Branching processes and random trees

Svante Janson

Modern perspectives of branching in probability
Münster, 27 September, 2017

Simply generated trees

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

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

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.

(Not if grows w_k grows too rapidly, such as $k!$.)

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.

(Not if grows w_k grows too rapidly, such as $k!$.)

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.
Example: $(k+1)^{-3}/\zeta(3)$.

III. Simply generated tree, not equivalent to any Galton–Watson tree.
Example: $w_k = k!$.

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

Some of these results can be extended 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

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.

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 $\Phi(t) < \infty$.
Denote the mean $\sum_k k p_k$ of this distribution by $\Psi(t)$.

Let

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

In words:

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

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. $\nu = \rho = 0$.

In particular, $\nu = 0 \iff \rho = 0$.

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

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 $\Psi(\tau) = 1$.

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).

Fringe trees

More generally:

Given a tree T, let T_v be the fringe tree at v, i.e., the subtree rooted at v, and let T^* be the fringe tree at a uniformly random node v.

Theorem

Let $\mathcal{T}_n^*$, be the random fringe tree of $\mathcal{T}_n$. Then, as $n \to \infty$, $\mathcal{T}_n^*$ converges in distribution to the (unconditioned) Galton–Watson tree $\mathcal{T}$ with offspring distribution π, i.e., for any fixed (finite) tree T,

$$\mathbb{P}(\mathcal{T}_n^* = T) \to \mathbb{P}(\mathcal{T} = T).$$

For $\mu = 1$, i.e., critical Galton–Watson trees, explicit in Aldous (1991), referring to Kolchin (1986).

Extended fringe trees

Even more generally:

Define the extended fringe tree T^{**} by adding also the mother of v, with its descendents, the grandmother, and so on, i.e., by considering T "shifted" with centre at the random node v.

Theorem

The extended fringe tree $\mathcal{T}_n^{**}$ converges to a random tree $\hat{\widehat{\mathcal{T}}}$ constructed as follows:

(i). *If $\mu = 1$ (critical case), add an infinite spine backwards from the root of $\widehat{\mathcal{T}}$; let each node in the spine be special (with a $\widehat{\xi}$ offspring distribution), and add independent forward $\mathcal{T}$ Galton–Watson trees to all their children.*

(ii). *If $\mu < 1$ (subcritical case), add a spine backwards with special nodes, until the first node with an infinite number of children appears; then continue backwards and add more special nodes until another node with an infinite number of children appears; stop and discard the last node. Then add independent forward Galton–Watson trees as above.*

Quenched version

Let $n_T(\mathcal{T}_n)$ be the number of subtrees of $\mathcal{T}_n$ that are isomorphic to T.

Theorem

Assume $\mu := \mathbb{E}\,\xi = 1$ *and* $\operatorname{Var}\xi < \infty$.

(i). *For any fixed tree* T,

$$\frac{n_T(\mathcal{T}_n)}{n} = \mathbb{P}(\mathcal{T}_n^* = T \mid \mathcal{T}_n) \xrightarrow{\mathrm{p}} \mathbb{P}(\mathcal{T} = T).$$

(ii).

$$\frac{n_T(\mathcal{T}_n) - n\,\mathbb{P}(\mathcal{T} = T)}{\sqrt{n}} \xrightarrow{\mathrm{d}} N(0, \gamma^2)$$

for some $\gamma^2 = \gamma_T^2 < \infty$.

Part II – general CMJ branching processes

A Crump–Mode–Jagers process is a branching process in continuous time, where each individual has a random number N of children (with $0 \le N \le \infty$), born at times when the individual itself has ages $\xi_1 \le \xi_2 \dots$; these are also random (and may be dependent in any way). (Technically, best seen as a point process.)

Different individuals have i.i.d. life stories.

Let $\mathcal{T}_\infty$ be the complete family tree of the process, starting with a single individual born at time 0, and let $\mathcal{T}_t$ be the subtree of individuals born up to time t.

We are interested in cases when $\mathcal{T}_\infty$ is infinite but each $\mathcal{T}_t$ a.s. is finite. Thus assume $\mathbb{E}\, N > 1$ (supercritical case) and assume for simplicity $N \ge 1$.

Let $Z_t := |\mathcal{T}_t|$, the number of individuals at time t.

More generally, a *characteristic* of an individual is a random function ϕ of the age $t \geq 0$; we assume $\phi(t) \geq 0$ and $\phi \in D[0, \infty)$. Let, where σ_x is the time individual x is born,

$$Z_t^\phi := \sum_{x : \sigma_x \leq t} \phi_x(t - \sigma_x),$$

the total characteristic of all individuals existing at time t.

Known results (Crump, Mode, Jagers, Nerman, et al):

Assume some technical conditions.

- There exists $\alpha > 0$ (the Malthusian parameter), such that
$$e^{-\alpha t} Z_t \xrightarrow{\text{a.s.}} W$$
for some random variable $W > 0$.
- More generally,
$$e^{-\alpha t} Z_t^{\phi} \xrightarrow{\text{a.s.}} m_\phi W$$
for a constant $m_\phi > 0$.
- Hence
$$Z_t^{\phi} / Z_t \xrightarrow{\text{a.s.}} m_\phi.$$

Fix a characteristic ψ.

Define $\tau_n := \inf\{t : Z_t^{\psi} \geq n\}$ and $T_n := \mathcal{T}_{\tau(n)}$.

Main example : $\psi = 1$.
T_n has n nodes (if birth times are a.s. distinct).

Fringe trees

Theorem

(i). *(Annealed version.) The random fringe tree T_n^* converges in distribution to the random tree $\overline{\mathcal{T}} = \mathcal{T}_\tau$, where $\tau \sim \mathrm{Exp}(\alpha)$ is a random time, independent of $\mathcal{T}$.*

(ii). *(Quenched version.) For every finite tree T,*

$$\mathbb{P}(T_n^* = T \mid T_n) = \frac{|\{v : T_{n;v} = T\}|}{|T_n|} \xrightarrow{\text{a.s.}} \mathbb{P}(\overline{\mathcal{T}} = T).$$

Extended fringe trees

Define a sin-tree $\tilde{T}$ as follows:

- Start with a copy of the branching process, starting with o born at time 0.
- Give o an infinite line of ancestors, $o^{(1)}, o^{(2)}, \dots$, each having a modified life history where one child is distinguished, and called *heir*, and the probability is weighted by a factor $e^{-\alpha\xi}$, where ξ is the time the heir is born.
- Let the heir of $o^{(i)}$ be $o^{(i-1)}$. This defines the (negative) birth times of the ancestors. Let all other children of the ancestors start new copies of $\mathcal{T}$.

Theorem

(i). *(Annealed.) The extended fringe tree of* T_n *converges in distribution to* $\tilde{T}$.

(ii). *(Quenched.) This holds also conditioned on* T_n*, a.s.*

Random recursive tree

Example

Children born with independent Exp(1) waiting times, i.e., according to a Poisson process with rate 1. The branching process is the Yule process.

T_n is the random recursive tree. The next node is added as a child to a uniformly chosen node.

General preferential attachment tree

Example

Let $(w_k)_0^\infty$ be a sequence of weights with $w_k \geq 0$ and $w_0 > 0$.

Grow a tree by choosing the mother of each new node randomly with probability proportional to w_d where d is the outdegree (number of existing children).

T_n where the waiting time for child k is $\mathrm{Exp}(w_{k-1})$.

General preferential attachment tree

Example

Let $(w_k)_0^\infty$ be a sequence of weights with $w_k \geq 0$ and $w_0 > 0$.

Grow a tree by choosing the mother of each new node randomly with probability proportional to w_d where d is the outdegree (number of existing children).

T_n where the waiting time for child k is $\mathrm{Exp}(w_{k-1})$.

Standard case: $w_k = k + 1$.

General preferential attachment tree

Example

Let $(w_k)_0^\infty$ be a sequence of weights with $w_k \geq 0$ and $w_0 > 0$.

Grow a tree by choosing the mother of each new node randomly with probability proportional to w_d where d is the outdegree (number of existing children).

T_n where the waiting time for child k is $\mathrm{Exp}(w_{k-1})$.

Standard case: $w_k = k + 1$.

Linear case: $w_k = \chi k + \rho$.

Binary search tree

Example

Each individual gets two children, one left and one right; each after an Exp(1) time (independent).

m-ary search tree (with external nodes)

Example

$m \geq 2$ fixed.

A newborn has 0 "keys". It get $m - 1$ keys after independent waiting times $Y_1, \ldots, Y_{m-1}$ with $Y_i \sim \text{Exp}(i)$. When the last key arrives, m children are born.

$\psi(t)$ is the number of keys at time t.

Fragmentation trees

Example

Start with an object of mass $x_0 > 0$; break it ito b pieces with masses $V_1 x_0, \ldots, V_b x_0$, where $(V_1, \ldots, V_b)$ is a random vector with $V_i \geq 0$ and $\sum_i V_i = 1$. Continue recursively with each piece of mass $\geq x_1$, using a new copy of $(V_1, \ldots, V_b)$ each time.

Regard the fragments of masses $\geq x_1$ seen during the process as nodes in the *fragmentation tree*.

CMJ process: An individual has b children, born at times $\xi_1, \ldots, \xi_b$ with $\xi_i := -\log V_i$.

The fragmentation tree is the tree $\mathcal{T}_{\log(x_0/x_1)}$.