Applications and Applied Mathematics: An International Journal (AAM)

Volume 15 | Issue 2 | Article 5

12-2020

Statistics of Branched Populations Split into Different Types

Thierry E. Huillet
CY Cergy Paris Université

Follow this and additional works at: https://digitalcommons.pvamu.edu/aam

Part of the Biology Commons, Other Physical Sciences and Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation

Huillet, Thierry E. (2020). Statistics of Branched Populations Split into Different Types, Applications and Applied Mathematics: An International Journal (AAM), Vol. 15, Iss. 2, Article 5.
Available at: https://digitalcommons.pvamu.edu/aam/vol15/iss2/5

This Article is brought to you for free and open access by Digital Commons @PVAMU. It has been accepted for inclusion in Applications and Applied Mathematics: An International Journal (AAM) by an authorized editor of Digital Commons @PVAMU. For more information, please contact hvkoshy@pvamu.edu.

Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Vol. 15, Issue 2 (December 2020), pp. 764 – 800

Applications and Applied
Mathematics:
An International Journal
(AAM)

Statistics of Branched Populations Split into Different Types

Thierry E. Huillet

Laboratoire de Physique Théorique et Modélisation
CY Cergy Paris Université / CNRS UMR-8089
Site de Saint Martin, 2 Avenue Adolphe-Chauvin
95302 Cergy-Pontoise, France
Thierry.Huillet@cyu.fr

Received: January 3, 2019; Accepted: May 30, 2020

Abstract

Some population is made of n individuals that can be of P possible species (or types) at equilibrium. How are individuals scattered among types? We study two random scenarios of such species abundance distributions. In the first one, each species grows from independent founders according to a Galton-Watson branching process. When the number of founders P is either fixed or random (either Poisson or geometrically-distributed), a question raised is: given a population of n individuals as a whole, how does it split into the species types? This model is one pertaining to forests of Galton-Watson trees. A second scenario that we will address in a similar way deals with forests of increasing trees. Underlying this setup, the creation/annihilation of clusters (trees) is shown to result from a recursive nucleation/aggregation process as one additional individual is added to the total population.

Keywords: Branching processes; Galton-Watson trees; Increasing trees; Poisson and geometric forests of trees; Combinatorial probability; Canonical and grand-canonical species abundance distributions; Singularity analysis of large forests

MSC 2010 No.: 60J80, 92D25

1. Introduction

In the following, we will make the following tacit identifications. A branching tree is understood as a branched population of individuals of a given type or species, at equilibrium. The type of a branching tree is the one carried by its founder. By equilibrium, we mean the steady state after all productive individuals in the tree have exhausted their lifetimes. The node of a tree is thus an individual of this population. An internal node is a node (individual) that was once productive. A leaf is a node (individual) that went sterile. The size of this tree (its total progeny) is the number of individuals (nodes) constituting this population at equilibrium. We shall consider both Galton-Watson trees (Harris 1963) and increasing trees (Bergeron at al. 1992) with zero or one lifetimes.

Central to our purpose are also forests of trees. The number of trees in a forest corresponds to the number of species in a population made of many species. Given a forest (population) with n nodes (individuals) in total at equilibrium, how many constitutive trees (species) is it made of and what are their sizes? We shall consider the cases when there are either a fixed number p of species or when this number is random, say P, with P either Poisson or geometrically distributed. In this manuscript, we shall consider both forests of Galton-Watson trees and forests of increasing trees. Species abundance distribution problems of a similar flavor have been addressed in Engen (1974) and Engen (1978).

We start with Galton-Watson trees with a single founder for which the probability generating function $\Phi\left(z\right)$ of the total progeny is given by a functional equation. We compute the probability mass function of its total progeny when the branching mechanism is either binomial, Poisson or negative binomial distributed. In the supercritical cases for which there is a chance to observe a giant tree, an expansion of the extinction probability can be derived from such considerations. When dealing with Galton-Watson trees forests with a fixed number p of trees, we compute the canonical species occupancy distributions (with or without repetitions) given n nodes in total in a forest with p trees. We show that these probability mass functions are explicit in the three examples discussed above, as well as the typical mass function of each species. For general Galton-Watson trees, only a large n asymptotic estimate of these probabilities is available, as a result of the probability generating function $\Phi\left(z\right)$ of the total progeny having generically a (branch point) power singularity of order $-1/2$: the probability generating function $\Phi\left(z\right)$ of the total progeny takes a finite value at its singularity branch point.

Then, we move to a grand canonical situation when the number of trees P is random, either (a) Poisson or (b) geometrically distributed. A key issue in this setup is the computation of the joint probability to observe a forest (population) with p trees (species) and n nodes (individuals) in total. This allows to compute the conditional probability to observe a forest made of P_n trees, given it has n nodes in total.

We show that in the Poisson forest case (a), P_n converges in distribution, in the large population regime $n \to \infty$, to some P_∞ which is shifted Poisson distributed. The distribution of the number of size$-m$ trees are also investigated in this limit.

When dealing with the geometric forest case (b), we show that a phase transition takes place in the following sense: in a subcritical regime, P_n converges in distribution, as $n \to \infty$, to some P_∞ which is shifted "squared-geometric" distributed, whereas in a (super-) critical regime, $n^{-1}P_n$ converges almost surely, as $n \to \infty,$ to some fixed fraction. The origin of this critical behavior results from $\Phi(z)$ being finite at its singularity point.

In a second part of this manuscript, we address similar issues when the trees under concern are now random increasing trees. These are weighted random versions of enumerative increasing trees, first introduced in Bergeron et al. (1992). Increasing trees are those labeled trees whose nodes are labeled in increasing order from root to leaves; being more constrained, they are less likely to be observed than their free Galton-Watson counterparts. The probability generating function $\Phi(z)$ of the total progeny in an increasing tree setup is now amenable to the solution of a nonlinear differential equation. We compute the probability mass function of its total progeny when the branching mechanism is either binomial, Poisson or negative binomial distributed, in which cases the nonlinear differential equation can be solved explicitly. In the first two cases, the (polynomial and exponential) branching mechanisms have no singularity at finite distance of the origin and this induces a power singularity of positive order in the first case and logarithmic singularity in the second case for $\Phi(z)$: the probability generating functions of the total progeny diverges at its induced singularity point located at some finite distance of the origin. In the third negative binomial case, the branching mechanism has a singularity at finite distance of the origin and this induces a power singularity of negative order for $\Phi(z)$: the probability generating function of the total progeny is finite at its induced singularity point. In the increasing tree setup, the type of the singularity of $\Phi(z)$ depends on the branching mechanism, in contrast with the universal behavior (with power singularity of order $-1/2$) observed for Galton-Watson trees.

When dealing with increasing trees forests with a fixed number p of trees, we can compute the large n canonical species occupancy distributions (with or without repetitions) given n nodes in total in a forest with p trees.

When dealing with random forests with P increasing trees, we show that the joint probabilities to observe a forest (population) with p trees (species) and n nodes (individuals) in total can be obtained by recurrence when considering the transition from n to $n+1$. More specifically, suppose a given population of n individuals that are of p different types. A new individual pops in and either nucleates (forms a new species by itself) or aggregates q out of the p species to form a new population with $n+1$ individuals and $p-q+1$ species. This scenario of clusters formation indicates that the creation/deletion of clusters (trees) can be obtained from a recursive nucleation/aggregation model as one additional individual is added to the total population. It is specific to forests of increasing trees.

Due to the greatest variability of the singularity type of $\Phi(z)$ in the increasing tree context, we show that:

- for the binomial with degree d branching mechanism: the number P_n of trees in a forest with n nodes in total grows like $n^{1/d}$ in the Poisson forest case (a), whereas it always grows like n in the geometric forest case (b).

- for the Poisson branching mechanism: the number P_n of trees in a forest with n nodes in total grows like $\log n$ in the Poisson forest case (a), whereas it always grows like n in the geometric forest case (b).

- for the negative binomial branching mechanism: in the Poisson case (a), P_n converges in distribution, as $n \to \infty$, to some P_∞ which is shifted Poisson distributed, whereas, when dealing with the geometric forest case (b), a phase transition takes place: in a subcritical regime, P_n converges in distribution, as $n \to \infty$, to some P_∞ which is shifted 'squared-geometric' distributed, whereas in a (super-) critical regime, $n^{-1} P_n$ converges almost surely, as $n \to \infty$, to some fixed fraction. This critical behavior also results from $\Phi(z)$ being finite at its singularity point in this case.

The main constructive tools are Lagrange inversion formula and asymptotic singularity analysis of the coefficients of probability generating functions with power-logarithmic singularities of given orders (see the short Appendix).

2. Forests of Galton-Watson trees

In a discrete Galton-Watson (GW) process, some founder gives birth to a random number M of offspring at the next generation, each daughter proceeding similarly independently of its sisters, and so on till some daughter is found sterile. Let then

$$\mathbf{E}\left(z^M\right) = \phi(z), \tag{1}$$

be the branching probability generating function (pgf) of this GW process with $\phi(0) \neq 0$ (non-extinction possible). We assume ϕ (with $\phi(1) = 1$) has convergence radius $z_+ > 1$ (possibly with $z_+ = \infty$) and we let $\mu = \mathbf{E}(M) = \phi'(1) < \infty$ and $\sigma^2 = \text{Var}(M) < \infty$. We also let $\pi_k = \mathbf{P}(M = k)$. We avoid the trivial case where $\phi(z)$ is an affine function of z. Starting from a single founder, the birth and death process is iterated indefinitely, ending up in a population with a total random amount $\overline{N}(1)$ of descendants (the asymptotic total progeny of founder 1 at equilibrium). A productive individual in the tree will produce $M > 0$ offspring at each step, including itself, meaning that a productive individual passes to the next generation and generates $M - 1$ daughters in such a reproduction event. If $M = 1$, the reproduction event is reduced to itself in a self-regeneration process. As a result, the only individuals that one "sees" after the whole lifetime of the tree was exhausted are the leaves (corresponding to nodes at which the death event $M = 0$ takes place). Clearly then, denoting $\overline{N}_m(1)$ the number of nodes of the tree with outdegree m, the number of leaves $\overline{N}_0(1)$ with outdegree 0 obey

$$\overline{N}_0(1) = 1 + \sum_{m \geq 1} (m-1) \overline{N}_m(1),$$

together with of course $\overline{N}(1) = \sum_{m \geq 0} \overline{N}_m(1)$.

2.1. GW tree with a single founder and the survival probability (Harris 1963)

The GW process is subcritical if $\mu < 1$, critical if $\mu = 1$ and supercritical if $\mu > 1$. As indicated above, we let $\overline{N}(1)$ be the asymptotic **total progeny** of a single founder at equilibrium. Then, with $\overline{\pi}_n = \mathbf{P}\left(\overline{N}(1) = n\right)$ and $\mathbf{E}\left(z^{\overline{N}(1)}\right) = \sum_{n\geq 1} \overline{\pi}_n z^n$, the pgf $\Phi(z) = \mathbf{E}\left(z^{\overline{N}(1)}\right)$ obeys the functional equation of a Lagrangian distribution (see Consul and Famoye (2006)),

$$\Phi(z) = z\phi(\Phi(z)), \text{ with } \Phi(0) = 0. \tag{2}$$

We have $\Phi(1) = \mathbf{P}\left(\overline{N}(1) < \infty\right) = \rho_e$, the extinction probability of $N_t(1)$. It obeys

$$\rho_e = \phi(\rho_e), \tag{3}$$

with $\rho_e = 1$ if $\mu \leq 1$, $\overline{\rho}_e = 1 - \rho_e > 0$ if $\mu > 1$. If $\mu < 1$, $m = \mathbf{E}\overline{N}(1) = 1/(1-\mu) < \infty$, otherwise if $\mu \geq 1$, $m = \infty$. In the critical case when $\mu = 1$, $\overline{N}(1) < \infty$ with probability 1 but $m = \mathbf{E}\overline{N}(1) = \infty$ as a result of $\overline{N}(1)$ displaying heavy tails. In the supercritical case when $\mu > 1$, $m = \infty$ because with some positive probability $\overline{\rho}_e$, the tree is a giant tree with infinitely many nodes or branches (one more node than branches in a tree corresponding to the root).

Whenever one deals with a supercritical situation with $\rho_e = \Phi(1) < 1$, defining the pgf of $\widetilde{N}(1) = \overline{N}(1) \mid \overline{N}(1) < \infty$ to be

$$\widetilde{\Phi}(z) = \frac{\Phi(z) - \Phi(1)}{1 - \Phi(1)},$$

we have

$$\widetilde{\Phi}(z) = z\widetilde{\phi}\left(\widetilde{\Phi}(z)\right) \text{ and } \widetilde{\phi}(z) = \frac{\phi(z) - \rho_e}{1 - \rho_e},$$

where $\widetilde{\phi}(z)$ is the modified subcritical branching mechanism obeying $\widetilde{\mu} = \widetilde{\phi}'(1) = \phi'(\rho_e) < 1$. Conditioning a supercritical tree on being finite is amenable to a subcritical tree problem so with extinction probability 1. But this requires the computation of ρ_e which can be quite involved in general. Indeed however (with $[z^n] f(z)$ denoting the coefficient in front of z^n in the power-series expansion of $f(z)$ at 0), by Lagrange inversion formula

$$\overline{\pi}_n = [z^n]\Phi(z) = \mathbf{P}\left(\overline{N}(1) = n\right) = \frac{1}{n}\left[z^{n-1}\right]\phi(z)^n, \tag{4}$$

so that

$$\rho_e = \mathbf{P}\left(\overline{N}(1) < \infty\right) = \sum_{n\geq 1} \mathbf{P}\left(\overline{N}(1) = n\right) = \sum_{m\geq 0} \frac{1}{m+1}[z^m]\phi(z)^{m+1},$$

is the power series expansion of the extinction probability ρ_e in the supercritical case. There is an estimate of ρ_e when the GW process is nearly supercritical (μ slightly above 1). Let $\overline{\rho}_e = 1 - \rho_e$ be the survival probability and $f(z) = \phi(z) - z$, with

$$f(1) = 0, f'(1) = \mu - 1 \text{ and } f''(1) = \mathbf{E}(M(M-1)) = \sigma^2 + \mu^2 - \mu \underset{\mu \sim 1^+}{\sim} \sigma^2.$$

We have

$$\rho_e = \phi(\rho_e) \Leftrightarrow f(1 - \overline{\rho}_e) = 0.$$

As a result of

$$f\left(1-x\right)\sim f\left(1\right)-xf'\left(1\right)+\frac{1}{2}x^{2}f''\left(1\right),$$

we get the small survival probability estimate $\overline{\rho}_e \sim 2\left(\mu-1\right)/\sigma^2$ when the GW process is nearly supercritical. As a function of $\mu-1$, $\overline{\rho}_e$ is continuous at 0 ($\overline{\rho}_e=0$ if $\mu-1\leq 0$), but with a discontinuous slope. As $\mu\to\infty$, clearly $\overline{\rho}_e\to 1$.

A full power-series expansion of $\overline{\rho}_e$ in terms of $\mu-1>0$ can also be obtained as follows: define $\overline{\phi}\left(z\right)$ by $\phi\left(z\right)=1+\mu\left(z-1\right)+\overline{\phi}\left(1-z\right)$, so with $\overline{\phi}\left(0\right)=0$. The equation $\rho_e=\phi\left(\rho_e\right)$ becomes

$$\frac{\overline{\phi}\left(\overline{\rho}_e\right)}{\overline{\rho}_e}=\mu-1.$$

Lagrange inversion formula gives $\overline{\rho}_e=\sum_{n\geq 1}\rho_n\left(\mu-1\right)^n$ with

$$\rho_n=\frac{1}{n}\left[x^{n-1}\right]\left(\frac{\overline{\phi}\left(x\right)}{x^2}\right)^{-n}.$$

Note $\rho_1=2/\phi''\left(1\right)$ with $\phi''\left(1\right)\sim\sigma^2$ when μ is slightly above 1. To the first order in $\mu-1$, we recover $\overline{\rho}_e\sim 2\left(\mu-1\right)/\sigma^2$. The second-order coefficient is found to be $\rho_2=4/3\cdot\phi'''\left(1\right)/\phi''\left(1\right)^3$. Let us check these formulas on an explicit example.

Example 2.1.

If $\phi\left(z\right)=1/\left(1+\mu\left(1-z\right)\right)$, with $\mu>1$, the fixed point $\rho_e=1/\mu$ is explicitly found. Here $\overline{\phi}\left(x\right)/x^2=\mu^2/\left(1+\mu x\right)$ with $\rho_n=\mu^{-(n+1)}$. Thus, consistently, $\overline{\rho}_e=\sum_{n\geq 1}\rho_n\left(\mu-1\right)^n=1-1/\mu$ and, owing to $\phi''\left(1\right)=2\mu^2$ and $\phi'''\left(1\right)=6\mu^3$, $\rho_2=\mu^{-3}=4/3\cdot\phi'''\left(1\right)/\phi''\left(1\right)^3$.

2.2. GW tree with a fixed number $p>1$ of founders: canonical species abundance distributions

Let $\overline{N}\left(p\right)=\overline{N}_\infty\left(p\right)$ be now the total progeny of p independent founders, so with $\mathbf{E}\left(z^{\overline{N}(p)}\right)=\Phi\left(z\right)^p$. We view the progenies of each of the p founders as the progenies of p distinct species. Let $\left(\overline{N}_{n,p}\left(1\right),...,\overline{N}_{n,p}\left(p\right)\right)$ be the vector of the progenies given p distinct species and $\overline{N}\left(p\right)=n$ individuals in total. Then, with $\left(n_1,...,n_p\right)$ positive integers summing to n,

$$\mathbf{P}\left(\overline{N}_{n,p}\left(1\right)=n_1,...,\overline{N}_{n,p}\left(p\right)=n_p\right)=\frac{\prod_{q=1}^{p}\left[z_q^{n_q}\right]\Phi\left(z_q\right)}{\left[z^n\right]\Phi\left(z\right)^p}=\frac{\prod_{q=1}^{p}\overline{\pi}_{n_q}}{\mathbf{P}\left(\overline{N}\left(p\right)=n\right)}, \tag{5}$$

gives the joint species abundance distribution among types when n and p are fixed. This probability mass function is exchangeable. The typical one-dimensional marginal distribution in particular read ($n_1\in\left\{1,...,n-p+1\right\}$):

$$\begin{aligned}\mathbf{P}\left(\overline{N}_{n,p}\left(1\right)=n_1\right)&=\frac{\left[z^{n_1}\right]\Phi(z)\left[z^{n-n_1}\right]\Phi(z)^{p-1}}{\left[z^n\right]\Phi(z)^p}\\&=\frac{\mathbf{P}\left(\overline{N}(1)=n_1\right)\mathbf{P}\left(\overline{N}(p-1)=n-n_1\right)}{\mathbf{P}\left(\overline{N}(p)=n\right)},\end{aligned} \tag{6}$$

clearly with $\mathbf{E}\left(\overline{N}_{n,p}\left(1\right)\right)=n/p$.

A second important sampling formula is the following, dealing with repetitions: let $(P_{n,p}(m); m = 1, ..., n)$ be the number of species with m representatives in a population with total size n and p distinct species (the number of size$-m$ trees). Then, for all $p_1, .., p_n \geq 0$, non-negative integers, obeying $\sum_{m=1}^{n-(p-1)} p_m = p$, $\sum_{m=1}^{n-(p-1)} m p_m = n$ (the number of such p's is the number of partitions of n into p summands),

$$\mathbf{P}\left(P_{n,p}(1) = p_1, ..., P_{n,p}(n) = p_n\right) = \frac{\binom{p}{p_1 \cdots p_n} \prod_{m=1}^{n} \overline{\pi}_m^{p_m}}{[z^n]\, \Phi(z)^p}. \tag{7}$$

In all cases, we are left with the problem of computing, say the normalizing denominator of Equation (7), which can be obtained by Lagrange inversion formula as:

$$[z^n]\, \Phi(z)^p = \mathbf{P}\left(\overline{N}(p) = n\right) = \frac{p}{n}\left[z^{n-p}\right] \phi(z)^n \text{ for all } n \geq p. \tag{8}$$

Both abundance distributions (5) and (7) are explicit whenever one is able to compute $\mathbf{P}\left(\overline{N}(p) = n\right) = [z^n]\, \Phi(z)^p$ for all $n, p \leq n$, in particular $\overline{\pi}_m = \mathbf{P}\left(\overline{N}(1) = m\right) = [z^m]\, \Phi(z)$. This will be the case in the examples of the forthcoming section.

Remark 2.1.

Note in passing that $\mathbf{P}\left(\overline{N}(p) < \infty\right) = \rho_e(p) = \rho_e^p$. So

$$\mathbf{P}\left(\overline{N}(p) < \infty\right) = \sum_{n \geq p} \mathbf{P}\left(\overline{N}(p) = n\right) = p \sum_{m \geq 0} \frac{1}{m+p}\, [z^m]\, \phi(z)^{m+p},$$

is the power series expansion of the extinction probability $\rho_e(p)$ of a population with p species in the supercritical case.

2.3. Three explicit examples

- Let $M \sim$bin(d, α), $\alpha \in (0,1)$ and d integer ≥ 2, so with

$$\phi(z) = (1 - \alpha + \alpha z)^d, \tag{9}$$

with mean $\mu = d\alpha$. By Stirling formula, with $n \geq p$, with $z_c = \frac{(d-1)^{d-1}}{\alpha(1-\alpha)^{d-1} d^d} \geq 1$,

$$\mathbf{P}\left(\overline{N}(p) = n\right) = \frac{p}{n}\left[z^{n-p}\right] \phi(z)^n = \frac{p}{n}\binom{nd}{n-p} \alpha^{n-p} (1-\alpha)^{n(d-1)+p}$$
$$\sim \frac{1}{\sqrt{2\pi}} \frac{p}{n^{3/2}} \left(\frac{d-1}{d}\right)^{-1/2} \left(\frac{1-\alpha}{\alpha(d-1)}\right)^p z_c^{-n},$$

when p is fixed and $n \to \infty$. We note $z_c = w(\mu - 1)/\mu$, where $w(x) = (1 + x/(1-d))^{1-d}$, $w(x) \sim 1 + x$ as $x \to 0^+$, precising how z_c depends on $\mu - 1$. The function $z_c(\mu)$ attains its minimum 1 if $\mu = d\alpha = 1$ ($\alpha = 1/d$). If $\mu = d\alpha = 1$, therefore, (corresponding to the critical case)

$$\mathbf{P}\left(\overline{N}(p) = n\right) \underset{n \to \infty}{\sim} \frac{p}{\sqrt{2\pi(1-\alpha)}} n^{-3/2},$$

a pure power law (with tail index $1/2$) and with no geometric cutoff. The tail exponent being smaller than one, $\overline{N}\left(p\right)$ has no finite mean as expected. Note that this probability is proportional to p. Finally,

$$\begin{aligned}\rho_e\left(p\right) &= \mathbf{P}\left(\overline{N}\left(p\right) < \infty\right) = \textstyle\sum_{n\geq p}\mathbf{P}\left(\overline{N}\left(p\right) = n\right)\\ &= \textstyle\sum_{n\geq p}\frac{p}{n}\binom{nd}{n-p}\alpha^{n-p}\left(1-\alpha\right)^{n(d-1)+p}\\ &= p\left(1-\alpha\right)^{pd}\textstyle\sum_{m\geq 0}\frac{1}{m+p}\binom{(m+p)d}{m}\left(\alpha\left(1-\alpha\right)^{d-1}\right)^{m},\end{aligned}$$

is the series expansion of the extinction probability. This probability is 1 if $\alpha = \alpha_c = 1/d$. When $\alpha - \alpha_c$ is small and positive (the just supercritical case), with $\overline{\rho}_e\left(p\right) = 1 - \rho_e\left(p\right)$: $\overline{\rho}_e\left(p\right) \sim 2p\left(\mu - 1\right)/\sigma^2$. In the present case, $\overline{\rho}_e\left(p\right) \sim p\frac{2d^2}{d-1}\left(\alpha - \alpha_c\right)$.

Remark 2.2.

(i) The latter model is related to the Flory-Stockmayer binomial model (randomly branched polymers with degree-$(d+1)$ functional monomers, see Flory (1941a), Flory (1941b)), if $d = 2$, Stockmayer (1943) for any integer d and also Simkin and Roychowdhury (2011)): in this model with one founder $p = 1$, the $\Phi\left(z\right)$ obtained above from the bin(d, α) generating model ϕ is in fact the pgf of first generation polymers. Here, each monomer with d functional units (arms) is identified to a node of a Galton-Watson tree. Independently of one another, each of the d functional units has a probability α to be attached to a second generation functional unit and so on. At generation 0 however, a seed monomer with full $d+1$ functional units gives birth to a random number (so with distribution bin$(d+1, \alpha)$) of first generation such polymers, all with pgf $\Phi\left(z\right)$. The true size of the Flory branched polymer has thus distribution given by the modified pgf

$$\Phi\left(z\right) \to z\left(1 - \alpha + \alpha\Phi\left(z\right)\right)^{d+1}.$$

This translates the fact that the seed monomer can have up to $d+1$ activated functional units whereas all its descendants only up to d, the first and subsequent generation trees growing away from the seed monomer, thereby presenting only d possible free arms. In the supercritical case with $d\alpha > 1$, there is a positive probability that the Flory tree (polymer) is a giant one with infinitely many monomers (the gelation transition).

(ii) With $\Phi\left(z\right)$ solving $\Phi\left(z\right) = z\left(1 - \alpha + \alpha\Phi\left(z\right)\right)^d$ and defining

$$\Phi_d\left(z\right) = \Phi\left(z^d\right)^{1/d},$$

we get that $\Phi_d\left(z\right)$ solves

$$\Phi_d\left(z\right) = z\left(1 - \alpha + \alpha\Phi_d\left(z\right)^d\right) = z\phi_d\left(\Phi_d\left(z\right)\right),$$

corresponding to the branching mechanism $\phi_d\left(z\right) = 1 - \alpha + \alpha z^d$. So $\Phi_d\left(z\right)$ is the pgf of the total progeny, say $\overline{N}_d\left(1\right)$, of a branching process whose offspring per capita is either d with probability α or 0 (with probability $1-\alpha$), so all or nothing. We clearly have $\mathbf{P}\left(\overline{N}_d\left(1\right) = n\right) > 0$ only for those $n = md + 1$, $m \geq 0$ (the number of tree branches being multiple of d), so $\Phi_d\left(z^{1/d}\right)$ is well-defined together with $\Phi\left(z\right) = \Phi_d\left(z^{1/d}\right)^d$. We can deduce the main properties of the new model generated by ϕ_d from the previous one generated by the binomial ϕ.

• Poisson model: M ~Poi(μ), mean $\mu > 0$. This model occurs in the Press-Schechter description of gravitational clustering (Sheth (1996)). Consider now the branching mechanism

$$\phi\left(z\right) = e^{-\mu(1-z)}. \tag{10}$$

This pgf can be obtained while putting $\alpha = \mu/d$ and letting $d \to \infty$ in the binomial model ϕ (the Poisson limit of the binomial distribution). Here

$$\begin{aligned}\left[z^{n}\right]\Phi\left(z\right)^{p} = \mathbf{P}\left(\overline{N}\left(p\right) = n\right) &= \frac{p}{n}\left[z^{n-p}\right]e^{-\mu n(1-z)}\\ &= \frac{p}{n}\frac{\left(\mu n\right)^{n-p}}{\left(n-p\right)!}e^{-\mu n},\end{aligned}$$

a Borel-Tanner distribution (Borel if $p = 1$) (see Tanner (1961)). By Stirling expansion, with $n \geq p$:

$$\mathbf{P}\left(\overline{N}\left(p\right) = n\right) \sim \frac{p}{\sqrt{2\pi}}\mu^{-p}\left(\frac{e^{\mu-1}}{\mu}\right)^{-n}n^{-3/2},$$

which in the critical case $\mu = 1$ reduces to $\mathbf{P}\left(\overline{N}\left(p\right) = n\right) \sim \frac{p}{\sqrt{2\pi}}n^{-3/2}$. We note $z_{c} = e^{\mu-1}/\mu = w\left(\mu-1\right)/\mu$, where $w\left(x\right) = e^{x}$, $w\left(x\right) \sim 1 + x$ as $x \to 0$.

Remark 2.3.

As an illustration of Equation (6) in this Poisson model context, with $n_1 \in \{1, ..., n-p+1\}$:

$$\begin{aligned}\mathbf{P}\left(\overline{N}_{n,p}\left(1\right) = n_1\right) &= \frac{\left[z^{n_1}\right]\Phi\left(z\right)\left[z^{n-n_1}\right]\Phi\left(z\right)^{p-1}}{\left[z^{n}\right]\Phi\left(z\right)^{p}}\\ &= \frac{p-1}{pn_1}\binom{n-p}{n_1-1}\left(\frac{n_1}{n-n_1}\right)^{n_1-1}\left(\frac{n-n_1}{n}\right)^{n-p-1},\end{aligned}$$

is the typical abundance distribution of species 1.

• Negative binomial model for M: Let $(\theta, 1/\lambda)$ be the shape and scale parameters of a Gamma random variable on the real half-line. Suppose M is now given as a Gamma($\theta, 1/\lambda$)-Poisson mixture. With $\lambda = \alpha/\beta, \theta > 0$, α, β such that $\alpha + \beta = 1$ and $\left[\theta\right]_{k} = \Gamma\left(\theta+k\right)/\Gamma\left(\theta\right) = \theta\left(\theta+1\right)...\left(\theta+k-1\right)$,

$$\begin{aligned}&\phi\left(z\right) = \left(1+\lambda\left(1-z\right)\right)^{-\theta} = \left(\tfrac{1-\alpha z}{\beta}\right)^{-\theta},\\ &\pi_{k} = \left[z^{k}\right]\phi\left(z\right) = \beta^{\theta}\left[\theta\right]_{k}z_{+}^{k}/k! \underset{k\to\infty}{\sim} \beta^{\theta}k^{\theta-1}z_{+}^{k}/\Gamma\left(\theta\right),\end{aligned} \tag{11}$$

where $z_{+} = 1/\alpha$. M has mean $\mu = \theta\lambda$. When $\theta = 1$, the mixing distribution is Exp($1/\lambda$) distributed and $\phi\left(z\right)$ is the pgf of a geometric random variable with success parameter α.

By Stirling formula, with $n \geq p$,

$$\begin{aligned}&\mathbf{P}\left(\overline{N}\left(p\right) = n\right) = \tfrac{p}{n}\left[z^{n-p}\right]\phi\left(z\right)^{n} = \tfrac{p}{n}\beta^{n\theta}\alpha^{n-p}\tfrac{\left[n\theta\right]_{n-p}}{(n-p)!}\\ &\sim \tfrac{p}{\sqrt{2\pi}}\left(\alpha\left(\theta+1\right)\right)^{-p}\left(\tfrac{\theta+1}{\theta}\right)^{-1/2}\left(\tfrac{\theta^{\theta}}{\alpha\beta^{\theta}\left(\theta+1\right)^{\theta+1}}\right)^{-n}n^{-3/2}.\end{aligned}$$

We note $z_c = \theta^\theta / \left(\alpha\beta^\theta \left(\theta+1\right)^{\theta+1}\right) = w\left(\mu-1\right)/\mu$, where $w\left(x\right) = \left(1 + x/\left(\theta+1\right)\right)^{\theta+1}$, $w\left(x\right) \sim 1 + x$ as $x \to 0$.

If $\mu = \theta\lambda = \theta\alpha/\beta = 1$ (critical case)

$$\mathbf{P}\left(\overline{N}\left(p\right) = n\right) \underset{n\to\infty}{\sim} \frac{p}{\sqrt{2\pi\left(1+\lambda\right)}} n^{-3/2},$$

a pure power law with no geometric cutoff. A striking feature is that it has the same shape as in the binomial case, up to a constant. Consequently,

$$\begin{aligned}\rho_e\left(p\right) &= \mathbf{P}\left(\overline{N}\left(p\right) < \infty\right) = \textstyle\sum_{n\geq p} \mathbf{P}\left(\overline{N}\left(p\right) = n\right) \\ &= \tfrac{p}{\alpha^p} \textstyle\sum_{n\geq p} \tfrac{1}{n} \left(\alpha\beta^\theta\right)^n \tfrac{\left[n\theta\right]_{n-p}}{\left(n-p\right)!} \\ &= p\beta^{p\theta} \textstyle\sum_{m\geq 0} \tfrac{\left(\alpha\beta^\theta\right)^m}{m+p} \tfrac{\left[n\theta\right]_m}{m!},\end{aligned}$$

is the series expansion of the extinction probability. This probability is 1 if $\theta = \beta/\alpha$ (or $\alpha = 1/\left(1+\theta\right)$) showing that

$$\sum_{m\geq 0} \frac{z^m}{m+p} \frac{\left[n\theta\right]_m}{m!} = \frac{1}{p\left(z\left(1+\theta\right)\right)^p},$$

if the latter series converges. Thus, $\mathbf{P}\left(\overline{N}\left(p\right) < \infty\right) = 1$ if $\theta \leq \beta/\alpha$ and if $\theta > \beta/\alpha$:

$$\mathbf{P}\left(\overline{N}\left(p\right) < \infty\right) = \frac{p\beta^{p\theta}}{p\left(\alpha\beta^\theta\left(1+\theta\right)\right)^p} = \left(\alpha\left(1+\theta\right)\right)^{-p} = \rho_e\left(p\right).$$

With $\theta_c = \beta/\alpha$, if $\theta - \theta_c$ is small,

$$\rho_e\left(p\right) = 1 - p\alpha\left(\theta-\theta_c\right) + pO\left(\left(\theta-\theta_c\right)^2\right).$$

2.4. GW tree with a random number P of founders

So far, we assumed that the number of species p was known. We shall now randomize p, so deal with a random number P of founders with pgf $\phi_0\left(z\right) = \mathbf{E}\left(z^P\right)$, a classical issue in the grand canonical ensemble (Demetrius (1983)). We shall assume that P has a finite mean μ_0.

Let then $\mathcal{N} = \overline{N}_\infty\left(P\right)$ be the total progeny of a population with P founders. We have $\mathbf{E}\left(z^{\mathcal{N}}\right) = \Psi\left(z\right) = \phi_0\left(\Phi\left(z\right)\right)$ and $\mathbf{P}\left(\mathcal{N} < \infty\right) = \phi_0\left(\rho_e\right)$. By Lagrange inversion formula,

$$\left[z^n\right]\Psi\left(z\right) = \mathbf{P}\left(\mathcal{N} = n\right) = \frac{1}{n}\left[z^{n-1}\right]\left(\phi_0'\left(z\right)\phi\left(z\right)^n\right). \tag{12}$$

With $\pi_m^0 = \mathbf{P}\left(P = m\right)$, $\phi_0\left(z\right) = \sum_{m\geq 0}\pi_m^0 z^m$ entails $\phi_0'\left(z\right) = \sum_{m\geq 1} m\pi_m^0 z^{m-1} = z^{-1}\sum_{m\geq 1} m\pi_m^0 z^m = \mu_0 z^{-1}\phi_0^*\left(z\right)$, where $\phi_0^*\left(z\right)$ is the pgf of a size-biased version P^* of P :

$$\phi_0^*\left(z\right) = \mathbf{E}\left(z^{P_*}\right) = z\phi_0'\left(z\right)/\mu_0.$$

We let $\pi_m^* = m\pi_m^0/\mu_0$ the probability system of $P^* \geq 1$. So $\mathbf{P}\left(\mathcal{N} = 0\right) = \pi_0^0$ and for $n \geq 1$,

$$\left[z^n\right]\Psi\left(z\right) = \mathbf{P}\left(\mathcal{N} = n\right) = \frac{\mu_0}{n}\left[z^n\right]\left(\phi_0^*\left(z\right)\phi\left(z\right)^n\right).$$

Thus, as required, by the convolution formula and Equation (8),

$$\mathbf{P}(\mathcal{N}=n)=\frac{\mu_0}{n}\sum_{p=1}^{n}\pi_p^*\cdot\left[z^{n-p}\right]\phi(z)^n$$

$$=\mu_0\sum_{p=1}^{n}\frac{\pi_p^*}{p}\mathbf{P}\left(\overline{N}(p)=n\right)=\sum_{p=1}^{n}\pi_p^0\mathbf{P}\left(\overline{N}(p)=n\right).$$

2.5. General GW tree case: back to $p=1$

The above three models for ϕ are some of the rare ones for which $\overline{\pi}_n=\mathbf{P}\left(\overline{N}(1)=n\right)$ can be explicitly computed. However, for a general (aperiodic and different from an affine function, see Remark below) ϕ obeying: ϕ has convergence radius $z_+>1$ (possibly $z_+=\infty$) and $\pi_0>0$, a similar large n estimate can be obtained in general. For such ϕ's indeed, the unique positive real root to the equation

$$\phi(\tau)-\tau\phi'(\tau)=0, \tag{13}$$

exists, with $\rho_e=1<\tau<z_+$ if $\mu<1$, $\tau=1$ if $\mu=1$ and $\rho_e<\tau<1<z_+$ if $\mu>1$.

Remark 2.4.

When $\phi(z)=\overline{\alpha}+\alpha z$ is affine, the number τ below is rejected at ∞ and the following analysis of the corresponding $\Phi(z)$ is invalid. This case deserves a special treatment (see below).

The point $(\tau,\phi(\tau))$ is indeed the tangency point to the curve $\phi(z)$ of a straight line passing through the origin $(0,0)$. Let then $z_c=\tau/\phi(\tau)=1/\phi'(\tau)\geq 1$. The searched $\Phi(z)$ solves $\psi(\Phi(z))=z$, where $\psi(z)=z/\phi(z)$ obeys $\psi(\tau)=z_c$, $\psi'(\tau)=0$ and $\psi''(\tau)=-\frac{\tau\phi''(\tau)}{\phi(\tau)^2}<\infty$. Thus, $\psi(z)\sim z_c+\frac{1}{2}\psi''(\tau)(z-\tau)^2$ else $z\sim z_c+\frac{1}{2}\psi''(\tau)(\Phi(z)-\tau)^2$ (a branch-point singularity). It follows that $\Phi(z)$ displays a dominant power-singularity of order $-1/2$ at z_c with $\Phi(z_c)=\tau$ in the sense

$$\Phi(z)\underset{z\to z_c}{\sim}\tau-\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\left(1-z/z_c\right)^{1/2}. \tag{14}$$

By singularity analysis therefore (see Flajolet and Odlyzko (1990) and Flajolet and Sedgewick (1993) and the Appendix),

$$\mathbf{P}\left(\overline{N}(1)=n\right)=[z^n]\Phi(z)\underset{n\to\infty}{\sim}\sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}}n^{-3/2}z_c^{-n}, \tag{15}$$

to the dominant order in n. When $\mu=\phi'(1)\to 1$ (critical case) then both τ and $z_c\to 1$ and the above estimate boils down to a pure power-law with $[z^n]\Phi(z)\underset{n\to\infty}{\sim}\frac{1}{\sqrt{2\pi\phi''(1)}}n^{-3/2}$. It can more precisely be checked that when $|\mu-1|\ll 1$, $z_c^{-1}\sim 1-(\mu-1)^2$.

Note finally that with $F(\lambda)=\log\phi\left(e^{-\lambda}\right)$ the log-Laplace transform of M, $\tau>0$ is also the solution to $F'(-\log\tau)=1$.

A power-series expansion of τ and z_c in terms of the variable $\mu - 1$ can formally be obtained. Define $\overline{\phi}(z)$ by $\phi(z) = 1 + \mu(z-1) + \overline{\phi}(1-z)$, so with $\overline{\phi}(0) = 0$. With $\overline{\tau} = 1 - \tau$, the equation $\phi(\tau) - \tau\phi'(\tau) = 0$ giving τ becomes

$$\delta(\overline{\tau}) = \overline{\phi}(\overline{\tau}) + (1-\overline{\tau})\overline{\phi}'(\overline{\tau}) = \mu - 1.$$

By Lagrange inversion formula, we get:

1/ $\overline{\tau} = \overline{\tau}(\mu-1) = \sum_{n\geq1}\tau_n(\mu-1)^n$, where

$$\tau_n = \frac{1}{n}\left[x^{n-1}\right]\left(\frac{\delta(x)}{x}\right)^{-n}.$$

2/ $z_c = 1/\phi'(1-\overline{\tau}) = z(\overline{\tau}) = z_c(\mu-1) = \sum_{n\geq1}z_n(\mu-1)^n$, where

$$z_n = \frac{1}{n}\left[x^{n-1}\right]\left(z'(x)\frac{\delta(x)}{x}\right)^{-n}.$$

Example 2.2.

Let us briefly work out the explicit geometric case, where $\phi(z) = \beta/(1-\alpha z)$, with convergence radius $z_+ = 1/\alpha$. The rv M has mean $\mu = \alpha/\beta$ and variance $\sigma^2 = \alpha/\beta^2 = \mu/\beta$.

If $\mu < 1$ $(\alpha < 1/2)$: $\rho_e = 1 < \tau = 1/(2\alpha) < z_+ = 1/\alpha$. We have $\phi(\tau) = 2\beta$ and $z_c = 1/(4\alpha\beta) > 1$. Note $z_c < z_+$.

If $\mu = 1$ $(\alpha = 1/2)$: $\rho_e = 1 = \tau < z_+ = 2$. We have $\phi(\tau) = 1$ and $z_c = 1$.

If $\mu > 1$ $(\alpha > 1/2)$: $\rho_e = \beta/\alpha < \tau = 1/(2\alpha) < 1 < z_+ = 1/\alpha < 2$. We have $\phi(\tau) = 2\beta$ and $z_c = 1/(4\alpha\beta) > 1$. Note $z_c \lesseqgtr z_+$ if $\alpha \lesseqgtr 3/4$ and $\rho_e = 1/\mu$ with $\overline{\rho}_e \sim 2(\mu-1)/\sigma^2$ as $\mu \to 1_+$ $(\alpha \to (1/2)_+)$.

2.5.1. *The pure power-law case (geometric tilting)*

Define the tilted new pgf $\Phi_c(z) = \Phi(zz_c)/\Phi(z_c)$ and let $\overline{N}_c(1)$ be the random variable such that $\Phi_c(z) = \mathbf{E}\left(z^{\overline{N}_c(1)}\right)$. With $\phi_c(z) = z_c\phi(\tau z)/\tau = \phi(\tau z)/\phi(\tau)$ defining a new rescaled branching pgf, we have

$$\Phi_c(z) = z\phi_c(\Phi_c(z)). \tag{16}$$

Thus, $\overline{N}_c(1)$ is the tree size of a GW process with one single founder when the generating branching mechanism is $\phi_c(z)$. We note $\phi_c(1) = 1$, $\phi_c'(1) = z_c\phi'(\tau) = 1$ (a critical case with extinction probability $\rho_{e,c} = 1$, the smallest positive root of $\rho_{e,c} = \phi_c(\rho_{e,c})$) and the convergence radius of ϕ_c is $z_+/\tau > 1$. As a result, $\Phi_c(z) \underset{z\to1}{\sim} 1 - \tau^{-1}\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}(1-z)^{1/2}$, with singularity displaced to the left at 1, so that

$$\mathbf{P}\left(\overline{N}_c(1) = n\right) = [z^n]\,\Phi_c(z) \underset{n\to\infty}{\sim} \tau^{-1}\sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}}n^{-3/2}. \tag{17}$$

The geometric cutoff appearing in the probability mass of $\overline{N}(1)$ has been removed and we are left with a pure power-law case. This means that looking at the tree size pgf $\Phi_c(z)$ generated by the critical branching mechanism $\phi_c(z) = z_c\phi(\tau z)/\tau$, $\Phi_c(z)$ exhibits a power-singularity of order $-1/2$ at $z_c = 1$ so that the new tree size probability mass has pure power-law tails of order $1/2$. In particular $\mathbf{E}\left(\overline{N}_c(1)\right) = \infty$. In the explicit geometric example above, where $\phi(z) = \beta/(1-\alpha z)$, it can be checked that $\phi_c(z) = 1/(2-z)$; when dealing with $\phi(z) = (\beta/(1-\alpha z))^{\theta}$, $\phi_c(z) = (\theta/(\theta+1-z))^{\theta}$. Similarly, when dealing with the binomial pgf $\phi(z) = (1-\alpha+\alpha z)^d$, $\phi_c(z) = (1-1/d+z/d)^d$ and when dealing with the Poisson pgf $\phi(z) = e^{-\mu(1-z)}$, $\phi_c(z) = e^{-(1-z)}$.

Binomial (polynomial) and Poisson (exponential) models are examples of ϕ having convergence radius $z_+ = \infty$. For the negative binomial model, ϕ exhibits a power-singularity of positive order $\theta > 0$ at $z_+ = 1/\alpha$ with $1 < z_+ < \infty$, so with $\phi(z_+) = \infty$.

Here is now a family of ϕ's with a power-singularity of negative order $-\alpha$, $\alpha \in (0,1)$. Let $\alpha, \lambda \in (0,1)$ and $z_+ > 1$. Define the Sibuya pgf (see Sibuya (1979)),

$$h(z) = 1 - \lambda(1 - z/z_+)^{\alpha} \text{ and } \phi(z) = \frac{h(z)}{h(1)}.$$

It can be checked that this ϕ is a proper pgf with convergence radius z_+ and which is finite at $z = z_+ > 1$, with $\phi(z_+) = \frac{1}{h(1)} > 1$. Note that for $k \geq 1$,

$$\pi_k = \left[z^k\right]\phi(z) = \frac{\lambda}{h(1)}(-1)^{k-1}\binom{\alpha}{k}z_+^k \underset{k\to\infty}{\sim} \frac{\lambda\alpha}{h(1)}k^{-(\alpha+1)}z_+^k/\Gamma(1-\alpha).$$

The latter singularity expansion of Φ applies to this branching mechanism ϕ as well.

2.5.2. *Number of leaves (sterile individuals)*

In the branching population models just discussed it is important to control the number of leaves in the GW tree with a single founder because leaves are nodes (individuals) of the tree (population) that gave birth to no offspring (the frontier of the tree as sterile individuals), so responsible of its extinction. Leaves are nodes with outdegree zero, so let $\overline{N}^0(1)$ be the number of leaves in a GW tree with $\overline{N}(1)$ nodes. With $\Phi(z_0, z) = \mathbf{E}\left(z_0^{\overline{N}^0(1)}z^{\overline{N}(1)}\right)$ the joint pgf of $\left(\overline{N}^0(1), \overline{N}(1)\right)$, clearly solves the functional equation

$$\Phi(z_0, z) = z\left(\pi_0(z_0 - 1) + \phi(\Phi(z_0, z))\right). \tag{18}$$

With $\overline{N}_n^0(1) = \overline{N}^0(1) \mid \overline{N}(1) = n$, we have

$$\mathbf{E}\left(z_0^{\overline{N}_n^0(1)}\right) = \frac{[z^n]\Phi(z_0, z)}{[z^n]\Phi(1, z)},$$

where $\Phi(1,z) = \Phi(z)$. It is shown using this in Drmota (2009) (Th. 3.13, page 84) that, under our assumptions on ϕ,

$$\begin{array}{l} \frac{1}{n}\mathbf{E}\left(\overline{N}_n^0(1)\right) \underset{n\to\infty}{\to} m_0 = \frac{\pi_0}{\phi(\tau)}, \\ \frac{1}{n}\sigma^2\left(\overline{N}_n^0(1)\right) \underset{n\to\infty}{\to} \sigma_0^2 = \frac{\pi_0}{\phi(\tau)} - \frac{\pi_0^2}{\phi(\tau)^2} - \frac{\pi_0^2}{\tau^2\phi(\tau)^2\phi''(\tau)}, \\ \frac{\overline{N}_n^0(1) - m_0 n}{\sigma_0\sqrt{n}} \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,1). \end{array} \tag{19}$$

As $n \to \infty$, $\frac{1}{n}\overline{N}_n^0(1)$ converges in probability to $m_0 < 1$, the asymptotic fraction of nodes in a size n tree which are leaves. For the geometrically generated tree with $\phi(z) = \beta/(1-\alpha z)$, it can be checked that $m_0 = 1/2$, whereas for the Poisson generated tree with pgf $\phi(z) = e^{\mu(z-1)}$, $m_0 = e^{-1}$. For the negative binomial tree generated by $\phi(z) = (\beta/(1-\alpha z))^{\theta}$, $m_0 = (\theta/(\theta+1))^{\theta}$ and for the Flory tree generated by the pgf $\phi(z) = (1-\alpha+\alpha z)^d$, $m_0 = ((d-1)/d)^d$.

Almost sure convergence and large deviations: The functional equation solving $\Phi(z_0, z)$ may be put under the form

$$\Phi(z_0, z) = z\phi_{z_0}(\Phi(z_0, z)),$$

where $\phi_{z_0}(z) = \pi_0(z_0 - 1) + \phi(\Phi(z_0, z))$, with z_0 viewed as a parameter. Let $\tau(z_0)$ solve $\phi_{z_0}(\tau(z_0)) - \tau(z_0)\phi'_{z_0}(\tau(z_0)) = 0$, else

$$\pi_0(z_0 - 1) + \phi(\tau(z_0)) - \tau(z_0)\phi'(\tau(z_0)) = 0,$$

with $\tau(1) = \tau$. We have $([z^n]\Phi(1,z))^{1/n} \to z_c = 1/\phi'(\tau)$ and $([z^n]\Phi(z_0,z))^{1/n} \to z_c(z_0) = 1/\phi'(\tau(z_0))$, therefore

$$\mathbf{E}\left(z_0^{\overline{N}_n^0(1)}\right)^{1/n} = \left(\frac{[z^n]\Phi(z_0,z)}{[z^n]\Phi(1,z)}\right)^{1/n} \to a(z_0) = \frac{\phi'(\tau(1))}{\phi'(\tau(z_0))}.$$

This shows by Gärtner-Ellis theorem (Ellis (1985)), that, for all $\rho \in (0,1)$

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbf{P}\left(\frac{1}{n}\overline{N}_n^0(1) \to \rho\right) = f(\rho),$$

where, with $F(\lambda) = \log a\left(e^{-\lambda}\right)$, $f(\rho) = \inf_{\lambda\in\mathbb{R}}(\rho\lambda - F(\lambda)) < 0$. The function f is the large deviation rate function, as the Legendre transform of the concave function F. The value of ρ for which $f(\rho) = 0$ is $F'(0)$. We conclude that as $n \to \infty$

$$\frac{1}{n}\overline{N}_n^0(1) \overset{a.s.}{\to} \rho_* = F'(0). \tag{20}$$

Example 2.3.

With $\phi(z) = \beta/(1-\alpha z)$, $\tau(z_0) = \frac{z_0 - \sqrt{z_0}}{\alpha(z_0-1)}$, so with

$$1/\phi'(\tau(z_0)) = \frac{1}{\alpha\beta}\left(\sqrt{z_0}+1\right)^{-2} \text{ and } 1/\phi'(\tau(1)) = \frac{1}{4\alpha\beta},$$

leading to $\alpha\left(z_0\right)=4\left(\sqrt{z_0}+1\right)^{-2}$ and $F\left(\lambda\right)=\log 4-2\log\left(1+e^{-\lambda/2}\right)$ with $F'\left(0\right)=1/2$. So here $\frac{1}{n}\overline{N}_n^0\left(1\right)\overset{a.s.}{\to}1/2$ (not only in probability). One can check

$$f\left(\rho\right)=-\log 4-2\left(\rho\log\rho+\left(1-\rho\right)\log\left(1-\rho\right)\right),$$

is the Cramér's large deviation rate function for this example.

2.5.3. *Forests of trees: back to a random number P of trees (species)*

Let $\phi_0\left(z\right)$ with $\mu_0=\phi_0'\left(1\right)<\infty$ be such that $\Psi\left(z\right)=\phi_0\left(\Phi\left(z\right)\right)$ has itself a dominant power-singularity at z_c of order $-1/2$, so with $\Psi\left(z\right)\underset{z\to z_c}{\sim}\phi_0\left(\tau\right)+\phi_0'\left(\tau\right)\sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\left(1-z/z_c\right)^{1/2}$. Then, with $\mathcal{N}$ the total number of nodes of the forest

$$\mathbf{P}\left(\mathcal{N}=n\right)=\left[z^n\right]\Psi\left(z\right)\underset{n\to\infty}{\sim}\phi_0'\left(\tau\right)\sqrt{\frac{\phi\left(\tau\right)}{2\pi\phi''\left(\tau\right)}}n^{-3/2}z_c^{-n}.$$

Note that $z_c=1$ when $\tau=1$ and $\phi'\left(1\right)=1$ in which critical case $\left[z^n\right]\Psi\left(z\right)\underset{n\to\infty}{\sim}\mu_0\sqrt{\frac{1}{2\pi\phi''(1)}}n^{-3/2}$. For all branching mechanism ϕ with convergence radius $z_+>1$ and all ϕ_0 such that $\Psi\left(z\right)=\phi_0\left(\Phi\left(z\right)\right)$ still has a dominant singularity at z_c, the law of the size $\mathcal{N}$ of the forest of trees has a power-law factor which is $n^{-3/2}$, so independent of the model's details (a universality property). Note that z_c and the scaling constant in front of $n^{-3/2}z_c^{-n}$ are model-dependent though, both requiring the computation of τ.

2.6. Random Poisson or geometric number of founders

Trees are the connected components of some forest. Let $\Psi\left(z\right)=\phi_0\left(\Phi\left(z\right)\right)$, where $\phi_0\left(z\right)$ is either (a): $\phi_0\left(z\right)=e^{-\mu_0(1-z)}$ (Poissonian forest) or (b): $\phi_0\left(z\right)=\beta_0/\left(1-\alpha_0 z\right)$ (geometric forest). In case (a), the total number of nodes (individuals) in the forest (population) has a compound Poisson (infinitely divisible) distribution, whereas in case (b) this number is compound geometric distributed (geometrically infinitely divisible). Geometrically infinitely divisible rvs form a subclass of infinitely divisible rvs (see Aly and Bouzar (2000)).

In the sequel we shall address the following problems under both (a) and (b) forest models:

- given a random forest withy n nodes in total (a population with n individuals in total), how many trees (species) is it made of?

- given a random forest withy n nodes in total (a population with n individuals in total), what are the sizes of its constituting trees and how many trees (species) with given size (number of representatives) are they in the sample?

The answers to these questions follow from the computation of the joint probability $\mathbf{P}\left(\mathcal{N}=n,P=p\right)$ to observe a forest (population) with p trees (species) and n nodes (individuals).

2.6.1. *The numbers of trees in a random forest with n nodes*

Let us distinguish the two cases.

Case (a). With

$$\Phi\left(z\right)=\mathbf{E}\left(z^{\overline{N}(1)}\right), \tag{21}$$

we let $\overline{\pi}_{\bullet}=\left(\overline{\pi}_n\right)_{n\geq 1}$, where $\overline{\pi}_n=\mathbf{P}\left(\overline{N}\left(1\right)=n\right)$. The total size $\mathcal{N}$ of the forest is

$$\mathcal{N}=\sum_{p=1}^{P(\mu_0)}\overline{N}^{(p)}\left(1\right), \tag{22}$$

a compound Poisson random variable involving $\overline{N}^{(p)}\left(1\right)$ iid copies of $\overline{N}\left(1\right)$. We have

$$\Psi\left(z\right)=\mathbf{E}\left(z^{\mathcal{N}}\right)=e^{-\mu_0(1-\Phi(z))}, \tag{23}$$

and

$$\Psi\left(z\right)=e^{-\mu_0(1-\Phi(z))}=e^{-\mu_0}\left(1+\sum_{n\geq 1}\sigma_n\left(\mu_0\right)\frac{z^n}{n!}\right), \tag{24}$$

which defines $\sigma_n\left(\mu_0\right)$. In the development of $\Psi\left(z\right)$, $\sigma_n\left(\mu_0\right)$ is a degree-n polynomial in μ_0 with

$$\left[\mu_0^p\right]\sigma_n\left(\mu_0\right)=B_{n,p}\left(c_{\bullet}\right)=\frac{n!}{p!}\left[z^n\right]\Phi\left(z\right)^p,$$

known as the "exponential" Bell polynomial in the variables $c_{\bullet}=\bullet!\overline{\pi}_{\bullet}$. We have $B_{n,p}\left(c_{\bullet}\right)=0$ if $p>n$. With the boundary conditions

$$B_{n,0}\left(c_{\bullet}\right)=B_{0,p}\left(c_{\bullet}\right)=0,\ n,p\geq 1 \text{ and } B_{0,0}\left(c_{\bullet}\right)=1,$$

we get in particular

$$B_{n,1}\left(c_{\bullet}\right)=c_n=n!\overline{\pi}_n \text{ and } B_{n,n}\left(c_{\bullet}\right)=c_1^n=\overline{\pi}_1^n. \tag{25}$$

We also have (see Comtet (1970) and Pitman (2006)):

$$B_{n,p}\left(c_{\bullet}\right)=\sum^{*}\Omega_{c_{\bullet}}\left(p_1,..,p_n\right),\, p\leq n,$$

where the latter star-sum is over the integers $p_1,..,p_n\geq 0$ obeying

$$\sum_{m=1}^{n}p_m=p,\ \sum_{m=1}^{n}mp_m=n,$$

and

$$\Omega_{c_{\bullet}}\left(p_1,..,p_n\right)=n!\prod_{m=1}^{n}\frac{c_m^{p_m}}{p_m!m!^{p_m}}=n!\prod_{m=1}^{n}\frac{\overline{\pi}_m^{p_m}}{p_m!}. \tag{26}$$

Clearly $\Omega_{c_\bullet}(p_1,..,p_n)$ are the Boltzmann weights of the forest configurations with n nodes in total, p constituting trees and p_m trees of size m, $m=1,...,n$. We have

$$\Omega_{c_\bullet}(p_1,..,p_n)=\frac{n!}{p!}\binom{p}{p_1\cdots p_n}\prod_{m=1}^{n}\overline{\pi}_m^{p_m},$$

and the (canonical) probabilities to observe such configurations are (note their multinomial character and see Equation (7))

$$\frac{\Omega_{c_\bullet}(p_1,..,p_n)}{B_{n,p}(c_\bullet)}=\frac{\binom{p}{p_1\cdots p_n}\prod_{m=1}^{n}\overline{\pi}_m^{p_m}}{[z^n]\,\Phi(z)^p}.$$

It also holds that

$$\sigma_n(\mu_0)=\sum_{p=1}^{n}B_{n,p}(c_\bullet)\,\mu_0^p. \tag{27}$$

We thus have

$$\begin{array}{l}\mathbf{P}(\mathcal{N}=n)=[z^n]\,\Psi(z)=e^{-\mu_0}\sigma_n(\mu_0)\,,\\ \mathbf{P}(P=p)=e^{-\mu_0}\mu_0^p/p!\text{ (Poisson}\,(\mu_0)\,),\end{array} \tag{28}$$

so that the joint probability of $\mathcal{N}$ and P reads

$$\mathbf{P}(\mathcal{N}=n,P=p)=e^{-\mu_0}\mu_0^pB_{n,p}(c_\bullet)\,/n!. \tag{29}$$

It can be checked as required that:

$$\mathbf{P}(\mathcal{N}=n)=\sum_{p=1}^{n}\mathbf{P}(\mathcal{N}=n,P=p)=e^{-\mu_0}\sum_{p=1}^{n}\mu_0^pB_{n,p}(c_\bullet)\,/n!=e^{-\mu_0}\sigma_n(\mu_0)\,/n!\text{ and}$$

$$\begin{array}{l}\mathbf{P}(P=p)=\sum_{n\geq p}\mathbf{P}(\mathcal{N}=n,P=p)=e^{-\mu_0}\mu_0^p\sum_{n\geq p}B_{n,p}(c_\bullet)\,/n!\\ =\frac{e^{-\mu_0}\mu_0^p}{p!}\sum_{n\geq p}[z^n]\,\Phi(z)^p\mid_{z=1}=\frac{e^{-\mu_0}\mu_0^p}{p!}\Phi(1)^p\ \text{ (Poisson}\,(\mu_0)\,).\end{array}$$

Concerning the conditional rvs $P_n=(P=p\mid\mathcal{N}=n)$ and $\mathcal{N}_p=(\mathcal{N}=n\mid P=p)$

$$\begin{array}{l}\mathbf{P}(P_n=p)=\frac{\mu_0^pB_{n,p}(c_\bullet)}{\sigma_n(\mu_0)},\ p=1,...,n,\\ \mathbf{P}(N_p=n)=p!B_{n,p}(c_\bullet)\,/n!,\ n\geq p,\text{ independent of }\mu_0.\end{array} \tag{30}$$

From the definition of $\sigma_n(\mu_0)$ in Equation (27), we obtain,

$$\mathbf{E}\left(z^{P_n}\right)=\frac{\sigma_n(z\mu_0)}{\sigma_n(\mu_0)}\text{ and }\mathbf{E}\left(z^{N_p}\right)=\Phi(z)^p\,. \tag{31}$$

In particular N_p is, as required, the sum of p iid rvs with common pgf $\Phi(z)$. We have

$$[z^n]\,\Psi(z)=e^{-\mu_0}\sigma_n(\mu_0)\,/n!\underset{n\to\infty}{\sim}\phi_0'(\tau)\sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}}n^{-3/2}z_c^{-n}.$$

Owing to $\phi_0'(\tau)=\mu_0e^{\mu_0(\tau-1)}$, then $\sigma_n(\mu_0)\sim n!\mu_0e^{\mu_0\tau}\sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}}n^{-3/2}z_c^{-n}$ and we conclude from Equation (31) that

$$\mathbf{E}\left(z^{P_n}\right)=\frac{\sigma_n(z\mu_0)}{\sigma_n(\mu_0)}\underset{n\to\infty}{\sim}ze^{-\mu_0\tau(1-z)},$$

the pgf of a shifted Poisson$(\mu_0\tau)$ rv with mean $\mu_0\tau$. Thus, given a Poissonian forest with $\mathcal{N} = n$ nodes, the number of GW trees P_n of the forest obeys

$$P_n \underset{n\to\infty}{\overset{d}{\to}} 1 + \text{Poi}\left(\mu_0\tau\right). \tag{32}$$

Case (b). Assume $\phi_0\left(z\right) = \beta_0/\left(1-\alpha_0 z\right)$ (geometric), with $\mu_0 = \alpha_0/\beta_0$.

Proceeding similarly as in the Poisson case, we have $\Psi\left(z\right) = \beta_0\left(1+\sum_{n\geq 1}\sigma_n\left(\alpha_0\right)\frac{z^n}{n!}\right)$, where $\sigma_n\left(\alpha_0\right) = \sum_{p=1}^{n} p! B_{n,p}\left(c_\bullet\right)\alpha_0^p$. We find

$$\begin{aligned} &\mathbf{P}\left(\mathcal{N}=n, P=p\right) = \beta_0\alpha_0^p B_{n,p}\left(c_\bullet\right)p!/n!, \\ &\mathbf{P}\left(\mathcal{N}=n\right) = \beta_0\sigma_n\left(\alpha_0\right)/n! \text{ and } \mathbf{P}\left(P=p\right) = \beta_0\alpha_0^p, \end{aligned} \tag{33}$$

leading to $\mathbf{E}\left(z^{P_n}\right) = \frac{\sigma_n\left(z\alpha_0\right)}{\sigma_n\left(\alpha_0\right)}$. We have $\Psi\left(z\right) = \beta_0/\left(1-\alpha_0\Phi\left(z\right)\right)$ with a singularity at $z_c\left(\alpha_0\right)$ uniquely defined by $\Phi\left(z_c\left(\alpha_0\right)\right) = 1/\alpha_0$.

Recalling $\Phi\left(z\right)$ has a singularity at $z_c = \tau/\phi\left(\tau\right) \geq 1$ with $\Phi\left(z_c\right) = \tau$, three cases arise:

- (subcritical) If $z_c\left(\alpha_0\right) > z_c$, else $1/\alpha_0 > \tau$, then the dominant singularity of $\Psi\left(z\right)$ is still at z_c, the singularity of $\Phi\left(z\right)$. Then,

$$\left[z^n\right]\Psi\left(z\right) = \beta_0\sigma_n\left(\alpha_0\right)/n! \underset{n\to\infty}{\sim} \phi_0'\left(\tau\right)\sqrt{\frac{\phi\left(\tau\right)}{2\pi\phi''\left(\tau\right)}}n^{-3/2}z_c^{-n}.$$

Owing to $\phi_0'\left(\tau\right) = \alpha_0\beta_0/\left(1-\alpha_0\tau\right)^2 = \phi_0'\left(\alpha_0,\tau\right)$, we conclude

$$\mathbf{E}\left(z^{P_n}\right) = \frac{\sigma_n\left(z\alpha_0\right)}{\sigma_n\left(\alpha_0\right)} \underset{n\to\infty}{\sim} \frac{\phi_0'\left(\alpha_0 z,\tau\right)}{\phi_0'\left(\alpha_0,\tau\right)} = z\frac{\left(1-\alpha_0\tau\right)^2}{\left(1-\alpha_0\tau z\right)^2}.$$

In the pgf at the right-hand-side the factor $\left(1-\alpha_0\tau\right)^2/\left(1-\alpha_0\tau z\right)^2$ is the pgf of the sum S of two independent geometric random variables with same mean $\alpha_0\tau/\left(1-\alpha_0\tau\right)$. Thus, provided $1/\alpha_0 > \tau$, given a geometric forest with $\mathcal{N} = n$ nodes, the number P_n of its constituting trees obeys, after shifting by one:

$$P_n \underset{n\to\infty}{\overset{d}{\to}} 1 + S. \tag{34}$$

- (supercritical) If $z_c\left(\alpha_0\right) < z_c$, else $1/\alpha_0 < \tau$, then the singularity of $\Psi\left(z\right)$ is shifted to the left of z_c, at $z_c\left(\alpha_0\right)$, with $\Psi\left(z\right) \underset{z\to z_c\left(\alpha_0\right)}{\sim} \beta_0\left(1-z/z_c\left(\alpha_0\right)\right)^{-1}$. The nature of the singularity of Ψ is dictated by that of ϕ_0. Thus,

$$\left[z^n\right]\Psi\left(z\right) = \beta_0\sigma_n\left(\alpha_0\right)/n! \underset{n\to\infty}{\sim} \beta_0 z_c\left(\alpha_0\right)^{-n},$$

and, with $a_{\alpha_0}\left(z\right) = z_c\left(\alpha_0\right)/z_c\left(z\alpha_0\right)$,

$$\mathbf{E}\left(z^{P_n}\right) = \frac{\sigma_n\left(z\alpha_0\right)}{\sigma_n\left(\alpha_0\right)} \underset{n\to\infty}{\sim} a_{\alpha_0}\left(z\right)^n,$$

else $\mathbf{E}\left(z^{P_n}\right)^{1/n} \underset{n\to\infty}{\to} a_{\alpha_0}(z)$. Thus, given a supercritical geometric forest with $\mathcal{N} = n$ nodes, the mean number of its constituting trees P_n grows like n and

$$\frac{1}{n}P_n \underset{n\to\infty}{\overset{a.s.}{\to}} p_*, \text{ where } p_* = a'_{\alpha_0}(1). \tag{35}$$

Note that $z_c(\alpha_0)$ can in principle be obtained by Lagrange inversion formula as

$$z_c(\alpha_0) = \sum_{n\geq 1} \frac{\alpha_0^{-n}}{n}\left[z^{n-1}\right]\left(\frac{\Phi(z)}{z}\right)^{-n}.$$

When α_0 crosses the critical value $1/\tau$ ($\mu_0 > 1/(\tau-1)$), there is a drastic qualitative change in the large n behavior of P_n, which is reminiscent of a phase transition.

- (critical) In the critical case $1/\alpha_0 = \tau$, the two singular expansions of $\Phi(z)$ and of $\phi_0(z)$ at z_c should be composed. With $\phi_0(z) = (1-1/\tau)/(1-z/\tau)$, and $\Phi(z) \sim \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}(1-z/z_c)^{1/2}$, $z_c = \tau/\phi(\tau) = 1/(\alpha_0\phi(1/\alpha_0)) = z_c(\alpha_0)$, we get

$$\Psi(z) \sim (\tau-1)\sqrt{\frac{\phi''(\tau)}{2\phi(\tau)}}(1-z/z_c)^{-1/2},$$

and $[z^n]\Psi(z) = \beta_0\sigma_n(\alpha_0)/n! \underset{n\to\infty}{\sim} (\tau-1)\sqrt{\frac{\phi''(\tau)}{2\pi\phi(\tau)}}n^{-1/2}z_c^{-n}$. Thus, $\mathbf{E}\left(z^{P_n}\right)^{1/n} \underset{n\to\infty}{\to} a(z) = \frac{z_c(\alpha_0)}{z_c(z\alpha_0)} = \frac{z\phi(1/(\alpha_0 z))}{\phi(1/\alpha_0)}$.

An explicit example: Suppose $\phi(z) = \pi_0 + \pi_2 z^2$ (the binary branching mechanism with $\pi_0 + \pi_2 = 1$). Then,

$$\Phi(z) = \frac{1-\sqrt{1-4\pi_0\pi_2 z^2}}{2\pi_2 z},$$

with, for each n even, $\mathbf{P}\left(\overline{N}(1) = n\right) = 0$ and for each $n = 2m+1$, odd:

$$[z^n]\Phi(z) = \mathbf{P}\left(\overline{N}(1) = n\right) = \frac{1}{n}\left[z^{n-1}\right]\phi(z)^n = \frac{1}{n}\sqrt{\frac{\pi_0}{\pi_2}}\binom{n}{\frac{n-1}{2}}\left(\sqrt{\pi_0\pi_2}\right)^n.$$

The equation $\Phi(z_c(\alpha_0)) = 1/\alpha_0$ yields

$$z_c(\alpha_0) = \frac{\alpha_0}{1-\pi_0(1-\alpha_0^2)}.$$

On the other hand, $\tau = \sqrt{\pi_0/\pi_2}$ leading to $z_c = 1/\left(2\sqrt{\pi_0\pi_2}\right)$. Here, $z_c(\alpha_0) > z_c$ if and only if $\alpha_0 < \sqrt{\pi_2/\pi_0}$.

- (supercritical) If $z_c(\alpha_0) < z_c$ (else $\alpha_0 > \sqrt{\pi_2/\pi_0}$), then

$$a_{\alpha_0}(z) = \frac{z_c(\alpha_0)}{z_c(z\alpha_0)} = \frac{1-\pi_0\left(1-(z\alpha_0)^2\right)}{z\left(1-\pi_0\left(1-\alpha_0^2\right)\right)},$$

with $a'_{\alpha_0}(1) = \left(\pi_0\alpha_0^2 - (1-\pi_0)\right)/\left(1-\pi_0\left(1-\alpha_0^2\right)\right) \in (0,1)$.

- (critical) If $z_c\left(\alpha_0\right)=z_c$ (else $\alpha_0=\sqrt{\pi_2/\pi_0}$), then

$$a\left(z\right)=\frac{z_c\left(\alpha_0\right)}{z_c\left(z\alpha_0\right)}=\frac{z\phi\left(1/\left(\alpha_0 z\right)\right)}{\phi\left(1/\alpha_0\right)}=\frac{z+z^{-1}}{2}=\cosh\left(\log z\right).$$

2.6.2. *Trees' sizes in a random forest with n nodes*

Introduce

$$\Phi_{\mu}\left(z\right)=\sum_{n\geq 1}\mu_n\overline{\pi}_n z^n,$$

where the μ_n's marks the size-n trees. Note $\Phi_{\mu}\left(z\right)=\Phi\left(z\right)+\sum_{n\geq 1}\left(\mu_n-1\right)\overline{\pi}_n z^n$.

Case (a).

As above, μ_0 will mark the number of trees in the forest.

With $\mathbf{P}=\left(P_1,...,P_m,...\right)$ the vector of size-m trees satisfying $P=\sum_{m\geq 1}P_m$ and $\mathcal{N}=\sum_{m\geq 1}mP_m$, we have,

$$\Psi\left(z,\mu\right)=\mathbf{E}\left(z^{\mathcal{N}}\mu^{\mathbf{P}}\right)=e^{-\mu_0\left(1-\Phi_{\mu}\left(z\right)\right)}, \tag{36}$$

and, upon introducing $\sigma_n\left(\mu_0,\mu\right)$

$$\Psi\left(z,\mu\right)=e^{-\mu_0}\left(1+\sum_{n\geq 1}\sigma_n\left(\mu_0,\mu\right)\frac{z^n}{n!}\right). \tag{37}$$

Note that, as required:

If $\mu_n=\mu^n$, $\Phi_{\mu}\left(z\right)=\Phi\left(\mu z\right)$ and $\Psi\left(z,\mu\right)=\mathbf{E}\left(z^{\mathcal{N}}\mu^{\sum_{m\geq 1}mP_m}\right)=\mathbf{E}\left(\left(z\mu\right)^{\mathcal{N}}\right)$.

If $\mu_n=\mu$, $\Phi_{\mu}\left(z\right)=\mu\Phi\left(z\right)$ and $\Psi\left(z,\mu\right)=\mathbf{E}\left(z^{\mathcal{N}}\mu^{\sum_{m\geq 1}P_m}\right)=\mathbf{E}\left(z^{\mathcal{N}}\mu^{P}\right)$.

In the development of $\Psi\left(z,\mu\right)$, $\sigma_n\left(\mu_0,\mu\right)$ is a degree-n polynomial in μ_0 with $\left[\mu_0^p\right]\sigma_n\left(\mu_0,\mu\right)=B_{n,p}\left(\left(\mu c\right)_{\bullet}\right)$, the exponential Bell polynomial now in the variables

$$\left(\mu c\right)_{\bullet}=\mu_{\bullet}\bullet!\overline{\pi}_{\bullet}=\mu_{\bullet}c_{\bullet}\text{ with }c_{\bullet}=\left(1c\right)_{\bullet}=\bullet!\overline{\pi}_{\bullet}.$$

We have $B_{n,p}\left(\left(\mu c\right)_{\bullet}\right)=0$ if $p>n$ and (see Comtet (1970) and Pitman (2006)):

$$B_{n,p}\left(\left(\mu c\right)_{\bullet}\right)=\sum^{*}\Omega_{\left(\mu c\right)_{\bullet}}\left(p_1,..,p_n\right), p\leq n,$$

where the latter star-sum is over the integers $p_1,..,p_n\geq 0$ obeying

$$\sum_{m=1}^{n}p_m=p,\ \sum_{m=1}^{n}mp_m=n,$$

and

$$\Omega_{(\mu c)_\bullet}(p_1, .., p_n) = n! \prod_{m=1}^{n} \frac{\overline{\pi}_m^{p_m}}{p_m!} \mu_m^{p_m}. \tag{38}$$

Then,

$$\sigma_n(\mu_0, \mu) = \sum_{p=1}^{n} B_{n,p}\left((\mu c)_\bullet\right) \mu_0^p, \tag{39}$$

where $B_{n,p}\left((\mu c)_\bullet\right) = \frac{n!}{p!}[z^n]\,\Phi_\mu(z)^p$. From the definition of $\sigma_n(\mu_0, \mu)$ in Equation (39), with $\mathbf{P}_n = (P_1, ..., P_n \mid \mathcal{N} = n)$, we obtain,

$$\mathbf{E}\left(z^{P_n}\mu^{\mathbf{P}_n}\right) = \frac{\sigma_n(z\mu_0, \mu)}{\sigma_n(\mu_0, \mathbf{1})} \text{ and } \mathbf{E}\left(\mu^{\mathbf{P}_n}\right) = \frac{\sigma_n(\mu_0, \mu)}{\sigma_n(\mu_0, \mathbf{1})}. \tag{40}$$

Remark 2.5.

Defining $\mathbf{P}_{n,p} = (P_1, ..., P_n \mid \mathcal{N} = n, P = p)$, the tree sizes when both n and p are held fixed

$$\mathbf{E}\left(\mu^{\mathbf{P}_{n,p}}\right) = \frac{[\mu_0^p]\,\sigma_n(\mu_0, \mu)}{[\mu_0^p]\,\sigma_n(\mu_0, \mathbf{1})} = \frac{B_{n,p}\left((\mu c)_\bullet\right)}{B_{n,p}(c_\bullet)}, \tag{41}$$

consistently with Equation (7).

With $\tau(\mu) = \tau + \sum_{n\geq 1}(\mu_n - 1)\overline{\pi}_n z_c^n$, we have

$$\Phi_\mu(z) = \Phi(z) + \sum_{n\geq 1}(\mu_n - 1)\overline{\pi}_n z^n \underset{z\to z_c}{\sim} \tau(\mu) + \sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}}\left(1 - z/z_c\right)^{1/2},$$

so that

$$[z^n]\,\Psi(z, \mu) = e^{-\mu_0}\sigma_n(\mu_0, \mu)/n! \underset{n\to\infty}{\sim} \phi_0'(\tau(\mu))\sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}} n^{-3/2} z_c^{-n}.$$

Owing to $\phi_0'(\tau) = \mu_0 e^{\mu_0(\tau-1)} = \phi_0'(\mu_0, \tau)$ and observing $\tau(\mathbf{1}) = \tau$, we conclude

$$\begin{aligned}\mathbf{E}\left(z^{P_n}\mu^{\mathbf{P}_n}\right) &= \tfrac{\sigma_n(z\mu_0,\mu)}{\sigma_n(\mu_0,\mathbf{1})} \underset{n\to\infty}{\sim} e^{\mu_0(z-1)}\tfrac{\phi_0'(z\mu_0,\tau(\mu))}{\phi_0'(\mu_0\tau)} = e^{\mu_0(z-1)}\tfrac{\mu_0 z e^{\mu_0 z(\tau(\mu)-1)}}{\mu_0 e^{\mu_0(\tau(\mathbf{1})-1)}} \\ &= z e^{\mu_0(z\tau(\mu)-\tau)} = z e^{\mu_0\tau(z-1)} e^{\mu_0 z(\tau(\mu)-\tau)} = z e^{\mu_0\tau(z-1)} \textstyle\prod_{m=1}^{n} e^{\mu_0 z(\mu_m-1)\overline{\pi}_m z_c^m}.\end{aligned} \tag{42}$$

In particular ($z = 1$),

$$\mathbf{E}\left(\mu^{\mathbf{P}_n}\right) \sim \prod_{m=1}^{n} e^{\mu_0(\mu_m-1)\overline{\pi}_m z_c^m} \text{ and } \mathbf{E}\left(\mu_m^{P_n(m)}\right) \sim e^{-\mu_0\overline{\pi}_m z_c^m(1-\mu_m)}. \tag{43}$$

This means that given a Poissonian forest with $\mathcal{N} = n$ nodes, the numbers of its size-m trees $P_n(m)$, $m = 1, ..., n$, are asymptotically independent. Furthermore, for each $m = 1, ..., n$, $P_n(m)$ obeys

$$P_n(m) \underset{n\to\infty}{\overset{d}{\to}} \text{Poi}\left(\mu_0\overline{\pi}_m z_c^m\right), \tag{44}$$

a Poisson rv with mean $\lambda_m = \mu_0 \overline{\pi}_m z_c^m$. When m is large close to n, recalling $\overline{\pi}_m = \mathbf{P}\left(\overline{N}\left(1\right) = m\right) \underset{m\to\infty}{\sim} \sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}} m^{-3/2} z_c^{-m}$, the mean value λ_m approaches $\mu_0 \sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}} m^{-3/2}$ so that the larger m, the smaller λ_m. Of concrete interest is also the probability that there are no size-m trees in the Poissonian forest. This event occurs with probability $\exp\left(-\mu_0 \overline{\pi}_m z_c^m\right)$ and it approaches 1 when m is large close to n. Large size-m trees have a small average size λ_m because with a large probability they are not there.

Case (b). Recalling $\Phi\left(z\right)$ has a singularity at $z_c = \tau/\phi\left(\tau\right) \geq 1$ with $\Phi\left(z_c\right) = \tau$, three cases arise:

- (subcritical) If $z_c\left(\alpha_0\right) > z_c$, else $1/\alpha_0 > \tau$, then the singularity of $\Psi\left(z\right)$ is still at z_c, the singularity of $\Phi\left(z\right)$. Then,

$$\left[z^n\right]\Psi\left(z,\mu\right) = \beta_0 \sigma_n\left(\alpha_0,\mu\right)/n! \underset{n\to\infty}{\sim} \phi_0'\left(\tau\left(\mu\right)\right)\sqrt{\frac{\phi\left(\tau\right)}{2\pi\phi''\left(\tau\right)}} n^{-3/2} z_c^{-n}.$$

Owing to $\phi_0'\left(\tau\right) = \alpha_0\beta_0/\left(1-\alpha_0\tau\right)^2 = \phi_0'\left(\alpha_0,\tau\right)$, we conclude

$$\mathbf{E}\left(z^{P_n}\mu^{\mathbf{P}_n}\right) = \frac{\sigma_n\left(z\alpha_0,\mu\right)}{\sigma_n\left(\alpha_0,\mathbf{1}\right)} \underset{n\to\infty}{\sim} \frac{\phi_0'\left(z\alpha_0,\tau\left(\mu\right)\right)}{\phi_0'\left(\alpha_0\tau\right)} = z\frac{\left(1-\alpha_0\tau\right)^2}{\left(1-\alpha_0 z\tau\left(\mu\right)\right)^2}.$$

Putting $\mu = \mathbf{1}$, $\tau\left(\mu\right) = \tau$ which yields back $\mathbf{E}\left(z^{P_n}\right) = z\left(1-\alpha_0\tau\right)^2/\left(1-\alpha_0 z\tau\right)^2$. Putting $z = 1$ and $\mu_n = 1$ except for μ_m :

$$\mathbf{E}\left(\mu_m^{P_n(m)}\right) \sim \frac{\left(1-\alpha_0\tau\right)^2}{\left(1-\alpha_0\left(\tau+\left(\mu_m-1\right)\overline{\pi}_m z_c^m\right)\right)^2} = \frac{1}{\left(1+\frac{\alpha_0\overline{\pi}_m z_c^m}{1-\alpha_0\tau}\left(1-\mu_m\right)\right)^2}.$$

The pgf at the right-hand-side is the one of the sum S_m of two independent geometric random variables with mean $\alpha_0\overline{\pi}_m z_c^m/\left(1-\alpha_0\tau\right)$. Thus, provided $1/\alpha_0 > \tau$, given a geometric forest with $\mathcal{N} = n$ nodes, the number $P_n\left(m\right)$ of its size-m trees obeys:

$$P_n\left(m\right) \underset{n\to\infty}{\overset{d}{\to}} S_m. \tag{45}$$

Note that $\mathbf{E}\left(\mu^{\mathbf{P}_n}\right) \neq \prod_{m=1}^n \mathbf{E}\left(\mu_m^{P_n(m)}\right)$ so that the $P_n\left(m\right)$, $m = 1,...,n$, are no longer asymptotically independent.

- (supercritical) With $\Phi_\mu\left(z\right) = \Phi\left(z\right) + \sum_{n\geq 1}\left(\mu_n - 1\right)\overline{\pi}_n z^n$, we have $\Psi\left(z,\mu\right) = \beta_0/\left(1-\alpha_0\Phi_\mu\left(z\right)\right)$ with a singularity at $z_c\left(\alpha_0,\mu\right)$ uniquely defined by $\Phi_\mu\left(z_c\left(\alpha_0,\mu\right)\right) = 1/\alpha_0$.

$\Phi_\mu\left(z\right)$ has a singularity at $z_c \geq 1$ with $\Phi_\mu\left(z_c\right) = \tau\left(\mu\right) = \tau + \sum_{n\geq 1}\left(\mu_n-1\right)\overline{\pi}_n z_c^n < \infty$. If $z_c\left(\alpha_0,\mu\right) < z_c$, else $1/\alpha_0 < \tau\left(\mu\right)$, then the singularity of $\Psi\left(z,\mu\right)$ is shifted to the left of z_c, at $z_c\left(\alpha_0,\mu\right)$, with $\Psi\left(z,\mu\right) \sim \beta_0\left(1-z/z_c\left(\alpha_0,\mu\right)\right)^{-1}$ as $z \to z_c\left(\alpha_0,\mu\right)$. The nature of its singularity is dictated by that of ϕ_0. Thus,

$$\left[z^n\right]\Psi\left(z,\mu\right) = \beta_0\sigma_n\left(\alpha_0,\mu\right)/n! \underset{n\to\infty}{\sim} \beta_0 z_c\left(\alpha_0,\mu\right)^{-n},$$

and, with $a_{\alpha_0}(z,\mu) = z_c(\alpha_0, \mathbf{1}) / z_c(z\alpha_0, \mu)$

$$\mathbf{E}\left(z^{P_n}\mu^{\mathbf{P}_n}\right) = \frac{\sigma_n(z\alpha_0,\mu)}{\sigma_n(\alpha_0,\mathbf{1})} \underset{n\to\infty}{\sim} a_{\alpha_0}(z,\mu)^n .$$

This means that $\mathbf{E}\left(z^{P_n}\mu^{\mathbf{P}_n}\right)^{1/n} \underset{n\to\infty}{\to} a_{\alpha_0}(z,\mu)$ and $\mathbf{E}\left(\mu^{\mathbf{P}_n}\right)^{1/n} \underset{n\to\infty}{\to} a_{\alpha_0}(\mu) = z_c(\alpha_0,\mathbf{1})/z_c(\alpha_0,\mu)$. Thus, given a supercritical geometric forest with $\mathcal{N} = n$ nodes, the numbers of its size-m trees $\mathbf{P}_n = (P_n(m), m = 1, ..., n)$ grows like n and $\frac{1}{n}\mathbf{P}_n \underset{n\to\infty}{\to} \mathbf{p}_*$ a.s., where $\mathbf{p}_* = \nabla_\mu a_{\alpha_0}(1)$.

2.7. Relation to enumerative combinatorics

In this section, we emphasize that the probabilistic formulation of GW trees is intimately related to the one arising in enumerative combinatorics.

Let $g(\theta) = 1/(1-\theta)$ and consider the solution of the functional equation

$$t(\theta) = \theta g(t(\theta)), t(0) = 0,$$

namely $t(\theta) = \left(1 - \sqrt{1-4\theta}\right)/2$. The numbers

$$t_n = n! [\theta^n] t(\theta) = \frac{n!}{n}\left[\theta^{n-1}\right] g(\theta)^n = \frac{(2n-2)!}{(n-1)!},$$

constitute the number of all rooted plane (ordered) trees with n labeled nodes, known as Catalan numbers (Comtet (1970)).

Let $g(\theta) = 1 + \sum_{k\geq 1} g_k\theta^k$, where $g_k \geq 0$ (and $g_k > 0$ for at least some $k \geq 2$) now constitute a system of weights. Consider now

$$t(\theta) = \theta g(t(\theta)), t(0) = 0 \text{ with } t(\theta) = \sum_{n\geq 1}\frac{t_n}{n!}\theta^n.$$

Let $\mathcal{P}_n$ be the set of all rooted ordered trees with n labeled nodes. With $N_k(\tau_n)$ the number of nodes in $\tau_n \in \mathcal{P}_n$ with outdegree k, we now have

$$t_n = n! [\theta^n] t(\theta) = \sum_{\tau_n\in\mathcal{P}_n}\prod_{k=0}^{n-1} g_k^{N_k(\tau_n)}.$$

For all θ such that $t(\theta) < \infty$, with $\overline{N}(1) = |\tau_n|$, we can define

$$\mathbf{P}\left(\overline{N}(1) = n\right) = \frac{\theta^n[\theta^n] t(\theta)}{\sum_{n\geq 1}\theta^n[\theta^n] t(\theta)} = \frac{\theta^n[\theta^n] t(\theta)}{t(\theta)},$$

the exponentially-tilted probability to draw a rooted ordered tree of size n. We have

$$\Phi(z) = \mathbf{E}\left(z^{\overline{N}(1)}\right) = \sum_{n\geq 1} z^n \mathbf{P}\left(\overline{N}(1) = n\right) = \frac{\sum_{n\geq 1} z^n\theta^n[\theta^n] t(\theta)}{t(\theta)} = \frac{t(\theta z)}{t(\theta)},$$

where $\Phi(z)$ (obeying $\Phi(1) = 1$ as in a subcritical case) solves

$$\Phi(z) = z\phi_\theta(\Phi(z)), \ \Phi(0) = 0.$$

Here $\phi_\theta(z) = g(t(\theta) z)/g(t(\theta))$ is a branching pgf for all values of θ for which $g(t(\theta)) = t(\theta)/\theta < \infty$, coinciding with the ones for which $t(\theta) < \infty$, namely $\theta \in (0, \theta_c)$, where $\theta_c = \inf(\theta > 0 : t(\theta) = \infty) \leq \infty$. It has mean $\mu_\theta = \phi'_\theta(1) = t(\theta) g'(t(\theta))/g(t(\theta)) = t \log g(t)' \mid_{t=t(\theta)} < 1$. Note that $m_\theta = \Phi'(1) = \mathbf{E}\left(\overline{N}(1)\right)$ is the mean value of $\overline{N}(1)$ and there is a one-to-one Legendre-like correspondence between θ and $m_\theta = 1/(1-\mu_\theta) > 0$.

Remark 2.6.

- It can be checked that the probability sequence $\overline{\pi}_n = \mathbf{P}\left(\overline{N}(1) = n\right)$ minimizes the Kullback-Leibler divergence of $\theta_n = [\theta^n] t(\theta) = t_n/n!$ with respect to $\overline{\pi}_n$, viz

$$K(\{\overline{\pi}_n\} \parallel \{\theta_n\}) = \sum_{n\geq 1} \overline{\pi}_n \log\left(\frac{\overline{\pi}_n}{\theta_n}\right),$$

under the constraints $\sum_{n\geq 1} \overline{\pi}_n = 1$ and $\sum_{n\geq 1} n\overline{\pi}_n = m$. Indeed, using Lagrange multipliers

$$\overline{\pi}_n = \frac{t_n \theta^n}{t(\theta)},$$

where θ and m are related by $m_\theta = \theta t'(\theta)/t(\theta) = 1/(1-\mu_\theta)$ where $\mu_\theta = \phi'_\theta(1) = t(\theta) g'(t(\theta))/g(t(\theta)) = \theta g'(t(\theta))$.

- So far we dealt with a single tree generating function $t(\theta)$. While considering the generating functions

(a) $e^{\gamma t(\theta)}$,
(b) $1/(1-\gamma t(\theta))$,

we deal with forests with γ marking the number of its constituting trees. In case (a) the labeled trees are unordered (indistinguishable) while in case (b) they are ordered (distinguishable). This is why in the probabilistic setup we considered (a) Poissonian forests and (b) geometric forests (see Sheth (1996) and Pitman (1998)).

3. Forests of rooted labeled ordered and increasing trees

3.1. Combinatorial increasing trees

We start with the notion of increasing trees first introduced in Bergeron et al. (1992).

Let $g(\theta) = 1/(1-\theta)$ and consider the solution to the ordinary differential equation

$$t'(\theta) = g(t(\theta)), t(0) = 0,$$

namely $t(\theta) = 1 - \sqrt{1-2\theta}$. The numbers

$$t_n = n! [\theta^n] t(\theta) = (2n-3)!! = 2^{-(n-1)} \frac{(2n-2)!}{(n-1)!},$$

constitute the number of all rooted plane (ordered) increasing trees with n nodes as those trees whose nodes can be labeled in increasing order for all path from root to leaves (see Bergeron et al. (1992) and Drmota (2009)). Due to the increasingness constraint, such trees are less numerous than the ones of simply generated trees (compare the latter numbers with the Catalan numbers).

Let $g(\theta) = 1 + \sum_{k\geq 1} g_k \theta^k$, where $g_k \geq 0$ (and $g_k > 0$ for at least some $k \geq 2$) constitute a system of weights. Consider now

$$t'(\theta) = g(t(\theta)), t(0) = 0 \text{ with } t(\theta) = \sum_{n\geq 1} \frac{t_n}{n!}\theta^n.$$

Let $\mathcal{P}_n$ be the set of all rooted ordered and increasing trees with n nodes. With $N_k(\tau_n)$ the number of nodes in $\tau_n \in \mathcal{P}_n$ with outdegree k, we now have

$$t_n = n! [\theta^n] t(\theta) = \sum_{\tau_n \in \mathcal{P}_n} \prod_{k=0}^{n-1} g_k^{N_k(\tau_n)}.$$

For all $\theta < \theta_c$ such that $t(\theta) < \infty$, with $\overline{N}(1) = |\tau_n|$, we define

$$\mathbf{P}\left(\overline{N}(1) = n\right) = \frac{\theta^n [\theta^n] t(\theta)}{\sum_{n\geq 1} \theta^n [\theta^n] t(\theta)} = \frac{\theta^n [\theta^n] t(\theta)}{t(\theta)},$$

the exponentially-tilted probability to draw a rooted weighted ordered increasing tree of size n. We have

$$\Phi(z) = \mathbf{E}\left(z^{\overline{N}(1)}\right) = \sum_{n\geq 1} z^n \mathbf{P}\left(\overline{N}(1) = n\right) = \frac{\sum_{n\geq 1} z^n \theta^n [\theta^n] t(\theta)}{t(\theta)} = \frac{t(\theta z)}{t(\theta)},$$

where $\Phi(z)$ (obeying $\Phi(1) = 1$) solves

$$\Phi'(z) = m_\theta \phi_\theta (\Phi(z)), \ \Phi(0) = 0.$$

Here $m_\theta = \theta g(t(\theta))/t(\theta)$ and $\phi_\theta(z) = g(t(\theta) z)/g(t(\theta))$ is a pgf for all values of θ for which $g(t(\theta)) = t'(\theta) < \infty$, coinciding with the ones for which $t(\theta) < \infty$, namely $\theta \in (0, \theta_c)$, where $\theta_c = \inf(\theta > 0 : t(\theta) = \infty) \leq \infty$. Note that $m_\theta = \Phi'(1) = \mathbf{E}\left(\overline{N}(1)\right)$ is the mean value of $\overline{N}(1)$ and there is a one-to-one Legendre-like correspondence between $\theta < \theta_c$ and $m_\theta = \theta \log t(\theta)' > 0$. Here also, $\overline{\pi}_n = \mathbf{P}\left(\overline{N}(1) = n\right)$ minimizes the Kullback-Leibler divergence of $\theta_n = [\theta^n] t(\theta) = t_n/n!$ with respect to $\overline{\pi}_n$.

3.2. Random increasing trees

From now on, we will therefore consider the tree pgf $\Phi(z)$ solution to

$$\Phi'(z) = m\phi(\Phi(z)), \text{ with boundary conditions } \Phi(0) = 0 \text{ and } \Phi(1) = 1, \tag{46}$$

where $\phi(z)$ is any branching mechanism with finite mean $\mu = \phi'(1) > 0$. Stated differently, the searched $\Phi(z)$ solves $\Phi(0) = 0$ and

$$z = \frac{1}{m} \int_0^{\Phi(z)} \frac{dz'}{\phi(z')}, \text{ with } m = \int_0^1 \frac{dz'}{\phi(z')}. \tag{47}$$

We have $\Phi''(1) = m^2\mu$ leading to

$$\sigma^2\left(\overline{N}(1)\right) = \Phi''(1) + \Phi'(1) - \Phi'(1)^2 = m\left(1 + m(\mu - 1)\right) > 0.$$

Note that $\mu > 1$ entails $\sigma^2\left(\overline{N}(1)\right) > \mathbf{E}\left(\overline{N}(1)\right)$ (overdispersion of the tree size $\overline{N}(1)$), whereas $\mu < 1$ entails $\sigma^2\left(\overline{N}(1)\right) < \mathbf{E}\left(\overline{N}(1)\right)$ (underdispersion) and $m < 1/(1-\mu)$. In this setup, $\mu > 1$ is not in contradiction with $\Phi(1) = 1$ (the a.s. finiteness of the tree), so μ is no longer restricted to $\mu < 1$ to produce a finite tree. In contrast with GW trees, random increasing trees with $\mu < \infty$ go extinct with probability 1. Note also $\mathbf{P}\left(\overline{N}(1) = 1\right) = \Phi'(0) = m\phi(\Phi(0)) = m\phi(0) = m\pi_0$.

3.3. Three explicit examples

As in the GW setup, the increasing trees with binomial, Poisson and negative binomial generators can be explicitly solved.

- Binomial: $\phi(z) = (\overline{\alpha} + \alpha z)^d$, $d \geq 2$ integer with $\mu = d\alpha$, $\overline{\alpha} = 1 - \alpha$. With $m = \left(\overline{\alpha}^{-(d-1)} - 1\right)/\left(\alpha(d-1)\right) = \frac{d\left((d/(d-\mu))^{d-1} - 1\right)}{\mu(d-1)}$,

$$\begin{aligned}\Phi(z) &= \frac{1}{\alpha}\left(\left(1 + \alpha m(d-1)(1-z)\right)^{-1/(d-1)} - \overline{\alpha}\right)\\ &= \frac{\overline{\alpha}}{\alpha}\left(\left(1 - \alpha m(d-1)\overline{\alpha}^{d-1}z\right)^{-1/(d-1)} - 1\right)\\ &= \frac{\overline{\alpha}}{\alpha}\left(\left(1 - z/z_c\right)^{-1/(d-1)} - 1\right).\end{aligned}$$

It has a power-singularity of order $1/(d-1)$ at $z_c = 1/\left(\alpha m(d-1)\overline{\alpha}^{d-1}\right) = 1/\left(1 - \overline{\alpha}^{d-1}\right) > 1$. Note $\Phi(z) \to \infty$ as $z \to z_c$. We have

$$\overline{\pi}_n = \mathbf{P}\left(\overline{N}(1) = n\right) = \frac{\overline{\alpha}}{\alpha}\frac{[1/(d-1)]_n}{n!}z_c^{-n}.$$

Thus, for large n,

$$[z^n]\,\Phi(z) \sim \frac{\overline{\alpha}}{\alpha\Gamma\left(1/(d-1)\right)}n^{-(1-1/(d-1))}z_c^{-n}.$$

The exponent of the power-law factor is $1 - 1/(d-1)$.

Remark 3.1.

The affine case with $\phi(z) = \overline{\alpha} + \alpha z$ (binomial with $d = 1$), is also instructive. The pgf $\Phi(z)$ of increasing trees generated by this ϕ is

$$\Phi(z) = \frac{\overline{\alpha}}{\alpha}\left(e^{mz} - 1\right) = \frac{\overline{\alpha}}{\alpha}\left(\overline{\alpha}^{-z} - 1\right),$$

where $m = -\log\overline{\alpha}$. It is the pgf of a Poisson(m) rv conditioned on being ≥ 1.

This branching mechanism (which was excluded for simply generated GW trees), would have led to $\Phi(z) = \overline{\alpha}z/(1 - \alpha z)$ in the GW setup; this is the pgf of a Geometric distribution with success probability $\overline{\alpha}$. It has no branch-point singularity.

• Poisson: $\phi\left(z\right)=e^{-\mu\left(1-z\right)}$, $\mu>0$. With $m=\left(e^{\mu}-1\right)/\mu$,

$$\begin{aligned}\Phi\left(z\right)&=1-\frac{1}{\mu}\log\left(1-\mu m\left(z-1\right)\right)=-\frac{1}{\mu}\log\left(1-\frac{\mu mz}{1+\mu m}\right)\\&=-\frac{1}{\mu}\log\left(1-z/z_{c}\right).\end{aligned}$$

It has a pure logarithmic singularity at $z_{c}=\left(1+\mu m\right)/\mu m=1/\left(1-e^{-\mu}\right)>1$. Note $\Phi\left(z\right)\rightarrow\infty$ as $z\rightarrow z_{c}$. Thus,

$$\overline{\pi}_{n}=\left[z^{n}\right]\Phi\left(z\right)=\mu^{-1}n^{-1}\cdot z_{c}^{-n},\ n\geq1,$$

a Fisher log-series probability mass function.

• Negative binomial: $\phi\left(z\right)=\left(\left(1-\alpha z\right)/\beta\right)^{-\theta}$, $\theta>0$ with $\mu=\theta\alpha/\beta$. with $m=\frac{\beta}{\alpha}\frac{\beta^{-\left(\theta+1\right)}-1}{1+\theta}=\frac{\theta}{\mu}\frac{\left(1+\mu/\theta\right)^{\theta+1}-1}{\theta+1}$,

$$\begin{aligned}\Phi\left(z\right)&=\frac{1}{\alpha}\left(1-\left(1-\left(\theta+1\right)m\alpha\beta^{\theta}z\right)^{1/\left(\theta+1\right)}\right)\\&=\frac{1}{\alpha}\left(1-\left(1-z/z_{c}\right)^{1/\left(\theta+1\right)}\right).\end{aligned}$$

It has a power-singularity of order $-1/\left(\theta+1\right)$ at $z_{c}=1/\left(\alpha m\left(\theta+1\right)\beta^{\theta}\right)=1/\left(1-\beta^{\theta+1}\right)>1$. Note $\infty>\Phi\left(z_{c}\right)=\frac{1}{\alpha}>1$.

Thus, for large n,

$$\left[z^{n}\right]\Phi\left(z\right)\sim\frac{-1}{\alpha\Gamma\left(-1/\left(\theta+1\right)\right)}n^{-\left(1+1/\left(\theta+1\right)\right)}z_{c}^{-n}.$$

When β approaches 0 ($\mu\rightarrow\infty$), $z_{c},\alpha\rightarrow1$ and

$$\left[z^{n}\right]\Phi\left(z\right)\underset{n\rightarrow\infty}{\sim}\frac{-1}{\Gamma\left(-1/\left(\theta+1\right)\right)}n^{-\left(1+1/\left(\theta+1\right)\right)},$$

a pure power-law.

The exponent of the power-law factor is model-dependent and universality of the power-law exponent $3/2$ is lost: it is respectively $1-1/\left(d-1\right)$, 1 or $1+1/\left(\theta+1\right)$ for the binomial, Poisson and negative binomial cases. These exponents are dictated by singularity analysis, namely: one minus the order of the power-singularity of $\Phi\left(z\right)$ at z_{c}.

For increasing weighting trees whose pgfs obey $\Phi'\left(z\right)=m\phi\left(\Phi\left(z\right)\right)$, $\Phi\left(0\right)=0$, $\Phi\left(1\right)=1$, we have

$$z=\frac{1}{m}\int_{0}^{\Phi\left(z\right)}\frac{dz'}{\phi\left(z'\right)},\ \text{with}\ m=\int_{0}^{1}\frac{dz'}{\phi\left(z'\right)}.$$

Two cases arises:

$*$ If ϕ is entire, then Φ is singular at

$$z_{c}=\frac{1}{m}\int_{0}^{\infty}\frac{dz'}{\phi\left(z'\right)},\ \text{with}\ z_{c}-z=\frac{1}{m}\int_{\Phi\left(z\right)}^{\infty}\frac{dz'}{\phi\left(z'\right)},\tag{48}$$

giving the behavior of $\Phi\left(z\right)$ near z_c. Here, $\Phi\left(z_c\right)=\infty$. This behavior is explicit when $1/\phi$ is integrable.

Example 3.1.

- Binomial: $\phi\left(z\right)=\left(\overline{\alpha}+\alpha z\right)^d$. Here, $z_c=1/\left(1-\overline{\alpha}^{d-1}\right)$ and $\Phi\left(z\right)$ has a power-singularity of order $1/\left(d-1\right)$ with $\Phi\left(z_c\right)=\infty$.

- (Poisson): $\phi\left(z\right)=e^{-\mu(1-z)}$, $\mu>0$. With $z_c=\left(1+\mu m\right)/\mu m=1/\left(1-e^{-\mu}\right)>1$,

$$\Phi\left(z\right)=-\frac{1}{\mu}\log\left(1-z/z_c\right),\text{ with }\Phi\left(z_c\right)=\infty.$$

The power order of the singularity is 0 and $\Phi\left(z\right)$ has a pure logarithmic singularity at z_c of order 1. Note $\Phi\left(z_c\right)=\infty$.

$*$ If ϕ is singular at $z=z_+$, then Φ is singular at

$$z_c=\frac{1}{m}\int_0^{z_+}\frac{dz'}{\phi\left(z'\right)}, \tag{49}$$

leading to

$$z_c-z=\frac{1}{m}\int_{\Phi(z)}^{z_+}\frac{dz'}{\phi\left(z'\right)},\text{ with }\Phi\left(z_c\right)=z_+. \tag{50}$$

Example 3.2.

- (negative binomial): $\phi\left(z\right)=\left(\left(1-\alpha z\right)/\beta\right)^{-\theta}$, for which $z_+=1/\alpha$, consistently gives $z_c=1/\left(1-\beta^{\theta+1}\right)$ and

$$\Phi\left(z\right)=z_+-z_+\left(\frac{mz_c\left(\theta+1\right)\beta^{\theta}}{z_+}\right)^{1/(\theta+1)}\left(1-z/z_c\right)^{1/(\theta+1)},$$

with a power-singularity of order $-1/\left(\theta+1\right)$ at z_c and $\Phi\left(z_c\right)=z_+<\infty$ (note that the order $-1/2$ found for generic Galton-Watson trees is obtained when $\theta=1$ only).

- (Sibuya) When $\phi\left(z\right)=2-\sqrt{\left(2-z\right)}$, in the Sibuya class with $z_+=2$, with $z_c=\frac{2}{m}\left(2\log\left(2+\sqrt{2}\right)-\sqrt{2}\right)$,

$$\Phi\left(z\right)=2-H\left(\frac{mz_c}{2}\left(1-z/z_c\right)\right)^2,$$

where H is the inverse function of $G\left(z\right)=-2\log\left(1-z/2\right)-z$. Thus, for some explicit constants C_1,C_2, to the dominant orders, $\Phi\left(z\right)\underset{z\to z_c}{\sim}2-C_1\left(1-z/z_c\right)-C_2\left(1-z/z_c\right)^{3/2}$. Thus, $\Phi'\left(z\right)$ displays a power-singularity of order $-1/2$ at z_c, leading to $\left[z^n\right]\Phi\left(z\right)\approx n^{-5/2}z_c^n$.

For Sibuya with $\alpha\in\left(0,1\right)$, $\Phi\left(z\right)\underset{z\to z_c}{\sim}2-C_1\left(1-z/z_c\right)-C_2\left(1-z/z_c\right)^{1+\alpha}$, so $\Phi'\left(z\right)$ displays a power-singularity of order $-\alpha$ at z_c and

$$\left[z^n\right]\Phi'\left(z\right)\approx n^{-(\alpha+1)}z_c^n,\text{ else }\left[z^n\right]\Phi\left(z\right)\approx n^{-(\alpha+2)}z_c^n.$$

Remark 3.2.

When dealing with a forest of increasing trees with a fixed number p of trees, both abundance distributions (5) and (7) also hold and are explicit whenever one is able to compute $\mathbf{P}\left(\overline{N}\left(p\right)=n\right)=\left[z^{n}\right]\Phi\left(z\right)^{p}$ for all $n,p\leq n$, in particular $\overline{\pi}_{m}=\mathbf{P}\left(\overline{N}\left(1\right)=m\right)=\left[z^{m}\right]\Phi\left(z\right)$, but now with the $\Phi\left(z\right)$ defined by (46). Large n estimates of $\left[z^{n}\right]\Phi\left(z\right)^{p}$ can be found in the binomial, Poisson and negative binomial generators of increasing trees. For instance, Equation (41) involving Bell polynomials holds with $c_{\bullet}=\bullet!\overline{\pi}_{\bullet}$ and $\overline{\pi}_{\bullet}=\left[z^{\bullet}\right]\Phi\left(z\right)$, explicit in the large n asymptotics for the three examples considered.

We shall now consider the case with a random number P of trees, either Poisson or geometric.

3.4. A recursive nucleation/aggregation model of trees creation/deletion

For random forests of increasing weighted trees with P trees, Equation (29) holds with $c_{\bullet}=\bullet!\overline{\pi}_{\bullet}$ and $\overline{\pi}_{\bullet}=\left[z^{\bullet}\right]\Phi\left(z\right)$, where $\Phi\left(z\right)$ now solves (46). It turns out however that these joint probabilities $\mathbf{P}\left(\mathcal{N}=n,P=p\right)$ can be obtained by recurrence, suggesting that the creation/deletion of clusters (trees) can be obtained from a recursive nucleation/aggregation model as one individual is added.

Case (a). Random Poissonnian forest of increasing trees.

$$\Psi\left(z\right)=\mathbf{E}\left(z^{\mathcal{N}}\right)=e^{-\mu_{0}\left(1-\Phi\left(z\right)\right)}=e^{-\mu_{0}}\left(1+\sum_{n\geq1}\sigma_{n}\left(\mu_{0}\right)\frac{z^{n}}{n!}\right),$$

$$\sigma_{n}\left(\mu_{0}\right)=\sum_{p=1}^{n}B_{n,p}\left(c_{\bullet}\right)\mu_{0}^{p},$$

$$\mathbf{P}\left(\mathcal{N}=n,P=p\right)=e^{-\mu_{0}}\mu_{0}^{p}B_{n,p}\left(c_{\bullet}\right)/n!.$$

From $\Phi'\left(z\right)=m\phi\left(\Phi\left(z\right)\right)$ we get that $\Sigma\left(z\right)=e^{\mu_{0}\Phi\left(z\right)}$ obeys (recall $\pi_{k}=\mathbf{P}\left(M=k\right)$)

$$\begin{aligned}\Sigma'\left(z\right)&=m\mu_{0}\phi\left(\Phi\left(z\right)\right)\Sigma\left(z\right)=m\mu_{0}\textstyle\sum_{k\geq0}\pi_{k}\Phi\left(z\right)^{k}\Sigma\left(z\right)\\&=m\mu_{0}\textstyle\sum_{k\geq0}\pi_{k}\partial_{\mu_{0}}^{(k)}\Sigma\left(z\right).\end{aligned}$$

Thus,

$$\sigma_{n+1}\left(\mu_{0}\right)=m\mu_{0}\sum_{k\geq0}\pi_{k}\partial_{\mu_{0}}^{(k)}\sigma_{n}\left(\mu_{0}\right).$$

With $p=1,...,n+1$, we obtain the recurrence

$$B_{n+1,p}\left(c_{\bullet}\right)=\sum_{q=p-1}^{n}a_{q,p}B_{n,q}\left(c_{\bullet}\right)\text{ with }a_{q,p}=m\left(q\right)_{p-1}\pi_{q-(p-1)},$$

or equivalently,

$$\left(n+1\right)\mathbf{P}\left(\mathcal{N}=n+1,P=p\right)=\sum_{q=p-1}^{n}b_{q,p}\mathbf{P}\left(\mathcal{N}=n,P=q\right),\tag{51}$$

$$b_{q,p} = a_{q,p}\mu_0^{p-q} = m\mu_0^{p-q}\left(q\right)_{p-1}\pi_{q-(p-1)}.$$

In particular,

$$\begin{aligned}\mathbf{P}\left(\mathcal{N}=n, P=n\right) &= e^{-\mu_0}\mu_0^n B_{n,n}\left(c_\bullet\right)/n! = e^{-\mu_0}\left(\mu_0\overline{\pi}_1\right)^n/n!,\\ \mathbf{P}\left(\mathcal{N}=n, P=1\right) &= e^{-\mu_0}B_{n,1}\left(c_\bullet\right)/n! = e^{-\mu_0}\overline{\pi}_n,\end{aligned}$$

are respectively the probabilities to observe a Poissonian forest with n singleton root trees and a Poissonian forest made of one single tree with n nodes.

The number $b_{p-1,p}/\left(n+1\right)$ is the nucleation *speciation* probability, i.e., the probability that, when adding one *individual* or node (the transition $n \to n+1$), it forms a new cluster or *species* (as a singleton node tree) by itself. We have $b_{p-1,p}/\left(n+1\right) = m\mu_0\pi_0/\left(n+1\right)$, where we recognize the term $\overline{\pi}_1 = m\pi_0$.

The numbers $b_{q,p}/\left(n+1\right)$, $q \geq p$, are the aggregation probabilities, i.e., the merging probability that this new *individual* will connect $q-(p-1)$ clusters (trees), out of q possible, thereby ending up in the next step with a population made of p clusters (trees). Note that these aggregation probabilities are found to be proportional to $\pi_{q-(p-1)}$, the probability that these $q-(p-1)$ trees can have the new individual as a common ancestor. When an aggregation event takes places indeed, the new incoming individual $n+1$ becomes the common root to the $q-(p-1)$ trees that it aggregates (there are $\left(q\right)_{p-1}$ ways to chose them), thereby forming a new unique tree. Note that for a population with n individuals (labeled 1 to n) that has been formed recursively in this way, the nodes of each tree of the forest are labeled in decreasing order from root to leaves; upon a circular permutation of the n nodes labels, they can equivalently be arranged in increasing order.

Case (b). Random geometric forest of increasing trees.

In this model, the trees are assumed distinguishable. Proceeding similarly, with $\phi_0\left(z\right) = \beta_0/\left(1-\alpha_0 z\right)$, with $\sigma_n\left(\alpha_0\right) = \sum_{p=1}^{n} p! B_{n,p}\left(c_\bullet\right)\alpha_0^p$, we found

$$\mathbf{P}\left(\mathcal{N}=n, P=p\right) = \beta_0\alpha_0^p B_{n,p}\left(c_\bullet\right)p!/n!.$$

With $p = 1, ..., n+1$

$$B_{n+1,p}\left(c_\bullet\right) = \sum_{q=p-1}^{n} a_{q,p}B_{n,q}\left(c_\bullet\right) \text{ with } a_{q,p} = m\left(q\right)_{p-1}\pi_{q-(p-1)},$$

$$\left(n+1\right)\mathbf{P}\left(\mathcal{N}=n+1, P=p\right) = \sum_{q=p-1}^{n} b_{q,p}\mathbf{P}\left(\mathcal{N}=n, P=q\right), \tag{52}$$

$$b_{q,p} = a_{q,p}\frac{p!}{q!}\alpha_0^{p-q} = m\alpha_0^{p-q}\frac{p!}{q!}\left(q\right)_{p-1}\pi_{q-(p-1)}.$$

In particular, $\mathbf{P}\left(\mathcal{N}=n, P=n\right) = \beta_0\alpha_0^n B_{n,n}\left(c_\bullet\right) = \beta_0\left(\alpha_0\overline{\pi}_1\right)^n$, $\mathbf{P}\left(\mathcal{N}=n, P=1\right) = \beta_0\alpha_0 B_{n,1}\left(c_\bullet\right)/n! = \beta_0\alpha_0\overline{\pi}_n$ are respectively the probabilities to have a geometric forest with n singleton trees and a geometric forest made of one single tree with n nodes.

3.5. Number of trees in forests of increasing trees with a large number of nodes

We now investigate the number of species in a random forest of increasing trees when the total population size n goes large in the three examples: binomial, Poisson and negative binomial. Both cases P Poisson and P geometric are discussed.

I. Binomial increasing trees examples.

Case (a). Poissonian forests.

- Take a binomial increasing tree example with $d = 2$. With $z_c = 1/\alpha$:

$$\begin{aligned}\Phi(z) &= \tfrac{\overline{\alpha}}{\alpha}\left((1 - z/z_c)^{-1} - 1\right),\\ \Psi(z) &= \exp\left(-\mu_0\left(1 - \Phi(z)\right)\right) = e^{-\mu_0}\Sigma(z).\end{aligned}$$

With $K = \mu_0\overline{\alpha}/\alpha$, a saddle point analysis yields (see Flajolet and Sedgewick (1994), p. 29, and Hwang (1994), Th. 36, p. 147)

$$\sigma_n(\mu_0) = [z^n]\,\Sigma(z) \sim \frac{e^{K/2} n^{-3/4} e^{2\sqrt{Kn}}}{2\sqrt{\pi}K^{-1/4}} z_c^{-n}.$$

Therefore,

$$\mathbf{E}\left(z^{P_n}\right)^{1/\sqrt{n}} = \left(\frac{\sigma_n(z\mu_0)}{\sigma_n(\mu_0)}\right)^{1/\sqrt{n}} \underset{n\to\infty}{\to} a_{\mu_0}(z) = e^{2\sqrt{\mu_0\overline{\alpha}/\alpha}\left(\sqrt{z}-1\right)},$$

so that, almost surely,

$$\frac{1}{\sqrt{n}}P_n \underset{n\to\infty}{\to} a'_{\mu_0}(1) = \sqrt{\mu_0\overline{\alpha}/\alpha}. \tag{53}$$

We conjecture that if d is any integer with $d \geq 2$, the scaling factor in front of P_n is $n^{-1/d}$.

Case (b). Geometric forests.

With $z_c = 1/\left(1 - \overline{\alpha}^{d-1}\right)$:

$$\begin{aligned}\Phi(z) &= \tfrac{\overline{\alpha}}{\alpha}\left((1 - z/z_c)^{-1/(d-1)} - 1\right),\\ \Psi(z) &= \beta_0/\left(1 - \alpha_0\Phi(z)\right) = \beta_0\Sigma(z).\end{aligned}$$

The function $\Psi(z)$ always has a simple pole at $1 < z_*(\alpha_0) < z_c$ with

$$z_*(\alpha_0) = z_c\left(1 - \left(\frac{\alpha_0\overline{\alpha}}{\alpha + \alpha_0\overline{\alpha}}\right)^{d-1}\right).$$

Therefore,

$$\sigma_n(\alpha_0) = [z^n]\,\Sigma(z) \underset{n\to\infty}{\sim} -\alpha_0\Phi'(z_*)\left(1 - z/z_*(\alpha_0)\right)^{-1},$$

and

$$\mathbf{E}\left(z^{P_n}\right)^{1/n} = \left(\frac{\sigma_n(z\alpha_0)}{\sigma_n(\alpha_0)}\right)^{1/n} \underset{n\to\infty}{\to} a_{\alpha_0}(z) = \frac{z_*(\alpha_0)}{z_*(z\alpha_0)},$$

with

$$n^{-1}P_n \underset{n\to\infty}{\to} a'_{\alpha_0}(1) < 1, \text{ almost surely.} \tag{54}$$

II. Poissonian increasing trees examples.

Case (a). Poissonian forests.

- Take a Poissonian increasing tree example. With $z_c = 1/\left(1-e^{-\mu}\right)$,

$$\begin{aligned}&\Phi(z) = -\tfrac{1}{\mu}\log(1-z/z_c),\\ &\Psi(z) = \exp(-\mu_0(1-\Phi(z))) = e^{-\mu_0}\Sigma(z),\end{aligned}$$

where $\Sigma(z) = (1-z/z_c)^{-\mu_0/\mu}$. We get

$$\sigma_n(\mu_0) = [z^n]\,\Sigma(z) = [\mu_0/\mu]_n\, z_c^{-n}/n!.$$

Therefore,

$$\mathbf{E}\left(z^{P_n}\right) = \frac{\sigma_n(z\mu_0)}{\sigma_n(\mu_0)} = \frac{[\mu_0 z/\mu]_n}{[\mu_0/\mu]_n},$$

so that $P_n \overset{d}{=} 1+\sum_{m=1}^{n-1}B_m$, as a sum of independent Bernoulli rvs with success probabilities $(\mu_0/\mu)/(\mu_0/\mu+m)$. By strong law of large numbers, almost surely,

$$\frac{1}{\log n}P_n \underset{n\to\infty}{\to} \mu_0/\mu. \tag{55}$$

Remark 3.3.

Consider $\Phi(z,\mu_m) = (\mu_m-1)\,\overline{\pi}_m z^m + \Phi(z)$, where μ_m marks the size$-m$ trees. We have

$$\begin{aligned}\Psi(z,\mu_m) &= \exp(-\mu_0(1-\Phi(z,\mu_m))) = e^{-\mu_0}\Sigma(z,\mu_m)\\ &= e^{-\mu_0}e^{\mu_0(\mu_m-1)\overline{\pi}_m z^m}(1-z/z_c)^{-\mu_0/\mu},\end{aligned}$$

and $\sigma_n(\mu_0,\mu_m) = [z^n]\,\Sigma(z,\mu_m) = e^{\mu_0(\mu_m-1)\overline{\pi}_m z^m}[\mu_0/\mu]_n\, z_c^{-n}/n!$. Thus,

$$\mathbf{E}\left(z^{P_n}\mu_m^{P_n(m)}\right) = \frac{\sigma_n(z\mu_0,\mu_m)}{\sigma_n(\mu_0,1)} = e^{z\mu_0(\mu_m-1)\overline{\pi}_m z^m}\frac{[\mu_0 z/\mu]_n}{[\mu_0/\mu]_n}.$$

Taking $z=1$, we get $\mathbf{E}\left(\mu_m^{P_n(m)}\right) = e^{(\mu_m-1)\mu_0\overline{\pi}_m}$ showing that, given a forest with n nodes, the number of size$-m$ trees $P_n(m)$ is Poisson distributed with mean $\mu_0\overline{\pi}_m = (\mu_0/\mu)\,m^{-1}\cdot z_c^{-m}$, $m=1,...,n$.

Case (b). Geometric forests.

$$\begin{aligned}&\Phi(z) = -\tfrac{1}{\mu}\log(1-z/z_c),\\ &\Psi(z) = \beta_0/(1-\alpha_0\Phi(z)) = \beta_0\Sigma(z).\end{aligned}$$

The function $\Psi(z)$ always has a simple pole at $1 < z_*(\alpha_0) < z_c$ with

$$z_*(\alpha_0) = z_c\left(1-e^{-\mu/\alpha_0}\right).$$

Therefore,

$$\sigma_n\left(\alpha_0\right)=\left[z^n\right]\Sigma\left(z\right)\underset{n\to\infty}{\sim}-\alpha_0\Phi'\left(z_*\right)\left(1-z/z_*\left(\alpha_0\right)\right)^{-1},$$

and

$$\mathbf{E}\left(z^{P_n}\right)^{1/n}=\left(\frac{\sigma_n\left(z\alpha_0\right)}{\sigma_n\left(\alpha_0\right)}\right)^{1/n}\underset{n\to\infty}{\to}a_{\alpha_0}\left(z\right)=\frac{z_*\left(\alpha_0\right)}{z_*\left(z\alpha_0\right)}=\frac{1-e^{-\mu/\alpha_0}}{1-e^{-\mu/\left(z\alpha_0\right)}},$$

with

$$n^{-1}P_n\underset{n\to\infty}{\to}a'_{\alpha_0}\left(1\right)=\frac{\mu}{\alpha_0\left(e^{\mu/\alpha_0}-1\right)}<1,\text{ almost surely.}\tag{56}$$

III. Negative-binomial increasing trees examples.

Case (a). Poissonian forests.

- Take the negative-binomial increasing tree example. With $z_c=1/\left(1-\beta^{\theta+1}\right)>1$:

$$\begin{array}{l}\Phi\left(z\right)=\frac{1}{\alpha}\left(1-\left(1-z/z_c\right)^{1/(\theta+1)}\right),\\ \Psi\left(z\right)=\exp\left(-\mu_0\left(1-\Phi\left(z\right)\right)\right)=e^{-\mu_0}\Sigma\left(z\right).\end{array}$$

We get $\Sigma\left(z\right)=e^{\mu_0(1-1/\alpha)}e^{\mu_0/\alpha(1-z/z_c)^{1/(\theta+1)}}$ and

$$\sigma_n\left(\mu_0\right)=\left[z^n\right]\Sigma\left(z\right)\sim e^{\mu_0(1-1/\alpha)}\left(1+\frac{\mu_0}{\alpha}n^{-(\theta+1)}z_c^{-n}\right).$$

Therefore,

$$\mathbf{E}\left(z^{P_n}\right)=\left(\frac{\sigma_n\left(z\mu_0\right)}{\sigma_n\left(\mu_0\right)}\right)\underset{n\to\infty}{\to}ze^{\mu_0(1-1/\alpha)(z-1)},$$

so that

$$P_n\underset{n\to\infty}{\overset{d}{\to}}P_\infty,\tag{57}$$

where P_∞ is shifted Poisson($\mu_0\left(1-1/\alpha\right)$) distributed. It can easily be shown, as in the GW trees case, that if $P_n\left(m\right)$ denotes the number of size$-m$ trees in a large population with n individuals in total, that $P_n\left(m\right)$ converges in distribution to a Poisson random variable with mean $\left[z^m\right]\Phi\left(z\right)$.

Case (b). Geometric forests.

$$\begin{array}{l}\Phi\left(z\right)=\frac{1}{\alpha}\left(1-\left(1-z/z_c\right)^{1/(\theta+1)}\right),\\ \Psi\left(z\right)=\beta_0/\left(1-\alpha_0\Phi\left(z\right)\right)=\beta_0\Sigma\left(z\right).\end{array}$$

Because in this case $\Phi\left(z_c\right)=1/\alpha<\infty$, three cases arise, reminiscent of a phase transition:

- (subcritical) If $\alpha>\alpha_0$, then $\Psi\left(z\right)$ is still singular at z_c with

$$\Psi\left(z\right)\underset{z\to z_c}{\sim}\frac{\beta_0}{1-\alpha_0/\alpha}\left(1-\frac{\alpha_0}{\alpha-\alpha_0}\left(1-z/z_c\right)^{1/(\theta+1)}\right),$$

showing that $\sigma_n(\alpha_0) = [z^n]\,\Sigma(z) \underset{n\to\infty}{\sim} \frac{\alpha\alpha_0}{(\alpha-\alpha_0)^2} n^{-(1+1/(\theta+1))} z_c^{-n}/\Gamma(-1/(\theta+1))$ and

$$\mathbf{E}\left(z^{P_n}\right) = \left(\frac{\sigma_n(z\mu_0)}{\sigma_n(\mu_0)}\right) \underset{n\to\infty}{\to} z\left(\frac{1-\alpha_0/\alpha}{1-\alpha_0 z/\alpha}\right)^2 .$$

Therefore,

$$P_n \underset{n\to\infty}{\overset{d}{\to}} P_\infty, \tag{58}$$

where P_∞ is shifted squared-geometric distributed.

- (supercritical) If $\alpha < \alpha_0$, then $\Psi(z)$ has a simple pole at $1 < z_*(\alpha_0) < z_c$ with

$$z_*(\alpha_0) = z_c\left(1-(1-\alpha/\alpha_0)^{\theta+1}\right).$$

Therefore,

$$\sigma_n(\alpha_0) = [z^n]\,\Sigma(z) \underset{n\to\infty}{\sim} -\alpha_0\Phi'(z_*)\left(1-z/z_*(\alpha_0)\right)^{-1},$$

and

$$\mathbf{E}\left(z^{P_n}\right)^{1/n} = \left(\frac{\sigma_n(z\alpha_0)}{\sigma_n(\alpha_0)}\right)^{1/n} \underset{n\to\infty}{\to} a_{\alpha_0}(z) = \frac{z_*(\alpha_0)}{z_*(z\alpha_0)} = \frac{1-(1-\alpha/\alpha_0)^{\theta+1}}{1-(1-\alpha/(\alpha_0 z))^{\theta+1}},$$

with

$$n^{-1}P_n \underset{n\to\infty}{\to} a'_{\alpha_0}(1) = \frac{\alpha(\theta+1)(1-\alpha/\alpha_0)^{\theta}}{\alpha_0\left(1-(1-\alpha/\alpha_0)^{\theta+1}\right)} < 1, \tag{59}$$

almost surely.

- If $\alpha = \alpha_0$ (critical case), then

$$\Psi(z) = \beta_0(1-z/z_c)^{-1/(\theta+1)} = \beta_0\Sigma(z),$$

showing that $\sigma_n(\alpha_0) = [z^n]\,\Sigma(z) \underset{n\to\infty}{\sim} n^{-(1-1/(\theta+1))} z_c^{-n}/\Gamma(1/(\theta+1))$. In this case,

$$z_c = 1/\left(1-\beta^{\theta+1}\right) = 1/\left(1-(1-\alpha_0)^{\theta+1}\right) = z_c(\alpha_0),$$

and

$$\mathbf{E}\left(z^{P_n}\right)^{1/n} = \left(\frac{\sigma_n(z\alpha_0)}{\sigma_n(\alpha_0)}\right)^{1/n} \underset{n\to\infty}{\to} a_{\alpha_0}(z) = \frac{z_c(\alpha_0)}{z_c(z\alpha_0)} = \frac{1-(1-z\alpha_0)^{\theta+1}}{1-(1-\alpha_0)^{\theta+1}}.$$

This shows that

$$n^{-1}P_n \underset{n\to\infty}{\to} a'_{\alpha_0}(1) = \frac{\alpha_0(\theta+1)(1-\alpha_0)^{\theta}}{1-(1-\alpha_0)^{\theta+1}} < 1, \text{ almost surely.} \tag{60}$$

4. Conclusion

Let some population be made of n individuals that can be of P possible species (or types) at equilibrium. One question raised is: How are individuals scattered among types? In a first scenario for such species abundance distributions, each species grows from independent founders according to a Galton-Watson branching process. When the number of founders P is either fixed or random (either Poisson or geometrically-distributed), we address the above question. This model is one pertaining to forests of Galton-Watson trees. A second scenario that we will address in a similar way deals with forests of increasing trees. Underlying this setup, the creation/annihilation of clusters (trees) is shown to result from a recursive nucleation/aggregation process as one additional individual is added to the total population. To proceed, we use singularity analysis of generating functions.

Acknowledgment:

The author acknowledges partial support from the labex MME-DII (Modèles Mathématiques et Économiques de la Dynamique, de l' Incertitude et des Interactions), ANR11-LBX-0023-01. This work also benefited from the support of the Chair "Modélisation Mathématique et Biodiversité" of Veolia-Ecole Polytechnique-MNHN-Fondation X. Finally, this work was also funded by CY Initiative of Excellence (grant "Investissements d'Avenir" ANR- 16-IDEX-0008), Project "EcoDep" PSI-AAP2020-0000000013.

REFERENCES

Aly, E.E.A.A. and Bouzar, N. (2000). On geometric infinite divisibility and stability, Annals of the Institute of Statistical Mathematics, Vol. 52, pp. 790-799.

Bergeron, F., Flajolet, P. and Salvy, B. (1992). Varieties of increasing trees, CAAPs 92, ed. J.C. Raoult. Lecture Notes in Comput. Sci., Vol. 58, pp. 24-48.

Comtet, L. (1970). *Analyse Combinatoire*, Tomes 1 et 2, Presses Universitaires de France, Paris.

Consul, P.C. and Famoye, F. (2006). *Lagrangian Probability Distributions*, Birkhäuser, Boston.

Demetrius, L. (1983). Statistical mechanics and population biology, Journal of Statistical Physics, Vol. 30, No. 3, pp. 709-753.

Drmota, M. (2009). *Random Trees: An Interplay between Combinatorics and Probability* (First Edition), Springer Publishing Company, Incorporated.

Ellis, R.S. (1985). *Entropy, Large Deviations and Statistical Mechanics*, Grundlehren Math. Wiss. (Springer, New York).

Engen, S. (1974). On species frequency models, Biometrika, Vol. 61, pp. 263-270.

Engen, S. (1978). *Stochastic Abundance Models,* Monographs on Applied Probability and Statistics, Chapman and Hall, London.

Flajolet, P. and Odlyzko, A. (1990). Singularity analysis of generating functions, SIAM J. Discrete Math., Vol. 3, No. 2, pp. 216-240.

Flajolet, P. and Sedgewick, R. (1993). The average case analysis of algorithms: Complex asymptotics and generating functions, Research Report No 2026, INRIA-00074645.

Flajolet, P. and Sedgewick, R. (1994). The average case analysis of algorithms: Saddle point asymptotics, Research Report No 2376, INRIA-00074300.

Flory, P.J. (1941a). Molecular size distribution in three-dimensional polymers, I Gelation, Journal of the American Chemical Society, Vol. 63, pp. 3083-3090.

Flory, P.J. (1941b). Molecular size distribution in three-dimensional polymers, II Trifunctional branching units, Journal of the American Chemical Society, Vol. 63, pp. 3091-3096.

Harris, T.E. (1963). *The Theory of Branching Processes,* Die Grundlehren der Mathematischen Wissenschaften, Bd. 119 Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J.

Hwang H-K. (1994). *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques*, PhD Thesis, Ecole Polytechnique.

Pitman, J. (1998). Enumerations of trees and forests related to branching processes and random walks. In D. Aldous and J. Propp (Eds.), *Microsurveys in Discrete Probability*, No. 41 in DIMACS Ser. Discrete Math. Theoret. Comp. Sci, pp. 163-180, Providence RI.

Pitman, J. (2006). *Combinatorial Stochastic Processes,* In *Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 724, 2002, with a foreword by Jean Picard.* Lecture Notes in Mathematics, No. 1875. Springer-Verlag, Berlin.

Sheth, R.K. (1996). Galton-Watson branching processes and the growth of gravitational clustering, Mon. Not. R. Astron. Soc., Vol. 281, pp. 1277-1289.

Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions, Annals of the Institute of Statistical Mathematics, Vol. 31, pp. 373-390.

Simkin, M.V. and Roychowdhury, V.P. (2011). Re-inventing Willis, Physics Reports, Vol. 502, No. 1, pp. 1-35.

Stockmayer, W.H. (1943). Theory of molecular size distribution and gel formation in branched chain polymers, Journal of Chemical Physics, Vol. 11, pp. 45-55.

Tanner, J.C. (1961). A derivation of the Borel distribution, Biometrika, Vol. 48, pp. 222-224.

Appendix

We briefly recall a general transfer result of singularity analysis (see Flajolet and Odlyzko (1990)) for generating functions with power-logarithmic singularities of given orders a and b.

Let $\Phi\left(z\right)$ be any analytic function in the indented domain defined by

$$D = \left\{z : |z| \leq z_1,\ \ |\text{Arg}\left(z - z_c\right)| > \pi/2 - \eta\right\},$$

where z_c, $z_1 > z_c$, and η are positive real numbers. Assume that, with $\sigma\left(x\right) = x^a \log^b x$, a and b any real number (respectively the power and logarithmic singularity exponents or orders of Φ at

z_c), we have

$$\Phi\left(z\right)\sim\kappa_1+\kappa_2\sigma\left(\frac{1}{1-z/z_c}\right)\text{ as }z\to z_c\text{ in }D,\tag{61}$$

for some real constants $\kappa_1$and κ_2. Then,

- if $a\notin\{0,-1,-2,...\}$ the coefficients in the expansion of $\Phi\left(z\right)$ satisfy

$$\left[z^n\right]\Phi\left(z\right)\sim\kappa_1+\kappa_2z_c^{-n}\cdot\frac{\sigma\left(n\right)}{n}\frac{1}{\Gamma\left(a\right)}\text{ as }n\to\infty,\tag{62}$$

where $\Gamma\left(a\right)$ is the Euler function. $\Phi\left(z\right)$ presents a power-logarithmic singularity at $z=z_c$. If $b=0$, $\Phi\left(z\right)$ presents a pure power singularity at $z=z_c$ of order a (with power exponent a).

- if $a\in\{0,-1,-2,...\}$, the singularity $z=z_c$ is purely logarithmic and

$$\left[z^n\right]\Phi\left(z\right)\sim\kappa_1+\kappa_2bz_c^{-n}\cdot\frac{\sigma\left(n\right)}{n\cdot\log n}\left(\frac{1}{\Gamma}\right)'\left(a\right)\text{ as }n\to\infty,\tag{63}$$

involving the derivative of the reciprocal Euler function at a.

Thus, for power-logarithmic singularities with orders a and b, the asymptotics of the coefficients can be read from the singular behavior of the generating function under study.