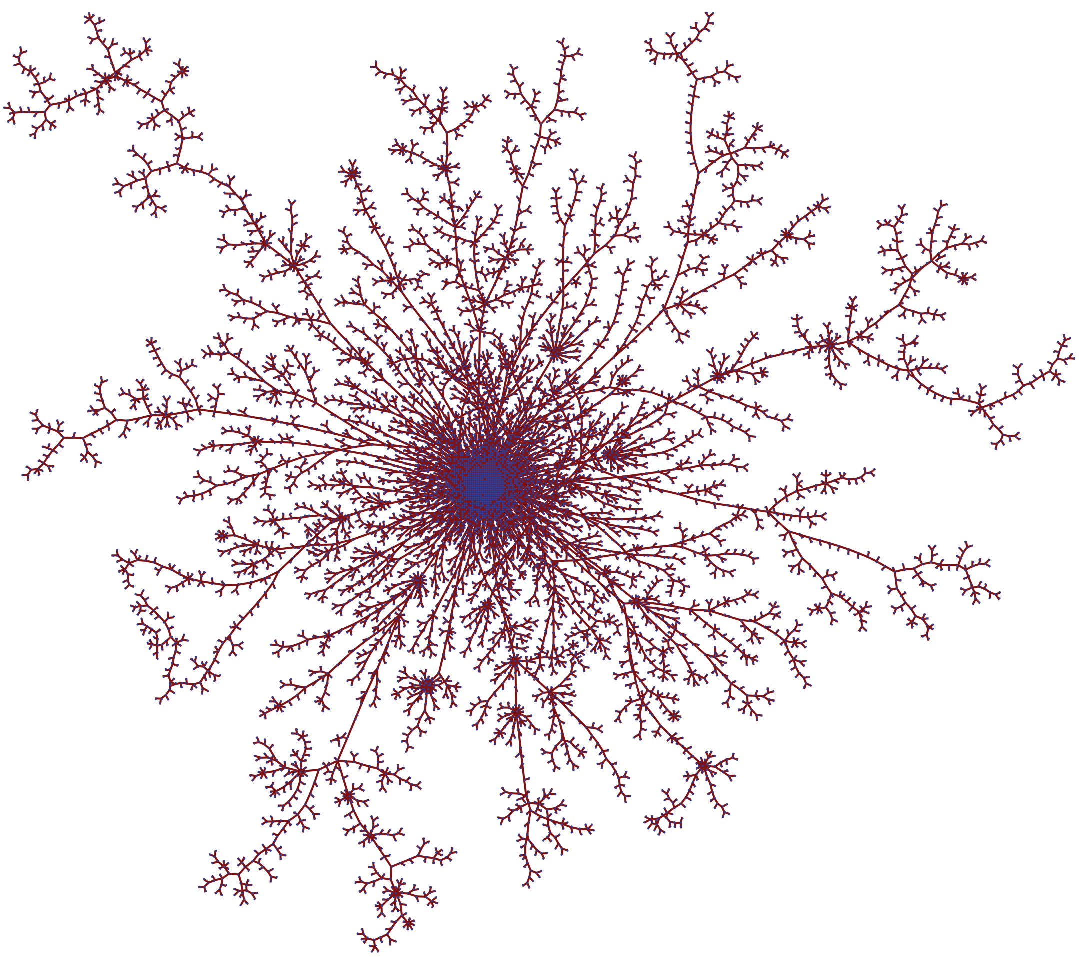
Condensation in random trees

RANDOM TREES AND GRAPHS SUMMER SCHOOL – JULY 2019 – CIRM

LECTURE NOTES – PRELIMINARY VERSION (LAST MODIFIED 3/07/2019)

Igor Kortchemski♠



♠CNRS & CMAP, École polytechnique.. igor.kortchemski@math.cnrs.fr

We study a particular family of random trees which exhibit a condensation phenomenon (identified by Jonsson & Stefánsson in 2011), meaning that a unique vertex with macroscopic degree emerges. This falls into the more general framework of studying the geometric behavior of large random discrete structures as their size grows. Trees appear in many different areas such as computer science (where trees appear in the analysis of random algorithms for instance connected with data allocation), combinatorics (trees are combinatorial objects by essence), mathematical genetics (as phylogenetic trees), in statistical physics (for instance in connection with random maps as we will see below) and in probability theory (where trees describe the genealogical structure of branching processes, fragmentation processes, etc.).

We shall specifically focus on Bienaymé–Galton–Watson trees (which is the simplest possible genealogical model, where individuals reproduce in an asexual and stationary way), whose offspring distribution is subcritical and is regularly varying. The main tool is to code these trees by integer-valued random walks with negative drift, conditioned on a late return to the origin. The study of such random walks, which is of independent interest, reveals a "one-big jump principle" (identified by Armendáriz & Loulakis in 2011), thus explaining the condensation phenomenon.

Section 1 gives some history and motivations for studying Bienaymé–Galton–Watson trees.

Section 2 defines Bienaymé–Galton–Watson trees.

Section 3 explains how such trees can be coded by random walks, and introduce several useful tools, such as cyclic shifts and the Vervaat transformation, to study random walks under a conditioning involving positivity constraints.

Section 4 contains exercises to manipulate connections between BGW trees and random walks, and to study ladder times of downward skip-free random walks.

Section 5 gives estimates, such as maximal inequalities, for random walks in order to establish a "one-big jump principle".

Section 6 transfers results on random walks to random trees in order to identity the condensation phenomenon.

The goal of these lecture notes is to be as most self-contained as possible.

Acknowledgments. Many thanks to Guillaume Conchon-Kerjan, Mickaël Maazoun and Cyril Marzouk for useful suggestions and comments.

Contents

1 History and motivations

Bienaymé–Galton–Watson processes. The origin of Bienaymé–Galton–Watson processes goes back to the middle of the 19th century, where they are introduced to estimate extinction probabilities of noble names. In 1875, Galton & Watson [69] use an approach based on generating functions. While the method is correct, a mistake appears in their work (they conclude that the extinction probability is always 1, see [11, Chapitre 9]), and one has to wait until 1930 for the first complete published proof by Steffensen [66].

However, in 1972, Heyde & Seneta [38] discover a note written by Bienaymé [15] dated from 1845, where Bienaymé correctly states that the extinction probability is equal to 1 if and only if the mean of the offspring distribution is at most 1. Some explanations are given, but there is no known published proof. Nonetheless it appears to be very plausible that Bienaymé had found a proof using generating functions (see [43] and [11, Chap 7] for a historical overview).

Since, there has been a large amount of work devoted to the study of long time asymptotics of branching processes; see [55, Section 12] and [10] for a description of results in this description.

1.1 Scaling limits

The birth of the Brownian tree. Starting from the second half of the 20th century, different scientific communities have been interested in the asymptotic behavior of random trees chosen either uniformly at random among a certain class, or conditioned to be "large". At the crossroads of probability, combinatorics and computer science, using generating functions and analytic combinatorics, various statistics of such trees have been considered, such as the maximal degree, the number of vertices with fixed degree, or the profile of the tree. See [26] for a detailed treatment.

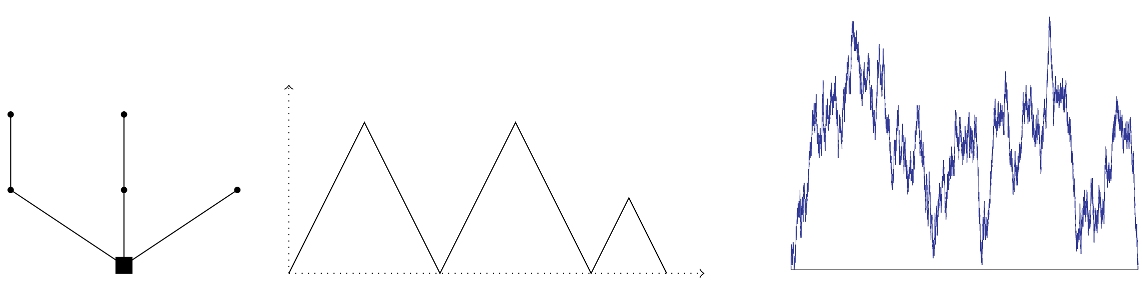


Figure 1: From left to right: a tree with 6 vertices, its associated contour function, and the contour function (appropriately scaled in time and space) of a large Bienaymé–Galton–Watson tree (with a critical finite variance offspring distribution) which converges in distribution to the brownian excursion.

In the early 1990's, instead of only considering statistics, Aldous suggested to study the convergence of large random trees (rooted and ordered, see Sec. 2.1 for a definition) globally. More precisely, Aldous [6] considered random trees as random compact subsets of the space ℓ_1 of summable sequences, and established in this framework that a random Bienaymé–Galton–Watson tree with Poisson parameter 1 offspring distribution, conditioned on having n vertices, converges in distribution, as $n \to \infty$, to a random compact subset called the *Continuum Random Tree* (in short, the CRT). A bit later, Aldous [7, 8], gave a simple construction of the CRT from a normalised Brownian excursion $\mathbb{e}$ (which can informally be viewed as a Brownian motion conditioned to be back at 0 at time 1 and conditioned to be positive on $(0, 1)$), and showed that the appropriately scaled contour function (see Fig. 1) of a random Bienaymé–Galton–Watson tree with critical (i.e. mean 1) finite variance offspring distribution, conditioned to have n vertices, converges (in distribution in the space of real valued continuous functions on $[0, 1]$ equipped with the topology of uniform convergence), as $n \to \infty$ to $\mathbb{e}$. For this reason, the CRT is usually called the Brownian tree, and appears as a *universal* in the sense that BGW trees with various offspring distributions converge to the same continuous object (see Fig. 2 for a picture of a large Bienaymé–Galton–Watson tree with a critical finite variance offspring distribution). We mention that the finite variance condition, reminiscent of the central limit theorem, is crucial.

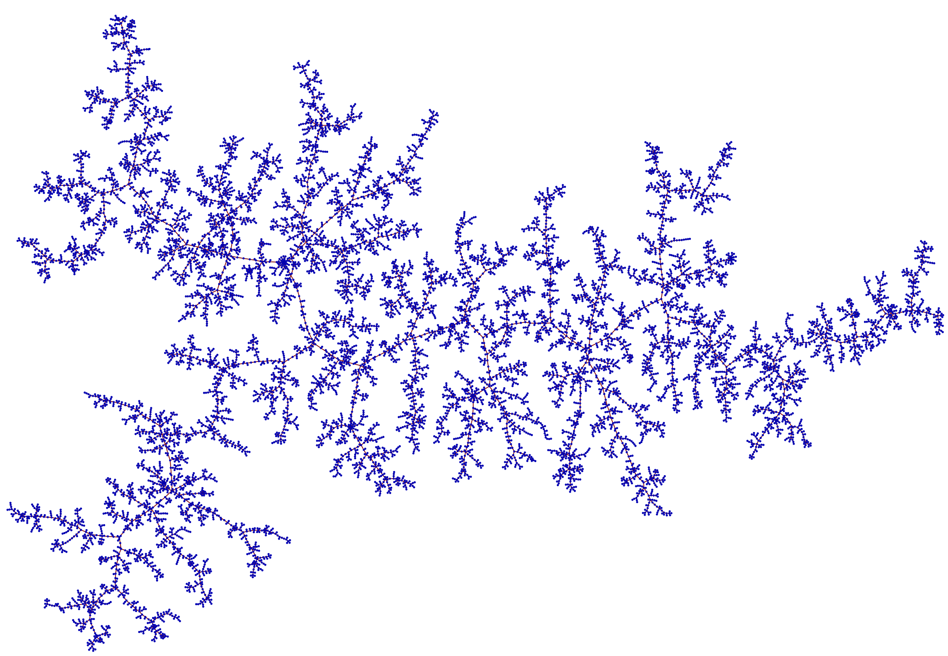


Figure 2: A realization of a large BGW tree with a critical finite variance offspring distribution, which approximates the Brownian CRT.

In 2003, Evans, Pitman & Winter [30] suggest to use the formalism of $\mathbb{R}$-trees, introduced earlier for geometric and algebraic purposes (see for instance [60]), and the Gromov–

Hausdorff topology, introduced by Gromov [35] to prove the so-called Gromov theorem of groups with polynomial growth. The Gromov–Hausdorff distance defines a topology on compact metric spaces (seen up to isometries), which allows to give a meaning to the notion of convergence of compact metric spaces. This point of view, which consists in viewing trees as compact metric spaces (by simply equipping their vertex set with the graph distance) and in studying their scaling limits, is now widely used and gives a natural and powerful framework for studying abstract converges of random graphs (in particular those who are not coded by excursion-type functions). By *scaling limits* we mean the study of limits of large random discrete objects seen in their globality, after suitable renormalisation.

Universality of the Brownian tree. In the last years, it has been realized that the Brownian tree is also the scaling limit of non-planar random trees [36, 58], non-rooted trees [62] but also of various models of random graphs with are not trees, such as stack triangulations [5], random dissections [20], random quadrangulations with a large boundary [14], random outerplanar maps [18, 68], random bipartite maps with one macroscopique face [40], brownian bridges in hyperbolic spaces [19] or subcritical random graphs [59]. See [67] for a combinatorial framework and further examples.

Stable Lévy trees. An important step in the generalization of Aldous' results was made by Le Gall & Le Jan [53], who considered the case where the offspring distribution μ is critical but infinite variance, under the assumption that μ has a heavy tail (more precisely, $\mu([n,\infty))$ is of order $n^{-\alpha}$ as $n \to \infty$, with $\alpha \in (1,2]$). In this setting, it was shown [28, 27, 29] that such a BGW_μ tree, conditioned on having n vertices and appropriately scaled, converges in distribution to another random limiting tree: the random α-stable random tree, who has roughly speaking vertices with large degrees (see Fig. 3 for simulations).

Stable trees (in particular of index $\alpha = 3/2$) play an important role in the theory of scaling limits of random planar maps [49, 23, 56], where one of the motivations is to give a precise sense to the notion of "canonical two-dimensional surface" [50] (see Fig. 4).

Other types of conditioning. Conditionings that involve other quantities than the total number of vertices have also been considered in the context of scaling limits, mostly in view of various applications. For instance, conditionings involving the height have been studied in [34, 52]. Other types of conditionings involving degrees has recently attracted attention: Rizzolo [64] introduced the conditioning on having a fixed number of vertices with given outdegrees (see also [46]), while Broutin & Marckert [17] and Addario-Berry [3] consider random trees with a given degree sequence.

Non-generic Bienaymé–Galton–Watson trees. Since the study of conditioned non-critical Bienaymé–Galton–Watson trees can often be reduced to critical ones (see Exercise 1), non-

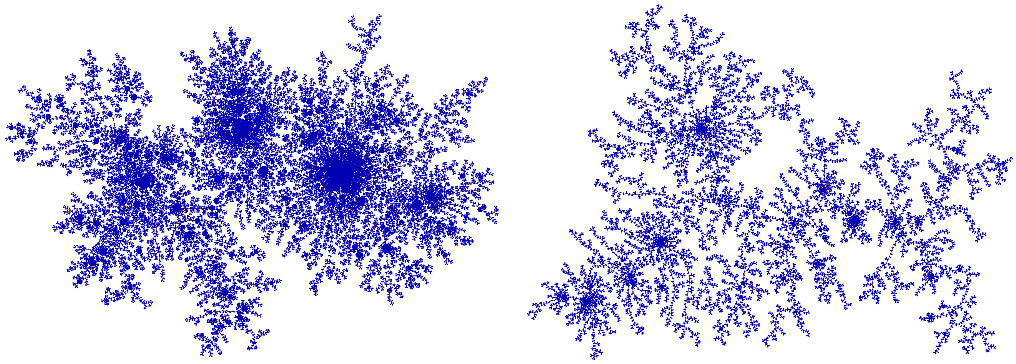


Figure 3: Left: a simulation of an approximation of an α-stable with $\alpha = 1.1$; right: a simulation of an approximation of an α-stable with $\alpha = 1.5$ (the smaller α is, the more vertices tend to have more offspring, which explains why the degrees seems to be bigger for α smaller.

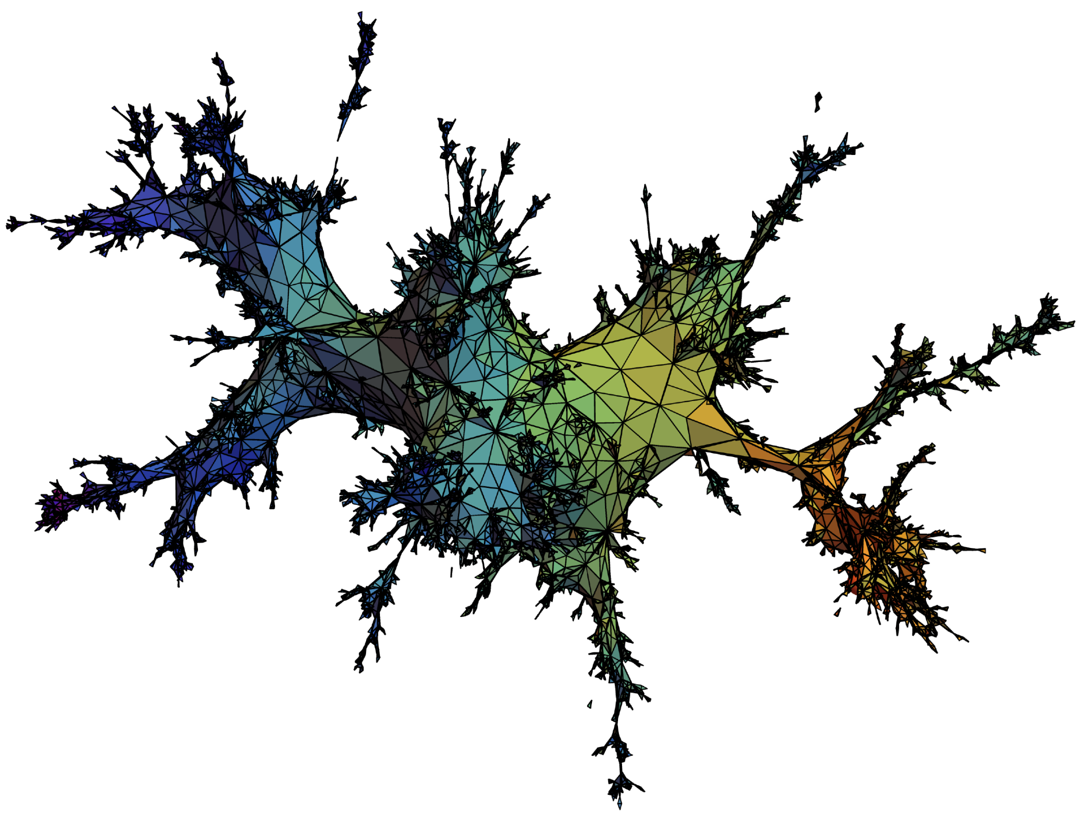


Figure 4: Simulation of a large random quadrangulation of the sphere.

critical Bienaymé–Galton–Watson trees have been set aside for a long time. However, Jonsson & Steffánsson [42] have recently considered the case non-generic trees, which are subcritical Bienaymé–Galton–Watson BGW_μ trees with $\mu(n) \sim c \cdot n^{-\beta-1}$ as $n \to \infty$, and have identified a new phenomenon, called *condensation*: a unique vertex with macroscopic degree, comparable to the size of the tree, emerges (see the figure on the first page). More precise results were then obtained in [47], which show that the second maximal degree is of order $n^{1/\min(2,\beta)}$ and also that in this case there are no nontrivial scaling limits.

Let us mention a recent result [48] concerning the case of critical Cauchy Bienaymé–Galton–Watson trees, where μ is critical and $\mu(n) = L(n)/n^2$ with L slowly varying. In such trees a condensation phenomenon also occurs, but at a slightly smaller scale.

In the recent years, it has been realized that BGW trees in which a condensation phenomenon occurs code a variety of random combinatorial structures such as random planar maps [4, 41, 63], outerplanar maps [65], supercritical percolation clusters of random triangulations [23] or minimal factorizations [32]. These applications are one of the motivations for the study of the fine structure of such large conditioned BGW trees.

Summary (scaling limits). Let us summarize the previously mentioned results, when we consider $\mathcal{T}_n$ a BGW_μ tree conditioned on having n vertices (as we will see in Exercise 1 below, the study of super critical offspring distributions can always be reduced to critical ones) :

- μ is critical and has finite variance. Then distances in $\mathcal{T}_n$ are of order $\sqrt{n}$ (up to a constant), and the scaling limit is the Brownian CRT [6, 7, 8].
- μ is critical, has infinite variance, and $\mu([n,\infty)) = L(n)/n^\alpha$, with L slowly varying and $1 < \alpha \leqslant 2$. Then distances in $\mathcal{T}_n$ are of order $n^{1/\alpha}$ (up to a slowly varying function), and the scaling limit is the α-stable tree [53, 27].
- μ is subcritical and $\mu(n) = L(n)/n^{1+\beta}$ with $\beta > 1$ and L slowly varying. Then condensation occurs: there is a unique vertex of degree of order n (up to a constant) [42], the other degrees are of order $n^{1/\min(2,\beta)}$ (up to a slowly varying constant), the height of the vertex with maximal degree converges in distribution and there are no nontrivial scaling limits [47].

  The goal of these lectures is precisely to study this case.
- μ is critical and $\mu(n) = L(n)/n^2$ with L slowly varying. Condensation occurs, but at a smaller scale, that is $n/L_1(n)$ (where L_1 is slowly varying), the other degrees are of order $n/L_2(n)$ (where L_2 is slowly varying, with $L_2 = o(L_1)$), the height of the vertex with maximal degree converges in probability to ∞ and there are no nontrivial scaling limits [48].

1.2 Local limits

Kesten [45] initiated the study of "local limits" of large random trees (under the conditioning on having height at least n). In this setting, one is interested in the asymptotic behavior of balls of fixed radius around the root as the size of the trees grows. Janson [39] and Abraham & Delmas [2, 1] described the local limits of $\mathcal{T}_n$, a BGW_μ tree conditioned on having n vertices (Jonsson & Stefánsson [42] introduced the so-called condensation tree with a finite spine), in full generality:

- μ is critical. Then $\mathcal{T}_n$ converges locally to a locally finite infinite random tree having an infinite spine (which is the so-called infinite BGW tree conditioned to survive).
- μ is subcritical and the radius of convergence of $\sum_i \mu(i) z^i$ is 1. Then $\mathcal{T}_n$ converges locally to an infinite random tree having a finite spine on top of which sits a vertex with infinite degree.

It is interesting to note that in the case where μ is critical and $\mu(n) = L(n)/n^2$ with L slowly varying, condensation occurs, but the height of the vertex with maximal degree converges in probability to ∞, thus explaining why the local limit is locally finite.

2 Bienaymé–Galton–Watson trees

2.1 Trees

Here, by *tree*, we will always mean plane tree (sometimes also called rooted ordered tree). To define this notion, we follow Neveu's formalism. Let $\mathcal{U}$ be the set of labels defined by

$$\mathcal{U} = \bigcup_{n=0}^{\infty} (\mathbb{N}^*)^n,$$

where, by convention, $(\mathbb{N}^*)^0 = \{\emptyset\}$. In other words, an element of $\mathcal{U}$ is a (possible empty) sequence $u = u_1 \cdots u_j$ of positive integers. When $u = u_1 \cdots u_j$ and $v = v_1 \cdots v_k$ are elements of $\mathcal{U}$, we let $uv = u_1 \cdots u_j v_1 \cdots v_k$ be the concatenation of u and v. In particular, $u\emptyset = \emptyset u = u$. Finally, a *plane tree* is a finite subset of $\mathcal{U}$ satisfying the following three conditions:

(i) $\emptyset \in \tau$,

(ii) if $v \in \tau$ and $v = uj$ for a certain $j \in \mathbb{N}^*$, then $u \in \tau$,

(iii) for every $u \in \tau$, there exists an integer $k_u(\tau) \geqslant 0$ such that for every $j \in \mathbb{N}^*$, $uj \in \tau$ if and only if $1 \leqslant j \leqslant k_u(\tau)$.

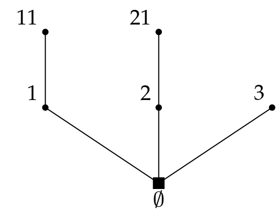


Figure 5: An example of a tree τ, where $\tau = \{\emptyset, 1, 11, 2, 21, 3\}$.

In the sequel, by *tree* we will always mean finite plane tree. We will often view the elements of τ as the individuals of a population whose τ is the genealogical tree and $\emptyset$ is the ancestor (the root). In particular, for $u \in \tau$, we say that $k_u(\tau)$ is the number of children of u, and write k_u when τ is implicit. The size of τ, denoted by $|\tau|$, is the number of vertices of τ. We denote by $\mathbb{A}$ the set of all trees and by $\mathbb{A}_n$ the set of all trees of size n.

2.2 Bienaymé–Galton–Watson trees

We now define a probability measure on $\mathbb{A}$ which describes, roughly speaking, the law of a random tree which describes the genealogical tree of a population where individuals have a random number of children, independently, distributed according to a probability measure μ, called the offspring distribution. Such models were considered by Bienaymé [15] and Galton & Watson [69], who were interested in estimating the probability of extinction of noble names.

We will always make the following assumptions on μ:

(i) $\mu = (\mu(i) : i \geqslant 0)$ is a probability distribution on $\{0, 1, 2, \ldots\}$,

(ii) $\sum_{k \geqslant 0} k\mu(k) \leqslant 1$,

(iii) $\mu(0) + \mu(1) < 1$.

Theorem 2.1. *Set, for every $\tau \in \mathbb{A}$,*

$$\mathbb{P}_\mu(\tau) = \prod_{u \in \tau} \mu(k_u). \tag{1}$$

Then $\mathbb{P}_\mu$ defines a probability distribution on $\mathbb{A}$.

Before proving this result, let us mention that in principle we should define the σ-field used for $\mathbb{A}$. Here, since $\mathbb{A}$ is countable, we simply take the set of all subsets of $\mathbb{A}$ as the σ-field, and we will never mention again measurability issues (one should however be careful when working with infinite trees).

Proof of Theorem 2.1. Set $c = \sum_{\tau \in \mathbb{A}} \mathbb{P}_\mu(\tau)$. Our goal is to show that $c = 1$.

Step 1. We decompose the set of trees according to the number of children of the root and write

$$c = \sum_{k \geqslant 0} \sum_{\tau \in \mathbb{A}, k_\emptyset = k} \mathbb{P}_\mu(\tau) = \sum_{k \geqslant 0} \sum_{\tau_1 \in \mathbb{A}, \ldots, \tau_k \in \mathbb{A}} \mu(k) \mathbb{P}_\mu(\tau_1) \cdots \mathbb{P}_\mu(\tau_k) = \sum_{k \geqslant 0} \mu(k) c^k.$$

Step 2. Set, for $0 \leqslant s \leqslant 1$, $f(s) = \sum_{k \geqslant 0} \mu(k) s^k - s$. Then $f(0) = \mu(0) > 0$, $f(1) = 0$, $f'(1) = (\sum_{i \geqslant 0} i\mu(i)) - 1 < 0$ and $f'' > 0$ on $[0,1]$. Therefore, the only solution of $f(s) = 0$ on $[0,1]$ is $s = 1$.

Step 3. We check that $c \leqslant 1$ by constructing a random variable whose "law" is $\mathbb{P}_\mu$. To this ender, consider a collection $(K_u : u \in \mathcal{U})$ of i.i.d. random variables with same law μ (defined on the same probability space). Then set

$$\mathcal{T} := \left\{ u_1 \cdots u_n \in \mathcal{U} : u_i \leqslant K_{u_1 u_2 \cdots u_{i-1}} \text{ for every } 1 \leqslant i \leqslant n \right\}.$$

(Intuitively, K_u represents the number of children of $u \in \mathcal{U}$ if u is indeed in the tree. Then $\mathcal{T}$ is a random plane tree, but possible infinite. But for a fixed tree $\tau \in \mathbb{T}$, we have

$$\mathbb{P}\left(\mathcal{T} = \tau\right) = \mathbb{P}\left(X_u = k_u(\tau) \text{ for every } u \in \tau\right) = \prod_{u \in \tau} \mu(k_u) = \mathbb{P}_\mu\left(\tau\right).$$

Therefore

$$c = \sum_{\tau \in \mathbb{A}} \mathbb{P}_\mu\left(\tau\right) = \sum_{\tau \in \mathcal{A}} \mathbb{P}\left(\mathcal{T} = \tau\right) = \mathbb{P}\left(\mathcal{T} \in \mathcal{A}\right) \leqslant 1.$$

By the first two steps, we conclude that $c = 1$ and this completes the proof. □

Remark 2.2. When $\sum_{i \geqslant 0} i\mu(i) > 1$, let us mention that it is possible to define a probability measure $\mathbb{P}_\mu$ on the set of all plane (not necessarily finite) trees in such a way that the formula (1) holds for finite trees. However, since we are only interested in finite trees, we will not enter such considerations.

In the sequel, by Bienaymé–Galton–Watson tree with offspring distribution μ (or simply BGW_μ tree), we mean a random tree (that is a random variable defined on some probability space taking values in $\mathbb{A}$) whose distribution is $\mathbb{P}_\mu$. We will alternatively speak of a BGW tree when the offspring distribution is implicit.

2.3 A particular case of Bienaymé–Galton–Watson trees

Goal. The goal of this lecture is to study the geometry of large subcritical BGW trees whose offspring distribution is regularly varying. Specifically, we shall consider BGW_μ trees conditioned on having n vertices, as $n \to \infty$, under the following assumptions: there exists $\beta > 1$ and a slowly varying function L (that is a function $L : \mathbb{R}_+ \to \mathbb{R}$ such that for every fixed $x > 0$, $L(ux)/L(u) \to 1$ as $u \to \infty$) such that

$$\sum_{i=0}^{\infty} i\mu(i) < 1 \qquad \text{and} \qquad \mu(n) = \frac{L(n)}{n^{1+\beta}} \quad \text{for } n \geqslant 1.$$

See Section 6.

3 Coding Bienaymé–Galton–Watson trees by random walks

The most important tool in the study of BGW trees is their coding by random walks, which are usually well understood. The idea of coding BGW trees by functions goes back to Harris [37], and was popularized by Le Gall & Le Jan [53] et Bennies & Kersting [12]. We start by explaining the coding of deterministic trees. We refer to [51] for further applications.

3.1 Coding trees

To code a tree, we first define an order on its vertices. To this end, we use the lexicographic order $\prec$ on the set $\mathcal{U}$ of labels, for which $v \prec w$ if there exists $z \in \mathcal{U}$ with $v = z(a_1, \dots, a_n)$, $w = z(b_1, \dots, b_m)$ and $a_1 < b_1$.

If $\tau \in \mathbb{A}$, let $u_0, u_1, \dots, u_{|\tau|-1}$ be the vertices of τ ordered in lexicographic order, an recall that k_u is the number of children of a vertex u.

Definition 3.1. The *Łukasiewicz path* $\mathcal{W}(\tau) = (\mathcal{W}_n(\tau), 0 \leqslant n \leqslant |\tau|)$ of τ is defined by $\mathcal{W}_0(\tau) = 0$ and, for $0 \leqslant n \leqslant |\tau| - 1$:

$$\mathcal{W}_{n+1}(\tau) = \mathcal{W}_n(\tau) + k_{u_n}(\tau) - 1.$$

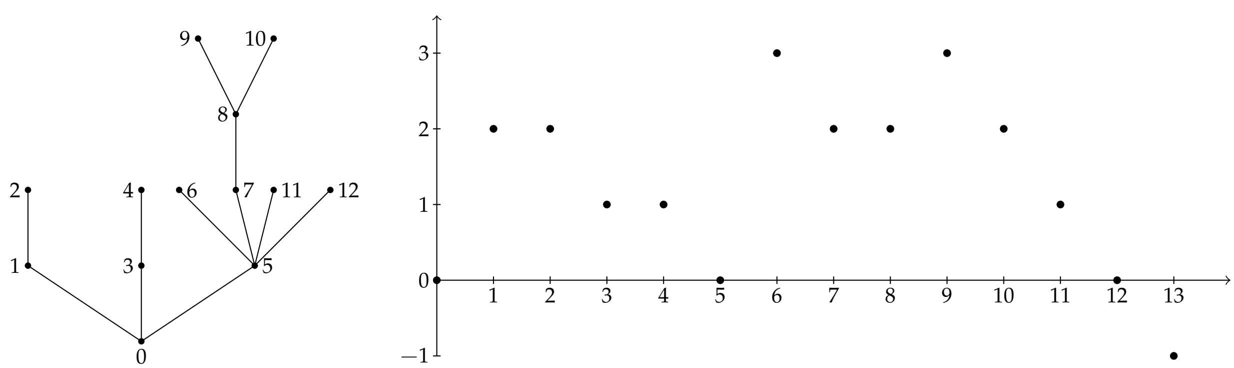


Figure 6: A tree (with its vertices numbered according to the lexicographic order) and its associated Łukasiewicz path.

See Fig. 6 for an example. Before proving that the Łukasiewicz path codes bijectively trees, we need to introduce some notation. For $n \geqslant 1$, set

$$\overline{\mathcal{S}}_n = \{(x_1, \ldots, x_n) \in \{-1, 0, 1, \ldots\} : \quad x_1 + \cdots + x_n = -1 \\ \text{and } x_1 + \cdots + x_j > -1 \text{ for every } 1 \leqslant j \leqslant n-1\}.$$

Proposition 3.2. *For every $n \geqslant 1$, the mapping Φ_n defined by*

$$\begin{array}{cccc} \Phi_n : & \mathbb{A}_n & \longrightarrow & \overline{\mathcal{S}}_n \\ & \tau & \longmapsto & \left(k_{u_{i-1}} - 1 : 1 \leqslant i \leqslant n\right) \end{array}$$

is a bijection.

For $\tau \in \mathbb{A}$, set $\Phi(\tau) = \Phi_{|\tau|}(\tau)$. Proposition 3.2 shows that the Łukasiewicz indeed bijectively codes trees (because the increments of the Łukasiewicz path of τ are the elements of $\Phi(\tau)$) and that $\mathcal{W}_{|\tau|}(\tau) = -1$.

Proof. For $k, n \geqslant 1$, set

$$\overline{\mathcal{S}}_n^{(k)} = \{(x_1, \ldots, x_n) \in \{-1, 0, 1, \ldots\} : \quad x_1 + \cdots + x_n = -k \\ \text{and } x_1 + \cdots + x_j > -k \text{ for every } 1 \leqslant j \leqslant n-1\}$$

so that $\overline{\mathcal{S}}_n^{(1)} = \overline{\mathcal{S}}_n$. If $\mathbf{x} = (x_1, \ldots, x_n) \in \overline{\mathcal{S}}_n^{(k)}$ and $\mathbf{y} = (y_1, \ldots, y_m) \in \overline{\mathcal{S}}_m^{(k')}$, we write $\mathbf{xy} = (x_1, \ldots, x_n, y_1, \ldots, y_m)$ for the concatenation of $\mathbf{x}$ and $\mathbf{y}$. In particular, $\mathbf{xy} \in \overline{\mathcal{S}}_{n+m}^{(k+k')}$. If $\mathbf{x} \in \overline{\mathcal{S}}^{(k)}$, we may write $\mathbf{x} = \mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_k$ with $\mathbf{x}_i \in \overline{\mathcal{S}}^{(1)}$ for every $1 \leqslant i \leqslant k$ in a unique way.

We now turn to the proof of Proposition 3.2. Fix $\tau \in \mathbb{A}_n$. We first check that $\Phi_n(\tau) \in \overline{\mathcal{S}}_n$. For every $1 \leqslant j \leqslant n$, we have

$$\sum_{i=1}^{j} \left(k_{u_{i-1}} - 1\right) = \sum_{i=1}^{j} k_{u_{i-1}} - j. \tag{2}$$

Note that the sum $\sum_{i=1}^{j} k_{u_{i-1}}$ counts the number of children of $u_0, u_1, \ldots, u_{j-1}$. If $j < n$, the vertices $u_1, \ldots, u_j$ are children of $u_0, u_1, \ldots, u_{j-1}$, so that the quantity (2) is positive. If $j = n$, the sum $\sum_{i=1}^{n} k_{u_{i-1}}$ counts vertices who have a parent, that is everyone except the root, so that this sum is $n-1$. Therefore, $\Phi_n(\tau) \in \overline{\mathcal{S}}_n$.

We next show that Φ_n is bijective by strong induction on n. For $n = 1$, there is nothing to do. Fix $n \geqslant 2$ and assume that Φ_j is a bijection fore very $j \in \{1, 2, \ldots, n-1\}$. Take $\mathbf{x} = (a, x_1, \ldots, x_{n-1}) \in \overline{\mathcal{S}}_n$. We have $\Phi_n(\tau) = \mathbf{x}$ if and only if $k_\emptyset(\tau) = a + 1$, and $(x_1, \ldots, x_{n-1})$ must be the concatenation of the images by Φ of the subtrees $\tau_1, \ldots, \tau_{a+1}$ attached on the children of $\emptyset$. But $(x_1, x_1, \ldots, x_{n-1}) \in \overline{\mathcal{S}}_{n-1}^{(a+1)}$, so $(x_1, \ldots, x_{n-1}) = \mathbf{x}_1 \cdots \mathbf{x}_{a+1}$

can be written as a concatenation of elements of $\overline{\mathcal{S}}^{(1)}$ in a unique way. Hence

$$\begin{aligned}\Phi_n(\tau) = \mathbf{x} \quad &\Longleftrightarrow \quad \Phi_{|\tau_i|}(\tau_i) = \mathbf{x}_i \text{ for every } i \in \{1, 2, \dots, a+1\} \\ &\Longleftrightarrow \quad \tau = \{\emptyset\} \cup \bigcup_{i=1}^{a+1} i\Phi^{-1}_{|\tau_i|}(\mathbf{x}_i),\end{aligned}$$

where we have used the induction hypothesis (since $|\tau_i| < |\tau|$). This completes the proof. □

Remark 3.3. For $0 \leqslant k \leqslant n-1$, the height of vertex u_k is given by $\mathrm{Card}(\{0 \leqslant i < k : \mathcal{W}_i(\tau) = \min_{[i,k]} \mathcal{W}\})$. Indeed, the elements of this set correspond to the indices of the ancestors of u_k.

3.2 Coding BGW trees by random walks

We will now identify the law of the Łukasiewicz path of a BGW tree. Consider the random walk $(W_n)_{n \geqslant 0}$ on $\mathbb{Z}$ such that $W_0 = 0$ with jump distribution given by $\mathbb{P}(W_1 = k) = \mu(k+1)$ for every $k \geqslant -1$. In other words, for $n \geqslant 1$, we may write

$$W_n = X_1 + \cdots + X_n,$$

where the random variables $(X_i)_{i \geqslant 1}$ are independent and identically distributed with $\mathbb{P}(X_1 = k) = \mu(k+1)$ for every $k \geqslant -1$. This random walk will play a crucial role in the sequel. Finally, for $j \geqslant 1$, set

$$\zeta = \inf\{n \geqslant 1 : W_n = -1\},$$

which is the first passage time of the random walk at -1 (which could a priori be infinite!).

Proposition 3.4. *Let $\mathcal{T}$ be a random BGW_μ tree. Then the random vectors (of random length)*

$$\big(\mathcal{W}_0(\mathcal{T}), \mathcal{W}_1(\mathcal{T}), \dots, \mathcal{W}_{|\mathcal{T}|}(\mathcal{T})\big) \qquad \text{and} \qquad (W_0, W_1, \dots, W_\zeta)$$

have the same distribution. In particular, $|\mathcal{T}|$ and ζ have the same distribution.

Proof. Fix $n \geqslant 1$ and integers $x_1, \dots, x_n \geqslant -1$. Set

$$\begin{aligned}A &= \mathbb{P}\left(\mathcal{W}_1(\mathcal{T}) = x_1, \mathcal{W}_2(\mathcal{T}) - \mathcal{W}_1(\mathcal{T}) = x_2, \dots, \mathcal{W}_n(\mathcal{T}) - \mathcal{W}_{n-1}(\mathcal{T}) = x_n\right), \\ B &= \mathbb{P}\left(W_1 = x_1, W_2 - W_1 = x_2, \dots, W_n - W_{n-1} = x_n\right).\end{aligned}$$

We shall show that $A = B$.

First of all, if $(x_1, \dots, x_n) \notin \overline{\mathcal{S}}_n$, then $A = B = 0$. Now, if $(x_1, \dots, x_n) \in \overline{\mathcal{S}}_n$, by Proposition 3.2 there exists a tree τ whose Łukasiewicz path is $(0, x_1, x_1 + x_2, \dots)$. Then, by (1),

$$A = \mathbb{P}(\mathcal{T} = \tau) = \prod_{u \in \tau} \mu(k_u) = \prod_{i=1}^{n} \mu(x_i + 1),$$

et

$$\begin{aligned} B &= \mathbb{P}\left(W_1 = x_1, W_2 - W_1 = x_2, \ldots, W_n - W_{n-1} = x_n, \zeta = n\right) \\ &= \mathbb{P}\left(W_1 = x_1, W_2 - W_1 = x_2, \ldots, W_n - W_{n-1} = x_n\right) \\ &= \prod_{i=1}^{n} \mu(x_i + 1). \end{aligned}$$

For the second equality, we have used the equality of events $\{W_1 = x_1, W_2 - W_1 = x_2, \ldots, W_n - W_{n-1} = x_n, \zeta = n\} = \{W_1 = x_1, W_2 - W_1 = x_2, \ldots, W_n - W_{n-1} = x_n\}$, which comes from the fact that $(x_1, \ldots, x_n) \in \overline{\mathcal{S}}_n$. Hence $A = B$, and this completes the proof. $\square$

Remark 3.5. If μ is an offspring distribution with mean m, we have $\mathbb{E}[W_1] = m - 1$. Indeed,

$$\mathbb{E}[W_1] = \sum_{i \geqslant -1} i\mu(i+1) = \sum_{i \geqslant 0} (i-1)\mu(i) = m - 1.$$

In particular, $(W_n)_{n \geqslant 0}$ is a centered random walk if and only if $m = 1$ (that is if the offspring distribution is critical).

3.3 The cyclic lemma

For $n \geqslant 1$ set

$$\mathcal{S}_n \coloneqq \{(x_1, \ldots, x_n) \in \{-1, 0, 1, \ldots\} : x_1 + \cdots + x_n = -1\},$$

and recall the notation

$$\overline{\mathcal{S}}_n = \{(x_1, \ldots, x_n) \in \mathcal{S}_n : x_1 + \cdots + x_j > -1 \text{ for every } 1 \leqslant j \leqslant n-1\}.$$

In the following, we identify an element of $\mathbb{Z}/n\mathbb{Z}$ with its unique representative in $\{0, 1, \ldots, n-1\}$. For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{S}_n$ and $i \in \mathbb{Z}/n\mathbb{Z}$, we set

$$\mathbf{x}^{(i)} = (x_{i+1}, \ldots, x_{i+n}),$$

where the addition of indices is considered modulo n. We say that $\mathbf{x}^{(i)}$ is obtained from $\mathbf{x}$ by a cyclic permutation. Note that $\mathcal{S}_n$ is stable by cyclic permutations.

Definition 3.6. For $\mathbf{x} \in \mathcal{S}_n$, set

$$I_{\mathbf{x}} = \left\{ i \in \mathbb{Z}/n\mathbb{Z} : \mathbf{x}^{(i)} \in \overline{\mathcal{S}}_n \right\}.$$

See Fig. 7 for an example.

Note that if $\mathbf{x} \in \mathcal{S}_n$ and $i \in \mathbb{Z}/n\mathbb{Z}$, then $\mathrm{Card}(I_{\mathbf{x}}) = \mathrm{Card}(I_{\mathbf{x}^{(i)}})$.

Theorem 3.7. *(Cyclic Lemma) For every* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{S}_n$*, we have* $\mathrm{Card}(I_{\mathbf{x}}) = 1$*. In addition, the unique element of* $I_{\mathbf{x}}$ *is the smallest element of* $\mathrm{argmin}_j(x_1 + \cdots + x_j)$.

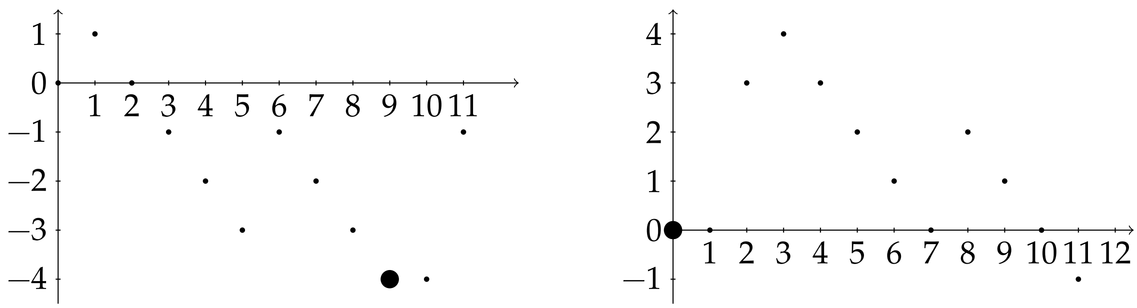


Figure 7: For $\mathbf{x} = (x_1, x_2, \dots)$, we represent $x_1 + \dots + x_i$ as a function of i. On the left, we take $\mathbf{x} = (1, -1, -1, -1, -1, 2, -1, -1, -1, 0, 3) \in \mathcal{S}_{11}$, where $I_{\mathbf{x}} = \{9\}$. On the right, we take $\mathbf{x}^{(9)}$, which is indeed an element of $\overline{\mathcal{S}}_{11}$.

Therefore, if $\mathbf{x} \in \cup_{n \geqslant 1} \mathcal{S}_n$, the set $I_{\mathbf{x}}$ depends on $\mathbf{x}$, but its cardinal does not depend on $\mathbf{x}$!

Proof. We start with an intermediate result: we check that $\mathrm{Card}(I_{\mathbf{x}})$ does not change if one concatenates

$$(a, \underbrace{-1, \dots, -1}_{a \text{ times}})$$

to the left of $\mathbf{x}$, for an integer $a \geqslant 1$. To this end, fix $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{S}_n$ and set

$$\widetilde{\mathbf{x}} = (a, \underbrace{-1, \dots, -1}_{a \text{ times}}, x_1, \dots, x_n).$$

First, it is clear that $0 \in I_{\widetilde{\mathbf{x}}}$ if and only if $0 \in I_{\mathbf{x}}$. Then, if $0 < j \leqslant n - 1$, we have

$$\widetilde{\mathbf{x}}^{(j+a+1)} = (x_{j+1}, \dots, x_n, a, -1, \dots, -1, x_1, \dots, x_j).$$

It readily follows that $j \in I_{\mathbf{x}}$ if and only if $j + a + 1 \in I_{\widetilde{\mathbf{x}}}$. Next, we check that if $0 < i \leqslant a + 1$, then $i \notin I_{\widetilde{\mathbf{x}}}$. Indeed, if $0 < i \leqslant a + 1$, then

$$\widetilde{\mathbf{x}}^{(i)} = (\underbrace{-1, \dots, -1}_{a-i+1 \text{ times}}, x_1, x_2, \dots, x_n, a, -1, \dots, -1).$$

The sum of the elements of $\widetilde{\mathbf{x}}^{(i)}$ up to element x_n is

$$x_1 + \dots + x_n - (a - i + 1) = -1 - (a - i + 1) \leqslant -1.$$

Hence $\widetilde{\mathbf{x}}^{(i)} \notin I_{\widetilde{\mathbf{x}}}$. This shows our intermediate result.

Let us now establish the Cyclic Lemma by strong induction on n. For $n = 1$, there is nothing to do, as the only element of $\mathcal{S}_n$ is $\mathbf{x} = (-1)$. Then consider an integer $n > 1$ such that the Cyclic Lemma holds for elements of $\mathcal{S}_j$ with $j = 1, \dots, n-1$. Take $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{S}_n$. Since $\mathrm{Card}(I_{\mathbf{x}})$ does not change under cyclic permutations of $\mathbf{x}$ and since there exists $i \in \{1, 2, \dots, n\}$ such that $x_i \geqslant 0$ (because $n > 1$), without loss of generality we may assume that $x_1 \geqslant 0$. Denote by $1 = i_1 < i_2 < \cdots < i_m$ the indices i such that $x_i \geqslant 0$ and set $i_{m+1} = n+1$ by convention. Then

$$-1 = \sum_{i=1}^{n} x_i = \sum_{j=1}^{m} \Big(x_{i_j} - (i_{j+1} - i_j - 1) \Big)$$

since $i_{j+1} - i_j - 1$ count the number of consecutive -1 that immediately follows x_{i_j}. Since this sum is negative, there exists $j \in \{1, 2, \ldots, m\}$ such that $x_{i_j} \leqslant i_{j+1} - i_j - 1$. Therefore x_{i_j} is immediately followed by at least x_{i_j} consecutive times -1. Then let $\widetilde{\mathbf{x}}$ be the vector obtained from $\mathbf{x}$ by suppressing x_{i_j} immediately followed by x_{i_j} times -1, so that $\mathrm{Card}(I_{\widetilde{\mathbf{x}}}) = \mathrm{Card}(I_{\mathbf{x}})$ by the intermediate result. Hence $\mathrm{Card}(I_{\mathbf{x}}) = 1$ by induction hypothesis.

The fact that the unique element of $I_{\mathbf{x}}$ is $\mathrm{argmin}_j(x_1 + \cdots + x_j)$ follows from the fact that this property is invariant under insertion of $(a, -1, \ldots, -1)$ for an integer $a \geqslant 1$ (where -1 is written a times). $\square$

Remark 3.8. The statement of Lemma 3.7 is actually valid in the more general setting where steps can take any integer value and not only in $\{-1, 0, 1, \ldots\}$.

In another direction, for $1 \leqslant k \leqslant n$, set

$$\mathcal{S}_n^{(k)} := \{(x_1, \ldots, x_n) \in \{-1, 0, 1, \ldots\} : x_1 + \cdots + x_n = -k\},$$

and

$$\overline{\mathcal{S}}_n^{(k)} = \{(x_1, \ldots, x_n) \in S_n^{(k)} : x_1 + \cdots + x_j > -k \text{ for every } 1 \leqslant j \leqslant n-1\}.$$

Then, for $\mathbf{x} \in \mathcal{S}_n^{(k)}$, we similarly define $I_{\mathbf{x}} = \{i \in \mathbb{Z}/n\mathbb{Z} : \mathbf{x}^{(i)} \in \overline{\mathcal{S}}_n^{(k)}\}$, then a simple adaptation of the proof of Theorem 3.7 shows that the following: for every $\mathbf{x} = (x_1, \ldots, x_n) \in S_n^{(k)} \in \mathcal{S}_n^{(k)}$, we have $\mathrm{Card}(I_{\mathbf{x}}) = k$. Also, if $m = \min\{x_1 + \cdots + x_i : 1 \leqslant i \leqslant n\}$ and $\zeta_i(\mathbf{x}) = \min\{j \geqslant 1 : x_1 + \cdots + x_j = m + i - 1\}$ for $1 \leqslant i \leqslant k$, then $I_{\mathbf{x}} = \{\zeta_1(\mathbf{x}), \ldots, \zeta_k(\mathbf{x})\}$.

3.4 Applications to random walks

In this section, we fix a random walk $(W_n = X_1 + \cdots + X_n)_{n \geqslant 0}$ on $\mathbb{Z}$ such that $W_0 = 0$, $\mathbb{P}(W_1 \geqslant -1) = 1$ and $\mathbb{P}(W_1 = 0) < 1$. We set

$$\zeta = \inf\{i \geqslant 0 : W_i = -1\}.$$

Definition 3.9. A function $F : \mathbb{Z}^n \to \mathbb{R}$ is said to be invariant under cyclic permutations if

$$\forall\, \mathbf{x} \in \mathbb{Z}^n, \quad \forall\, i \in \mathbb{Z}/n\mathbb{Z}, \quad F(\mathbf{x}) = F(\mathbf{x}^{(i)}).$$

Let us give several example of functions invariant by cyclic permutations. If $\mathbf{x} = (x_1, \ldots, x_n)$, one may take $F(\mathbf{x}) = \max(x_1, \ldots, x_n)$, $F(\mathbf{x}) = \min(x_1, \ldots, x_n)$, $F(\mathbf{x}) = x_1 x_2 \cdots x_n$, $F(\mathbf{x}) = x_1 + \cdots + x_n$, or more generally $F(\mathbf{x}) = x_1^\lambda + \cdots + x_n^\lambda$ avec $\lambda > 0$. If $A \subset \mathbb{Z}$,

$$F(\mathbf{x}) = \sum_{i=1}^{n} \mathbb{1}_{x_i \in A},$$

which counts the number of elements in A, is also invariant under cyclic permutations. If F is invariant under cyclic permutations and $g : \mathbb{R} \to \mathbb{R}$ is a function, then $g \circ F$ is also invariant under cyclic permutations. Finally, $F(x_1, x_2, x_3) = (x_2 - x_1)^3 + (x_3 - x_2)^3 + (x_1 - x_3)^3$ is invariant under cyclic permutations but not invariant under all permutations.

Proposition 3.10. *Let $F : \mathbb{Z}^n \to \mathbb{R}$ be a function invariant under cyclic permutations. Then for every integers $n \geqslant 1$ the following assertions hold.*

(i) $\mathbb{E}\left[F(X_1, \ldots, X_n)\mathbb{1}_{\zeta=n}\right] = \frac{1}{n}\mathbb{E}\left[F(X_1, \ldots, X_n)\mathbb{1}_{W_n=-1}\right]$,

(ii) $\mathbb{P}(\zeta = n) = \frac{1}{n}\mathbb{P}(W_n = -1)$.

The assertion (ii) is known as Kemperman's formula.

Proof. The second assertion follows from the first one simply by taking $F \equiv 1$. For (i), to simplify notation, set $\mathbf{X}_n = (X_1, \ldots, X_n)$. Note that the following equalities of events hold

$$\{W_n = -1\} = \{\mathbf{X}_n \in \mathcal{S}_n\} \qquad \text{and} \qquad \{\zeta = n\} = \left\{\mathbf{X}_n \in \overline{\mathcal{S}}_n\right\}.$$

In particular,

$$\mathbb{E}\left[F(X_1, \ldots, X_n)\mathbb{1}_{\zeta=n}\right] = \mathbb{E}\left[F(\mathbf{X}_n)\mathbb{1}_{\mathbf{X}_n \in \overline{\mathcal{S}}_n}\right].$$

Then write

$$\begin{aligned}
\mathbb{E}\left[F(\mathbf{X}_n)\mathbb{1}_{\mathbf{X}_n \in \overline{\mathcal{S}}_n}\right] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[F(\mathbf{X}_n^{(i)})\mathbb{1}_{\mathbf{X}_n^{(i)} \in \overline{\mathcal{S}}_n}\right] \quad (\text{since } \mathbf{X}_n^{(i)} \text{ and } \mathbf{X}_n \text{ have the same law}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[F(\mathbf{X}_n)\mathbb{1}_{\mathbf{X}_n^{(i)} \in \overline{\mathcal{S}}_n}\right] \quad (\text{invariance of } F \text{ by cyclic permutations}) \\
&= \frac{1}{n}\mathbb{E}\left[F(\mathbf{X}_n)\left(\sum_{i=1}^{n} \mathbb{1}_{\mathbf{X}_n^{(i)} \in \overline{\mathcal{S}}_n}\right)\right] \\
&= \frac{1}{n}\mathbb{E}\left[F(X_1, \ldots, X_n)\mathbb{1}_{\mathbf{X}_n \in \mathcal{S}_n}\right],
\end{aligned}$$

where the last equality is a consequence of the equality of the random variables

$$\sum_{i=1}^{n} \mathbb{1}_{\mathbf{X}_n^{(i)} \in \overline{\mathcal{S}}_n} = \mathbb{1}_{\mathbf{X}_n \in \mathcal{S}_n}$$

by the Cyclic Lemma. We conclude that

$$\mathbb{E}\left[F(\mathbf{X}_n)\mathbb{1}_{\mathbf{X}_n \in \overline{\mathcal{S}}_n}\right] = \frac{1}{n}\mathbb{E}\left[F(X_1, \ldots, X_n)\mathbb{1}_{W_n=-1}\right],$$

which is the desired result. □

Proposition 3.10 is useful to compute quantities conditionally on $\{\zeta = n\}$ which are invariant under cyclic permutations by instead working conditionally on $\{W_n = -1\}$, which is a simpler conditioning (since it only involves W_n).

Actually, Proposition 3.10 can be in extended in the sense that instead of working conditionally on $\{\zeta = n\}$, we may often work conditionally on $\{W_n = -1\}$. To this end, we need to define the Vervaat transform.

Definition 3.11. Let $n \in \mathbb{N}$, $(x_1, \ldots, x_n) \in \mathbb{Z}^n$ and let $\mathbf{w} = (w_i : 0 \leqslant i \leqslant n)$ be the associated walk defined by

$$w_0 = 0 \quad \text{and} \quad w_i = \sum_{j=1}^{i} x_j, \quad 1 \leqslant i \leqslant n.$$

We also introduce the first time at which $(w_i : 0 \leqslant i \leqslant n)$ reaches its overall minimum,

$$k_n := \min\{0 \leqslant i \leqslant n : w_i = \min\{w_j : 0 \leqslant j \leqslant n\}\},$$

so that $I_{\{x_1,\ldots,x_n\}} = \{k_n\}$. The Vervaat transform $\mathcal{V}(\mathbf{w}) := (\mathcal{V}(\mathbf{w})_i : 0 \leqslant i \leqslant n)$ of $\mathbf{w}$ is the walk obtained by reading the increments $(x_1, \ldots, x_n)$ from left to right in cyclic order, started from k_n. Namely,

$$\mathcal{V}(\mathbf{w})_0 = 0 \quad \text{and} \quad \mathcal{V}(\mathbf{w})_{i+1} - \mathcal{V}(\mathbf{w})_i = x_{k_n + i \mod [n]}, \quad 0 \leqslant i < n,$$

see Figure 7 for an illustration.

We keep the notation $\mathbf{X}_n = (X_1, \ldots, X_n)$.

Proposition 3.12.

(i) The law of $\mathbf{X}_n$ conditionally given $\{\zeta = n\}$ is equal to the law of $\mathbf{X}_n^{(I_n)}$ conditionally given $\{W_n = -1\}$, where I_n is the unique element of $I_{\mathbf{X}_n}$.

(ii) Conditionally given $\{W_n = -1\}$, I_n follows the uniform distribution on $\{0, 1, \ldots, n-1\}$, and I_n and $\mathbf{X}_n^{(I_n)}$ are independent.

In other words, to construct a random variable following the conditional law of $\mathbf{X}$ given $\{\zeta = n\}$, one can start with a random variable following the conditional law of $\mathbf{X}$ given $\{W_n = -1\}$ and apply the Vervaat transform.

Proof of Proposition 3.12. Fix $\mathbf{x} \in \overline{\mathcal{S}}_n$ (it is important to take $\mathbf{x} \in \overline{\mathcal{S}}_n$ and not only $\mathbf{x} \in \mathcal{S}_n$).

Since the events $\{\mathbf{X}_n = \mathbf{x}, \zeta = n\}$ et $\{\mathbf{X}_n = \mathbf{x}, W_n = -1\}$ are equal, we have

$$\begin{aligned}
\mathbb{P}\left(\mathbf{X}_n = \mathbf{x}, \zeta = -1\right) = \mathbb{P}\left(\mathbf{X}_n = \mathbf{x}, W_n = -1\right) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\left(\mathbf{X}_n^{(i)} = \mathbf{x}, W_n = -1\right) \\
&= \frac{1}{n}\mathbb{E}\left[\mathbb{1}_{W_n=-1}\left(\sum_{i=1}^{n} \mathbb{1}_{\mathbf{X}_n^{(i)}=\mathbf{x}}\right)\right] \\
&= \frac{1}{n}\mathbb{E}\left[\mathbb{1}_{W_n=-1}\mathbb{1}_{\mathbf{X}_n^{(I_n)}=\mathbf{x}}\right] \\
&= \frac{1}{n}\mathbb{P}\left(\mathbf{X}_n^{(I_n)} = \mathbf{x}, W_n = -1\right). \qquad (3)
\end{aligned}$$

We divide by the equality $\mathbb{P}\left(\zeta = n\right) = \frac{1}{n}\mathbb{P}\left(W_n = -1\right)$ to get

$$\mathbb{P}\left(\mathbf{X}_n = \mathbf{x} \middle| \zeta = n\right) = \mathbb{P}\left(\mathbf{X}_n^{(I_n)} = \mathbf{x} \middle| W_n = -1\right),$$

which shows (i).

For (ii), fix $k \in \{0, 1, \ldots, n-1\}$ and $\mathbf{x} \in \overline{\mathbb{S}}_n$. Since the events $\{I_n = k, \mathbf{X}_n^{(I_n)} = \mathbf{x}, W_n = -1\}$ and $\{\mathbf{X}_n^{(k)} = \mathbf{x}, W_n = -1\}$ are equal (because $\mathrm{Card}(I_{\mathbf{X}_n}) = 1$ when $W_n = -1$ by the cyclic lemma), we have

$$\begin{aligned}
\mathbb{P}\left(I_n = k, \mathbf{X}_n^{(I_n)} = \mathbf{x}, W_n = -1\right) &= \mathbb{P}\left(\mathbf{X}_n^{(k)} = \mathbf{x}, W_n = -1\right) \\
&= \mathbb{P}\left(\mathbf{X}_n = \mathbf{x}, W_n = -1\right) \qquad (\mathbf{X}_n^{(k)} \text{ have } \mathbf{X}_n \text{ the same law}) \\
&= \frac{1}{n}\mathbb{P}\left(\mathbf{X}_n^{(I_n)} = \mathbf{x}, W_n = -1\right) \qquad (\text{by } (3))
\end{aligned}$$

we divise this equality by $\mathbb{P}\left(W_n = -1\right)$, and get that

$$\mathbb{P}\left(I_n = k, \mathbf{X}_n^{(I_n)} = \mathbf{x} \middle| W_n = -1\right) = \frac{1}{n} \cdot \mathbb{P}\left(\mathbf{X}^{(I_n)} = \mathbf{x} \middle| W_n = -1\right).$$

By summing over all the possible $\mathbf{x} \in \overline{\mathbb{S}}_n$ we get that $\mathbb{P}\left(I_n = k \middle| W_n = -1\right) = \frac{1}{n}$ (which is indeed the uniform law on $\{0, 1, \ldots, n-1\}$) and then

$$\mathbb{P}\left(I_n = k, \mathbf{X}_n^{(I_n)} = \mathbf{x} \middle| W_n = -1\right) = \mathbb{P}\left(I_n = k \middle| W_n = -1\right) \cdot \mathbb{P}\left(\mathbf{X}^{(I_n)} = \mathbf{x} \middle| W_n = -1\right),$$

which completes the proof. $\square$

4 Exercise session

4.1 Exercises

Exercise 1 (Exponential tilting). We say that two offspring distributions (that is probability measures on $\mathbb{Z}_+$) are *equivalent* if there exist $a, b > 0$ such that $\widetilde{\mu}(i) = ab^i\mu(i)$ for every $i \geqslant 0$.

(1) Let μ and $\widetilde{\mu}$ be two equivalent offspring distributions. Let $\mathcal{T}_n$ be a BGW_μ random tree conditioned on having n vertices (here and after, we always assume that conditionings are non-degenerate) and let $\widetilde{\mathcal{T}}_n$ be a $\mathrm{BGW}_{\widetilde{\mu}}$ random tree conditioned on having n vertices. Show that $\mathcal{T}_n$ and $\widetilde{\mathcal{T}}_n$ have the same distribution.

(2) Let μ be an offspring distribution with infinite mean and such that $\mu(0) > 0$. Can one find a critical offspring distribution equivalent to μ?

(3) Can one always find a critical offspring distribution equivalent to any offspring distribution?

(4) Find a critical offspring distribution μ such that a BGW_μ random tree conditioned on having n vertices follows the uniform distribution on the set of all plane trees with n vertices.

(5) Find a critical offspring distribution ν such that a BGW_ν random tree conditioned on having n leaves follows the uniform distribution on the set of all plane trees with n leaves having no vertices with only one child.

NB: a leaf is a vertex with no children.

Exercise 2. Let μ be a subcritical offspring distribution (with mean $m < 1$) and let $\mathcal{T}$ be a BGW_μ random tree. Denote by $\mathsf{Height}(\mathcal{T})$ the last generation of $\mathcal{T}$. Show that $\mathbb{P}(\mathsf{Height}(\mathcal{T}) \geqslant n) \leqslant m^n$.

In the following exercises, $(X_i)_{i\geqslant 1}$ is a sequence of i.i.d. integer valued random variables such that $\mathbb{P}(X_1 \geqslant -1) = 1$ and $\mathbb{P}(X_1 > 0) > 0$ (to avoid trivial cases). We set $W_n = X_1 + \cdots + X_n$ and $\zeta = \inf\{n \geqslant 1 : W_n = -1\}$. Finally, we denote by X a random variable having the same law as X_1.

Exercise 3. Assume that $\mathbb{E}[X_1] = -c \leqslant 0$ The goal of this exercise is to show that $\mathbb{P}(\forall n \geqslant 1, W_n < 0) = c$.

(1) Show that $\mathbb{E}[\zeta] = \frac{1}{c}$.

Set $T_1 = \inf\{n \geqslant 1 : W_n \geqslant 0\} \in \mathbb{N} \cup \{\infty\}$ and, by induction, $T_{k+1} = \inf\{n > T_k : W_n \geqslant W_{T_k}\}$ (the sequence (T_k) is called the sequence of weak ladder times of the random walk).

(2) Show that $\mathbb{P}(\zeta > n) = \mathbb{P}(n \in \{T_1, T_2, T_3, \ldots\})$ for $n \geqslant 1$.

(3) Conclude.

Exercise 4.

(1) Assume that $\mathbb{E}[X_1] = -c \leqslant 0$. Show that

$$\sum_{n=1}^{\infty} \frac{1}{n} \mathbb{P}(W_n = -1) = 1, \qquad \sum_{n=1}^{\infty} \mathbb{P}(W_n = -1) = \frac{1}{c}.$$

(2) Show that for every $0 \leqslant \lambda \leqslant 1$,

$$\sum_{n \geqslant 1} \frac{(\lambda n)^{n-1} e^{-\lambda n}}{n!} = 1, \qquad \sum_{n=1}^{\infty} \frac{(\lambda n)^{n-1} e^{-\lambda n}}{(n-1)!} = \frac{1}{1-\lambda}.$$

Exercise 5. Show that for $0 \leqslant s \leqslant 1$:

$$\sum_{n=1}^{\infty} \mathbb{P}(\zeta = n) s^n = 1 - \exp\left(-\sum_{n=1}^{\infty} \frac{s^n}{n} \mathbb{P}(W_n < 0) \right)$$

and

$$\sum_{n=0}^{\infty} \mathbb{P}(\zeta > n) s^n = \exp\left(\sum_{n=1}^{\infty} \frac{s^n}{n} \mathbb{P}(W_n \geqslant 0) \right).$$

Hint. For $r \geqslant 1$, set $\zeta_r = \inf\{i \geqslant 1 : W_i = -r\} \in \mathbb{N} \cup \{\infty\}$. You may use the following extension of the cyclic lemma: for $n \geqslant 1$,

$$\mathbb{P}(\zeta_r = n) = \frac{r}{n} \mathbb{P}(W_n = -r).$$

Exercise 6 (Open problem: first hitting time for Cauchy random walks)**.** Assume that

$$\mathbb{E}[X] = 0, \qquad \mathbb{P}(X \geqslant n) \underset{n \to \infty}{\sim} \frac{L(n)}{n}$$

for a slowly varying function L (meaning that for every fixed $t > 0$, $L(tx)/L(x) \to 1$ as $x \to \infty$. Let $(a_n : n \geqslant 1)$ be a sequence $n\mathbb{P}(X \geqslant a_n) \to 1$ and set $b_n = n\mathbb{E}\left[X \mathbb{1}_{|X| \leqslant a_n}\right]$. Do we have

$$\mathbb{P}(\zeta \geqslant n) \underset{n \to \infty}{\sim} \frac{nL(|b_n|)}{b_n^2} \quad ?$$

4.2 Solutions

Solution of Exercise 1.

Remark. The technique of "exponential tilting" in branching processes goes back to at least Kennedy [44]. See [39, Section 4] for a detailed exposition in the slightly more general context of size-conditioned simply generated trees.

(1) The idea is to introduce $\mathcal{T}$ and $\widetilde{\mathcal{T}}$, respectively two (nonconditioned) BGW_μ and $\mathrm{BGW}_{\widetilde{\mu}}$ random trees. Fix a tree T with n vertices. Then, if k_u denotes the number of children of a vertex $u \in T$, by definition:

$$\mathbb{P}\left(\widetilde{\mathcal{T}} = T\right) = \prod_{u \in T} \widetilde{\mu}(k_u) = \prod_{u \in T} \left(ab^{k_u}\mu(k_u)\right) = a^n b^{n-1} \prod_{u \in T} \mu(k_u) = a^n b^{n-1} \mathbb{P}\left(\mathcal{T} = T\right), \tag{4}$$

where we have used the simple deterministic fact that $\sum_{u \in T} k_u = n-1$. The key is that the quantity $a^n b^{n-1}$ does not depend on T. Indeed, by summing over all trees T with n vertices, we get that

$$\mathbb{P}\left(|\widetilde{\mathcal{T}}| = n\right) = a^n b^{n-1} \mathbb{P}\left(|\mathcal{T}| = n\right). \tag{5}$$

The desired result follows by dividing (4) by (5).

(2) Set $F(z) = \sum_{i \geqslant 0} \mu(i) z^i$, so that $F'(1) = \infty$. The question is whether we can find an offspring distribution of the form $\widetilde{\mu}(i) = ab^i \mu(i)$ for every $i \geqslant 0$ with mean 1. The generating function of $\widetilde{\mu}$ is

$$\widetilde{F}(z) = \frac{F(bz)}{F(b)},$$

so that the mean of $\widetilde{\mu}$ is $\widetilde{F}'(1) = \frac{bF'(b)}{F(b)}$. This quantity is continuous in b, tends to 0 as $b \to 0$ and is equal to ∞ for $b = 1$. The desired result follows by continuity.

Note that this argument works for any supercritical offspring distribution.

(3) No, for instance if μ is subcritical and $F(z) = \sum_{i \geqslant 0} \mu(i) z^i$ has radius of convergence equal to 1. Indeed, one checks that

$$G : b \mapsto \frac{bF'(b)}{F(b)} = \frac{\sum_{k \geqslant 0} k\mu(k) b^k}{\sum_{k \geqslant 0} \mu(k) b^k}$$

is increasing. Indeed,

$$G'(b) = \frac{\sum_{k \geqslant 1} k^2 \mu(k) b^{k-1}}{\sum_{k \geqslant 0} \mu(k) b^k} - \frac{\left(\sum_{k \geqslant 0} k\mu(k) b^k\right)\left(\sum_{k \geqslant 0} k\mu(k) b^{k-1}\right)}{\left(\sum_{k \geqslant 0} \mu(k) b^k\right)^2}.$$

The key is then to observe that $bG'(b)$ is the variance of the offspring distribution $\widetilde{\mu}$ with generating function G, so that $G'(b) > 0$. Since $G(1) < 1$ and G has radius of convergence equal to 1, one cannot find $0 \leqslant b \leqslant 1$ such that $G(b) = 1$.

However, if μ is supercritical and $\mu(0) > 0$, question (2) shows that it is possible. If μ is supercritical and $\mu(0) = 0$, this is not possible since the only critical offspring distribution μ with $\mu(0) = 0$ is $\mu(1) = 1$.

(4) Inspired by Question (1), it is natural to try to find μ of the form $\mu(i) = ab^i$ for $i \geqslant 0$. Since we want μ to be critical, a small calculation yields $\mu(i) = 2^{-i-1}$ for $i \geqslant 0$. The same calculation as in (1) shows that if $\mathcal{T}_n$ be a BGW_μ random tree, then $\mathbb{P}(\mathcal{T}_n = T)$ for a fixed tree T with n vertices does not depend on n, so that $\mathcal{T}_n$ follows the uniform distribution on the set of all plane trees with n vertices.

(5) The idea it to look for ν of the form $\nu(0) = a$, $\nu(1) = 0$, $\nu(i) = b^{i-1}$ for $i \geqslant 2$. Indeed, if $\mathcal{T}$ is a nonconditioned BGW_ν random tree and T a tree with n leaves and such that no vertices have only one child,

$$\mathbb{P}\left(\widetilde{\mathcal{T}} = T\right) = \prod_{u \in T} \widetilde{\mu}(k_u) = a^n \prod_{u \in T, k_u \geqslant 2} \left(b^{k_u - 1}\right) = a^n b^{n-1},$$

where we have used the simple deterministic fact that $\sum_{u \in T, k_u \geqslant 2} = n - 1$. Since this quantity does not depend on T, the same reasoning as in Question (1) shows that such a BGW_ν tree conditioned on having n leaves follows the uniform distribution on the set of all plane trees with n leaves having no vertices with only one child.

It remains to choose a, b such that ν is a critical offspring distribution. A small calculation yields

$$\nu(0) = 2 - \sqrt{2}, \qquad \nu(1) = 0, \qquad \nu(i) = \left(\frac{2-\sqrt{2}}{2}\right)^{i-1} \qquad \text{for } i \geqslant 2.$$

Remark. This was used in to study random Schröder bracketings of words [61] and to study uniform dissections of polygons [21].

Solution of Exercise 2. The idea is to use the Bienaymé–Galton–Watson process associated with $\mathcal{T}$. Specifically, let $(X_j^{(n)})_{j,n \geqslant 1}$ be a sequence of i.i.d. random variables with law μ (defined on the same probability space). Define recursively $(Z_n : n \geqslant 0)$ as follows:

$$Z_0 = 1, \qquad \text{and for every } n \geqslant 1, \quad Z_{n+1} = \sum_{j=1}^{Z_n} X_j^{(n)}.$$

Then $\max\{i : Z_i \neq 0\}$ has the same distribution as $\mathsf{Height}(\mathcal{T})$. Since $\max\{i : Z_i \neq 0\} \geqslant n$ implies that $Z_n \geqslant 1$, we get

$$\mathbb{P}(\mathsf{Height}(\mathcal{T}) \geqslant n) \leqslant \mathbb{P}(Z_n \geqslant 1).$$

The next idea is to notice that $\mathbb{P}(Z_n \geqslant 1) \leqslant \mathbb{E}[Z_n]$ and that $\mathbb{E}[Z_n]$ can be readily calculated. Indeed, by definition of Z_{n+1} we have $\mathbb{E}[Z_{n+1} | Z_n] = mZ_n$, which gives $\mathbb{E}[Z_n] = m^n$ for $n \geqslant 0$. The desired result follows.

Remark. See [54] for asymptotic equivalents on the tail of $\mathsf{Height}(\mathcal{T})$.

Solution of Exercise 3.

(1) The idea is to use the connection with Bienaymé–Galton–Watson processes. If μ is the offspring distribution defined by $\mu(n) = \mathbb{P}(X = n+1)$ for $n \geqslant -1$, the coding of a BGW_μ by its Lukasiewicz path (Proposition 3.4) entails that ζ has the same distribution as the total size of a BGW_μ tree. But if $(Z_n)_{n \geqslant 0}$ denotes the BGW process as in the solution of Exercise 2, we see that the total size of a BGW_μ tree has the same distribution as $\sum_{n \geqslant 0} Z_n$. We conclude that

$$\mathbb{E}[\zeta] = \sum_{n=0}^{\infty} \mathbb{E}[Z_n] = \sum_{n=0}^{\infty} m^n = \frac{1}{1-m} = \frac{1}{c}$$

since $-c = \mathbb{E}[X_1] = m - 1$.

(2) The idea is to use a time-reversal argument. Specifically, fix $n \geqslant 1$ and for $0 \leqslant i \leqslant n$ define $W_i^{[n]} = W_n - W_{n-i}$ (this amounts to considering the walk obtained by reading the jumps of $(W_0, W_1, \ldots, W_n)$ backwards). Extend $W^{[n]}$ by defining $W_{n+i}^{[n]} = W_n^{[n]}) + X'_1 + X'_2 + \cdots + X'_i$ for $i \geqslant 1$, with $(X'_i)_{i \geqslant 1}$ i.i.d. with law X_1 and independent of $(X_i)_{i \geqslant 1}$. Define $(T_k^{[n]})_{k \geqslant 1}$ as $(T_k)_{k \geqslant 1}$ but by replacing W with $W^{[n]}$. Then

$$\{\zeta > n\} = \{n \in \{T_1^{[n]}, T_2^{[n]}, T_3^{[n]}, \ldots\}\},$$

see Figure 8.

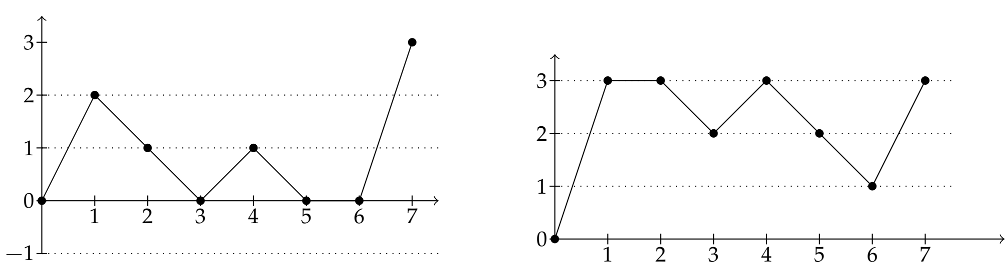


Figure 8: Left: an example of $(W_i : 0 \leqslant i \leqslant 7)$ such that $\zeta > 7$; Right: its associated time-reversed path $(W_i^{[7]} : 0 \leqslant i \leqslant 7)$ (obtained by reading the jumps from right to left).

The desired result follows from the fact that $(W_i)_{i \geqslant 0}$ and $(W_i^{[n]})_{i \geqslant 0}$ have the same distribution, and therefore $(T_i)_{i \geqslant 1}$ and $(T_i^{[n]})_{i \geqslant 1}$ as well have the same distribution, so that

$$\mathbb{P}(n \in \{T_1^{[n]}, T_2^{[n]}, T_3^{[n]}, \ldots) = \mathbb{P}(n \in \{T_1, T_2, T_3, \ldots).$$

(3) By (2), we have $\mathbb{P}(\zeta > n) = \sum_{k=1}^{\infty} \mathbb{P}(T_k = n)$. By summing over $n \geqslant 1$, we get

$$\mathbb{E}[\zeta] - 1 = \sum_{n=1}^{\infty} \mathbb{P}(\zeta > n) = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{P}(T_k = n) = \sum_{k=1}^{\infty} \mathbb{P}(T_k < \infty) = \sum_{k=1}^{\infty} \mathbb{P}(T_1 < \infty)^k,$$

where the last equality follows from the strong Markov property. By (1), we get that

$$\frac{1}{c} - 1 = \frac{\mathbb{P}(T_1 < \infty)}{1 - \mathbb{P}(T_1 < \infty)},$$

so that $\mathbb{P}(T_1 < \infty) = 1 - c$. It remains to observe that $\mathbb{P}(\forall n \geqslant 1, W_n < 0) = \mathbb{P}(T_1 = \infty)$.

Solution of Exercise 4.

(1) The idea is to use the connection with Bienaymé–Galton–Watson processes as in Exercice 4. If μ is the offspring distribution defined by $\mu(n) = \mathbb{P}(X = n+1)$ for $n \geqslant -1$. We have seen that $\mathbb{P}(\zeta_1 = n) = \frac{1}{n}\mathbb{P}(W_n = -1)$. Therefore

$$\sum_{n=1}^{\infty} \frac{1}{n}\mathbb{P}(W_n = -1) = \sum_{n=1}^{\infty} \mathbb{P}(\zeta_1 = n) = \mathbb{P}(\zeta_1 < \infty),$$

which is the probability that a BGW_μ is finite, and is therefore 1 since μ has expectation $1 - c \leqslant 1$.

For the second equality, write

$$\sum_{n=1}^{\infty} \mathbb{P}(W_n = -1) = \sum_{n=1}^{\infty} n\mathbb{P}(\zeta_1 = n) = \mathbb{E}[\zeta_1] = \frac{1}{c}$$

by the first question of Exercise 4.

(2) We apply the first question with $X_1 = \mathsf{Poisson}(\lambda) - 1$. In particular, $W_n + n$ is distributed according to a Poisson random variable of parameter λn. Therefore, by (1):

$$1 = \sum_{n=1}^{\infty} \frac{1}{n}\mathbb{P}(W_n = -1) = \sum_{n=1}^{\infty} \frac{1}{n}\mathbb{P}(W_n + n = n - 1) = \sum_{n=1}^{\infty} \frac{1}{n} \cdot \frac{(\lambda n)^{n-1} e^{-\lambda n}}{(n-1)!},$$

which shows (1).

Similarly,

$$\frac{1}{1-\lambda} = \sum_{n=1}^{\infty} \mathbb{P}(W_n = -1) = \sum_{n=1}^{\infty} \mathbb{P}(W_n + n = n - 1) = \sum_{n=1}^{\infty} \frac{(\lambda n)^{n-1} e^{-\lambda n}}{(n-1)!}.$$

Solution of Exercise 5. We start with summing the identity of the hint over r:

$$\sum_{r=1}^{\infty} \frac{s^n}{r} \mathbb{P}\left(\zeta_r = n\right) = \sum_{r=1}^{\infty} \frac{s^n}{n} \mathbb{P}\left(W_n = -r\right) = \frac{\mathbb{P}(W_n < 0)}{n} s^n.$$

If $(\zeta^{(i)})_{i \geqslant 1}$ is an i.i.d. sequence of independent random variables distributed as ζ_1, we have

$$\sum_{n=1}^{\infty} \sum_{r=1}^{\infty} \frac{1}{r} \mathbb{P}\left(\zeta_r = n\right) s^n = \sum_{r=1}^{\infty} \frac{1}{r} \mathbb{E}\left[s^{\zeta^{(1)}+\zeta^{(2)}+\cdots+\zeta^{(r)}}\right] = \sum_{r=1}^{\infty} \frac{1}{r} \mathbb{E}\left[s^{\zeta_1}\right]^r = \ln\left(\frac{1}{1-\mathbb{E}\left[s^{\zeta_1}\right]}\right).$$

The first identity follows.

For the second one, write

$$\exp\left(-\sum_{n=1}^{\infty} \frac{s^n}{n} \mathbb{P}\left(W_n < 0\right)\right) = (1-s) \exp\left(\sum_{n=1}^{\infty} \frac{s^n}{n} \mathbb{P}\left(W_n \geqslant 0\right)\right).$$

The desired result follows by noting that

$$\frac{1}{1-s}\left(1 - \sum_{n=1}^{\infty} \mathbb{P}\left(\zeta = n\right) s^n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(\zeta = n\right) \frac{1-s^n}{1-s} = \sum_{n=0}^{\infty} \mathbb{P}\left(\zeta > n\right) s^n.$$

Remarks on Exercise 6. This has been established in [48] under the assumption that $\mathbb{P}\left(X = n\right) \sim L(n)/n^2$ as $n \to \infty$, see also [13, Theorem 3.4] for estimates under regularity conditions on L.

5 Estimates for centered, downward skip-free random walks in the domain of attraction of a stable law

In this Section, we establish estimates for a certain class of random walks, in view of applying them to study large subcritical BGW trees whose offspring distribution is regularly varying.

Assumptions. Let X be an integer-valued random variable such that:

- $\mathbb{P}(X \geqslant -1) = 1$ and $\mathbb{P}(X > 0) > 0$;
- $\mathbb{E}\left[X\right] = 0$;
- X is in the domain of attraction of an α-stable distribution, with $1 < \alpha \leqslant 2$.

The last assumption means that either X has finite variance (in which case $\alpha = 2$), or

$$\mu([n, \infty)) = \frac{L(n)}{n^\alpha} \tag{6}$$

for a slowly varying function L (that is for every fixed $x > 0$, $L(ux)/L(u) \to 1$ as $u \to \infty$; for instance ln is slowly varying, and so is any function converging to a positive limit at ∞). We refer to [16] for details slowly varying functions and to [31, IX.8,XVII.5] for background on domains of attraction.

In order to unify the finite and infinite variance cases for $\alpha = 2$, it is useful to rely on the following equivalent characterization: X in the domain of attraction of an α-stable distribution, with $1 < \alpha \leqslant 2$, if there exists a slowly varying function L_0 such that

$$\mathbb{E}\left[X^2 \mathbb{1}_{X \leqslant n}\right] = L_0(n) n^{2-\alpha}. \tag{7}$$

We also choose a scaling sequence (b_n) such that

$$\frac{n L_0(b_n)}{b_n^\alpha} \quad \underset{n\to\infty}{\longrightarrow} \quad \begin{cases} \dfrac{1}{(2-\alpha)\Gamma(-\alpha)} & \text{for } \alpha < 2 \\ 2 & \text{for } \alpha = 2. \end{cases} \tag{8}$$

In particular, b_n is of order $n^{1/\alpha}$ up to a slowly varying function (if $\mu([n, \infty)) \sim c/n^\alpha$, then $b_n \sim c' n^{1/\alpha}$ for a certain constant c'). Its importance comes from the fact that if $(X_i)_{i \geqslant 1}$ are i.i.d. distributed as X and $S_n = X_1 + \cdots + X_n$, then we have the convergence in distribution

$$\frac{S_n}{b_n} \quad \overset{(d)}{\underset{n\to\infty}{\longrightarrow}} \quad Y_\alpha, \tag{9}$$

where Y_α is an α-stable spectrally positive random variable normalized so that $\mathbb{E}[e^{-\lambda Y_\alpha}] = e^{\lambda^\alpha}$ for every $\lambda \geqslant 0$. In particular, for $\alpha = 2$, Y_2 is a multiple of standard Gaussian distribution.

Remark 5.1. When $1 < \alpha < 2$, we may take $L(n) = \frac{2-\alpha}{\alpha} L_0(n)$ and when $\alpha = 2$ we have $L(n) = o(L_0(n))$. If X has finite variance we have $L_0(n) \to \mathsf{Var}(X)$, so that we may take $b_n = \sqrt{\mathsf{Var}(X) n/2}$.

In the sequel, we shall use, sometimes without notice, the following useful result concerning slowly varying function (sometimes called the Potter bounds, see [16, Theorem 1.5.4]. For every $A > 1, \delta > 0$, there exists $M > 0$ such that for evey $x, y \geqslant M$:

$$\frac{L(y)}{L(x)} \leqslant A \max\left(\left(\frac{y}{x}\right)^\delta, \left(\frac{y}{x}\right)^{-\delta}\right).$$

5.1 A maximal inequality

Our goal is to establish the following inequality, where $S_n = X_1 + \cdots + X_n$ with $(X_i)_{i \geqslant 1}$ i.i.d. satisfying the assumptions in the beginning of Section 5.

Proposition 5.2. *There exists a constant* $C > 0$ *such that for every* $n \geqslant 1$, $x > 0$ *and* $c \geqslant 1$ *we have*

$$\mathbb{P}\left(S_n > x b_n, X_1 \leqslant c b_n, \ldots, X_n \leqslant c b_n\right) \leqslant C \exp\left(-\frac{x}{c}\right).$$

In our setting, a rather short proof can be given by adapting the proof of [33, Theorem 2]. See [24, Lemma 2.1] for greater generality.

Proof. The idea is to introduce the "truncated" random walk $\widetilde{S}_n$ defined by

$$\widetilde{S}_n = \sum_{i=1}^{n} X_i \mathbb{1}_{X_i \leqslant c b_n}.$$

Indeed, we have $\mathbb{P}\left(S_n > x b_n, X_1 \leqslant c b_n, \ldots, X_n \leqslant c b_n\right) \leqslant \mathbb{P}(\widetilde{S}_n \geqslant x b_n)$. To bound the latter quantity, we use a Chernoff bound. Specifically, set $\lambda_n = 1/(c b_n)$ and write

$$\mathbb{P}(\widetilde{S}_n \geqslant x b_n) \leqslant e^{-\lambda x b_n} \mathbb{E}\left[e^{\lambda_n \widetilde{S}_n}\right] = e^{-x/c} \mathbb{E}\left[e^{\lambda_n \widetilde{S}_1}\right]^n.$$

The idea is to write $\mathbb{E}\left[e^{\lambda_n \widetilde{S}_1}\right] = 1 + m_n + s_n$ with

$$m_n = \frac{1}{c b_n} \mathbb{E}\left[X \mathbb{1}_{X \leqslant c b_n}\right], \qquad s_n = \frac{1}{(c b_n)^2} \mathbb{E}\left[\frac{e^{X/(c b_n)} - 1 - X/(c b_n)}{(X/(c b_n))^2} X^2 \mathbb{1}_{X \leqslant c b_n}\right].$$

It suffices to check that $m_n = \mathcal{O}(1/n)$ and $s_n = \mathcal{O}(1/n)$ to finish the proof.

Estimation of m_n. Since $\mathbb{E}[X] = 0$, we have $\mathbb{E}\left[X \mathbb{1}_{X \leqslant c b_n}\right] = -\mathbb{E}\left[X \mathbb{1}_{X > c b_n}\right]$. But

$$\frac{u^{\alpha-1}}{L_0(u)} \cdot \mathbb{E}\left[X \mathbb{1}_{X \geqslant u}\right] \quad \underset{u \to \infty}{\longrightarrow} \quad \frac{2-\alpha}{\alpha-1},$$

see [31, Lemma in XVII.5]. Therefore,

$$m_n \quad \underset{n \to \infty}{\sim} \quad \frac{2-\alpha}{\alpha-1} \cdot \frac{L_0(c b_n)}{(c b_n)^\alpha},$$

where for $\alpha = 2$ the equivalent $\sim$ should be interpreted as a little-o. By (8), the last quantity is indeed $\mathcal{O}(1/n)$.

Estimation of s_n. Since the function $x \mapsto (e^x - 1 - x)/x^2$ is bounded on $[-1, 1]$, we have $s_n = \mathcal{O}(\mathbb{E}\left[X^2 \mathbb{1}_{X \leqslant c b_n}\right])/b_n^2$. Hence, by (7) and (8), $s_n = \mathcal{O}(L_0(b_n)/b_n^\alpha) = \mathcal{O}(1/n)$. This completes the proof. □

Remark 5.3. In the previous statement, note that is is important to take $c \geqslant 1$. Indeed, the statement is false in general for every $c > 0$. For example, if $\mathbb{P}(X = n) \sim \frac{1}{n^4}$ as $n \to \infty$, then $S_n/\sqrt{n}$ and $\max(X_1, \ldots, X_n)/n^{1/4}$ converge (jointly) in distribution to non-degenerate random variables. If Proposition 5.2 were true for every $c > 0$, we would have

$$\mathbb{P}(S_n > \sqrt{n}, X_1 \leqslant 2n^{1/4}, \ldots, X_n \leqslant 2n^{1/4}) \leqslant Ce^{-n^{1/4}/2} \quad \underset{n\to\infty}{\longrightarrow} \quad 0,$$

which is not the case.

5.2 A local estimate

We now establish a local estimate under the following assumptions.

Assumptions. Let X be an integer-valued random variable such that:

- $\mathbb{P}(X \geqslant -1) = 1$ and $\mathbb{P}(X > 0) > 0$;
- $\mathbb{E}[X] = 0$;
- We have
$$\mathbb{P}(X = n) \quad \underset{n\to\infty}{\sim} \quad \frac{L(n)}{n^{1+\beta}} \tag{10}$$
with $\beta > 1$.

The last condition implies that X is in the domain of attraction of a stable law of index $\min(2, \beta)$, so these assumptions are stronger than those made in the beginning of Section 5.

The following result is due to Doney [25], where for $n \geqslant 1$ we set $S_n = X_1 + \cdots + X_n$ with $(X_i)_{i \geqslant 1}$ i.i.d. satisfying the previous assumptions.

Theorem 5.4. *Fix $\epsilon > 0$. Uniformly for $m \geqslant \epsilon n$,*

$$\mathbb{P}(S_n = m) \quad \underset{n\to\infty}{\sim} \quad n \cdot \mathbb{P}(X_1 = m) \tag{11}$$

This result indicates a "one-big jump principle": for $m \geqslant \epsilon n$, intuitively speaking having $S_n = m$ amounts to having one big jump of size m among the n possible. The proof will confirm the intuition, and Theorem 5.6 below gives a quantitive statement in this direction. The question of finding the optimal m_n such that (11) holds uniformly for $m \geqslant m_n$ has attracted a lot of attention, see [24]

Proof. We follow the proof of [25], which follows the lines of [57]. Set

$$\ell_m = \frac{m}{(\ln(m))^3}.$$

The idea is to separate the cases depending on the number of jumps at least equal to ℓ_m by writing $\mathbb{P}(S_n = m) \leqslant nP_1^{m,n} + P_2^{m,n} + P_3^{m,n}$ with

$$P_1^{m,n} = \mathbb{P}\left(S_n = m, X_n \geqslant \ell_m, \max_{1\leqslant k\leqslant n-1} X_k < \ell_m\right), \quad P_2^{m,n} = \mathbb{P}\left(S_n = m, \max_{1\leqslant k\leqslant n} X_k < \ell_m\right)$$

and

$$P_3^{m,n} = \mathbb{P}\left(S_n = m, \bigcup_{1\leqslant i<k\leqslant n} \{X_i \geqslant \ell_m, X_k \geqslant \ell_m\}\right).$$

Estimation of $P_3^{m,n}$. Write

$$\begin{aligned} P_3^{m,n} &\leqslant \binom{n}{2}\mathbb{P}(S_n = m, X_1 \geqslant \ell_m, X_2 \geqslant \ell_m) \\ &\leqslant n^2 \sum_{i\geqslant 0} \mathbb{P}(S_{n-2} = i)\,\mathbb{P}(X_1 \geqslant \ell_m, X_2 \geqslant \ell_m, X_1 + X_2 = m - i) \\ &\leqslant n^2\mathbb{P}(X_1 \geqslant \ell_m) \sup_{j\geqslant \ell_m} \mathbb{P}(X_2 = j). \end{aligned}$$

By (10), the last quantity is asymptotic to

$$n^2 \cdot \frac{\beta L(\ell_m)}{\ell_m^{\beta}} \cdot \frac{L(\ell_m)}{\ell_m^{1+\beta}}.$$

Since $\beta > 1$, by definition of ℓ_m and using the fact that $m \geqslant \varepsilon n$ together with the Potter bounds, we get that $P_3^{m,n} = o(n\mathbb{P}(X_1 = m))$ uniformly in $m \geqslant \varepsilon n$.

Estimation of $P_2^{m,n}$. We use Proposition 5.2 to write

$$P_2^{m,n} \leqslant \mathbb{P}\left(S_n \geqslant m, \max_{1\leqslant k\leqslant n} X_k < \ell_m\right) \leqslant C\exp\left(-\frac{m}{\ell_m}\right) \leqslant C\exp\left(-\ln(m)^3\right),$$

which is $o(n\mathbb{P}(X_1 = m))$ uniformly in $m \geqslant \varepsilon n$.

Estimation of $P_1^{m,n}$. This is the more difficult part. The idea is to introduce another cutoff to take into account the small values of S_{n-1}. Specifically, set $\alpha = \min(\beta, 2)$ and $\alpha' = \frac{1+\alpha}{2\alpha} \in (1/\alpha, 1)$, so that $S_n/n^{\alpha'} \to 0$ in probability by (8) , and write $P_2^{m,n} \leqslant Q_1^{m,n} + Q_2^{m,n} + Q_3^{m,n}$ with

$$\begin{aligned} Q_1^{m,n} &= \mathbb{P}\left(S_n = m, X_n \geqslant \ell_m, \max_{1\leqslant k\leqslant n-1} X_k < \ell_m, S_{n-1} > \frac{m}{\ln(m)}\right) \\ Q_2^{m,n} &= \mathbb{P}\left(S_n = m, X_n \geqslant \ell_m, \max_{1\leqslant k\leqslant n-1} X_k < \ell_m, S_{n-1} < -n^{\alpha'}\right) \\ Q_3^{m,n} &= \mathbb{P}\left(S_n = m, X_n \geqslant \ell_m, \max_{1\leqslant k\leqslant n-1} X_k < \ell_m, -n^{\alpha'} \leqslant S_{n-1} \leqslant \frac{m}{\ln(m)}\right). \end{aligned}$$

To estimate $Q_1^{m,n}$, using Proposition 5.2 we have

$$Q_1^{m,n} \leqslant \mathbb{P}\left(S_{n-1} > \frac{m}{\ln(m)}, \max_{1\leqslant k\leqslant n-1} X_k < \ell_m\right) \leqslant C\exp\left(-\ln(m)^2\right)$$

which is $o(\mathbb{P}(X_1 = m))$ uniformly in $m \geqslant \varepsilon n$.

To estimate $Q_2^{m,n}$, write

$$Q_2^{m,n} \leqslant \sum_{j < -n^{\alpha'}} \mathbb{P}(S_{n-1} = j, X_n = m - j) \leqslant \mathbb{P}\left(S_{n-1} < -n^{\alpha'}\right) \sup_{j > n^{\alpha'}} \mathbb{P}(X_1 = m + j),$$

which is $o(\mathbb{P}(X_1 = m))$ uniformly in $m \geqslant \varepsilon n$.

To estimate $Q_3^{m,n}$, notice that $\mathbb{P}(X_n = m - j) \sim \mathbb{P}(S_1 = m)$ uniformly in $-n^{\alpha'} \leqslant j \leqslant \frac{m}{\ln(m)}$. Therefore, since $m - m/\ln(m) \geqslant \ell_m$ for m sufficiently large, we have

$$Q_3^{m,n} \quad \underset{n\to\infty}{\sim} \quad \mathbb{P}\left(\max_{1 \leqslant k \leqslant n-1} X_k < \ell_m, -n^{\alpha'} \leqslant S_{n-1} \leqslant \frac{m}{\ln(m)}\right) \mathbb{P}(S_1 = m).$$

Since $\mathbb{P}\left(-n^{\alpha'} \leqslant S_{n-1} \leqslant m/\ln(m)\right) \to 1$, it suffices to check that $\mathbb{P}(\max_{1 \leqslant k \leqslant n-1} X_k < \ell_m) \to 1$. This readily follows from the fact that $n\mathbb{P}(X_1 \geqslant \ell_m) \to 0$. $\square$

Remark 5.5. If one only assumes that $\mathbb{P}(X \geqslant -1) = 1$, $\mathbb{P}(X > 0) > 0$, $\mathbb{E}[X] = 0$ and $\mathbb{P}(X \geqslant n) = \frac{L(n)}{n^\beta}$ with $\beta > 1$, one can show that for every $\epsilon > 0$, uniformly for $m \geqslant \epsilon n$, $\mathbb{P}(S_n \geqslant m) \sim n \cdot \mathbb{P}(X_1 \geqslant m)$ as $n \to \infty$ (this is one of the main results of [57]).

5.3 A one big jump principle

We keep the notation and assumptions of Section 5.2 and fix a sequence (x_n) such that $\liminf_{n\to\infty} x_n/n > 0$. We establish here that, conditionally given $S_n = x_n$, a one-big jump principle appears.

We start with some notation. Let

$$V_n := \inf\left\{1 \leqslant j \leqslant n : X_j = \max\{X_i : 1 \leqslant i \leqslant n\}\right\}$$

be the first index of the maximal element of $(X_1, \ldots, X_n)$. Let $(X_1^{(n)}, \ldots, X_{n-1}^{(n)})$ be a random variable distributed as $(X_1, \ldots, X_{V_n - 1}, X_{V_n + 1} \ldots, X_n)$ under $\mathbb{P}(\,\cdot\,|S_n = x_n)$.

The following result is due to Armendáriz & Loulakis [9]:

Theorem 5.6. *We have*

$$\sup_{A \in \mathcal{B}(\mathbb{R}^{n-1})} \left|\mathbb{P}\left(\left(X_i^{(n)} : 1 \leqslant i \leqslant n-1\right) \in A\right) - \mathbb{P}((X_i : 1 \leqslant i \leqslant n-1) \in A)\right| \quad \underset{n\to\infty}{\longrightarrow} \quad 0.$$

Roughly speaking, this results states that the conditioning $S_n = x_n$ affects only the maximum jump in the limit, and the other jumps become asymptotically independent.

Let us explain the main conceptual idea of the proof first. If (μ_n) and $(\widetilde{\mu}_n)$ are two sequences of probability measure (with μ_n "complicated" and $\widetilde{\mu}_n$ "simpler"), to show that $\sup_A |\mu_n(A) - \widetilde{\mu}_n(A)| \to 0$, it is enough to:

– find an event E_n typical for $\widetilde{\mu}_n$, in the sense that $\widetilde{\mu}_n(E_n) \to 1$;

– show that $\sup_A |\mu_n(A \cap E_n) - \widetilde{\mu}_n(A \cap E_n) \to 0$ (which is simpler because we work on the event E_n).

Proof. For every $A \in \mathcal{B}(\mathbb{R}^{n-1})$, note that $\mathbb{P}((X_1, \dots, X_{V_n-1}, X_{V_n+1} \dots, X_n) \in A, S_n = x_n)$ is bounded for n sufficiently large from below by the probability of the event

$$\bigcup_{i=1}^{n} \left\{ (X_1, \dots, X_{i-1}, X_{i+1} \dots, X_n) \in A, \Big| \sum_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} X_j \Big| \leqslant K b_n, \max_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} X_j < x_n - K b_n, S_n = x_n \right\},$$

where $K > 0$ is an arbitrary constant and the events are disjoint. By cyclic invariance of the law of $(X_1, \dots, X_n)$, we get that $\mathbb{P}((X_1, \dots, X_{V_n-1}, X_{V_n+1} \dots, X_n) \in A, S_n = x_n)$ is bounded from below by

$$n\mathbb{P}\left((X_1, \dots, X_{n-1}) \in A, \Big| S_{n-1} \Big| \leqslant K b_n, \max_{1 \leqslant j \leqslant n-1} X_j < x_n - K b_n, S_n = x_n \right).$$

Let us introduce the event

$$G_n(K) \coloneqq \left\{ \Big| S_{n-1} \Big| \leqslant K b_n, \max_{1 \leqslant j \leqslant n-1} X_j < x_n - K b_n \right\}.$$

To simplify notation, set $\Delta = [0, 1)$. By (10), observe that

$$\mathbb{P}(X_1 \in x_n - k_n + \Delta) \quad \underset{n \to \infty}{\sim} \quad \mathbb{P}(X_1 \in x_n + \Delta)$$

uniformly in k_n satisfying $|k_n| \leqslant K b_n$. Moreover, by Theorem 5.4 we have that $\mathbb{P}(S_n = x_n) \sim n\mathbb{P}(X_1 \in x_n)$. Therefore, there exists a sequence $\varepsilon_n \to 0$ such that

$$\begin{aligned} \mathbb{P}\left(\left(X_1^{(n)}, \dots, X_{n-1}^{(n)} \right) \in A \right) &\geqslant (1 - \varepsilon_n) \mathbb{P}\left((X_1, \dots, X_{n-1}) \in A, G_n(K) \right) \\ &\geqslant (1 - \varepsilon_n) \left(\mathbb{P}((X_1, \dots, X_{n-1}) \in A) - \mathbb{P}\left(\overline{G_n(K)} \right) \right). \end{aligned}$$

Hence

$$\begin{aligned} &\mathbb{P}\left(\left(X_1^{(n)}, \dots, X_{n-1}^{(n)} \right) \in A \right) - \mathbb{P}((X_1, \dots, X_{n-1}) \in A) \\ &\qquad\qquad \geqslant -\varepsilon_n \mathbb{P}((X_1, \dots, X_{n-1}) \in A) - (1 - \varepsilon_n) \mathbb{P}\left(\overline{G_n(K)} \right). \end{aligned}$$

By writing this inequality with $\overline{A}$ instead of A, we get that

$$\left| \mathbb{P}\left(\left(X_1^{(n)}, \dots, X_{n-1}^{(n)} \right) \in A \right) - \mathbb{P}((X_1, \dots, X_{n-1}) \in A) \right| \leqslant \varepsilon_n + \mathbb{P}\left(\overline{G_n(K)} \right).$$

It therefore remains to check that

$$\limsup_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\overline{G_n(K)} \right) = 0.$$

To this end, first notice that by (9)

$$\mathbb{P}\left(\left|S_{n-1}\right| > Kb_n\right) \quad \xrightarrow[n\to\infty]{} \quad \mathbb{P}(Y > K),$$

for a certain finite random variable Y. In particular, we have $\mathbb{P}(Y > K) \to 0$ as $K \to \infty$. Second, write

$$\mathbb{P}\left(\max_{1\leqslant j\leqslant n-1} X_j \geqslant x_n - Kb_n\right) = 1 - (1 - \mathbb{P}(X_1 \geqslant x_n - Kb_n)))^{n-1}.$$

But $(n-1)\mathbb{P}(X_1 \geqslant x_n - Kb_n) \to 0$, hence the result. □

The following corollary justifies the denomination "one-big jump principle" (because $b_n = o(x_n)$).

Corollary 5.7. *Denote by* Δ_n *the maximal element of* $(X_1, \dots, X_n)$ *conditionally given* $S_n = x_n$*, and by* $\Delta_n^{(2)}$ *its second maximal element. Then:*

(i) we have $\dfrac{\Delta_n}{x_n} \quad \xrightarrow[n\to\infty]{(\mathbb{P})} \quad 1.$

(ii) We have

$$\frac{\Delta_n - x_n}{b_n} \quad \xrightarrow[n\to\infty]{(d)} \quad -Y_\alpha;$$

(iii) Let (b'_n) *such that* $\beta\frac{nL(b'_n)}{(b'_n)^\beta} \to 1$*. We have for every* $u > 0$*,*

$$\mathbb{P}\left(\frac{\Delta_n^{(2)}}{b'_n} \leqslant u\right) \quad \xrightarrow[n\to\infty]{} \quad \exp\left(-u^{-\beta}\right).$$

Observe that b'_n is of order $n^{1/\beta}$ (up to a slowly varying function). Also, by Remark 5.1, for $1 < \beta < 2$, b_n/b'_n converges to a positive constant, while for $\beta = 2$, $b'_n = o(b_n)$.

Proof. First of all, (i) is a simple consequence of (ii): since $\liminf_{n\to\infty} x_n/n > 0$, we have $b_n = o(x_n)$.

For (ii), conditionally given $S_n = x_n$, we have $\Delta_n = x_n - X_1^{(n)} - \dots - X_{n-1}^{(n)}$. The desired result then follows from Theorem 5.6 and (9).

For (iii), by Theorem 5.6, it suffices to show that

$$\mathbb{P}\left(\frac{\max(X_1, \dots, X_n)}{b'_n} \leqslant u\right) \quad \xrightarrow[n\to\infty]{} \quad \exp\left(-u^{-\beta}\right).$$

To this end, write

$$\mathbb{P}(\max(X_1, \dots, X_n) \leqslant ub'_n) = (1 - \mathbb{P}(X_1 \geqslant ub'_n))^n = \exp(n \ln(1 - \mathbb{P}(X_1 \geqslant ub'_n))).$$

Since L is slowly varying,

$$n\mathbb{P}(X_1 \geqslant ub'_n) \quad \underset{n\to\infty}{\sim} \quad n\frac{\beta L(b'_n)}{u^\beta (b'_n)^\beta}.$$

The desired result follows by definition of b'_n. □

6 Application: condensation in subcritical Bienaymé–Galton–Watson trees

We now turn to our application concerning subcritical BGW trees whose offspring distribution is regularly varying.

Assumptions. We assume here that μ is an offspring distribution such that:

- $m \coloneqq \sum_{i=0}^{\infty} i\mu(i) < 1$;
- we have $\mu(n) = \frac{L(n)}{n^{1+\beta}}$, with L slowly varying and $\beta > 1$.

We set $\alpha = \min(\beta, 2)$, denote by $(X_i)_{i \geqslant 1}$ i.i.d. random variables with distribution given by $\mathbb{P}(X_1 = i) = \mu(i+1)$ for $i \geqslant -1$, and finally set $W_n = X_1 + \cdots + X_n$ as well as $\zeta = \inf\{n \geqslant 1 : W_n = -1\}$. Observe that $\mathbb{E}[X_1] = m - 1 < 0$.

As in the beginning of Section 5, the assumptions on μ entail that $\mathbb{E}\left[X_1^2 \mathbb{1}_{X_1 \leqslant n}\right] = L_0(n) n^{2-\alpha}$ for a certain slowly varying function L_0, and we consider here as well a scaling sequence (b_n) such that

$$\frac{nL_0(b_n)}{b_n^{\alpha}} \quad \xrightarrow[n\to\infty]{} \quad \begin{cases} \dfrac{1}{(2-\alpha)\Gamma(-\alpha)} & \text{for } \alpha < 2 \\ 2 & \text{for } \alpha = 2. \end{cases},$$

so that

$$\frac{W_n + (1-m)n}{b_n} \quad \xrightarrow[n\to\infty]{(d)} \quad Y_\alpha,$$

where Y_α is an α-stable spectrally positive random variable normalized so that $\mathbb{E}[e^{-\lambda Y_\alpha}] = e^{\lambda^\alpha}$ for every $\lambda \geqslant 0$. Recall that b_n is of order $n^{1/\alpha}$ (up to a slowly varying function) and that when μ has finite variance $\sigma^2 \in (0, \infty)$, one may take $b_n = \sigma\sqrt{n/2}$.

Finally, we let $\mathcal{T}_n$ be a BGW_μ tree conditioned on having n vertices (to avoid periodicity issues, we assume that this conditioning is non-degenerate for n sufficiently large).

Let $u_\star(\mathcal{T}_n)$ be the vertex with maximal degree of $\mathcal{T}_n$ (if there are several vertices with maximum degree, choose the first one in lexicographical order, but we will see that this vertex is unique with high probability) and denote by $\Delta(\mathcal{T}_n)$ its outdegree. Let also $\Delta^{(2)}(\mathcal{T}_n)$ be the maximal outdegree of the remaining vertices.

We investigate the condensation phenomenon in two directions. First, we establish a law of large numbers and a central limit type result for $\Delta(\mathcal{T}_n)$ (Theorem 6.1). Second, we study the asymptotic behavior of the height of $u_\star(\mathcal{T}_n)$ (Theorem 6.2) and show that it converges in distribution to a geometric random variable of parameter $1 - m$. Both results combine the coding of $\mathcal{T}_n$ by the Łukasiewicz path with the previously established "one big jump principle" (Theorem 5.6).

Theorem 6.1. *The following assertions hold:*

(i) we have $\dfrac{\Delta(\mathcal{T}_n)}{(1-m)n} \quad \xrightarrow[n\to\infty]{(\mathbb{P})} \quad 1.$

(ii) We have

$$\frac{\Delta(\mathcal{T}_n)-(1-m)n}{b_n} \quad \xrightarrow[n\to\infty]{(d)} \quad -Y_\alpha.$$

(iii) Let (b'_n) *such that* $\beta\frac{nL(b'_n)}{(b'_n)^\beta} \to 1$*. We have for every* $u>0$*,*

$$\mathbb{P}\left(\frac{\Delta^{(2)}(\mathcal{T}_n)}{b'_n} \leqslant u\right) \quad \xrightarrow[n\to\infty]{} \quad \exp\left(-u^{-\beta}\right),$$

with the quantity on the right-hand side being interpreted as 1 *for* $\alpha=2$*.*

When $\mu(n) \sim c/n^{1+\beta}$ as $n\to\infty$ (that is when $L = c+o(1)$), the first assertion is due to Jonsson & Stefánsson [42] and the others to Janson [39]. The general case is treated in [47].

Denote by $|u_\star(\mathcal{T}_n)|$ the height of $u_\star(\mathcal{T}_n)$. The following result was established in [47].

Theorem 6.2. *For every* $i \geqslant 0$*, we have*

$$\mathbb{P}\left(|u_\star(\mathcal{T}_n)|=i\right) \quad \xrightarrow[n\to\infty]{} \quad (1-m)m^i.$$

6.1 Approximating the Łukasiewicz path

Denote by $(W_i^{(n)} : 0 \leqslant i \leqslant n)$ the random walk $(W_i : i \geqslant 0)$ under the conditional probability $\mathbb{P}(\,\cdot\,|\zeta = n)$, which has the same distribution as the Łukasiewicz path of $\mathcal{T}_n$ (see Section 3). The first step to prove Theorems 6.1 and 6.2 is to show that $W^{(n)}$ can be well approximated by a path constructed in a simple way.

For every $n \geqslant 1$, define the random process $Z^{(n)} \coloneqq (Z_i^{(n)} : 0 \leqslant i \leqslant n)$ by

$$Z^{(n)} \coloneqq \mathcal{V}(W_0, W_1, \dots, W_{n-1}, -1), \tag{12}$$

where $\mathcal{V}$ denotes the Vervaat transform (see Definition 3.11). The next result shows that $(Z_i^{(n)} : 0 \leqslant i \leqslant n)$ is a good approximation of $(W_i^{(n)} : 0 \leqslant i \leqslant n)$ as n goes to infinity.

Theorem 6.3. *We have*

$$\sup_{A\in\mathcal{B}(\mathbb{R}^{n+1})}\left|\mathbb{P}\left(\left(W_i^{(n)} : 0 \leqslant i \leqslant n\right) \in A\right) - \mathbb{P}\left(\left(Z_i^{(n)} : 0 \leqslant i \leqslant n\right) \in A\right)\right| \quad \xrightarrow[n\to\infty]{} \quad 0.$$

Proof. Throughout the proof, we let $B^{(n)} \coloneqq (B_i^{(n)} : 0 \leqslant i \leqslant n)$ be a bridge of length n, that is, a process distributed as $(W_i : 0 \leqslant i \leqslant n)$ under $\mathbb{P}(\,\cdot\,|W_n = -1)$.

Denote by $(X_1^{(n)}, \dots, X_{n-1}^{(n)})$ the jumps of $B^{(n)}$ with the first maximal jump removed. We apply Theorem 5.6 with the centered random walk $S_n = W_n + (1-m)n$ with increments $(X_i + 1 - m)_{1 \leqslant i \leqslant n-1}$, conditionally given $S_n = x_n$ with $x_n = -1 + (1-m)n$ (note that (S_n) is not integer-valued, but it is a simple matter to see that the results carry through):

$$\sup_{A \in \mathcal{B}(\mathbb{R}^{n-1})} \left| \mathbb{P}\left(\left(X_i^{(n)} + 1 - m \right)_{1 \leqslant i \leqslant n-1} \in A \right) - \mathbb{P}\left((X_i + 1 - m)_{1 \leqslant i \leqslant n-1} \in A \right) \right| \quad \underset{n \to \infty}{\longrightarrow} \quad 0,$$

which implies

$$\sup_{A \in \mathcal{B}(\mathbb{R}^{n-1})} \left| \mathbb{P}\left(\left(X_i^{(n)} \right)_{1 \leqslant i \leqslant n-1} \in A \right) - \mathbb{P}\left((X_i)_{1 \leqslant i \leqslant n-1} \in A \right) \right| \quad \underset{n \to \infty}{\longrightarrow} \quad 0. \tag{13}$$

For every $0 \leqslant i < n$, we denote by $b_i^{(n)} := B_{i+1}^{(n)} - B_i^{(n)}$ the i-th increment of the bridge. We will need the first time at which $(B_i^{(n)} : 0 \leqslant i \leqslant n)$ reaches its largest jump, defined by

$$V_n^b := \inf \left\{ 0 \leqslant i < n : b_i^{(n)} = \max \left\{ b_j^{(n)} : 0 \leqslant j < n \right\} \right\}.$$

Without loss of generality, we assume that the largest jump of $B^{(n)}$ is reached once ((13) entails that this happens with probability tending to 1 as $n \to \infty$ since $\max(X_1, \dots, X_n)/b_n$ converges in distribution). We finally introduce the shifted bridge $R^{(n)} := (R_i^{(n)} : 0 \leqslant i \leqslant n)$, obtained by reading the jumps of the bridge $B^{(n)}$ from left to right starting from V_n^b. Namely, we set

$$R_0^{(n)} = 0 \quad \text{and} \quad r_i^{(n)} := R_{i+1}^{(n)} - R_i^{(n)} = b_{V_n^b + i + 1 \mod [n]}^{(n)}, \quad 0 \leqslant i < n,$$

see Figure 9 for an illustration.

Since V_n^b is independent of $(b_0^{(n)}, \dots, b_{V_n^b - 1}^{(n)}, b_{V_n^b + 1}^{(n)}, \dots, b_{n-1}^{(n)})$, we have

$$\left(r_i^{(n)} : 0 \leqslant i < n-1 \right) = \left(b_{V_n^b + i + 1 \mod [n]}^{(n)} : 0 \leqslant i < n-1 \right) \overset{(d)}{=} \left(X_1^{(n)}, \dots, X_{n-1}^{(n)} \right).$$

Hence, by (13),

$$\sup_{A \in \mathcal{B}(\mathbb{R}^{n-1})} \left| \mathbb{P}\left(\left(R_i^{(n)} \right)_{1 \leqslant i \leqslant n-1} \in A \right) - \mathbb{P}\left((W_i)_{1 \leqslant i \leqslant n-1} \in A \right) \right| \quad \underset{n \to \infty}{\longrightarrow} \quad 0.$$

We now use the Vervaat transform. By construction, $\mathcal{V}(R^{(n)}) = \mathcal{V}(B^{(n)})$ (see Figure 9), and $\mathcal{V}(B^{(n)})$ has the same distribution as the excursion $(W_i^{(n)} : 0 \leqslant i \leqslant n)$ by Lemma 3.12. Since $R_n^{(n)} = -1$ and $Z^{(n)} = \mathcal{V}(W_0, \dots, W_{n-1}, -1)$ by definition, this concludes the proof. □

6.2 Proof of the results

In order to prove Theorems 6.1 and 6.2, the main idea is to establish them for a modified tree $\mathcal{T}^{(n)}$ whose Lukasiewicz path $Z^{(n)}$ is defined by

$$Z^{(n)} := \mathcal{V}(W_0, W_1, \dots, W_{n-1}, -1).$$

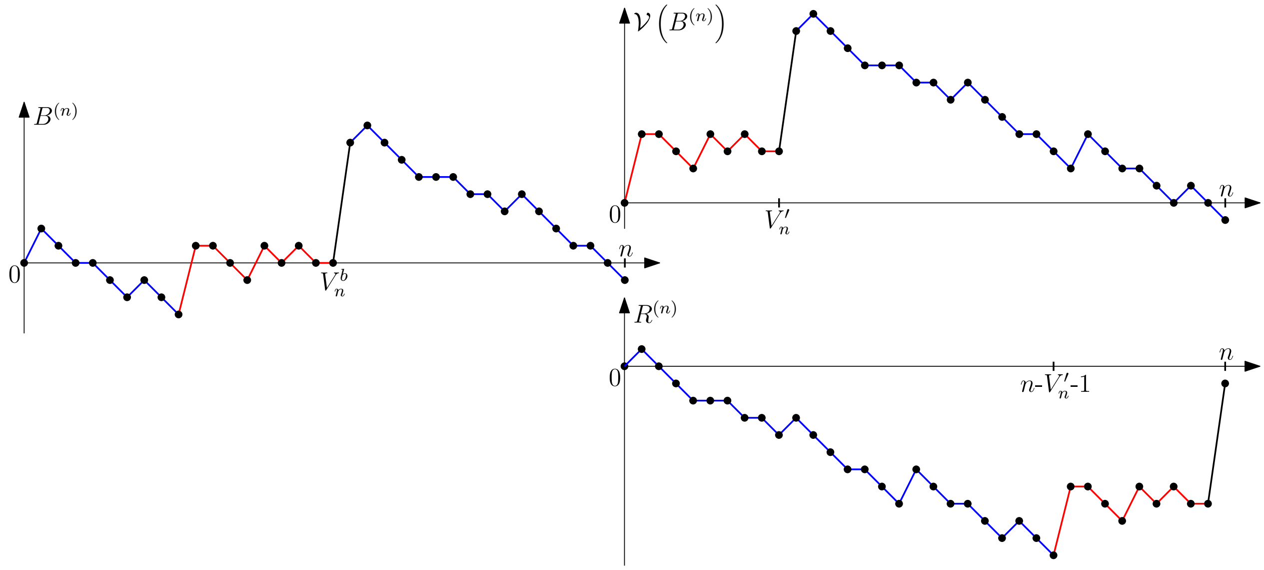


Figure 9: The bridge $B^{(n)} = (B_i^{(n)} : 0 \leqslant i \leqslant n)$ with the location V_n^b of its (first) maximal jump, its Vervaat transform $\mathcal{V}(B^{(n)})$ with the location V_n' of its (first) maximal jump, and the shifted bridge $R^{(n)} = (R_i^{(n)} : 0 \leqslant i \leqslant n)$ with the location of its first overall minimum.

Indeed, this is possible thank to Theorem 6.3.

First observe that, with probability tending to 1 as $n \to \infty$, the maximum jump of $(W_0, W_1, \dots, W_{n-1})$ is of order $b_n = o(n)$ and since W_n is of order $-(1-m)n$, the last jump of $(W_0, W_1, \dots, W_{n-1}, -1)$ is of order $(1-m)n$.

Proof of Theorem 6.1. By the previous observation, we may assume that the maximum jump of $(W_0, W_1, \dots, W_{n-1}, -1)$ is the last one. Then

$$\Delta(\mathcal{T}^{(n)}) = -W_{n-1}, \qquad \Delta^{(2)}(\mathcal{T}_n) = \max_{1 \leqslant i \leqslant n} (W_i - W_{i-1}),$$

and the desired result follows by the same calculations as in the proof of Corollary 5.7. $\square$

Proof of Theorem 6.2. Observe that

$$|u_\star(\mathcal{T}^{(n)})| = \mathrm{Card}\left(\left\{0 \leqslant i \leqslant n-1 : W_i = \min_{i \leqslant j \leqslant n-1} W_j\right\}\right),$$

see Figure 10.

By time reversal at time $n-1$ (see the solution of Exercice 4 for a similar argument), $|u_\star(\mathcal{T}^{(n)})|$ has the same distribution as

$$\mathrm{Card}(\{0 \leqslant i \leqslant n-1 : W_i = \max_{0 \leqslant i \leqslant n-1} W_i\}).$$

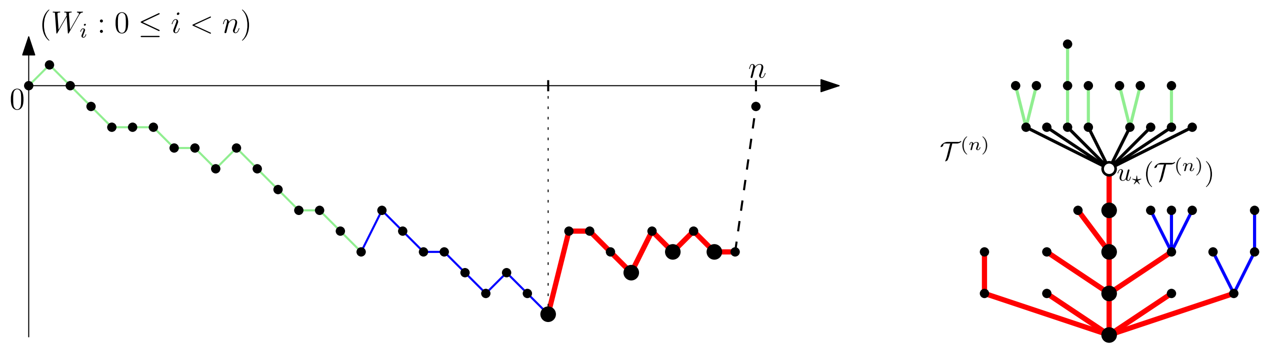


Figure 10: Left: the path $(W_1, W_1, \ldots, W_{n-1}, 1)$. Right: the tree $\mathcal{T}^{(n)}$ whose Lukasiewicz path is $\mathcal{V}(W_1, W_1, \ldots, W_{n-1}, 1)$. The ancestors of $u_\star(\mathcal{T}^{(n)})$ (in bold) correspond to times $0 \leqslant i \leqslant n-1$ such that $W_i = \min_{i \leqslant j \leqslant n-1} W_j$ (in bold).

Therefore, as $n \to \infty$, $|u_\star(\mathcal{T}^{(n)})|$ converges in distribution to the number of weak ladder times of $(W_i)_{i \geqslant 0}$ (which is almost surely finite, since the random walk has negative drift). We saw in the solution of Exercice 4 that if T_k denotes the k-th weak ladder time, $\mathbb{P}(T_k < \infty) = \mathbb{P}(T_1 < \infty)^k$ and that $\mathbb{P}(T_1 < \infty) = 1 - \mathbb{E}[W_1] = m$. The desired result follows. □

Remark 6.4. It is also possible to show that the height of $\mathcal{T}_n$ behaves logarithmically; more precisely,

$$\frac{\mathsf{Height}(\mathcal{T}_n)}{\ln(n)} \xrightarrow[n \to \infty]{(\mathbb{P})} \frac{1}{\ln(1/m)}.$$

see [47, Theorem 4]. Intuitively speaking, $\mathcal{T}_n$ looks like a finite spine on top of which are grafted $(1-m)n$ asymptotically independent BGW_μ trees, for which the tail of the height decreases exponentially fast.

References

[1] R. Abraham and J.-F. Delmas, *Local limits of conditioned Galton-Watson trees: the condensation case*, Electron. J. Probab., 19 (2014), pp. no. 56, 29.

[2] ——, *Local limits of conditioned Galton-Watson trees: the infinite spine case*, Electron. J. Probab., 19 (2014), pp. no. 2, 19.

[3] L. Addario-Berry, *Tail bounds for the height and width of a random tree with a given degree sequence*, Random Structures Algorithms, 41 (2012), pp. 253–261.

[4] ——, *A probabilistic approach to block sizes in random maps*, arxiv:1503.08159, (2015).

[5] M. Albenque and J.-F. Marckert, *Some families of increasing planar maps*, Electron. J. Probab., 13 (2008), pp. no. 56, 1624–1671.

[6] D. Aldous, *The continuum random tree. I*, Ann. Probab., 19 (1991), pp. 1–28.

[7] ——, *The continuum random tree. II. An overview*, in Stochastic analysis (Durham, 1990), vol. 167 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge, 1991, pp. 23–70.

[8] ——, *The continuum random tree III*, Ann. Probab., 21 (1993), pp. 248–289.

[9] I. ARMENDÁRIZ AND M. LOULAKIS, *Conditional distribution of heavy tailed random variables on large deviations of their sum*, Stochastic Process. Appl., 121 (2011), pp. 1138–1147.

[10] K. B. ATHREYA AND P. E. NEY, *Branching processes*, Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[11] N. BACAËR, *A short history of mathematical population dynamics*, Springer-Verlag London, Ltd., London, 2011.

[12] J. BENNIES AND G. KERSTING, *A random walk approach to Galton-Watson trees*, J. Theoret. Probab., 13 (2000), pp. 777–803.

[13] Q. BERGER, *Notes on random walks in the cauchy domain of attraction*, Probability Theory and Related Fields, (2018).

[14] J. BETTINELLI, *Scaling limit of random planar quadrangulations with a boundary*, Ann. Inst. Henri Poincaré Probab. Stat., 51 (2015), pp. 432–477.

[15] J. BIENAYMÉ, *De la loi de multiplication et de la durée des familles: probabilités,*, Société philomathique de Paris, (1845).

[16] N. H. BINGHAM, C. M. GOLDIE, AND J. L. TEUGELS, *Regular variation*, vol. 27 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1987.

[17] N. BROUTIN AND J.-F. MARCKERT, *Asymptotics of trees with a prescribed degree sequence and applications*, Random Structures Algorithms, 44 (2014), pp. 290–316.

[18] A. CARACENI, *The scaling limit of random outerplanar maps*, Ann. Inst. H. Poincaré Probab. Statist. (à paraître) et arxiv:1405.1971.

[19] X. CHEN AND G. MIERMONT, *Long Brownian bridges in hyperbolic spaces converge to Brownian trees*, Electron. J. Probab., 22 (2017), pp. Paper No. 58, 15.

[20] N. CURIEN, B. HAAS, AND I. KORTCHEMSKI, *The CRT is the scaling limit of random dissections*, Random Structures Algorithms, 47 (2015), pp. 304–327.

[21] N. CURIEN AND I. KORTCHEMSKI, *Random non-crossing plane configurations: a conditioned Galton-Watson tree approach*, Random Structures Algorithms, 45 (2014), pp. 236–260.

[22] ——, *Percolation on random triangulations and stable looptrees*, Probab. Theory Related Fields, 163 (2015), pp. 303–337.

[23] ——, *Percolation on random triangulations and stable looptrees*, Probab. Theory Related Fields, 163 (2015), pp. 303–337.

[24] D. DENISOV, A. B. DIEKER, AND V. SHNEER, *Large deviations for random walks under subexponentiality: the big-jump domain*, Ann. Probab., 36 (2008), pp. 1946–1991.

[25] R. A. DONEY, *A large deviation local limit theorem*, Math. Proc. Cambridge Philos. Soc., 105 (1989), pp. 575–577.

[26] M. DRMOTA, *Random trees*, SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[27] T. DUQUESNE, *A limit theorem for the contour process of conditioned Galton-Watson trees*, Ann. Probab., 31 (2003), pp. 996–1027.

[28] T. DUQUESNE AND J.-F. LE GALL, *Random trees, Lévy processes and spatial branching processes*, Astérisque, (2002), pp. vi+147.

[29] ———, *Probabilistic and fractal aspects of Lévy trees*, Probab. Theory Related Fields, 131 (2005), pp. 553–603.

[30] S. N. EVANS, J. PITMAN, AND A. WINTER, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Related Fields, 134 (2006), pp. 81–126.

[31] W. FELLER, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971.

[32] V. FÉRAY AND I. KORTCHEMSKI, *Random minimal factorizations and random trees*, In preparation, (2017).

[33] D. H. FUK AND S. V. NAGAEV, *Probabilistic inequalities for sums of independent random variables*, Teor. Verojatnost. i Primenen., 16 (1971), pp. 660–675.

[34] J. GEIGER AND G. KERSTING, *The Galton-Watson tree conditioned on its height.*, in Probability theory and mathematical statistics. Proceedings of the 7th international Vilnius conference, Vilnius, Lithuania, August, 12–18, 1998, Vilnius: TEV; Utrecht: VSP, 1999, pp. 277–286.

[35] M. GROMOV, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math., (1981), pp. 53–73.

[36] B. HAAS AND G. MIERMONT, *Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees*, Ann. of Probab., 40 (2012), pp. 2589–2666.

[37] T. E. HARRIS, *First passage and recurrence distributions*, Trans. Amer. Math. Soc., 73 (1952), pp. 471–486.

[38] C. C. HEYDE AND E. SENETA, *Studies in the history of probability and statistics. XXXI. The simple branching process, a turning point test and a fundamental inequality: a historical note on I. J. Bienaymé*, Biometrika, 59 (1972), pp. 680–683.

[39] S. JANSON, *Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation*, Probab. Surv., 9 (2012), pp. 103–252.

[40] S. JANSON AND S. Ö. STEFÁNSSON, *Scaling limits of random planar maps with a unique large face*, Ann. Probab., 43 (2015), pp. 1045–1081.

[41] S. JANSON AND S. O. STEFÁNSSON, *Scaling limits of random planar maps with a unique large face*, Ann. Probab., 43 (2015), pp. 1045–1081.

[42] T. JONSSON AND S. O. STEFÁNSSON, *Condensation in nongeneric trees*, J. Stat. Phys., 142 (2011), pp. 277–313.

[43] D. G. KENDALL, *The genealogy of genealogy: branching processes before (and after) 1873*, Bull. London Math. Soc., 7 (1975), pp. 225–253. With a French appendix containing Bienaymé's paper of 1845.

[44] D. P. KENNEDY, *The Galton-Watson process conditioned on the total progeny*, J. Appl. Probability, 12 (1975), pp. 800–806.

[45] H. KESTEN, *Subdiffusive behavior of random walk on a random cluster*, Ann. Inst. H. Poincaré Probab. Statist., 22 (1986), pp. 425–487.

[46] I. KORTCHEMSKI, *Invariance principles for Galton-Watson trees conditioned on the number of leaves*, Stochastic Process. Appl., 122 (2012), pp. 3126–3172.

[47] ———, *Limit theorems for conditioned non-generic Galton-Watson trees*, Ann. Inst. Henri Poincaré Probab. Stat., 51 (2015), pp. 489–511.

[48] I. KORTCHEMSKI AND L. RICHIER, *Condensation in critical Cauchy Bienaymé-Galton-Watson trees*, Ann. Appl. Probab., 29 (2019), pp. 1837–1877.

[49] M. A. KRIKUN, *Uniform infinite planar triangulation and related time-reversed critical branching process*, Journal of Mathematical Sciences, 131 (2005), pp. 5520–5537.

[50] J.-F. LE GALL, *Random geometry on the sphere*, Proceedings of the ICM 2014.

[51] ———, *Random trees and applications*, Probability Surveys, (2005).

[52] ———, *Itô's excursion theory and random trees*, Stochastic Process. Appl., 120 (2010), pp. 721–749.

[53] J.-F. LE GALL AND Y. LE JAN, *Branching processes in Lévy processes: the exploration process*, Ann. Probab., 26 (1998), pp. 213–252.

[54] R. LYONS, R. PEMANTLE, AND Y. PERES, *Conceptual proofs of* L *log* L *criteria for mean behavior of branching processes*, Ann. Probab., 23 (1995), pp. 1125–1138.

[55] R. LYONS AND Y. PERES, *Probability on Trees and Networks*, Current version available at http://mypage.iu.edu/ rdlyons/, In preparation.

[56] J. MILLER AND S. SHEFFIELD, *An axiomatic characterization of the Brownian map*, arXiv:1506.03806.

[57] S. V. NAGAEV, *On the asymptotic behavior of probabilities of one-sided large deviations*, Teor. Veroyatnost. i Primenen., 26 (1981), pp. 369–372.

[58] K. PANAGIOTOU AND B. STUFLER, *Scaling limits of random Pólya trees*, Probab. Theory Related Fields, 170 (2018), pp. 801–820.

[59] K. PANAGIOTOU, B. STUFLER, AND K. WELLER, *Scaling Limits of Random Graphs from Subcritical Classes*, Ann. Probab. (à paraître).

[60] F. PAULIN, *The Gromov topology on* **R**-*trees*, Topology Appl., 32 (1989), pp. 197–221.

[61] J. PITMAN AND D. RIZZOLO, *Schröder's problems and scaling limits of random trees*, Trans. Amer. Math. Soc., 367 (2015), pp. 6943–6969.

[62] L. RAMZEWS AND B. STUFLER, *Simply generated unrooted plane trees*, ALEA Lat. Am. J. Probab. Math. Stat., 16 (2019), pp. 333–359.

[63] L. RICHIER, *Limits of the boundary of random planar maps*, Probab. Theory Related Fields (to appear), (2018).

[64] D. RIZZOLO, *Scaling limits of Markov branching trees and Galton-Watson trees conditioned on the number of vertices with out-degree in a given set*, Ann. Inst. Henri Poincaré Probab. Stat., 51 (2015), pp. 512–532.

[65] S. O. STEFÁNSSON AND B. STUFLER, *Geometry of large Boltzmann outerplanar maps*, Preprint available on arxiv, arXiv:1710.04460, (2017).

[66] J. F. STEFFENSEN, *Om sandsynligheden for at afkommet uddør*, Matematisk tidsskrift. B, (1930), pp. 19–23.

[67] B. STUFLER, *Limits of random tree-like discrete structures*, arXiv preprint arXiv:1612.02580, (2016).

[68] ———, *Scaling limits of random outerplanar maps with independent link-weights*, Ann. Inst. Henri Poincaré Probab. Stat., 53 (2017), pp. 900–915.

[69] H. W. WATSON AND F. GALTON, *On the probability of the extinction of families.*, The Journal of the Anthropological Institute of Great Britain and Ireland, 4 (1875), pp. 138–144.