a $\frac{3}{2}$-stable random variable with Laplace transform

$$\mathbb{E}\, e^{-tX} = \exp\Bigl(\frac{2^{5/2}}{3} t^{3/2}\Bigr), \qquad \operatorname{Re} t \geqslant 0. \tag{19.183}$$

For $j \geqslant 2$, $Y_{(j)} = O_{\mathrm{p}}(n^{2/3})$, and $n^{-2/3}Y_{(j)} \xrightarrow{\mathrm{d}} W_j$ where W_2 has the Fréchet distribution

$$\mathbb{P}(W_2 \leqslant x) = \exp\Bigl(-\frac{2^{3/2}}{3\sqrt{\pi}} x^{-3/2}\Bigr), \qquad x \geqslant 0. \tag{19.184}$$

and, more generally, W_j has the density function (19.117) *with $c' = (2/\pi)^{1/2}$ and $\alpha = 3/2$.*

Note that the exponents $\frac{3}{2}$, 1 and $\frac{5}{2}$ in (19.150), (19.170) and (19.177) appear as coefficients of $\log\log n$ in (19.156), (19.174) and (19.179), respectively.

Proof. (i): This is very similar to the proofs of Theorems 19.44 and 19.45. We use $w_k = k^{k-2}/k!$. Then, as for rooted forests and (19.150) above, $w_{k+1}/w_k \to e$ as $k \to \infty$. Further, τ is given by (12.38), and thus $q := \tau e$ is given by (19.178). It follows, cf. (19.154) and (19.177), that $\pi_{k(n)} = \Theta(1/n)$ for

$$k(n) = \frac{\log n - \frac{5}{2}\log\log n}{\log(1/q)} + O(1), \tag{19.185}$$

and then (19.69) yields

$$\begin{aligned} N &\sim \frac{n w_{k(n)} e^{-k(n)}}{\Phi(\tau)(1-q)} \sim n \frac{\lambda^2}{2\sqrt{2\pi}(\lambda-1)(1-q)} k(n)^{-5/2} \\ &\sim \frac{\lambda^2 \log^{5/2}(1/q)}{2\sqrt{2\pi}(\lambda-1)(1-q)} n \log^{-5/2} n. \end{aligned} \tag{19.186}$$

Hence Theorem 19.19(ii) yields (19.179).

(ii): We use the equivalent probability weight sequence $(\widetilde{w}_k)$ given by (19.176). By (19.177), it satisfies the assumptions in Example 19.27 with

$\alpha = 3/2$ and $c = (2/\pi)^{1/2}$; thus (19.181) follows from (19.91), and (19.97) in Remark 19.28 applies.

(iii): We use again the probability weight sequence $(\widetilde{w}_k)$ and apply Theorem 19.34. We have $c' = c = (2/\pi)^{1/2}$ by (19.176), and thus $c'\Gamma(-3/2) = c'\frac{4}{3}\Gamma(1/2) = 2^{5/2}/3$ and $c'/\alpha = 2^{3/2}/(3\sqrt{\pi})$. □

Example 19.50. The random unrooted unlabelled forest (with labelled trees) in Example 12.11 is described by another weight sequence that satisfies $w_k \sim ck^{-5/2}\rho^{-k}$ as $k \to \infty$, and we thus obtain a result similar to Theorem 19.49, although the parameters differ (they can be obtained from the generating function of the number of unlabelled trees); in particular, the phase transition appears when λ is $\nu \approx 2.0513$, see Bernikovich and Pavlov [12] for details.

We do not know any corresponding results for random completely unlabelled forests (n unlabelled trees consisting of m unlabelled nodes); as said in Example 12.11, they cannot be described by balls-in-boxes.

20. Large nodes in simply generated trees with $\nu < 1$

In the tree case with $\nu < 1$, the results in Section 19.6 show condensation in the form of one or, sometimes, several nodes with very large degree, together making up the "missing mass" of about $(1-\nu)n$. On the other hand, Theorem 7.1 shows concentration in a somewhat different form, with a limit tree $\widehat{\mathcal{T}}$ having exactly one node of infinite degree. This node corresponds to a node with very large degree in $\mathcal{T}_n$ for n large but finite. How large is the degree? Why do we only see one node with very large degree in Theorem 7.1, but sometimes several nodes with large degrees above (Examples 19.37 and 19.38)?

The latter question is easily answered: recall that the convergence in Theorem 7.1 means convergence of the truncated trees ("left balls") $T_n^{[m]}$, see Lemma 6.3; thus we only see a small part of the tree close to the root, and the two pictures above are reconciled: if m is large but fixed, then in the set $V(T) \cap V^{[m]}$ of nodes, there is with probability close to 1 exactly one node with very large degree. (There may be several nodes with very large degree in the tree, but for any fixed m, w.h.p. at most one of them is in $V^{[m]}$.) Of course, to make this precise, we would have to define "very large", for example as below using a sequence Ω_n growing slowly to ∞ as in Lemma 19.32, but we are at the moment satisfied with an intuitive description.

To see how large the "very large" degree is, let us first look at the root. Lemma 15.7 says that the distribution of the root degree is the size-biased distribution of Y_1. We can write (15.7) as

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = d) = \frac{d}{n-1}\sum_{i=1}^{n}\mathbb{P}(Y_i = d) = \frac{d}{n-1}\sum_{j=1}^{n}\mathbb{P}(Y_{(j)} = d); \qquad (20.1)$$

hence the distribution of the root degree can be described by: sample $(Y_1, \dots, Y_n)$ and then take $Y_{(1)}$ with probability $Y_{(1)}/(n-1)$, $Y_{(2)}$ with probability $Y_{(2)}/(n-1)$, ...

In particular, if $Y_{(1)} = (1-\nu)n + o_{\mathrm{p}}(n)$, then (20.1) implies

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = Y_{(1)}) = 1 - \nu + o(1), \tag{20.2}$$

and comparing with Theorem 7.10 we see that w.h.p. either the root degree is small (more precisely, $O_{\mathrm{p}}(1)$), or it is the maximum outdegree $Y_{(1)}$. However, we also see that if $Y_{(1)}$ is not $(1-\nu)n + o_{\mathrm{p}}(n)$, then this conclusion does not hold; for example, in Example 19.38 for n in the subsequence (2^i) where (19.149) holds for each fixed j,

$$\mathbb{P}(d^+_{\mathcal{T}_n}(o) = 2^{-j}n) \to 2^{-j}. \tag{20.3}$$

In the case $\nu = 0$, we only have to consider the root, since the node with infinite degree in $\widehat{\mathcal{T}}$ always is the root, but for $0 < \nu < 1$, the node with infinite degree in $\widehat{\mathcal{T}}$ may be somewhere else. We shall see that it corresponds to a node in $\mathcal{T}_n$ with a large degree having (asymptotically) the same distribution as the root degree just considered, conditioned to be "large".

To make this precise, let $\Omega_n \to \infty$ be a fixed sequence which increases so slowly that Lemma 19.32(ii) holds. We say that an outdegree $d^+(v)$ is *large* if it is greater than Ω_n; we then also say that the node v is large. (Note that by Lemma 19.32(ii), w.h.p. at least one large node exists.) For each n, let $\widetilde{D}_n$ by a random variable whose distribution is the size-biased distribution of a large outdegree, i.e. of $(Y_1 \mid Y_1 > \Omega_n)$:

$$\mathbb{P}(\widetilde{D}_n = k) = \frac{k\,\mathbb{P}(Y_1 = k)}{\sum_{l>\Omega_n} l\,\mathbb{P}(Y_1 = l)} = \frac{k\,\mathbb{E}\,N_k}{\sum_{l>\Omega_n} l\,\mathbb{E}\,N_l} = \frac{k\,\mathbb{E}\,N_k}{(1-\nu+o(1))n}, \tag{20.4}$$

for $k > \Omega_n$ and $\mathbb{P}(\widetilde{D}_n = k) = 0$ otherwise. Equivalently, in view of Lemma 15.7, $\widetilde{D}_n$ has the distribution of the root degree $d^+_{\mathcal{T}_n}(o)$ conditioned to be greater than Ω_n. See also (20.1), and note that if $Y_{(1)} = (1-\nu)n + o_{\mathrm{p}}(n)$, then $\widetilde{D}_n \overset{\mathrm{d}}{\approx} Y_{(1)}$, i.e., we may take $\widetilde{D}_n = Y_{(1)}$ w.h.p.; in this case (but not otherwise) we thus have $\widetilde{D}_n = (1-\nu)n + o_{\mathrm{p}}(n)$.

Note that if Ω'_n is another such sequence, similarly defining a random variable $\widetilde{D}'_n$, then $\sum_{l>\Omega_n} l\,\mathbb{P}(Y_1 = l) \sim (1-\nu)n \sim \sum_{l>\Omega'_n} l\,\mathbb{P}(Y_1 = l)$, and it follows that $\widetilde{D}_n \overset{\mathrm{d}}{\approx} \widetilde{D}'_n$; hence the choice of Ω_n will not matter below.

We claim that, w.h.p., the infinite outdegree in $\widehat{\mathcal{T}}$ corresponds to an outdegree $\widetilde{D}_n$ in $\mathcal{T}_n$. To formalise this, recall from Section 6 that we may consider our trees as subtrees of the infinite tree U_∞ with node set V_∞, and that the convergence of trees defined there means convergence of each $d^+(v)$, see (6.6). Let $\widehat{\mathcal{T}}$ be the random infinite tree defined in Section 5; we are in case (T2), and thus $\widehat{\mathcal{T}}$ has a single node v with outdegree $d^+_{\widehat{\mathcal{T}}}(v) = \infty$. We assume that $\widetilde{D}_n$ and $\widehat{\mathcal{T}}$ are independent, and define the modified degree sequence

$$\widetilde{d}^+_{\widehat{\mathcal{T}}}(v) := \begin{cases} d^+_{\widehat{\mathcal{T}}}(v), & d^+_{\widehat{\mathcal{T}}}(v) < \infty, \\ \widetilde{D}_n, & d^+_{\widehat{\mathcal{T}}}(v) = \infty. \end{cases} \tag{20.5}$$

We thus change the single infinite value to the finite $\widetilde{D}_n$, leaving all other values unchanged. (Note that $\widetilde{d}^+_{\widehat{\mathcal{T}}}(v)$ may depend on n, since $\widetilde{D}_n$ does.) We then have the following theorem.

Theorem 20.1. *For any finite set of nodes $v_1, \dots, v_\ell \in V_\infty$,*

$$\big(d^+_{\mathcal{T}_n}(v_1), \dots, d^+_{\mathcal{T}_n}(v_\ell)\big) \overset{\mathrm{d}}{\approx} \big(\widetilde{d}^+_{\widehat{\mathcal{T}}}(v_1), \dots, \widetilde{d}^+_{\widehat{\mathcal{T}}}(v_\ell)\big). \tag{20.6}$$

Proof. Let $\varepsilon > 0$, and let v^* denote the unique node in $\widehat{\mathcal{T}}$ with $d^+_{\widehat{\mathcal{T}}}(v^*) = \infty$. By increasing the set $\{v_1, \dots, v_\ell\}$, we may assume that it equals $V^{[m]}$ (see Section 6) for some m, and that m is so large that $\mathbb{P}(v^* \in V^{[m]}) > 1 - \varepsilon$. We may then find $K < \infty$ such that

$$\mathbb{P}\big(d^+_{\widehat{\mathcal{T}}}(v) \in (K, \infty) \text{ for some } v \in V^{[m]}\big) < \varepsilon.$$

Since $\mathcal{T}_n \overset{\mathrm{d}}{\longrightarrow} \widehat{\mathcal{T}}$ by Theorem 7.1, we may by the Skorohod coupling theorem [69, Theorem 4.30] assume that the random trees are coupled such that $\mathcal{T}_n \to \widehat{\mathcal{T}}$ a.s., and thus $d^+_{\mathcal{T}_n}(v) \to d^+_{\widehat{\mathcal{T}}}(v)$ a.s. for every v. Then, for large n, with probability $> 1 - 3\varepsilon$, $v^* \in V^{[m]}$, $d^+_{\mathcal{T}_n}(v) = d^+_{\widehat{\mathcal{T}}}(v) = \widetilde{d}^+_{\widehat{\mathcal{T}}}(v) \leqslant K$ for all $v \in V^{[m]} \setminus \{v^*\}$, and $d^+_{\mathcal{T}_n}(v^*) \to d^+_{\widehat{\mathcal{T}}}(v^*) = \infty$. We may assume that $\Omega_n \to \infty$ so slowly that furthermore $\mathbb{P}(d^+_{\mathcal{T}_n}(v^*) \leqslant \Omega_n) \leqslant \varepsilon$. (Recall that we may change Ω_n without affecting the result (20.6).)

Let n be so large that also $\Omega_n > m$ and $\Omega_n > K$. It follows from Lemma 15.9 that for each choice of $v' \in V^{[m]}$ and numbers $d(v)$ for $v \in V^{[m]} \setminus v'$, and $k > \Omega_n$,

$$\begin{aligned}\mathbb{P}\big(d^+_{\mathcal{T}_n}(v) = d(v) \text{ for } v \in V^{[m]} \setminus \{v'\} \text{ and } d^+_{\mathcal{T}_n}(v') = k\big)& \\ = \big(k + O(1)\big) C(\{d(v)\}, v', n)\, \mathbb{P}(Y_1 = k)&\end{aligned}$$

for some constant $C(\{d(v)\}, v', n) \geqslant 0$ not depending on k; hence, by (20.4),

$$\begin{aligned}\mathbb{P}\big(d^+_{\mathcal{T}_n}(v') = k \mid d^+_{\mathcal{T}_n}(v) = d(v) \text{ for } v \in V^{[m]} \setminus \{v'\} \text{ and } d^+_{\mathcal{T}_n}(v') > \Omega_n\big)& \\ = \big(1 + o(1)\big) \frac{k\, \mathbb{P}(Y_1 = k)}{\sum_{k > \Omega_n} k\, \mathbb{P}(Y_1 = k)} = \big(1 + o(1)\big)\, \mathbb{P}(\widetilde{D}_n = k).&\end{aligned}$$

There is only a finite number of choices of v' and $(d(v))_{v \in V^{[m]} \setminus \{v'\}}$, and it follows that we may choose the coupling of $\mathcal{T}_n$ and $\widehat{\mathcal{T}}$ above such that also $d^+_{\mathcal{T}_n}(v^*) = \widetilde{D}_n$ w.h.p.; thus, with probability $> 1 - 4\varepsilon + o(1)$, $d^+_{\mathcal{T}_n}(v) = \widetilde{d}^+_{\widehat{\mathcal{T}}}(v)$ for all $v \in V^{[m]}$.

The result follows since $\varepsilon > 0$ is arbitrary. $\square$

We give some variations of this result, where we replace $\widetilde{d}^+_{\widehat{\mathcal{T}}}(v)$ by the degree sequences of some random trees obtained by modifying $\widehat{\mathcal{T}}$. (Note that $\widetilde{d}^+_{\widehat{\mathcal{T}}}(v)$ is not the degree sequence of a tree.)

First, let $\widehat{\mathcal{T}}_{1n}$ be the random tree obtained by pruning the tree $\widehat{\mathcal{T}}$ at the node v^* with infinite outdegree, keeping only the first $\widetilde{D}_n$ children of v^*. Then $\widehat{\mathcal{T}}_{1n}$ is a locally finite tree, and, in fact, it is a.s. finite. The random tree $\widehat{\mathcal{T}}_{1n}$ can be constructed as $\widehat{\mathcal{T}}$ in Section 5, starting with a spine, and then adding independent Galton–Watson trees to it, but now the number of children of a node in the spine is given by a finite random variable $\check{\xi}_n$ with the distribution

$$\mathbb{P}(\check{\xi}_n = k) = \mathbb{P}(\widehat{\xi} = k) + \mathbb{P}(\widehat{\xi} = \infty)\,\mathbb{P}(\widetilde{D}_n = k) = k\pi_k + (1-\nu)\,\mathbb{P}(\widetilde{D}_n = k). \tag{20.7}$$

The nodes not in the spine (the normal nodes) have offspring distribution (π_k) as before. (This holds also for the following modifications.)

The spine in $\widehat{\mathcal{T}}_{1n}$ stops when we obtain $\widehat{\xi} = \infty$, but we may also define another random tree $\widehat{\mathcal{T}}_{2n}$ by continuing the spine to infinity; this defines a random infinite but locally finite tree having an infinite spine; each node in the spine has a number of children with the distribution in (20.7), and the spine continues with a uniformly randomly chosen child. Equivalently, $\widehat{\mathcal{T}}_{2n}$ can be defined by a Galton–Watson process with normal and special nodes as in Section 5, but with the offspring distribution for special nodes changed from (5.2) to (20.7).

Finally, let $\widehat{Y}_n$ by a random variable with the size-biased distribution of Y_1:

$$\mathbb{P}(\widehat{Y}_n = k) = \frac{k\,\mathbb{P}(Y_1 = k)}{(n-1)/n} = \frac{k\,\mathbb{E}\,N_k}{n-1}, \tag{20.8}$$

recalling that $\sum_k kN_k = n-1$; cf. (20.1) and (20.4). (Thus $\widehat{Y}_n \overset{\mathrm{d}}{=} d^+_{\mathcal{T}_n}(o)$ by Lemma 15.7 and (20.1).) Define the infinite, locally finite random tree $\widehat{\mathcal{T}}_{3n}$ by the same Galton–Watson process again, but now with offspring distribution $\widehat{Y}_n$ for special nodes. (This does not involve $\widetilde{D}_n$ or Ω_n.) Thus $\widehat{\mathcal{T}}_{3n}$ also has an infinite spine.

We then have the following version of Theorem 20.1, where we also use the metric δ_1 on $\mathfrak{T}_{\mathrm{lf}}$ defined by

$$\delta_1(T_1, T_2) := 1/\sup\bigl\{m \geqslant 1 : d^+_{T_1}(v) = d^+_{T_2}(v) \text{ for } v \in V^{[m]}\bigr\}. \tag{20.9}$$

Theorem 20.2. *For $j = 1, 2, 3$, and any finite set of nodes $v_1, \dots, v_\ell \in V_\infty$,*

$$\bigl(d^+_{\mathcal{T}_n}(v_1), \dots, d^+_{\mathcal{T}_n}(v_\ell)\bigr) \overset{\mathrm{d}}{\approx} \bigl(\widetilde{d}^+_{\widehat{\mathcal{T}}_{jn}}(v_1), \dots, \widetilde{d}^+_{\widehat{\mathcal{T}}_{jn}}(v_\ell)\bigr). \tag{20.10}$$

Equivalently, there is a coupling of $\mathcal{T}_n$ and $\widehat{\mathcal{T}}_{jn}$ such that $\delta_1(\mathcal{T}_n, \widehat{\mathcal{T}}_{jn}) \to 0$ as $n \to \infty$.

Proof. If $\Omega_n > m$ we have $\widetilde{D}_n > m$ and then the branches of $\widehat{\mathcal{T}}$ pruned to make $\widehat{\mathcal{T}}_{1n}$ are all outside $V^{[m]}$, and thus $d^+_{\widehat{\mathcal{T}}_{1n}} = \widetilde{d}^+_{\widehat{\mathcal{T}}}$ defined in (20.5) for all $v \in V^{[m]}$. Thus the result for $\widehat{\mathcal{T}}_{1n}$ follows from Theorem 20.1.

Next, for any given m, and for any endpoint x of the spine of $\widehat{\mathcal{T}}_{1n}$, the probability that the continuation in $\widehat{\mathcal{T}}_{2n}$ of the spine contains some node in $V^{[m]}$ is less than $m/\Omega_n = o(1)$; thus, w.h.p. $\widehat{\mathcal{T}}_{1n}$ and $\widehat{\mathcal{T}}_{2n}$ are equal on any $V^{[m]}$.

Finally, Lemma 20.3 below implies that we can couple $\widehat{\mathcal{T}}_{2n}$ and $\widehat{\mathcal{T}}_{3n}$ such that they w.h.p. agree on each $V^{(m)}$; then $\widehat{\mathcal{T}}_{1n}$ and $\widehat{\mathcal{T}}_{3n}$ are w.h.p. equal on each $V^{[m]}$. $\square$

Lemma 20.3. $\check{\xi}_n \overset{\mathrm{d}}{\approx} \widehat{Y}_n$.

Proof. For each fixed k, $\mathbb{P}(\check{\xi}_n = k) = k\pi_k$ as soon as $\Omega_n > k$, and $\mathbb{P}(\widehat{Y}_n = k) \to k\pi_k$ by (20.8) and Theorem 11.7. Hence,

$$|\,\mathbb{P}(\check{\xi}_n = k) - \mathbb{P}(\widehat{Y}_n = k)| \to 0. \tag{20.11}$$

By (20.7), (20.4) and (20.8), uniformly for $k > \Omega_n$,

$$\mathbb{P}(\check{\xi}_n = k) = k\pi_k + (1-\nu)\frac{k\,\mathbb{P}(Y_1 = k)}{1-\nu+o(1)} = k\pi_k + \bigl(1+o(1)\bigr)\,\mathbb{P}(\widehat{Y}_n = k);$$

hence

$$\sum_{k>\Omega_n} |\,\mathbb{P}(\check{\xi}_n = k) - \mathbb{P}(\widehat{Y}_n = k)| \leqslant \sum_{k>\Omega_n} \bigl(k\pi_k + o(1)\,\mathbb{P}(\widehat{Y}_n = k)\bigr) = \sum_{k>\Omega_n} k\pi_k + o(1). \tag{20.12}$$

Further, for any fixed K,

$$\sum_{k=K+1}^{\Omega_n} \bigl(\mathbb{P}(\check{\xi}_n = k) - \mathbb{P}(\widehat{Y}_n = k)\bigr)_+ \leqslant \sum_{k=K+1}^{\Omega_n} \mathbb{P}(\check{\xi}_n = k) = \sum_{k=K+1}^{\Omega_n} k\pi_k. \tag{20.13}$$

Using Lemma 19.5(vii) together with (20.11) for $k \leqslant K$, (20.12) and (20.13) we obtain

$$d_{\mathrm{TV}}(\check{\xi}_n, \widehat{Y}_n) = \sum_{k=1}^{\infty} \bigl(\mathbb{P}(\check{\xi}_n = k) - \mathbb{P}(\widehat{Y}_n = k)\bigr)_+ \leqslant \sum_{k=K+1}^{\infty} k\pi_k + o(1). \tag{20.14}$$

Since K is arbitrary and $\sum_1^\infty k\pi_k < \infty$, it follows that $d_{\mathrm{TV}}(\check{\xi}_n, \widehat{Y}_n) \to 0$. $\square$

21. Further results and problems

21.1. **Level widths.** Let, as in Remark 5.6, $l_k(T)$ denote the number of nodes with distance k to the root in a rooted tree T.

If $\nu \geqslant 1$, then $\widehat{\mathcal{T}}$ is a locally finite tree so all level widths $l_k(\widehat{\mathcal{T}})$ are finite. It follows easily from the characterisation of convergence in Lemma 6.2 that, in this case, the functional l_k is continuous at $\widehat{\mathcal{T}}$, and thus Theorem 7.1 implies (see Billingsley [15, Corollary 1, p. 31])

$$l_k(\mathcal{T}_n) \xrightarrow{\mathrm{d}} l_k(\widehat{\mathcal{T}}) < \infty \tag{21.1}$$

for each $k \geqslant 0$.

On the other hand, if $\nu < 1$, then $\widehat{\mathcal{T}}$ has a node with infinite outdegree; this node has a random distance $L-1$ to the root, where L as in Section 5 is the length of the spine, and thus $l_L(\widehat{\mathcal{T}}) = \infty$.

In the case $0 < \nu < 1$, we have $\pi_0 < 1$ and $\mathbb{P}(\xi \geqslant 1) = 1 - \pi_0 > 0$, so for any j, there is a positive probability that the Galton–Watson tree $\mathcal{T}$ has height at least j, and it follows that of the infinitely many copies of $\mathcal{T}$ that start in generation L, a.s. infinitely many will survive at least until generation $L + j$. Consequently, a.s., $l_k(\widehat{\mathcal{T}}) = \infty$ for all $k \geqslant L$, while $l_k(\widehat{\mathcal{T}}) < \infty$ for $k < L$. It follows easily from Lemma 6.3, that in this case too, for each $k \geqslant 0$, the mapping $l_k : \mathfrak{T} \to \overline{\mathbb{N}}_0$ is continuous at $\widehat{\mathcal{T}}$. Consequently,

$$l_k(\mathcal{T}_n) \xrightarrow{\mathrm{d}} l_k(\widehat{\mathcal{T}}) \leqslant \infty, \qquad k = 0, 1, \dots, \tag{21.2}$$

with $\mathbb{P}(l_k(\widehat{\mathcal{T}}) < \infty) = \mathbb{P}(L > k) = \nu^k$. (Recall that $\mu = \nu$ in this case by (7.2).)

When $\nu = 0$, however, (21.2) does not always hold. By Example 5.1, $\widehat{\mathcal{T}}$ is an infinite star, with $l_1(\widehat{\mathcal{T}}) = \infty$ and $l_k(\widehat{\mathcal{T}}) = 0$ for all $k \geqslant 2$. By Theorem 7.10, $l_1(\mathcal{T}_n) = d^+_{\mathcal{T}_n}(o) \xrightarrow{\mathrm{d}} \widehat{\xi} \overset{\mathrm{d}}{=} l_1(\widehat{\mathcal{T}})$, so (21.2) holds for $k = 1$ (and trivially for $k = 0$) in the case $\nu = 0$ too (with $l_1(\widehat{\mathcal{T}}) = \infty$). However, by Example 10.8, if $w_k = k!$, then $l_2(\mathcal{T}_n) \xrightarrow{\mathrm{d}} \mathrm{Po}(1)$, so $l_2(\mathcal{T}_n)$ does not converge to $l_2(\widehat{\mathcal{T}}) = 0$. Similarly, by Example 10.9, if $j \geqslant 2$ and $w_k = k!^\alpha$ with $0 < \alpha < 1/(j-1)$, then the number of paths of length j attached to the root in $\mathcal{T}_n$ tends to ∞ (in probability), so $l_j(\mathcal{T}_n) \xrightarrow{\mathrm{p}} \infty$, while $l_j(\widehat{\mathcal{T}}) = 0$.

Turning to moments, we have for the expectation, by (5.8), $\mathbb{E}\, l_k(\widehat{\mathcal{T}}) = \infty$ if $0 < \nu < 1$ or $\sigma^2 = \infty$; in this case (21.1)–(21.2) and Fatou's lemma yield $\mathbb{E}\, l_k(\mathcal{T}_n) \to \mathbb{E}\, l_k(\widehat{\mathcal{T}}) = \infty$.

If $\nu \geqslant 1$ and $\sigma^2 < \infty$, then (5.7) yields $\mathbb{E}\, l_k(\widehat{\mathcal{T}}) = 1 + k\sigma^2 < \infty$. In this case, for each fixed k, the random variables $l_k(\mathcal{T}_n)$, $n \geqslant 1$, are uniformly integrable, and thus (21.1) implies $\mathbb{E}\, l_k(\mathcal{T}_n) \to \mathbb{E}\, l_k(\widehat{\mathcal{T}})$, see Janson [59, Section 10]. (In the case $\nu > 1$, this was shown already by Meir and Moon [85].) Consequently, for any $\mathbf{w}$ with $\rho > 0$ and any fixed k,

$$\mathbb{E}\, l_k(\mathcal{T}_n) \to \mathbb{E}\, l_k(\widehat{\mathcal{T}}) \leqslant \infty. \tag{21.3}$$

(When $\rho = 0$, this is not always true, by the examples above.)

For higher moments, there remains a small gap. Let $r \geqslant 1$. When $0 < \nu < 1$, (21.3) trivially implies $\mathbb{E}\, l_k(\mathcal{T}_n)^r \to \mathbb{E}\, l_k(\widehat{\mathcal{T}})^r = \infty$, so suppose $\nu \geqslant 1$. Then, by (5.2), $\mathbb{E}\,\widehat{\xi}^r = \mathbb{E}\,\xi^{r+1}$, so if $\mathbb{E}\,\xi^{r+1} = \infty$, then $\mathbb{E}\, l_1(\widehat{\mathcal{T}})^r = \infty$; moreover, each $l_k(\widehat{\mathcal{T}})$, $k \geqslant 1$, stochastically dominates $\widehat{\xi}$ (consider the offspring of the k:th node on the spine), and thus $\mathbb{E}\, l_k(\widehat{\mathcal{T}})^r = \infty$ for every $k \geqslant 1$. Consequently, again immediately by Fatou's lemma and (21.2), $\mathbb{E}\, l_k(\mathcal{T}_n)^r \to \mathbb{E}\, l_k(\widehat{\mathcal{T}})^r = \infty$. The only interesting case is thus when $\mathbb{E}\,\xi^{r+1} < \infty$. If $r \geqslant 1$ is an integer, it was shown in [59, Theorem 1.13] that $\mathbb{E}\,\xi^{r+1} < \infty$ implies that $\mathbb{E}\, l_k(\mathcal{T}_n)^r$, $n \geqslant 1$, are uniformly bounded for each $k \geqslant 1$. We conjecture that, moreover, $l_k(\mathcal{T}_n)^r$, $n \geqslant 1$, are uniformly integrable, which by (21.1) would yield the following:

Conjecture 21.1. *For every integer $r \geqslant 1$ and every $k \geqslant 1$, if $\nu > 0$, then*

$$\mathbb{E}\, l_k(\mathcal{T}_n)^r \to \mathbb{E}\, l_k(\widehat{\mathcal{T}})^r \leqslant \infty. \tag{21.4}$$

We further conjecture that this holds also for non-integer $r > 0$.

One thus has to consider the case $\mathbb{E}\,\xi^{r+1} < \infty$ only, and the result from [59] implies that (21.4) holds if $\mathbb{E}\,\xi^{r+2} < \infty$, since then $\mathbb{E}\, l_k(\mathcal{T}_n)^{\lfloor r \rfloor+1}$ are uniformly bounded.

21.2. **Asymptotic normality.** In Theorem 7.11, we proved that N_d, the number of nodes of outdegree d in the random tree $\mathcal{T}_n$, satisfies $N_d/n \xrightarrow{\mathrm{p}} \pi_d$.

In our case Iα ($\nu > 1$ or $\nu = 1$ and $\sigma^2 < \infty$), Kolchin [76, Theorem 2.3.1] gives the much stronger result that the random variable N_d is asymptotically normal, for every $d \geqslant 0$:

$$\frac{N_d - n\pi_d}{\sqrt{n}} \xrightarrow{\mathrm{d}} N(0, \sigma_d^2), \tag{21.5}$$

with

$$\sigma_d^2 := \pi_d\Bigl(1 - \pi_d - \frac{(d-1)^2\pi_d}{\sigma^2}\Bigr). \tag{21.6}$$

(In fact, Kolchin [76] gives a local limit theorem which is a stronger version of (21.5).)

Under the assumption $\xi^3 < \infty$, Janson [55, Example 3.4] gave another proof of (21.5), and showed further joint convergence for different d, with asymptotic covariances, using $I_k := \mathbf{1}\{\xi = k\}$,

$$\sigma_{kl}^2 = \operatorname{Cov}(I_k, I_l) - \frac{\operatorname{Cov}(I_k,\xi)\operatorname{Cov}(I_l,\xi)}{\operatorname{Var}\xi} = \pi_k\delta_{kl} - \pi_k\pi_l - \frac{(k-1)(l-1)\pi_k\pi_l}{\sigma^2}. \tag{21.7}$$

Moreover, Janson [55] showed that if $\mathbb{E}\,|\xi|^r < \infty$ for every r (which in particular holds when $\nu > 1$ since then $\tau < \rho$ and ξ has some exponential moment), then convergence of all moments and joint moments holds in (21.5); in particular

$$\mathbb{E}\, N_k = n\pi_k + o(n) \quad \text{and} \quad \operatorname{Cov}(N_k, N_l) = n\sigma_{kl}^2 + o(n). \tag{21.8}$$

In the case $\nu > 1$, Minami [89] and Drmota [33, Section 3.2.1] have given other proofs of the (joint) asymptotic normality using the saddle point method; Drmota [33] shows further the stronger moment estimates

$$\mathbb{E}\, N_d = n\pi_d + O(1) \quad \text{and} \quad \operatorname{Var} N_d = n\sigma_d^2 + O(1). \tag{21.9}$$

Problem 21.2. *Do these results hold in the case $\nu = 1$, $\sigma^2 < \infty$ without extra moment conditions? Do they extend to the case $\nu = 1$, $\sigma^2 = \infty$? What happens when $0 \leqslant \nu < 1$?*

Problem 21.3. *Extend this to the more general case of balls-in-boxes as in Theorem 11.4. (We guess that the case $0 < \lambda < \nu$ is easy by the methods in the references above, in particular [55] and [33, Section 3.2.1], but we have not checked the details.)*

Problem 21.4. *Extend this to the subtree counts in Theorem 7.12.*

21.3. **Height and width.** We have studied the random trees $\mathcal{T}_n$ without any scaling. Since our mode of convergence really means that we consider only a finite number of generations at a time, we are really looking at the base of the tree, with the first generations. The results in this paper thus do not say anything about, for example, the height and width of $\mathcal{T}_n$. (Recall that if T is a rooted tree, then the height $H(T) := \max\{k : l_k(T) > 0\}$, the maximum distance from the root, and the width $W(T) := \max_k\{l_k(T)\}$, the largest size of a generation.) However, there are other known results.

In the case $\nu \geqslant 1$, $\sigma^2 < \infty$ (the case Iα in Section 8), it is well-known that both the height $H(\mathcal{T}_n)$ and the width $W(\mathcal{T}_n)$ of $\mathcal{T}_n$ typically are of order $\sqrt{n}$; more precisely,

$$H(\mathcal{T}_n)/\sqrt{n} \xrightarrow{\mathrm{d}} 2\sigma^{-1}X, \tag{21.10}$$

$$W(\mathcal{T}_n)/\sqrt{n} \xrightarrow{\mathrm{d}} \sigma X, \tag{21.11}$$

where X is some strictly positive random variable (in fact, X equals the maximum of a standard Brownian excursion and has what is known as a theta distribution), see e.g. Kolchin [76], Aldous [4], Chassaing, Marckert and Yor [25], Janson [59] and Drmota [33]. There are also results for a single level giving an asymptotic distribution for $l_{k(n)}(T_n)/\sqrt{n}$ when the level $k(n) \sim a\sqrt{n}$ for some $a > 0$, see Kolchin [76, Theorem 2.4.5].

Since the variance σ^2 appears as a parameter in these results, we cannot expect any simple extensions to the case $\sigma^2 = \infty$, and even less to the case $0 \leqslant \nu < 1$. Nevertheless, we conjecture that (21.10) and (21.11) extend formally at least to the case $\nu = 1$ and $\sigma^2 = \infty$:

Conjecture 21.5. *If $\nu = 1$ and $\sigma^2 = \infty$, then $H(\mathcal{T}_n)/\sqrt{n} \xrightarrow{\mathrm{p}} 0$.*

Conjecture 21.6. *If $\nu = 1$ and $\sigma^2 = \infty$, then $W(\mathcal{T}_n)/\sqrt{n} \xrightarrow{\mathrm{p}} \infty$.*

Problem 21.7. *Does $\nu < 1$ imply that $H(\mathcal{T}_n)/\sqrt{n} \xrightarrow{\mathrm{p}} 0$?*

Problem 21.8. *Does $\nu < 1$ imply that $W(\mathcal{T}_n)/\sqrt{n} \xrightarrow{\mathrm{p}} \infty$?*

Furthermore, still in the case $\nu \geqslant 1$, $\sigma^2 < \infty$, Addario-Berry, Devroye and Janson [1] have shown sub-Gaussian tail estimates for the height and width

$$\mathbb{P}(H(\mathcal{T}_n) \geqslant x\sqrt{n}) \leqslant Ce^{-cx^2}, \tag{21.12}$$

$$\mathbb{P}(W(\mathcal{T}_n) \geqslant x\sqrt{n}) \leqslant Ce^{-cx^2}, \tag{21.13}$$

uniformly in all $x \geqslant 0$ and $n \geqslant 1$ (with some positive constants C and c depending on $\boldsymbol{\pi}$ and thus on $\mathbf{w}$). In view of (21.11), we cannot expect (21.13) to hold when $\sigma^2 = \infty$ (or when $\nu < 1$), but we see no reason why (21.12) cannot hold; (21.10) suggests that $H(\mathcal{T}_n)$ typically is smaller when $\sigma^2 = \infty$.

Problem 21.9. *Does* (21.12) *hold for any weight sequence* $\mathbf{w}$ *(with C and c depending on* $\mathbf{w}$*, but not on x or n)?*

It follows from (21.10)–(21.11) and (21.12)–(21.13) that $\mathbb{E}\, H(\mathcal{T}_n)/\sqrt{n}$ and $\mathbb{E}\, W(\mathcal{T}_n)/\sqrt{n}$ converge to positive numbers. (In fact, the limits are $\sqrt{2\pi}/\sigma$ and $\sqrt{\pi/2}\,\sigma$, see e.g. Janson [61], where also joint moments are computed.)

Problem 21.10. *What are the growth rates of* $\mathbb{E}\, H(\mathcal{T}_n)$ *and* $\mathbb{E}\, W(\mathcal{T}_n)$ *when* $\sigma^2 = \infty$ *or* $\nu < 1$*?*

21.4. **Scaled trees.** The results (21.10)–(21.11), as well as many other results on various asymptotics of $\mathcal{T}_n$ in the case $\nu \geqslant 1$, $\sigma^2 < \infty$, can be seen as consequences of the convergence of the tree $\mathcal{T}_n$, after rescaling in a suitable sense in both height and width by $\sqrt{n}$, to the *continuum random tree* defined by Aldous [3, 4, 5], see also Le Gall [80]. (The continuum random tree is not an ordinary tree; it is a compact metric space.) This has been extended to the case $\sigma^2 = \infty$ when $\boldsymbol{\pi}$ is in the domain of attraction of a stable distribution, see e.g. Duquesne [34] and Le Gall [80, 81]; the limit is now a different random metric space called a *stable tree.*

Problem 21.11. *Is there some kind of similar limiting object in the case* $\nu < 1$ *(after suitable scaling)?*

21.5. **Random walks.** Simple random walk on the infinite random tree $\widehat{\mathcal{T}}$ has been studied by many authors in the critical case $\nu \geqslant 1$, in particular when $\sigma^2 < \infty$, see e.g. Kesten [74], Barlow and Kumagai [9], Durhuus, Jonsson and Wheater [35], Fujii and Kumagai [43], but also when $\sigma^2 = \infty$, see Croydon and Kumagai [30] (assuming attraction to a stable law).

A different approach is to study simple random walk on $\mathcal{T}_n$ and study asymptotics os $n \to \infty$. For example, by rescaling the tree one can obtain convergence to a process on the continuum random tree (when $\sigma^2 < \infty$) or stable tree (assuming attraction to a stable law), see Croydon [28, 29].

For $\nu < 1$, the simple random walk on $\widehat{\mathcal{T}}$ does not make sense, since the tree has a node with infinite degree. Nevertheless, it might be interesting to study simple random walk on $\mathcal{T}_n$ and find asymptotics of interesting quantities as $n \to \infty$.

21.6. **Multi-type conditioned Galton–Watson trees.** It seems likely that there are results similar to the ones in Section 7 for multi-type Galton–Watson trees conditioned on the total size, or perhaps on the number of nodes of each type, and for corresponding generalizations of simply generated random trees. We are not aware of any such results, however, and leave this as an open problem. (See Kurtz, Lyons, Pemantle and Peres [78] for related results that presumably are useful.)

22. Different conditionings for Galton–Watson trees

One of the principal objects studied in this paper is the conditioned Galton–Watson tree $(\mathcal{T} \mid |\mathcal{T}| = n)$, i.e. a Galton–Watson tree $\mathcal{T}$ conditioned on its total size being n; we then let $n \to \infty$. This is one way to consider very large Galton–Watson trees, but there are also other similar conditionings. For comparison, we briefly consider two possibilities; see further Kennedy [73] and Aldous and Pitman [6]. We denote the offspring distribution by ξ and its probability generating function by $\Phi(t)$.

22.1. Conditioning on $|\mathcal{T}| \geqslant n$.

If $\mathbb{E}\,\xi \leqslant 1$, i.e., in the subcritical and critical cases, $|\mathcal{T}| < \infty$ a.s. and thus $\mathcal{T}$ conditioned on $|\mathcal{T}| \geqslant n$ is a mixture of $(\mathcal{T} \mid |\mathcal{T}| = N) = \mathcal{T}_N$ for $N \geqslant n$. It follows immediately from Theorem 7.1 that $(\mathcal{T} \mid |\mathcal{T}| = N) \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ as $n \to \infty$.

If $\mathbb{E}\,\xi > 1$, i.e., in the supercritical case, on the other hand, the event $|\mathcal{T}| = \infty$ has positive probability, and the events $|\mathcal{T}| \geqslant n$ decrease to $|\mathcal{T}| = \infty$. Consequently,

$$(\mathcal{T} \mid |\mathcal{T}| \geqslant n) \xrightarrow{\mathrm{d}} (\mathcal{T} \mid |\mathcal{T}| = \infty), \tag{22.1}$$

a supercritical Galton–Watson tree conditioned on non-extinction.

Remark 22.1. When $\mathcal{T}$ is supercritical, the conditioned Galton–Watson tree $(\mathcal{T} \mid |\mathcal{T}| = \infty)$ in (22.1) can be constructed by a 2-type Galton–Watson process, somewhat similar to the construction of $\widehat{\mathcal{T}}$ in Section 5: Let $q := \mathbb{P}(|\mathcal{T}| < \infty) < 1$ be the extinction probability, which is given by $\Phi(q) = q$. Consider a Galton–Watson process $\overline{\mathcal{T}}$ with individuals of two types, *mortal* and *immortal*, where a mortal gets only mortal children while an immortal may get both mortal and immortal children. The numbers ξ' of mortal and ξ'' of immortal children are described by the probability generating functions

$$\mathbb{E}\, x^{\xi'} y^{\xi''} = \Phi_{\mathrm{m}}(x) := \Phi_q(x) = \Phi(qx)/q \tag{22.2}$$

for a mortal and

$$\mathbb{E}\, x^{\xi'} y^{\xi''} = \Phi_{\mathrm{i}}(x) := \frac{\Phi(qx + (1-q)y) - \Phi(qx)}{1-q} \tag{22.3}$$

for an immortal (with the children coming in random order). Note that the subtree started by a mortal is subcritical (since $\Phi'_{\mathrm{m}}(1) = \Phi'(q) < 1$, cf. (4.9)), and thus a.s. finite, while every immortal has at least one immortal child (since $\Phi_{\mathrm{i}}(x, 0) = 0$) and thus the subtree started by an immortal is infinite. It is easily verified that $\mathcal{T}$ conditioned on non-extinction equals this random tree $\overline{\mathcal{T}}$ started with an immortal, while $\mathcal{T}$ conditioned on extinction equals $\overline{\mathcal{T}}$ started with a mortal. (See Athreya and Ney [8, Section I.12], where this is stated in a somewhat different form.)

One important difference from $\widehat{\mathcal{T}}$ is that $\overline{\mathcal{T}}$ does not have a single spine; started with an immortal it has a.s. an uncountable number of infinite paths from the root.

Note that $\widehat{\mathcal{T}}$ in the critical case can be seen as a limit case of this construction. If we let $q \nearrow 1$, which requires that we really consider a sequence of different distributions with generating functions $\Phi^{(n)}(t) \to \Phi(t)$, then taking the limits in (22.2)–(22.3) gives for the limiting critical distribution the offspring generating functions $\Phi_{\mathrm{m}}(x) = \Phi(x)$ and $\Phi_{\mathrm{i}}(x, y) = y\Phi'(x)$, which indeed are the generating functions for the offspring distributions in Section 5 in the critical case (with mortal = normal and immortal = special), since $\mathbb{E}\, x^{\widehat{\xi}-1}y = y\Phi'(x) = \Phi_{\mathrm{i}}(x, y)$ by (5.4).

22.2. **Conditioning on $H(\mathcal{T}) \geqslant n$.** To condition on the height $H(\mathcal{T})$ being at least n is the same as conditioning on $l_n(\mathcal{T}) > 0$, i.e., that the Galton–Watson process survives for at least n generations.

If $\mathbb{E}\,\xi > 1$, i.e., in the supercritical case, the events $l_n(\mathcal{T}) > 0$ decrease to $|\mathcal{T}| = \infty$. Consequently,

$$(\mathcal{T} \mid H(\mathcal{T}) \geqslant n) = (\mathcal{T} \mid l_n(\mathcal{T}) > 0) \xrightarrow{\mathrm{d}} (\mathcal{T} \mid |\mathcal{T}| = \infty), \tag{22.4}$$

exactly as when conditioning on $|\mathcal{T}| \geqslant n$ in (22.1). By Remark 22.1, the limit equals $\overline{\mathcal{T}}$, started with an immortal.

In the subcritical and critical cases, the following result, proved by Kesten [74] (at least for $\mathbb{E}\,\xi = 1$, see also Aldous and Pitman [6]), shows convergence to the size-biased Galton–Watson tree $\mathcal{T}^*$ in Remark 5.7.

Theorem 22.2. *Suppose that $\mu := \mathbb{E}\,\xi \leqslant 1$. Then, as $n \to \infty$,*

$$(\mathcal{T} \mid H(\mathcal{T}) \geqslant n) = (\mathcal{T} \mid l_n(\mathcal{T}) > 0) \xrightarrow{\mathrm{d}} \mathcal{T}^*. \tag{22.5}$$

Proof. Let $r_n := \mathbb{P}(l_n(\mathcal{T}) > 0)$, the probability of survival for at least n generations. Then $r_n \to 0$ as $n \to \infty$. Fix $\ell > 0$ and a tree T with height ℓ. Conditioned on $\mathcal{T}^{(\ell)} = T$, the remainder of the tree consists of $l_\ell(T)$ independent branches, each distributed as $\mathcal{T}$, and thus, for $n > \ell$,

$$\begin{aligned}\mathbb{P}(\mathcal{T}^{(\ell)} = T \mid H(\mathcal{T}) \geqslant n) &= \frac{\mathbb{P}(\mathcal{T}^{(\ell)} = T \text{ and } H(\mathcal{T}) \geqslant n)}{\mathbb{P}(H(\mathcal{T}) \geqslant n)} \\ &= \frac{\mathbb{P}(\mathcal{T}^{(\ell)} = T)\bigl(1 - (1 - r_{n-\ell})^{l_\ell(T)}\bigr)}{\mathbb{P}(H(\mathcal{T}) \geqslant n)}.\end{aligned} \tag{22.6}$$

Let $\mathfrak{T}_{\mathrm{f}}^{(\ell)}$ be the set of finite trees of height ℓ. Summing (22.6) over $T \in \mathfrak{T}_{\mathrm{f}}^{(\ell)}$ yields 1, and thus

$$\mathbb{P}(H(\mathcal{T}) \geqslant n) = \sum_{T \in \mathfrak{T}_{\mathrm{f}}^{(\ell)}} \mathbb{P}(\mathcal{T}^{(\ell)} = T)\bigl(1 - (1 - r_{n-\ell})^{l_\ell(T)}\bigr) \tag{22.7}$$

Dividing by $r_{n-\ell}$, and noting that for any $N \geqslant 1$, $\bigl(1-(1-r)^N\bigr)/r \nearrow N$ as $r \searrow 0$, we find by monotone convergence

$$\begin{aligned}\frac{\mathbb{P}(H(\mathcal{T}) \geqslant n)}{r_{n-\ell}} &= \sum_{T\in\mathfrak{T}_{\mathrm{f}}^{(\ell)}} \mathbb{P}(\mathcal{T}^{(\ell)} = T)\frac{1-(1-r_{n-\ell})^{l_\ell(T)}}{r_{n-\ell}} \\ &\to \sum_{T\in\mathfrak{T}_{\mathrm{f}}^{(\ell)}} \mathbb{P}(\mathcal{T}^{(\ell)} = T) l_\ell(T) = \mathbb{E}\, l_\ell(\mathcal{T}) = \mu^\ell. \end{aligned} \tag{22.8}$$

Hence, by (22.6) and (5.11),

$$\begin{aligned}\mathbb{P}(\mathcal{T}^{(\ell)} = T \mid H(\mathcal{T}) \geqslant n) &\sim \frac{\mathbb{P}(\mathcal{T}^{(\ell)} = T) l_\ell(T) r_{n-\ell}}{\mathbb{P}(H(\mathcal{T}) \geqslant n)} \\ &\to \frac{\mathbb{P}(\mathcal{T}^{(\ell)} = T) l_\ell(T)}{\mu^\ell} = \mathbb{P}(\mathcal{T}^{*(\ell)} = T). \end{aligned} \tag{22.9}$$

Thus, $(\mathcal{T} \mid H(\mathcal{T}) \geqslant n)^{(\ell)} \xrightarrow{\mathrm{d}} \mathcal{T}^{*(\ell)}$, and the result follows by (6.9). ◻

Note that if $\mathbb{E}\,\xi = 1$, then $\mathcal{T}^* = \widehat{\mathcal{T}}$, see Remark 5.7, so the limits in Theorems 7.1 and 22.2 of $\mathcal{T}$ conditioned on $|\mathcal{T}| = n$ and $H(\mathcal{T}) \geqslant n$ have the same limit. However, in the subcritical case $\mathbb{E}\,\xi < 1$, $\mathcal{T}^* \neq \widehat{\mathcal{T}}$; moreover, $\mathcal{T}^*$ differs also from the limit in Theorem 7.1, which is $\widehat{\mathcal{T}}$ for a conjugated distribution, and the same is true in the supercritical case. Hence, as remarked by Kennedy [73], conditioning on $|\mathcal{T}| = n$ and $H(\mathcal{T}) \geqslant n$ give similar results (in the sense that the limits as $n \to \infty$ are the same) in the critical case, but quite different results in the subcritical and supercritical cases. Similarly, conditioning on $|\mathcal{T}| \geqslant n$ and $H(\mathcal{T}) \geqslant n$ give quite different results in the subcritical case. Aldous and Pitman [6] remarks that the two different limits as $n \to \infty$ both can be intuitively interpreted as "$\mathcal{T}$ conditioned on being infinite", which shows that one has to be careful with such interpretations.

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, The continuum random tree III. *Ann. Probab.* **21** (1993), no. 1, 248–289.
[6] 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.
[7] R. Arratia, A. D. Barbour & S. Tavaré, *Logarithmic Combinatorial Structures: a Probabilistic Approach*, EMS, Zürich, 2003.
[8] K. B. Athreya & P. E. Ney, *Branching Processes.* Springer-Verlag, Berlin, 1972.
[9] M. T. Barlow & T. Kumagai, Random walk on the incipient infinite cluster on trees. *Illinois J. Math.* **50** (2006), no. 1–4, 33–65.
[10] D. Beihoffer, J. Hendry, A. Nijenhuis & S. Wagon, Faster algorithms for Frobenius numbers. *Electron. J. Combin.* **12** (2005), R27.
[11] J. Bennies & G. Kersting, A random walk approach to Galton–Watson trees. *J. Theoret. Probab.* **13** (2000), no. 3, 777–803.
[12] E. S. Bernikovich & Yu. L. Pavlov, On the maximum size of a tree in a random unlabelled unrooted forest. *Diskret. Mat.* **23** (2011), no. 1, 3–20 (Russian). English transl.: *Discrete Math. Appl.* **21** (2011), no. 1, 1–21.
[13] P. Bialas & Z. Burda, Phase transition in fluctuating branched geometry. *Physics Letters B* **384** (1996), 75–80.
[14] P. Bialas, Z. Burda & D. Johnston, Condensation in the backgammon model. *Nuclear Physics* **493** (1997), 505–516.
[15] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.
[16] N. H. Bingham, C. M. Goldie & J. L. Teugels, *Regular Variation.* Cambridge Univ. Press, Cambridge, 1987.
[17] C. W. Borchardt, Ueber eine der Interpolation entsprechende Darstellung der Eliminations-Resultante. *J. reine und angewandte Mathematik* **57** (1860), 111–121.
[18] É. Borel, Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Application au problème de l'attente à un guichet. *C. R. Acad. Sci. Paris* **214** (1942), 452–456.
[19] A. V. Boyd, Formal power series and the total progeny in a branching process. *J. Math. Anal. Appl.* **34** (1971), 565–566.
[20] V. E. Britikov, Asymptotic number of forests from unrooted trees. *Mat. Zametki* **43** (1988), no. 5, 672–684, 703 (Russian). English transl.: *Math. Notes* **43** (1988), no. 5–6, 387–394.
[21] R. Carr, W. M. Y. Goh & E. Schmutz, The maximum degree in a random tree and related problems. *Random Struct. Alg.* **5** (1994), no. 1, 13–24.
[22] A. Cayley, A theorem on trees. *Quart. J. Math.* **23** (1889), 376–378.

[23] 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.

[24] P. Chassaing & G. Louchard, Phase transition for parking blocks, Brownian excursion and coalescence. *Random Struct. Alg.* **21** (2002), no. 1, 76–119.

[25] P. Chassaing, J.-F. Marckert & M. Yor, The height and width of simple trees. *Mathematics and Computer Science (Versailles, 2000)*, 17–30, Trends Math., Birkhäuser, Basel, 2000.

[26] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey & D. E. Knuth, On the Lambert W function. *Adv. Comput. Math.* **5** (1996), no. 4, 329–359.

[27] H. Cramér, Sur un noveau théorème-limite de la théorie des probabilités. *Les sommes et les fonctions de variables aléatoires*, Actualités Scientifiques et Industrielles 736, Hermann, Paris, 1938, pp. 5–23.

[28] D. Croydon, Convergence of simple random walks on random discrete trees to Brownian motion on the continuum random tree. *Ann. Inst. H. Poincaré Probab. Statist.* **44** (2008), no. 6, 987–1019.

[29] D. Croydon, Scaling limits for simple random walks on random ordered graph trees. *Adv. Appl. Probab.* **42** (2010), no. 2, 528–558.

[30] D. Croydon & T. Kumagai, Random walks on Galton–Watson trees with infinite variance offspring distribution conditioned to survive. *Electron. J. Probab.* **13** (2008), no. 51, 1419–1441.

[31] A. Dembo & O. Zeitouni, *Large Deviations Techniques and Applications.* 2nd ed., Springer, New York, 1998.

[32] 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.

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

[34] T. Duquesne, A limit theorem for the contour process of conditioned Galton–Watson trees. *Ann. Probab.* **31** (2003), no. 2, 996–1027.

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

[36] M. Dwass, The total progeny in a branching process and a related random walk. *J. Appl. Probab.* **6** (1969), 682–686.

[37] F. Eggenberger & G. Pólya, Über die Statistik verketteter Vorgänge. *Zeitschrift Angew. Math. Mech.* **3** (1923), 279–289.

[38] W. Feller, *An Introduction to Probability Theory and its Applications, Volume I*, 2nd ed., Wiley, New York, 1957.

[39] W. Feller, *An Introduction to Probability Theory and its Applications, Volume II*, 2nd ed., Wiley, New York, 1971.

[40] P. Flajolet & R. Sedgewick, *Analytic Combinatorics.* Cambridge Univ. Press, Cambridge, UK, 2009.

[41] S. Franz & F. Ritort, Dynamical solution of a model without energy barriers. *Europhysics Letters* **31** (1995), 507–512

[42] S. Franz & F. Ritort, Glassy mean-field dynamics of the backgammon model. *J. Stat. Phys.* **85** (1996), 131–150.

[43] I. Fujii & T. Kumagai, Heat kernel estimates on the incipient infinite cluster for critical branching processes. *Proceedings of German–Japanese Symposium in Kyoto 2006*, RIMS Kôkyûroku Bessatsu B6 (2008), pp. 8–95.

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

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

[46] B. V. Gnedenko & A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow–Leningrad, 1949 (Russian). English transl.: Addison-Wesley, Cambridge, Mass., 1954.

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

[48] G. R. Grimmett, *The Random-Cluster Model*, Springer, Berlin, 2006.

[49] A. Gut, *Probability: A Graduate Course*. Springer, New York, 2005.

[50] G. H. Hardy, J. E. Littlewood & G. Pólya, *Inequalities*. 2nd ed., Cambridge, at the University Press, 1952.

[51] T. E. Harris, A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* **56** (1960), 13–20.

[52] L. Holst, Two conditional limit theorems with applications. *Ann. Statist.* **7** (1979), no. 3, 551–557.

[53] L. Holst, A unified approach to limit theorems for urn models. *J. Appl. Probab.* **16** (1979), 154–162.

[54] I. A. Ibragimov & Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*. Nauka, Moscow, 1965 (Russian). English transl.: Wolters-Noordhoff Publishing, Groningen, 1971.

[55] S. Janson, Moment convergence in conditional limit theorems. *J. Appl. Probab.* **38** (2001), no. 2, 421–437.

[56] S. Janson, Asymptotic distribution for the cost of linear probing hashing. *Random Struct. Alg.* **19** (2001), no. 3–4, 438–471.

[57] S. Janson, Cycles and unicyclic components in random graphs. *Combin. Probab. Comput.* **12** (2003), 27–52.

[58] S. Janson, Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.* **110** (2004), no. 2, 177–245.

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

[60] S. Janson, Rounding of continuous random variables and oscillatory asymptotics. *Ann. Probab.* **34** (2006), no. 5, 1807–1826.

[61] S. Janson, On the asymptotic joint distribution of height and width in random trees, *Studia Sci. Math. Hungar.* **45** (2008), no. 4, 451–467.

[62] S. Janson, Probability asymptotics: notes on notation. Institute Mittag-Leffler Report 12, 2009 spring. `arXiv:1108.3924.`

[63] S. Janson, Stable distributions. Unpublished notes, 2011. `arXiv:1112.0220.`

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

[65] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[66] N. L. Johnson & S. Kotz, *Urn Models and their Application.* Wiley, New York, 1977.

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

[68] O. Kallenberg, *Random Measures.* Akademie-Verlag, Berlin, 1983.

[69] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

[70] N. I. Kazimirov, On some conditions for absence of a giant component in the generalized allocation scheme. *Diskret. Mat.* **14** (2002), no. 2, 107–118 (Russian). English transl.: *Discrete Math. Appl.* **12** (2002), no. 3, 291–302.

[71] N. I. Kazimirov, Emergence of a giant component in a random permutation with a given number of cycles. *Diskret. Mat.* **15** (2003), no. 3, 145–159 (Russian). English transl.: *Discrete Math. Appl.* **13** (2003), no. 5, 523–535.

[72] N. I. Kazimirov & Yu. L. Pavlov, A remark on the Galton–Watson forests. *Diskret. Mat.* **12** (2000), no. 1, 47–59 (Russian). English transl.: *Discrete Math. Appl.* **10** (2000), no. 1, 49–62.

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

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

[75] D. E. Knuth, *The Art of Computer Programming. Vol. 3: Sorting and Searching.* 2nd ed., Addison-Wesley, Reading, Mass., 1998.

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

[77] V. F. Kolchin, B. A. Sevast'yanov & V. P. Chistyakov, *Random Allocations.* Nauka, Moscow, 1976 (Russian). English transl.: Winston, Washington, D.C., 1978.

[78] 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.

[79] J.-L. Lagrange, Nouvelle méthode pour résoudre les équations littérales par le moyen des séries. *Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin*, **XXIV** (1770), 5–73.

[80] J.-F. Le Gall, Random trees and applications. *Probab. Surveys* **2** (2005), 245–311.

[81] J.-F. Le Gall, Random real trees. *Ann. Fac. Sci. Toulouse Math.* (6) **15** (2006), no. 1, 35–62.

[82] M. R. Leadbetter, G. Lindgren & H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York, 1983.

[83] T. Łuczak & B. Pittel, Components of random forests. *Combin. Probab. Comput.* **1** (1992), no. 1, 35–52.

[84] 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.

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

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

[87] A. Meir & J. W. Moon, On nodes of large out-degree in random trees. *Congr. Numer.* **82** (1991), 3–13.

[88] A. Meir & J. W. Moon, A note on trees with concentrated maximum degrees. *Utilitas Math.* **42** (1992), 61–64. Coorigendum: *Utilitas Math.* **43** (1993), 253.

[89] 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.

[90] J. W. Moon, On the maximum degree in a random tree. *Michigan Math. J.* **15** (1968), 429–432.

[91] J. Neveu, Arbres et processus de Galton–Watson. *Ann. Inst. H. Poincaré Probab. Statist.* **22** (1986), no. 2, 199–207.

[92] R. Otter, The number of trees. *Ann. of Math. (2)* **49** (1948), 583–599.

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

[94] Yu. L. Pavlov, The asymptotic distribution of maximum tree size in a random forest. *Teor. Verojatnost. i Primenen.* **22** (1977), no. 3, 523–533 (Russian). English transl.: *Th. Probab. Appl.* **22** (1977), no. 3, 509–520.

[95] Yu. L. Pavlov, The limit distributions of the maximum size of a tree in a random forest. *Diskret. Mat.* **7** (1995), no. 3, 19–32 (Russian). English transl.: *Discrete Math. Appl.* **5** (1995), no. 4, 301–315.

[96] Yu. L. Pavlov, *Random Forests.* Karelian Centre Russian Acad. Sci., Petrozavodsk, 1996 (Russian). English transl.: VSP, Zeist, The Netherlands, 2000.

[97] Yu. L. Pavlov, Limit theorems on sizes of trees in a random unlabelled forest. *Diskret. Mat.* **17** (2005), no. 2, 70–86 (Russian). English transl.: *Discrete Math. Appl.* **15** (2005), no. 2, 153–170.

[98] Yu. L. Pavlov & E. A. Loseva, Limit distributions of the maximum size of a tree in a random recursive forest. *Diskret. Mat.* **14** (2002), no. 1, 60–74 (Russian). English transl.: *Discrete Math. Appl.* **12** (2002), no. 1, 45–59.

[99] 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.

[100] F. Ritort, Glassiness in a model without energy barriers. *Physical Review Letters* **75** (1995), 1190–1193.

[101] W. Rudin, *Real and Complex Analysis.* McGraw-Hill, London, 1970

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

[103] R. P. Stanley, *Enumerative Combinatorics, Volume 2.* Cambridge Univ. Press, Cambridge, 1999.

[104] J. J. Sylvester, On the change of systems of independent variables, *Quart J. Math.* **1** (1857), 42–56.

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

[106] L. Takács, Ballots, queues and random graphs. *J. Appl. Probab.* **26** (1989), no. 1, 103–112.

[107] J.C. Tanner, A derivation of the Borel distribution. *Biometrika* **48** (1961), 222–224.

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

[109] Herbert S. Wilf, *generatingfunctionology.* 2nd ed., Academic Press, 1994.

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/