Trees, random allocations and condensation

Svante Janson

26th Nordic and 1st European-Nordic Congress of Mathematicians,
Lund, 12 June 2013

Simply generated trees

Trees are rooted and ordered (a.k.a. plane).

$\mathbf{w} = (w_k)_{k \geq 0}$ is a fixed *weight sequence*.

The *weight* of a finite tree T is

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

where $d^+(v)$ is the outdegree of v.

Trees with such weights are called *simply generated trees* and were introduced by Meir and Moon (1978).

We let $\mathcal{T}_n$ be the random simply generated tree obtained by picking a tree with n nodes at random with probability proportional to its weight.

Galton–Watson trees

Let $\sum_{k=0}^{\infty} w_k = 1$, so $(w_k)_1^{\infty}$ is a probability distribution on $\{0, 1, 2, \dots\}$ (a *probability weight sequence*).

Let ξ be a random variable with $\mathbb{P}(\xi = k) = w_k$.

Then the random tree $\mathcal{T}_n$ = the *conditioned Galton–Watson tree* with offspring distribution ξ.

(The random Galton–Watson tree defined by ξ conditioned on having exactly n vertices.)

Many kinds of random trees occuring in various applications can be seen as simply generated random trees and conditioned Galton–Watson trees.

Example $w_k = 1$ yields uniformly random *ordered trees* (*plane trees*).

Also $w_k = 2^{-k-1}$, a *Geometric distribution* $\mathrm{Ge}(1/2)$

Example $w_k = 1/k!$ yields uniformly random *labelled trees*.

Also $w_k = e^{-1}/k!$, a *Poisson distribution* $\mathrm{Po}(1)$.

Example $w_0 = 1$, $w_1 = 2$, $w_2 = 1$, $w_k = 0$ for $k \geq 3$ yields uniformly random *binary trees*.

Also $w_k = \binom{2}{k}\frac{1}{4}$, a *Binary distribution* $\mathrm{Bi}(2, 1/2)$.

Equivalent weights

Let $a, b > 0$ and change w_k to

$$\widetilde{w}_k := ab^k w_k.$$

Then the distribution of $\mathcal{T}_n$ is not changed.

In other words, the new weight sequence $(\widetilde{w}_k)$ defines the same simply generated random trees $\mathcal{T}_n$ as (w_k).

We say that weight sequence (w_k) and $(\widetilde{w}_k)$ are *equivalent*.

For many (w_k) there exists an equivalent probability weight sequence; in this case $\mathcal{T}_n$ can thus be seen as a conditioned Galton–Watson tree.

Moreover, in many cases this can be done such that the resulting probability distribution has mean 1. In such cases it thus suffices to consider the case of a probability weight sequence with mean $\mathbb{E}\,\xi = 1$; then $\mathcal{T}_n$ is a conditional critical Galton–Watson tree.

Thus, simply generated trees and (critical) conditioned Galton–Watson trees are almost the same

For many (w_k) there exists an equivalent probability weight sequence; in this case $\mathcal{T}_n$ can thus be seen as a conditioned Galton–Watson tree.

Moreover, in many cases this can be done such that the resulting probability distribution has mean 1. In such cases it thus suffices to consider the case of a probability weight sequence with mean $\mathbb{E}\,\xi = 1$; then $\mathcal{T}_n$ is a conditional critical Galton–Watson tree.

Thus, simply generated trees and (critical) conditioned Galton–Watson trees are almost the same

– BUT ONLY ALMOST !

Three types

Three types:

I. Critical Galton–Watson tree.

II. Subcritical Galton–Watson tree; not equivalent to any critical.

III. simply generated tree, not equivalent to any Galton–Watson tree.

Critical Galton–Watson trees form a nice and natural setting, with many known results (possibly with extra assumptions).

We extend some of these results to the general case, including cases II and III.

A theorem

Theorem

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

Then $\mathcal{T}_n \xrightarrow{\mathrm{d}} \widehat{\mathcal{T}}$ as $n \to \infty$, where $\widehat{\mathcal{T}}$ is an infinite modified Galton–Watson tree (see below).

The limit (in distribution) in the theorem is for a topology where convergence means convergence of outdegree for any fixed node; it thus really means local convergence close to the root.

(It is for this purpose convenient to regard the trees as subtrees of the infinite Ulam–Harris tree.)

Kennedy (1975), Aldous & Pitman (1998), Kolchin (1984), Jonsson & Stefánsson (2011), et al + J

Algebraic characterizations of the cases

Let

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

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

Let (for t such that $\Phi(t) < \infty$)

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

Let

$$\nu := \Psi(\rho) := \lim_{t \nearrow \rho} \Psi(t) \leq \infty.$$

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

$$\nu = \frac{\rho\Phi'(\rho)}{\Phi(\rho)} \leq \infty.$$

The three cases can be characterised as

I. $\nu \geq 1$. Then $0 < \rho \leq \infty$.

II. $0 < \nu < 1$. Then $0 < \rho < \infty$.

III. $\nu = \rho = 0$.

Thus $\nu = 0 \iff \rho = 0$.

If $\rho > 0$, then the probability weight sequences equivalent to (w_k) are

$$p_k = \frac{t^k w_k}{\Phi(t)}, \qquad k \geq 0,$$

where $t > 0$ and $\Phi(t) < \infty$. The mean is $\Psi(t)$.

ν is the supremum of the means of all probability weight sequences equivalent to (w_k).

If $\nu \geq 1$, let τ be the unique number in $[0, \rho]$ such that $\Psi(\tau) = 1$, i.e.

$$t\Phi'(t) = \Phi(t)$$

.

If $0 \leq \nu < 1$, let $\tau := \rho$.

In both cases, τ is the minimum point in $[0, \rho]$, or $[0, \infty)$, of $\Phi(t)/t$.

If $\nu \geq 1$, let τ be the unique number in $[0, \rho]$ such that $\Psi(\tau) = 1$, i.e.

$$t\Phi'(t) = \Phi(t)$$

.

If $0 \leq \nu < 1$, let $\tau := \rho$.

In both cases, τ is the minimum point in $[0, \rho]$, or $[0, \infty)$, of $\Phi(t)/t$.

Let

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

(π_k) is a probability weight sequence. Its mean is $\mu = \Psi(\tau)$. Its variance is

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

The three cases again

I. $\nu \geq 1$. Then $0 < \tau < \infty$ and $\tau \leq \rho \leq \infty$. 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 $\rho/\tau \geq 1$. (Exponential moment iff $\rho/\tau > 1$ iff $\nu > 1$.)

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.

The infinite limit tree

Let ξ be a random variable with distribution $(\pi_k)_{k=0}^{\infty}$:

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

Assume that $\mu := \mathbb{E}\,\xi = \sum_k k\pi_k \le 1$.

There are *normal* and *special* nodes. The root is special.

Normal nodes have offspring (outdegree) as copies of ξ.
Special nodes have offspring as copies of $\widehat{\xi}$, where

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

When a special node gets a finite number of children, one of its children is selected uniformly at random and is special.
All other children are normal.

(Based on Kesten ($\mu = 1$) + Jonsson & Stefánsson ($\mu < 1$).)

The spine

The special nodes form a path from the root; we call this path the *spine* of $\widehat{\mathcal{T}}$.

There are three cases:

I. $\mu = 1$ (the critical case).

$\widehat{\xi} < \infty$ a.s. Each special node has a special child and the spine is an infinite path. Each outdegree in $\widehat{\mathcal{T}}$ is finite, so the tree is infinite but locally finite.

The distribution of $\widehat{\xi}$ is the *size-biased* distribution of ξ, and $\widehat{\mathcal{T}}$ is the size-biased Galton–Watson tree defined by Kesten.

I. $\mu = 1$ (the critical case).

$\widehat{\xi} < \infty$ a.s. Each special node has a special child and the spine is an infinite path. Each outdegree in $\widehat{\mathcal{T}}$ is finite, so the tree is infinite but locally finite.

The distribution of $\widehat{\xi}$ is the *size-biased* distribution of ξ, and $\widehat{\mathcal{T}}$ is the size-biased Galton–Watson tree defined by Kesten.

Alternative construction: Start with the spine (an infinite path from the root). 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 $\{0, \ldots, k\}$.

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.

II. $0 < \mu < 1$ (the subcritical case).

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 $\mathrm{Ge}(1 - \mu)$,

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

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

II. $0 < \mu < 1$ (the subcritical case).

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 $\mathrm{Ge}(1 - \mu)$,

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

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

Alternative construction: Start with a spine of random length L. Attach 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.

III. $\mu = 0$ ($\rho = \nu = \tau = 0$. Not Galton–Watson tree.)

A degenerate special case of II.

A normal node has 0 children. A special node has ∞ children, all normal.

The root is the only special node. The spine has length $L = 1$.

The tree $\widehat{\mathcal{T}}$ is an infinite star. (No randomness.)

III. $\mu = 0$ ($\rho = \nu = \tau = 0$. Not Galton–Watson tree.)

A degenerate special case of II.

A normal node has 0 children. A special node has ∞ children, all normal.

The root is the only special node. The spine has length $L = 1$.

The tree $\widehat{\mathcal{T}}$ is an infinite star. (No randomness.)

Example

$w_k = k!$.

In the limit, $\mathcal{T}_n$ has Po(1) branches of length 2; all others have length 1.

Node degrees

Theorem

As $n \to \infty$,

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

Consequently,

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

where $\widehat{\xi}$ *is a random variable in* $\{0, 1, \ldots, \infty\}$.

Note that the sum $\sum_0^\infty d\pi_d = \mu$ of the limiting probabilities in may be less than 1; in that case we do not have convergence to a proper finite random variable.

If we instead take a random node, we obtain a different limit distribution, viz. (π_k).

Theorem

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

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

Consequently,

$$d^+_{\mathcal{T}_n}(v) \xrightarrow{\mathrm{d}} \xi,$$

When $\nu > 1$, this was proved by Otter (1949).

The maximum degree

Denote the maximum outdegree in the tree $\mathcal{T}_n$ by $Y_{(1)}$.

Ia: $\nu > 1$. $(0 < \tau < \rho \leq \infty.)$

A logarithmic bound due to Meir and Moon (1990):

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

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

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

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

Ia: $\nu > 1$. $(0 < \tau < \rho \le \infty.)$

A logarithmic bound due to Meir and Moon (1990):

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

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

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

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

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 $k(n)$. (No limit distribution exists.)

Ia: $\nu > 1$. $(0 < \tau < \rho \le \infty.)$

A logarithmic bound due to Meir and Moon (1990):

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

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

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

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

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 $k(n)$. (No limit distribution exists.)

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.

Iα: $\nu \geq 1$ and $\sigma^2 < \infty$.

$Y_{(1)}$ is asymptotically distributed as the maximum of n i.i.d. copies of ξ; this holds in the strong sense that the total variation distance tends to 0.

Since $\mathbb{E}\,\xi^2 < \infty$, this implies in particular

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

Iβ: $\nu \geq 1$ and $\sigma^2 = \infty$

Then

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

and this is (more or less) best possible.

II: $0 < \nu < 1$

If further (w_k) satisfies an asymptotic power-law $w_k \sim ck^{-\beta}$ as $k \to \infty$, then

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

while the second largest node degree $Y_{(2)} = o_{\mathrm{p}}(n)$. (Jonsson & Stefánsson)

II: $0 < \nu < 1$

If further (w_k) satisfies an asymptotic power-law $w_k \sim ck^{-\beta}$ as $k \to \infty$, then

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

while the second largest node degree $Y_{(2)} = o_{\mathrm{p}}(n)$. (Jonsson & Stefánsson)

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.

III: $\nu = \rho = 0$

This is similar to case II.

In some regular cases we have $Y_{(1)} = n + o_{\mathrm{p}}(n)$, and then necessarily $Y_{(2)} = o_{\mathrm{p}}(n)$.

But there are exceptions in other cases with an irregular weight sequence.

Balls-in-boxes

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

$$\mathcal{B}_{m,n} := \Big\{(y_1, \dots, y_n) : y_i \geq 0, \ \sum_{i=1}^{n} y_i = m\Big\},$$

where y_i counts the number of balls in box i.

The weight of an allocation $\mathbf{y} = (y_1, \dots, y_n)$ is

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

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

We can replace the weight sequence by an equivalent weight sequence for the balls-in-boxes model just as we did for the random trees above.

Example: probability weights

If (w_k) is a probability weight sequence, let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with the distribution (w_k).

Then, $B_{m,n}$ has the same distribution as $(\xi_1, \ldots, \xi_n)$ conditioned on $\sum_{i=1}^{n} \xi_i = m$.

(This construction of a random allocation $B_{m,n}$ is used by Kolchin (1984) and called the *general scheme of allocation*.)

Random allocations and trees

If T is a tree with $|T| = n$, then its degree sequence (in depth-first order, say) is an allocation in $\mathcal{B}_{n-1,n}$, with the same weight as the tree. Moreover, a converse holds by the following well-known lemma.

Lemma

If $(d_1, \ldots, d_n) \in \mathcal{B}_{n-1,n}$, then exactly one of the n cyclic shifts of $(d_1, \ldots, d_n)$ is the degree sequence of a tree T with $|T| = n$.

Other examples of random allocations:

Different types of random forests with a given number of components, with each component regarded as a box, and each vertex as a ball.

The classical Maxwell–Bolzmann, Bose–Einstein and Fermi–Dirac statistics in statistical mechanics.

Asymptotics for balls-in-boxes

Suppose that $n \to \infty$ and $m = m(n)$ with $m/n \to \lambda$ with $0 \le \lambda < \sup\{i : w_i > 0\} \le \infty$.

I. If $\lambda \le \nu$, let τ be the unique number in $[0, \rho]$ such that $\Psi(\tau) = \lambda$.

II. If $\lambda > \nu$, let $\tau := \rho$.

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

Asymptotics for balls-in-boxes

Suppose that $n \to \infty$ and $m = m(n)$ with $m/n \to \lambda$ with $0 \le \lambda < \sup\{i : w_i > 0\} \le \infty$.

I. If $\lambda \le \nu$, let τ be the unique number in $[0, \rho]$ such that $\Psi(\tau) = \lambda$.

II. If $\lambda > \nu$, let $\tau := \rho$.

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

Remark. For trees, $m = n - 1$ and thus $\lambda = 1$.

Let

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

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

$$\mu = \Psi(\tau) = \min(\lambda, \nu)$$

and variance $\sigma^2 = \tau\Psi'(\tau) \leq \infty$.

Theorem

Let $N_k(B_{m,n})$ be the number of boxes with exactly k balls in the allocation $B_{m,n}$.

For every $k \geq 0$,

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

Theorem

Let $N_k(B_{m,n})$ be the number of boxes with exactly k balls in the allocation $B_{m,n}$.

For every $k \geq 0$,

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

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 $\lambda = \nu$.

Condensation

There are roughly $n\pi_k$ boxes with k balls in a random allocation $B_{m,n}$. Summing this approximation over all k we would get n boxes (as we should) with a total number of balls $n\sum_{k=0}^{\infty} k\pi_k = n\mu = n\min(\lambda, \nu)$.

However, the total number of balls is $m \approx n\lambda$, so in the case $\lambda > \nu$, there are 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 for fixed k.

In physical terminology, this can be regarded as condensation of part of the mass (= balls).

The simplest case is that there is a single giant box with $\approx (\lambda - \nu)n$ balls. This happens in the important case of a power-law weight sequence: $w_k \sim ck^{-\beta}$ as $k \to \infty$ for some $c > 0$ (Jonsson & Stefánsson).

However, there are also other possibilities when the weight sequence is less regular.

Recall that for simply generated random trees, which correspond to balls-in-boxes with $\lambda = 1$, there is a related form of condensation when $\nu < \lambda = 1$; in this case the condensation appears as a node of infinite degree in the random limit tree $\widehat{\mathcal{T}}$ of type II or III.