

The exploration process of inhomogeneous continuum random trees, and an extension of Jeulin's local time identity

David Aldous*and Grégory Miermont†and Jim Pitman‡

March 8, 2004

Abstract

We study the inhomogeneous continuum random trees (ICRT) that arise as weak limits of birthday trees. We give a description of the *exploration process*, a function defined on $[0, 1]$ that encodes the structure of an ICRT, and also of its *width process*, determining the size of layers in order of height. These processes turn out to be transformations of bridges with exchangeable increments, which have already appeared in other ICRT related topics such as stochastic additive coalescence. The results rely on two different constructions of birthday trees from processes with exchangeable increments, on weak convergence arguments, and on general theory on continuum random trees.

Key words: Continuum random tree, exchangeable increments, exploration process, Lévy process, weak convergence.

Mathematics Subject Classification: 60C05, 60F17, 60G09, 60G51.

*Department of Statistics, University of California, 367 Evans Hall # 3860, Berkeley, CA 94720-3860. Research supported in part by N.S.F. Grant DMS-0203062

†DMA, Ecole Normale Supérieure, and LPMA, Université Paris 6. 45 rue d'Ulm, 75230 Paris Cedex 05, France

‡Department of Statistics, University of California, Berkeley. Research supported in part by N.S.F. Grant DMS-0071448

1 Introduction

This paper completes one circle of ideas (describing the inhomogeneous continuum random tree) while motivated by another (limits of non-uniform random \mathbf{p} -mappings which are essentially different from the uniform case limit). Along the way, a curious extension of Jeulin’s result on total local time for standard Brownian excursion will be established.

Consider a continuous function $f : [0, 1] \rightarrow [0, \infty)$ which is an “excursion” in the sense

$$f(0) = f(1) = 0; \quad f(u) > 0, \quad 0 < u < 1.$$

Use f to make $[0, 1]$ into the pseudo-metric space with distance

$$d(u_1, u_2) := (f(u_1) - \inf_{u_1 \leq u \leq u_2} f(u)) + (f(u_2) - \inf_{u_1 \leq u \leq u_2} f(u)), \quad u_1 \leq u_2. \quad (1)$$

After taking the quotient by identifying points of $[0, 1]$ that are at d -pseudo distance 0, this space is a *tree* in that between any two points there is a unique path; it carries a *length measure* induced by the distance d , and a *mass measure*, with unit total mass, induced from Lebesgue measure on $[0, 1]$. An object with these properties can be abstracted as a *continuum tree*. Using a random excursion function yields a *continuum random tree* (CRT): Aldous [2, 3]. The construction of a continuum random tree \mathcal{T} via a random function f , in this context called the *exploration process* of \mathcal{T} (in Le Gall et al. [26, 17], it is instead called *height process* while the term exploration process is used for a related measure-valued process), is not the only way of looking at a CRT; there are also

- (a) constructions via line-breaking schemes
- (b) descriptions via the spanning subtrees on k random points chosen according to mass measure
- (c) descriptions as weak or strong $n \rightarrow \infty$ limits of rescaled n -vertex discrete random trees.

As discussed in [2, 3] the fundamental example is the *Brownian CRT*, whose exploration process is twice standard Brownian excursion (this was implicit in Le Gall [25]), with line-breaking construction given in Aldous [1], spanning subtree description in Aldous [3] and Le Gall [24], and weak limit (for conditional Galton-Watson trees) behavior in [2, 3] (see Marckert and Mokkadem [27] for recent review). A more general model, the *inhomogeneous continuum random tree* (ICRT) \mathcal{T}^θ , arose in Camarri and Pitman [15] as a weak limit in a certain model (\mathbf{p} -trees) of discrete random trees. The definition and simplest description of \mathcal{T}^θ is via a line-breaking construction based on a Poisson point process in the plane (Aldous and Pitman [15, 9]), which we recall below. The spanning subtree description is set out in Aldous and Pitman [8], and the main purpose of this paper is to complete the description of \mathcal{T}^θ by determining its exploration process (Theorem 1).

1.1 Statement of results

The parameter space Θ of the ICRT \mathcal{T}^θ is defined [9] to consist of sequences $\theta = (\theta_0, \theta_1, \theta_2, \dots)$ such that

- (i) $\theta_0 \geq 0; \quad \theta_1 \geq \theta_2 \geq \dots \geq 0;$
- (ii) $\sum_i \theta_i^2 = 1;$
- (iii) if $\sum_{i=1}^\infty \theta_i < \infty$ then $\theta_0 > 0.$

We will often consider the *finite-length* subspace Θ_{finite} of Θ for which $\theta_i = 0 \forall i > I$, for some $I \geq 0$, calling I the *length* of θ . Note that $\theta \in \Theta_{\text{finite}}$ can be specified by specifying a decreasing sequence $(\theta_1, \dots, \theta_I)$ for which $\sum_{i=1}^I \theta_i^2 < 1$; then set $\theta_0 = \sqrt{1 - \sum_{i \geq 1} \theta_i^2} > 0$.

Let $\{(U_i, V_i), i \geq 1\}$ be a Poisson measure on the first octant $\{(x, y) : 0 \leq y \leq x\}$, with intensity θ_0^2 per unit area. For every $i \geq 1$ let also $(\xi_{i,j}, j \geq 1)$ be a Poisson process on the positive real line with intensity θ_i per unit length. The hypotheses on θ entail that the set of points $\{U_i, i \geq 1, \xi_{i,j}, i \geq 1, j \geq 2\}$ is discrete and can be ordered as $0 < \eta_1 < \eta_2 < \dots$, we call them *cutpoints*. It is easy to see that $\eta_{k+1} - \eta_k \rightarrow 0$ as $k \rightarrow \infty$. By convention let $\eta_0 = 0$. Given a cutpoint $\eta_k, k \geq 1$, we associate a corresponding *joinpoint* η_k^* as follows. If the cutpoint is of the form U_i , then $\eta_k^* = V_i$. If it is of the form $\xi_{i,j}$ with $j \geq 2$, we let $\eta_k^* = \xi_{i,1}$. The hypothesis $\theta_0 > 0$ or $\sum_{i \geq 1} \theta_i = \infty$ implies that joinpoints are a.s. everywhere dense in $(0, \infty)$.

The tree is then constructed as follows. Start with a branch $[0, \eta_1]$, and recursively, given the tree is constructed at stage J , add the line segment $(\eta_J, \eta_{J+1}]$ by branching its left-end to the joinpoint η_J^* (notice that $\eta_J^* < \eta_J$ a.s. so that the construction is indeed recursive as J increases). When all the branches are attached to their respective joinpoints, relabel the joinpoint corresponding to some $\xi_{j,1}$ as joinpoint j , and forget other labels (of the form η_i or η_i^*). We obtain a metric tree (possibly with marked vertices $1, 2, \dots$), whose completion we call \mathcal{T}^θ .

A heuristic description of the structure of the ICRT goes as follows. When $\theta_0 = 1$ and hence $\theta_i = 0$ for $i \geq 1$, the tree is the Brownian CRT, it has no marked vertex and it is a.s. binary, meaning that branchpoints have degree 3. It is the only ICRT for which the width process defined below is continuous, and for which no branchpoint has degree more than 3. When $\theta \in \Theta_{\text{finite}}$ has length $I \geq 1$, the structure looks like that of the CRT, with infinitely many branchpoints with degree 3, but there exist also exactly I branchpoints with infinite degree which we call *hubs*, and these are precisely the marked vertices $1, 2, \dots, I$ corresponding to the joinpoints $\xi_{1,1}, \dots, \xi_{I,1}$ associated to the Poisson processes with intensities $\theta_1, \dots, \theta_I$ defined above. The width process defined below has I jumps with respective sizes $\theta_1, \dots, \theta_I$, which occur at distinct times a.s. These jump-sizes can be interpreted as the *local time* of the different hubs – see remark following Theorem 2. When $\theta \notin \Theta_{\text{finite}}$, then the hubs become everywhere dense on the tree. Whether there exists branchpoints with degree 3 or not depends on whether $\theta_0 \neq 0$ or $\theta_0 = 0$. Also, the tree can become unbounded.

It turns out that the relevant exploration process is closely related to processes recently studied for slightly different purposes. The Brownian CRT in [7], and then the ICRT in [9], were used by Aldous and Pitman to construct versions of the *standard*, and then the *general, additive coalescent*, and its dual *fragmentation process*, which are Markov processes on the state space Δ of sequences $\{(x_1, x_2, \dots) : x_i \geq 0, \sum_i x_i = 1\}$. In [7, 9] the time- t state $\mathbf{X}(t)$ is specified as the vector of masses of tree-components in the forest obtained by randomly cutting the Brownian CRT or ICRT at some rate depending on t . Bertoin [12] gave the following more direct construction. Let $(B_s^{\text{exc}}, 0 \leq s \leq 1)$ be standard Brownian excursion. For fixed $t \geq 0$ consider the process of height-above-past-minimum of

$$B_s^{\text{exc}} - ts, \quad 0 \leq s \leq 1.$$

Then its vector of excursion lengths is Δ -valued, and this process (as t varies) can be identified

with the standard case of the additive coalescent. More generally, for $\theta \in \Theta$ consider the “bridge” process

$$\theta_0 B_s^{\text{br}} + \sum_{i=1}^{\infty} \theta_i (\mathbb{1}_{\{U_i \leq s\}} - s), \quad 0 \leq s \leq 1$$

where (U_i) are independent random variables with uniform law on $(0, 1)$. Use the *Vervaat transform* – relocate the space-time origin to the location of the infimum – to define an “excursion” process $X^\theta = (X_s^\theta, 0 \leq s \leq 1)$ which has positive but not negative jumps. Bertoin [13] used X^θ to construct the general additive coalescent, and Miermont [28] continued the study of fragmentation processes by this method. In this paper we use X^θ to construct a *continuous* excursion process Y^θ ; here is the essential idea. A jump of X^θ at time t_J defines an interval $[t_J, T_J]$ where $T_J := \inf\{t > t_J : X_t^\theta = X_{t_J-}^\theta\}$. Over that interval, replace X_s^θ by $X_{t_J-}^\theta + X_s^\theta - \inf_{t_J \leq u \leq s} X_u^\theta$. Do this for each jump, and let Y^θ be the resulting process. To write it in a more compact way, the formula

$$Y_s^\theta = m \left\{ \inf_{u \leq r \leq s} X_r^\theta : 0 \leq u \leq s \right\} \quad (2)$$

holds, where m is Lebesgue’s measure on \mathbb{R} . Details are given in section 2. We can now state our main result.

Theorem 1. *Suppose $\theta \in \Theta$ satisfies $\sum_i \theta_i < \infty$. Then the exploration process of the ICRT \mathcal{T}^θ is distributed as $\frac{2}{\theta_0^2} Y^\theta$.*

As will be recalled in Sect. 3, the precise meaning of this theorem is: let U_1, U_2, \dots be independent uniform variables on $[0, 1]$, independent of Y^θ , and as around (1), replacing f by $\frac{2}{\theta_0^2} Y^\theta$, endow $[0, 1]$ with a pseudo-distance d , so that the natural quotient gives a tree $\mathcal{T}^{2\theta_0^{-2}Y}$ where $Y = Y^\theta$. Then for every $J \in \mathbb{N}$, the subtree spanned by the root (the class of 0) and the (classes of) U_1, \dots, U_J has the same law as the tree \mathcal{T}_J^θ obtained by performing the stick-breaking construction until the J -th step. Since $(U_i, i \geq 1)$ is a.s. dense in $[0, 1]$ and by uniqueness of the metric completion, \mathcal{T}^θ and $\mathcal{T}^{2\theta_0^{-2}Y}$ indeed encode the same random topological space. We also note that our proofs easily extend to showing that the hub with extra label i is associated to the class of t_i or T_i , and this class is exactly $\{s \in [t_i, T_i] : Y_s^\theta = Y_{t_i}^\theta\}$. To avoid heavier notations, we will not take these extra labels into account from now on.

When $\sum_i \theta_i = \infty$, the exploration process of the ICRT, if it exists, can be obtained as a certain weak limit of processes of the form $\frac{2}{(\theta_0^n)^2} Y^{\theta^n}$ for approximating sequences $\theta^n \in \Theta_{\text{finite}}$, and in particular, when $\theta_0 > 0$ one guesses that the exploration process of \mathcal{T}^θ will still be $\frac{2}{\theta_0^2} Y^\theta$, but we will not concentrate on this in the present paper.

Remark. Formula (2) is inspired by the work of Duquesne and Le Gall [17], in which continuum random trees (“Lévy trees”) are built out of sample paths of Lévy processes. Our work suggest that there are many similarities between ICRTs and Lévy trees. In fact, Lévy trees turn out to be “mixings” of ICRTs in an analogous way that Lévy bridges are mixing of extremal bridges with exchangeable increments. This will be pursued elsewhere.

In principle Theorem 1 should be provable within the continuous-space context, but we do not see such a direct proof. Instead we use weak convergence arguments. As background, there are

many ways of coding discrete trees as walks. In particular, one can construct a Galton-Watson tree with offspring distribution ξ in terms of an excursion of the discrete-time integer-valued random walk with step distribution $\xi - 1$. In fact there are different ways to implement the same construction, which differ according to how one chooses to order vertices in the tree, and the two common choices are the *depth-first* and the *breadth-first* orders. In section 3 we give a construction of a random n -vertex \mathbf{p} -tree, based on using n i.i.d. uniform(0, 1) random variables to define an excursion-type function with drift rate -1 and with n upward jumps, and again there are two ways to implement the construction depending on choice of vertex order. These constructions seem similar in spirit to, but not exactly the same as, those used in the *server system* construction in [13] or the *parking process* construction in Chassaing and Louchard [16]. When $\theta \in \Theta_{\text{finite}}$, by analyzing asymptotics of the (appropriately rescaled) discrete excursion using depth-first order, in the asymptotic regime where convergence to the ICRT holds, we get weak convergence to the process Y^θ , and we show that this discrete excursion asymptotically agrees with $\theta_0^2/2$ times the discrete exploration process; we extend this to the case $\sum_i \theta_i < \infty$ by approximating the tree \mathcal{T}^θ by the tree \mathcal{T}^{θ^n} associated to the truncated sequence $(\theta_1, \dots, \theta_n, 0, \dots)$, and that is the proof of Theorem 1. It is a curious feature of the convergence of approximating \mathbf{p} -trees to \mathcal{T}^θ that the rescaled discrete approximation process converges to $\frac{2}{\theta_0^2} Y^\theta$ for a topology which is weaker than the usual Skorokhod topology. In the course of proving Theorem 1, we will give sufficient conditions for this stronger convergence to happen.

For any continuum tree with mass measure μ , we can define

$$\bar{W}(h) = \mu\{x : \text{ht}(x) \leq h\}, \quad h \geq 0$$

where the *height* $\text{ht}(x)$ of point x is just its distance to the root. If $\bar{W}(h) = \int_0^h W(y) dy$, $h \geq 0$ then $W(y)$ is the “width” or “height profile” of the tree (analogous to the size of a particular generation in a branching process model). The time-changed function $(W(\bar{W}^{-1}(u)), 0 \leq u \leq 1)$ can be roughly interpreted as the width of the layer of the tree containing vertex u , where vertices are labelled by $[0, 1]$ in breadth-first order. Parallel to (but simpler than) the proof of Theorem 1 sketched above, we show that excursions coding \mathbf{p} -trees using *breadth-first* order converge to X^θ , and agree asymptotically with the height profile (sizes of successive generations) of the \mathbf{p} -tree. In other words

Theorem 2. *Let $\theta \in \Theta$. For the ICRT \mathcal{T}^θ the width process $W(y) = W^\theta(y)$ exists, and*

$$(W^\theta((\bar{W}^\theta)^{-1}(u)), 0 \leq u \leq 1) \stackrel{d}{=} (X^\theta(u), 0 \leq u \leq 1).$$

Qualitatively, in breadth-first traversal of the ICRT, when we encounter a hub at some $0 < u < 1$ we expect the time-changed width function $W(\bar{W}^{-1}(u))$ to jump by an amount representing a “local time” measuring relative numbers of edges at that hub. Theorem 2 shows these jump amounts are precisely the θ -values of the hubs.

When $\sum_i \theta_i < \infty$, combining Theorems 1 and 2 gives a result whose statement does not involve trees:

Corollary 1. *Let $\theta \in \Theta$ satisfy $\sum_i \theta_i < \infty$. The process $\frac{2}{\theta_0^2} Y^\theta$ has an occupation density $(W^\theta(y), 0 \leq y < \infty)$ satisfying*

$$(W^\theta((\bar{W}^\theta)^{-1}(u)), 0 \leq u \leq 1) \stackrel{d}{=} (X^\theta(u), 0 \leq u \leq 1).$$

Note that the ‘‘Lamperti-type’’ relation between W^θ and X^θ is easily inverted as

$$(X_{L^{-1}(y)}^\theta, y \geq 0) \stackrel{d}{=} (W^\theta(y), y \geq 0), \quad (3)$$

where

$$L(t) := \int_0^t \frac{ds}{X_s^\theta} \in [0, \infty], \quad 0 \leq t \leq 1.$$

This provides a generalization of the following result of Jeulin [19] (see also Biane-Yor [14]), which from our viewpoint is the Brownian CRT case where $\theta_0 = 1$. Let $(l_u, 0 \leq u < \infty)$ be occupation density for $(B_s^{\text{exc}}, 0 \leq s \leq 1)$. Then

$$(\frac{1}{2}l_{u/2}, 0 \leq u < \infty) \stackrel{d}{=} (B_{L^{-1}(u)}^{\text{exc}}, 0 \leq u < \infty)$$

where $L(t) := \int_0^t \frac{1}{B_s^{\text{exc}}} ds$. One might not have suspected a possible generalization of this identity to jump processes without the interpretation provided by the ICRT.

Theorem 2 has the following other corollary:

Corollary 2. *For any $\theta \in \Theta$, the height $\sup_{v \in \mathcal{T}^\theta} \text{ht}(v)$ of the ICRT \mathcal{T}^θ has the same law as*

$$\int_0^1 \frac{ds}{X_s^\theta}.$$

1.2 Discussion

As formulated above, the purpose of this paper is to prove Theorems 1 and 2 concerning the ICRT. But we have further motivation. As ingredients of the proof, we take a known result (Proposition 1) on weak convergence of random \mathbf{p} -trees to the ICRT, and improve it to stronger and more informative versions (Propositions 2 and 3). The Theorems and these ingredients will be used in a sequel [5] studying asymptotics of random \mathbf{p} -mappings. By using Joyal’s bijection between mappings and trees, one can in a sense reduce questions of convergence of \mathbf{p} -mappings to convergence of random \mathbf{p} -trees. In particular, under a *uniform asymptotic negligibility* hypothesis which implies that the exploration process of \mathbf{p} -trees converges to Brownian excursion, one can use a ‘‘continuum Joyal functional’’ (which takes Brownian excursion to reflecting Brownian motion) to show [4] that the exploration process of the random \mathbf{p} -mappings converges to reflecting Brownian bridge. The results of the present paper give the limit exploration process Y^θ for more general sequences of \mathbf{p} -trees, and to deduce convergence of the associated random \mathbf{p} -mappings we need to understand how the continuum Joyal functional acts on Y^θ . This is the subject of the sequel [5].

2 Constructing X^θ and Y^θ

Let $\theta \in \Theta$, and consider a standard Brownian bridge B^{br} , and independent uniformly distributed random variables $(U_i, i \geq 1)$ in $[0, 1]$, independent of B^{br} . Define

$$X_t^{\text{br},\theta} = \theta_0 B_t^{\text{br}} + \sum_{i=1}^{\infty} \theta_i (\mathbb{1}_{\{U_i \leq t\}} - t), \quad 0 \leq t \leq 1. \quad (4)$$

From Kallenberg [20], the sum on the right converges a.s. uniformly on $[0, 1]$. Then $X^{\text{br},\theta}$ has exchangeable increments and infinite variation, and by Knight [23] and Bertoin [13] it attains its overall minimum at a unique location t_{\min} , which is a continuity point of $X^{\text{br},\theta}$. Consider the Vervaat transform X^θ of $X^{\text{br},\theta}$, defined by

$$X_t^\theta = X_{t+t_{\min}}^{\text{br},\theta} - X_{t_{\min}}^{\text{br},\theta}, \quad 0 \leq t \leq 1, \quad (5)$$

where the addition is modulo 1. Then X^θ is an excursion-type process with infinite variation, and a countable number of upward jumps with magnitudes equal to $(\theta_i, i \geq 1)$. See Figure 1. Write $t_j = U_j - t_{\min} \pmod{1}$ for the location of the jump with size θ_j in X^θ .

For each $j \geq 1$ such that $\theta_j > 0$, write $T_j = \inf\{s > t_j : X_s^\theta = X_{t_j-}^\theta\}$, which exists because the process X has no negative jumps. Notice that if for some $i \neq j$ one has $t_j \in (t_i, T_i)$, then one also has $T_j \in (t_i, T_i)$, so the intervals (t_i, T_i) are nested. Given a sample path of X^θ , for $0 \leq u \leq 1$ and $i \geq 1$ such that $\theta_i > 0$, let

$$R_i^\theta(u) = \begin{cases} \inf_{t_i \leq s \leq u} X_s^\theta - X_{t_i-}^\theta & \text{if } u \in [t_i, T_i] \\ 0 & \text{else.} \end{cases} \quad (6)$$

If $\theta_i = 0$ then let R_i^θ be the null process on $[0, 1]$. We then set

$$Y^\theta = X^\theta - \sum_{i \geq 1} R_i^\theta, \quad (7)$$

which is defined as the pointwise decreasing limit of $X^\theta - \sum_{1 \leq i \leq n} R_i^\theta$ as $n \rightarrow \infty$. See Figure 2. It is immediate that Y^θ is a non-negative process on $[0, 1]$. More precisely, for any $0 \leq u \leq s \leq 1$ and i such that $u \geq t_i$, $R_i^\theta(u)$ is equal to the magnitude of the jump (if any) accomplished at time t_i by the increasing process

$$\overleftarrow{X}_s^\theta(u) = \inf_{u \leq r \leq s} X_r^\theta, \quad 0 \leq u \leq s.$$

Since the Lebesgue measure of the range of an increasing function $(f(s), 0 \leq s \leq t)$ is $f(t) - f(0)$ minus the sum of sizes of jumps accomplished by f , we obtain that

$$Y_s^\theta = m\{\overleftarrow{X}_s^\theta(u) : 0 \leq u \leq s\} \quad 0 \leq s \leq 1, \quad (8)$$

where m is Lebesgue measure. This easily implies that Y^θ is a continuous (possibly null) process, and since the largest jump of $X^\theta - \sum_{1 \leq i \leq n} R_i^\theta$ is θ_{n+1} , which tends to 0 as $n \rightarrow \infty$, a variation of Dini's theorem implies that (7) holds in the sense of uniform convergence.

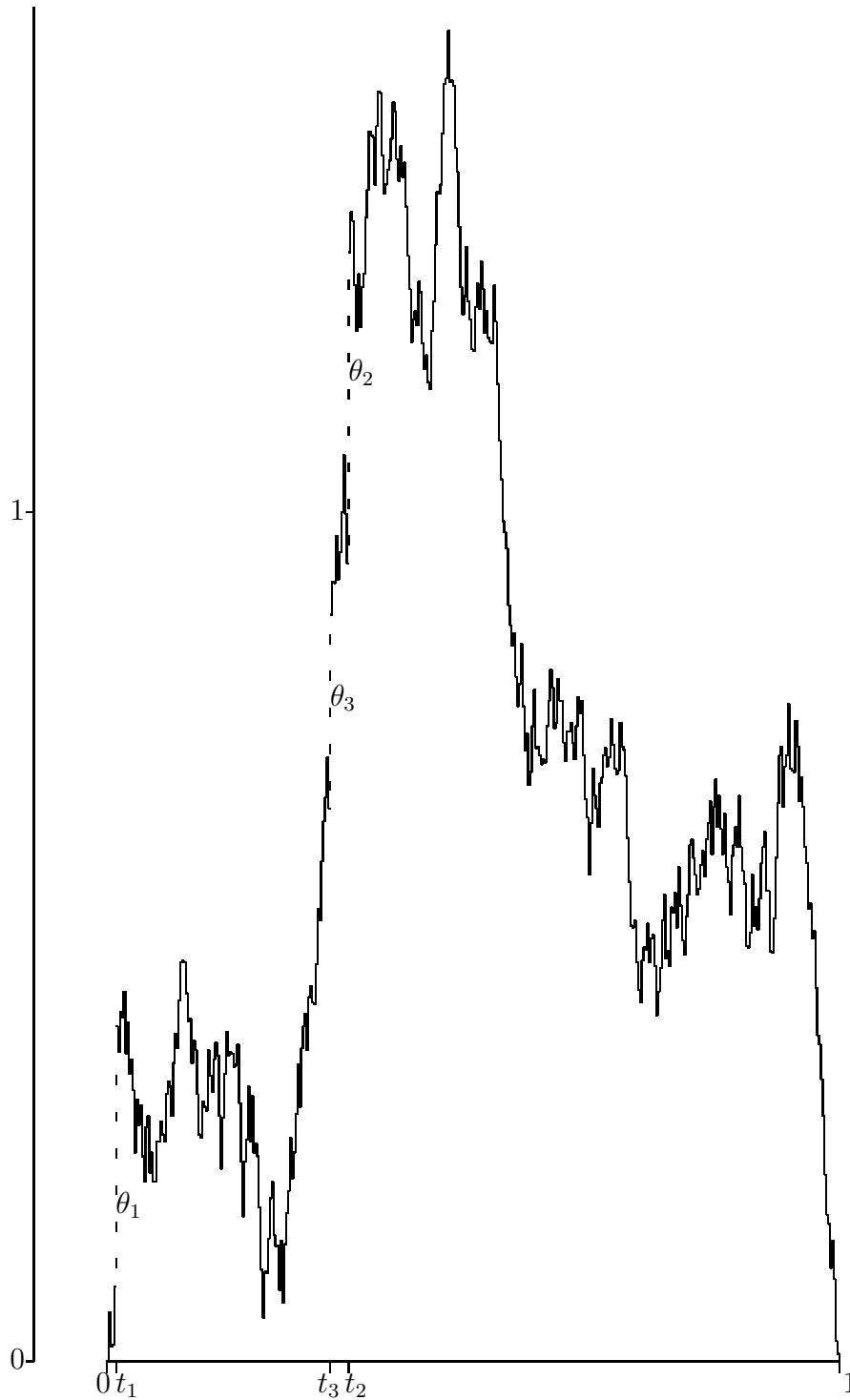


Figure 1: A realization of $(X_s^\theta, 0 \leq s \leq 1)$ with $I = 3$ and $(\theta_0, \theta_1, \theta_2, \theta_3) = (0.862, 0.345, 0.302, 0.216)$ ($I = 3$). The jumps are marked with dashed lines; the jump of height θ_i occurs at time t_i .

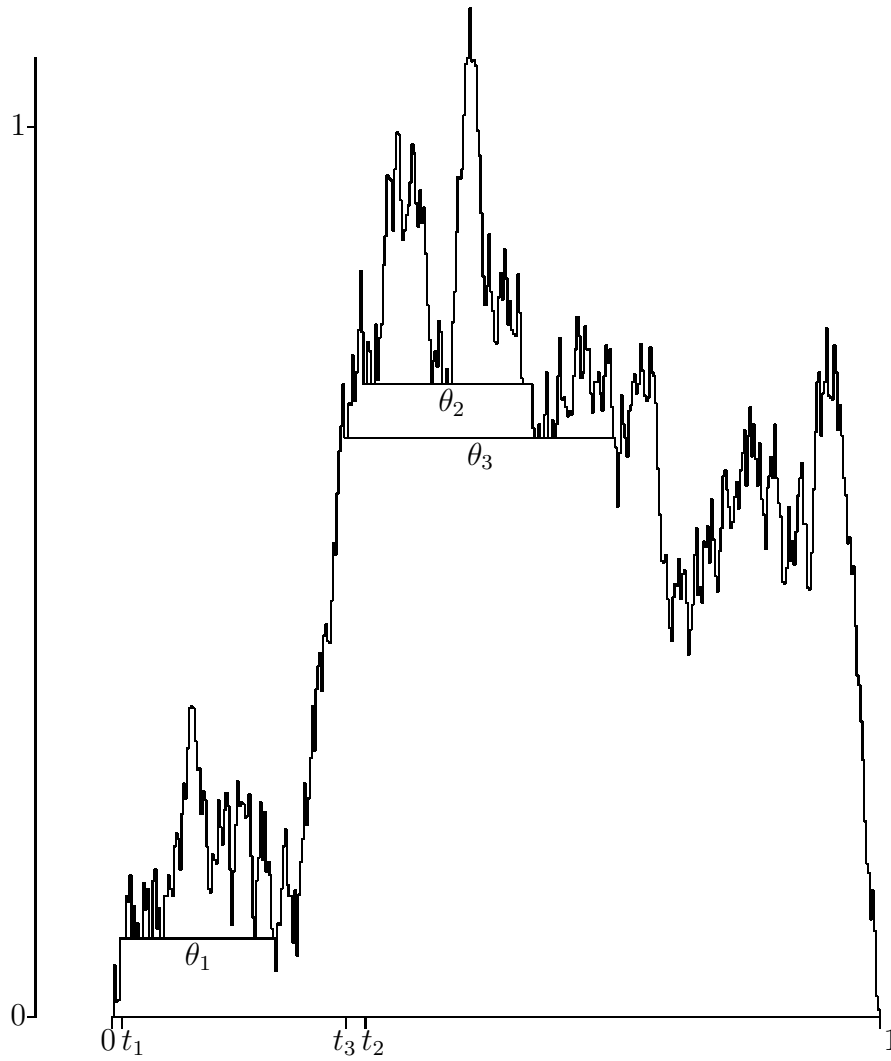


Figure 2: The process $(Y_s^\theta, 0 \leq s \leq 1)$ constructed from the process $(X_s^\theta, 0 \leq s \leq 1)$ in Figure 1. The “reflecting” portions of the path corresponding to jumps of X^θ are marked by the θ_i

The process Y^θ is an excursion-type process on $(0, 1)$. Moreover, since by classical properties of Brownian bridges the local infima of $X^{\text{br},\theta}$ are all distinct, the only local infima that Y^θ attains an infinite number of times are in the intervals $[t_i, T_i]$. Let us record some other sample path properties of Y^θ .

Lemma 1. *Suppose $\theta \in \Theta$ has length $I \in \mathbb{N} \cup \{\infty\}$ and $\theta_0 > 0$. Almost surely, the values $(X_{t_i-}^\theta, i \geq 1)$ taken by X^θ at its jump times are not attained at local minima of X^θ . Also, the times t_i are a.s. not right-minima of X^θ in the sense that there does not exist $\varepsilon > 0$ such that $X_s^\theta \geq X_{t_i}^\theta$ for $s \in [t_i, t_i + \varepsilon]$.*

Proof. Let $X_i^{\text{br},\theta}(s) = X_s^{\text{br},\theta} - \theta_i(\mathbb{1}_{\{U_i \leq s\}} - s)$, which is independent of U_i . The shifted process $X_i^{\text{br},\theta}(\cdot + t) - X_i^{\text{br},\theta}(t)$ (with addition modulo 1) has same law as $X_i^{\text{br},\theta}$ for every t , so the fact that 1 is not the time of a local extremum for $X_i^{\text{br},\theta}$ and that $|X_i^{\text{br},\theta}(1-t)|/t \rightarrow \infty$ as $t \rightarrow 0$ (e.g. by [21, Theorem 2.2 (i)] and time-reversal) implies by adding back $\theta_i(\mathbb{1}_{\{U_i \leq \cdot\}} - \cdot)$ to $X_i^{\text{br},\theta}$ that U_i is a.s. not a local minimum of $X^{\text{br},\theta}$. The statement about right-minima is obtained similarly, using the behavior of $X_i^{\text{br},\theta}$ at 0 rather than 1.

Next, since $X^{\text{br},\theta}$ is the sum of a Brownian bridge B^{br} and an independent process, the increments of $X^{\text{br},\theta}$ have continuous densities, as does the Brownian bridge (except of course the increment $X^{\text{br},\theta}(1) - X^{\text{br},\theta}(0) = 0$ a.s.). The probability that the minimum of $X^{\text{br},\theta}$ in any interval $[a, b]$ with distinct rational bounds not containing U_i equals $X_{U_i-}^{\text{br},\theta}$ is therefore 0. This finishes the proof. \square

The following lemma will turn out to be useful at the end of the proof of Theorem 1.

Lemma 2. *Let $\theta \in \Theta$ satisfy $\sum_i \theta_i < \infty$, and write $\theta^n = (\theta_0, \theta_1, \dots, \theta_n)$. Define X^{θ^n} as above, but where the sum defining X^{br,θ^n} is truncated at n . Last, define Y^{θ^n} as in (8) with X^{θ^n} instead of X^θ . Then Y^{θ^n} converges a.s. uniformly to Y^θ as $n \rightarrow \infty$.*

Proof. We want to estimate the uniform norm $\|Y^{\theta^n} - Y^\theta\|$, which by definition is $\|X^{\theta^n} - X^\theta - \sum_{i \geq 1} (R_i^{\theta^n} - R_i^\theta)\|$ with obvious notations. The first problem is that X^{br,θ^n} may not attain its overall infimum at the same time as $X^{\text{br},\theta}$, so that jump times for X^{θ^n} and X^θ may not coincide anymore. So, rather than using X^{θ^n} we consider $X'_n(s) = X^{\text{br},\theta^n}(s + t_{\min}) - X^{\text{br},\theta^n}(t_{\min})$ (with addition modulo 1) where t_{\min} is the time at which $X^{\text{br},\theta}$ attains its infimum. Then $X'_n \rightarrow X^\theta$ uniformly. Define $R'_{n,i}$ as in (6) but for the process X'_n and write $Y'_n = X'_n - \sum_{1 \leq i \leq n} R'_{n,i}$. Notice that Y'_n is just a slight space-time shift of Y^{θ^n} , so by continuity of Y^{θ^n} and Y^θ it suffices to show that $Y'_n \rightarrow Y^\theta$ uniformly. It is thus enough to show that $\|\sum_{1 \leq i \leq n} (R'_{n,i} - R_i^\theta)\| \rightarrow 0$ as $n \rightarrow \infty$. It is easy that for each $i \geq 1$, one has uniform convergence of $R'_{n,i}$ to R_i^θ . Therefore, it suffices to show that

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \left\| \sum_{k \leq i \leq n} R'_{n,i} \right\| = 0,$$

which is trivial because $\|R'_{n,i}\| \leq \theta_i$, and $\sum_i \theta_i < \infty$ by hypothesis.

Remark. Again, one guesses that the same result holds in the general $\theta_0 > 0$ case, so that the proof of Theorem 1 should extend to this case. However, the fact that $\sum_i \theta_i$ might be infinite does not *a priori* prevent vanishing terms of the sum $\sum_{1 \leq i \leq n} R'_{n,i}$ to accumulate, so the proof might become quite technical.

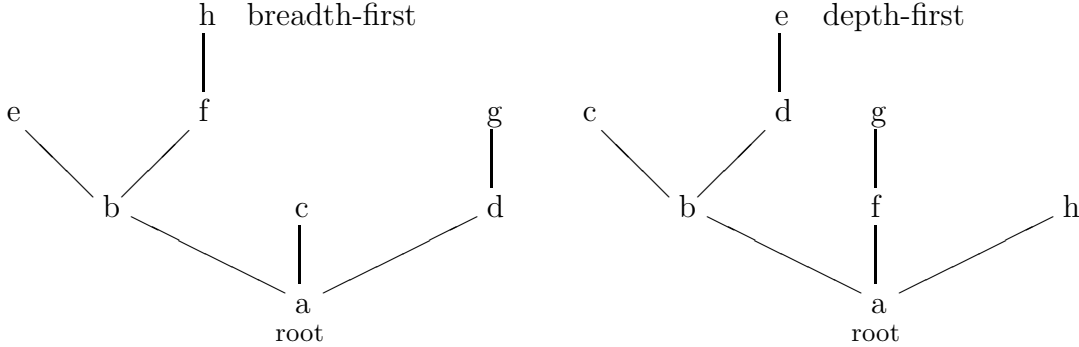


Figure 3: A planar tree, with the two orderings of vertices as a, b, c, d, e, f, g, h

3 Constructions of \mathbf{p} -trees and associated excursion processes

Write \mathbf{T}_n for the set of rooted trees \mathbf{t} on vertex-set $[n]$, where \mathbf{t} is directed towards its root. Fix a probability distribution $\mathbf{p} = (p_1, \dots, p_n)$. Recall that associated with \mathbf{p} is a certain distribution on \mathbf{T}_n , the \mathbf{p} -tree

$$P(\mathcal{T} = \mathbf{t}) = \prod_v p_v^{d_v}, \quad d_v \text{ in-degree of } v \text{ in } \mathbf{t}. \quad (9)$$

See [31] for systematic discussion of the \mathbf{p} -tree model. We shall define two maps $\psi_{\mathbf{p}} : [0, 1]^n \rightarrow \mathbf{T}_n$ such that, if (X_1, \dots, X_n) are independent $U(0, 1)$ then each $\psi_{\mathbf{p}}(X_1, \dots, X_n)$ has the distribution (9). The two definitions are quite similar, but the essential difference is that $\psi_{\mathbf{p}}^{\text{breadth}}$ uses a *breadth-first* construction whereas $\psi_{\mathbf{p}}^{\text{depth}}$ uses a *depth-first* construction.

3.1 The breadth-first construction

The construction is illustrated in Figure 4. Fix distinct $(x_1, \dots, x_n) \in [0, 1]^n$. Picture this as a configuration of particles on the circle of unit circumference, where particle i is at position x_i and has a “weight” p_i associated with it. Define

$$F^{\mathbf{p}}(u) = -u + \sum_i p_i \mathbb{1}_{\{x_i \leq u\}}, \quad 0 \leq u \leq 1. \quad (10)$$

There exists some particle v such that $F^{\mathbf{p}}(x_v -) = \inf_u F^{\mathbf{p}}(u)$: assume the particle is unique. Let $v = v_1, v_2, \dots, v_n$ be the ordering of particles according to the natural ordering of positions $x_{v_1} < x_{v_2} < \dots$ around the circumference of the circle. (In Figure 4 we have $v_1 = 4$ and the ordering is 4, 8, 2, 3, 7, 1, 5, 6). Write $y(1) = x_{v_1}$ and for $2 \leq j \leq n$ let $y(j+1) = y(j) + p_{v_j} \pmod{1}$. So $y(n+1) = y(1)$ and the successive intervals $[y(j), y(j+1)]$, $1 \leq j \leq n$ are adjacent and cover the circle. We assert

$$x_{v_j} \in [y(1), y(j)), \quad 2 \leq j \leq n. \quad (11)$$

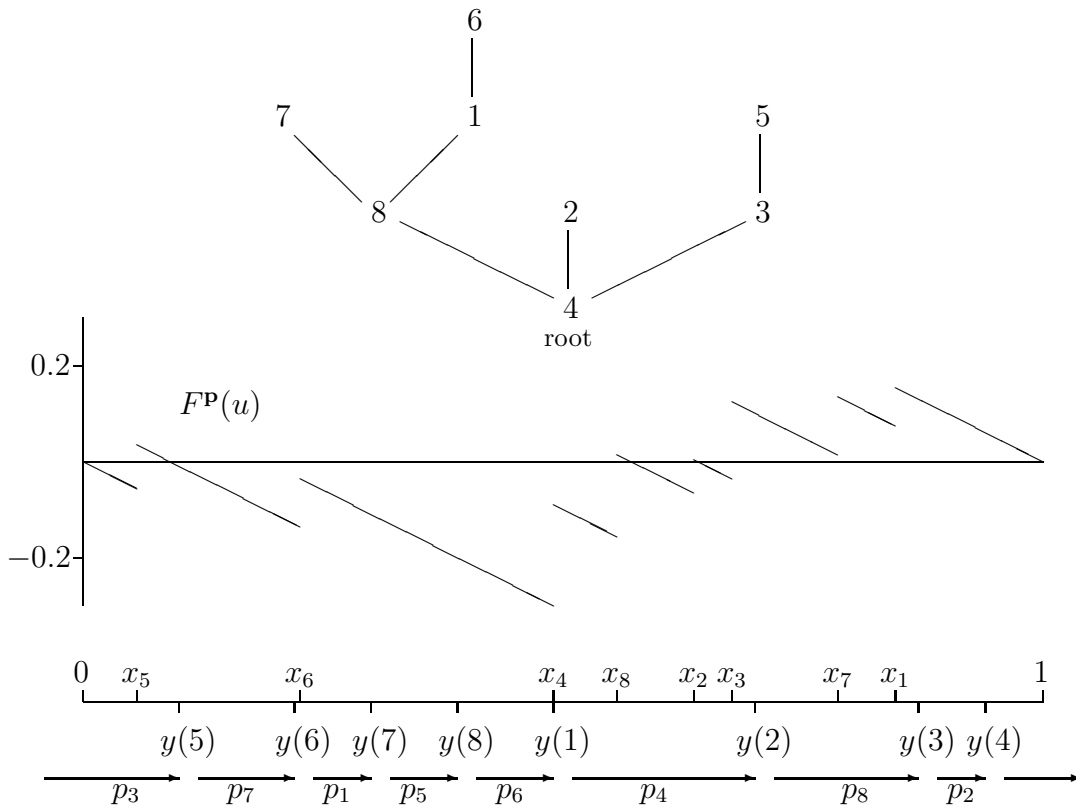


Figure 4: The construction of the tree $\psi_{\mathbf{p}}^{\text{breadth}}(x_1, \dots, x_8)$

To argue by contradiction, suppose this fails first for j . Then $[y(1), y(j))$, interpreted mod 1, contains particles v_1, \dots, v_{j-1} only. Since $y(j) - y(1) = p_{v_1} + \dots + p_{v_{j-1}}$ this implies $F^{\mathbf{P}}(y(j)-) = F^{\mathbf{P}}(y(1)-)$, contradicting uniqueness of the minimum.

We specify the tree $\psi_{\mathbf{p}}^{\text{breadth}}(x_1, \dots, x_n)$ by:

v_1 is the root

the children of v_j are the particles v with $x_v \in (y(j), y(j+1))$.

By (11), any child v_k of v_j has $k > j$, so the graph cannot contain a cycle. If it were a forest and not a single tree, then the component containing the root v_1 would consist of vertices v_1, \dots, v_j for some $j < n$. Then the interval $[y(1), y(j+1)]$ would contain only the particles v_1, \dots, v_j , contradicting (11) for $j+1$.

Thus the construction does indeed give a tree. From the viewpoint of this construction it would be natural to regard the tree as *planar* (or *ordered*: the d_v children of v are distinguished as first, second, etc) but we disregard order and view trees in \mathbf{T}_n as unordered.

Now consider the case where $(x_1, \dots, x_n) = (X_1, \dots, X_n)$ are independent $U(0, 1)$. Fix an unordered tree \mathbf{t} and write v_1 for its root. Fix an arbitrary $x_{v_1} \in (0, 1)$ and condition on $X_{v_1} = x_{v_1}$. Consider the chance that the construction yields the particular tree \mathbf{t} . For this to happen, the particles corresponding to the d_{v_1} children of v_1 must fall into the interval $[x_{v_1}, x_{v_1} + p_{v_1}]$, which has chance $p_{v_1}^{d_{v_1}}$. Inductively, for each vertex v an interval of length p_v is specified and it is required that d_v specified particles fall into that interval, which has chance $p_v^{d_v}$. So the conditional probability of constructing \mathbf{t} is indeed the probability in (9), and hence so is the unconditional probability.

Remark. Note that in the argument above we do not start by conditioning on $F^{\mathbf{P}}$ having its minimum at x_{v_1} , which would affect the distribution of the (X_i) .

We now derive an interpretation (13,14) of the function $F^{\mathbf{P}}$ at (10), which will be used in the asymptotic setting later. From now on we also suppose that for $j \geq 2$, $y(j)$ is not a jump time for $F^{\mathbf{P}}$ to avoid needing the distinction between $F^{\mathbf{P}}(y(j))$ and $F^{\mathbf{P}}(y(j)-)$; this is obviously true a.s. when the jump times are independent uniform, which will be the relevant case.

For $2 \leq j \leq n$, vertex v_j has some parent $v_{z(j)}$, where $1 \leq z(j) < j$. By induction on j ,

$$F^{\mathbf{P}}(y(j)) - F^{\mathbf{P}}(y(1)-) = \sum_{i: i > j, z(i) \leq j} p_{v_i}.$$

In words, regarding \mathbf{t} as ordered, the sum is over vertices i which are in the same generation as j but *later* than j ; and over vertices i in the next generation whose parents are *before* j or are j itself. For $h \geq 1$, write $t(h)$ for the number of vertices at height $\leq h-1$. The identity above implies

$$F^{\mathbf{P}}(y(t(h)+1)) - F^{\mathbf{P}}(y(1)-) = \sum_{v: \text{ht}(v)=h} p_v.$$

Also by construction

$$y(t(h)+1) - y(1) \bmod 1 = \sum_{v: \text{ht}(v) \leq h-1} p_v.$$

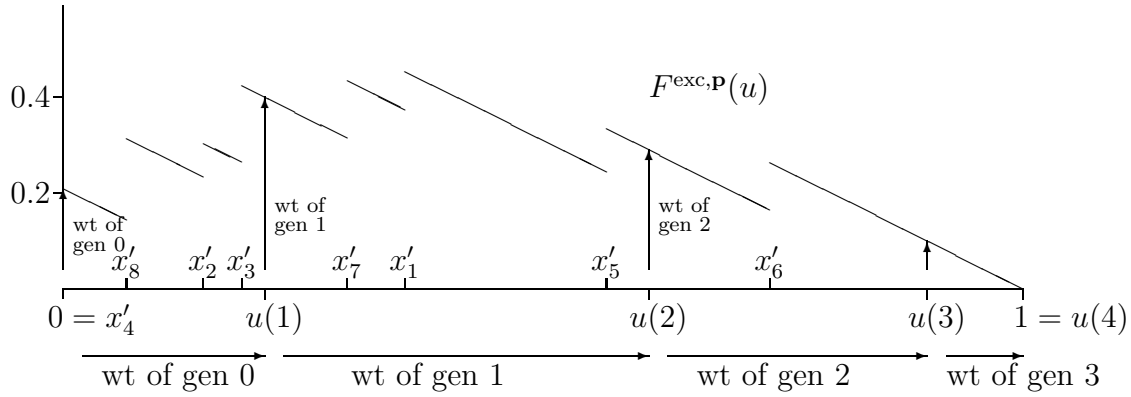


Figure 5: $F^{\text{exc}, \mathbf{P}}(\cdot)$ codes the weights of successive generations (wt of gen) of the \mathbf{p} -tree in Figure 4

We can rephrase the last two inequalities in terms of the “excursion” function

$$F^{\text{exc}, \mathbf{P}}(u) := F^{\mathbf{P}}(y(1) + u \bmod 1) - F^{\mathbf{P}}(y(1)-), \quad 0 \leq u \leq 1 \quad (12)$$

and of $u(h) := y(t(h) + 1) - y(1) \bmod 1$. Then

$$u(h) = \sum_{v: \text{ht}(v) \leq h-1} p_v \quad (13)$$

$$F^{\text{exc}, \mathbf{P}}(u(h)) = \sum_{v: \text{ht}(v)=h} p_v. \quad (14)$$

So the weights of successive generations are coded within $F^{\text{exc}, \mathbf{P}}(\cdot)$, as illustrated in Figure 5. Note that to draw Figure 5 we replace x_i by

$$x'_i := x_i - y(1) \bmod 1.$$

Remark. There is a queuing system interpretation to the breadth-first construction, which was pointed out to us by a referee. In this interpretation, the customer labelled i arrives at time x'_i and requires a total service time p_i . If customers are served according to the FIFO rule (first-in first-out) then $F^{\text{exc}, \mathbf{P}}(u)$ is the remaining amount of time needed to serve the customers in line at time u .

3.2 The depth-first construction

The construction is illustrated in Figure 6, using the same (x_i) and (p_i) as before, and hence the same $F^{\mathbf{P}}(u)$. In the previous construction we “examined” particles in the order v_1, v_2, \dots, v_n ; we defined $y(1) = v_1$ and inductively

- $y(j+1) = y(j) + p_{v_j} \bmod 1$

- the children of v_j are the particles v with $x_v \in (y(j), y(j+1))$.

In the present construction we shall examine particles in a different order w_1, w_2, \dots, w_n and use different $y'(j)$ to specify the intervals which determine the offspring of a parent. Start as before with $w_1 = v_1$ and $y'(1) = x_{w_1}$. Inductively set

- $y'(j+1) = y'(j) + p_{w_j} \bmod 1$
- the children of w_j are the particles v with $x_v \in (y'(j), y'(j+1))$.
- w_{j+1} is
 - the first child of w_j , if any; else
 - the next unexamined child of parent(w_j), if any; else
 - the next unexamined child of parent(parent(w_j)), if any; else
 - and so on.

Here “unexamined” means “not one of w_1, \dots, w_j ” and “next” uses the natural order of children of the same parent.

Figure 6 and the following paragraph talk through the construction in a particular example, using the same (x_i) and (p_i) as in Figure 4. Checking that $\psi_{\mathbf{p}}^{\text{depth}}(X_1, \dots, X_n)$ has distribution (9), i.e. is a random \mathbf{p} -tree, uses exactly the same argument as before.

As in Figure 4, the root of the tree is vertex 4 ($w_1 = 4$), and we set $y'(1) = x_4$. As before, $y'(2) = y'(1) + p_4$, and the children of the root are the vertices $\{8, 2, 3\}$ for which $x_v \in (y'(1), y'(2))$. As before, we next examine the first child $w_2 = 8$ of the root, set $y'(3) = y'(2) + p_8$, and let the children of 8 be the vertices $\{7, 1\}$ for which $x_v \in (y'(2), y'(3))$. At this stage the constructions differ. We next examine vertex 7, being the first child of vertex 8, by setting $y'(4) = y'(3) + p_7$; the children of vertex 8 are the vertices v with $x_v \in (y'(3), y'(4))$, and it turns out there are no such vertices. We continue examining vertices in the depth-first order 4, 8, 7, 1, 5, 2, 6, 3.

As with the breadth-first construction, the point of the depth-first construction is that the excursion function $F^{\text{exc}, \mathbf{p}}(\cdot)$ tells us something about the distribution of the tree. For each vertex v of $\psi_{\mathbf{p}}^{\text{depth}}(x_1, \dots, x_n)$ there is a path $\text{root} = y_0, y_1, \dots, y_j = v$ from the root to v . For each $0 \leq i < j$ the vertex y_{i+1} is a child of vertex y_i ; let $y_{i,1}, y_{i,2}, \dots$ be the later children of y_i , and let $y_{j,1}, y_{j,2}, \dots$ be all children of v . Write $\mathcal{N}(v) = \cup_{0 \leq i < j} \{y_{i,1}, y_{i,2}, \dots\}$.

In the u -scale of $F^{\text{exc}, \mathbf{p}}(u)$, we finish “examining” vertex w_i at time $y^*(i) := y'(i) - y'(1)$. For vertex $v = w_i$ set $e(v) = y^*(i)$. Then the relevant property of $F^{\text{exc}, \mathbf{p}}$ is

$$F^{\text{exc}, \mathbf{p}}(e(v)) = \sum_{w \in \mathcal{N}(v)} p_w, \quad \forall v. \quad (15)$$

See Figure 7 for illustration. As before, in Figure 7 the position of the jump of height p_i is moved from x_i to $x'_i := x_i - y'(1) \bmod 1$. At first sight, relation (15) may not look useful. But we shall see in section 6.2 that in the asymptotic regime the right side of (15) can be related to $\sum_w \text{ancestor of } v p_w$ which in turn relates to the height of v .

Remark. We might alternatively have defined the tree $\psi_{\mathbf{p}}^{\text{depth}}(x_1, \dots, x_n)$ in a way that would have been less suited for the forthcoming analysis, but which is worth mentioning. It is based on the LIFO-queuing system construction of Galton-Watson trees in Le Gall-Le Jan [26] which

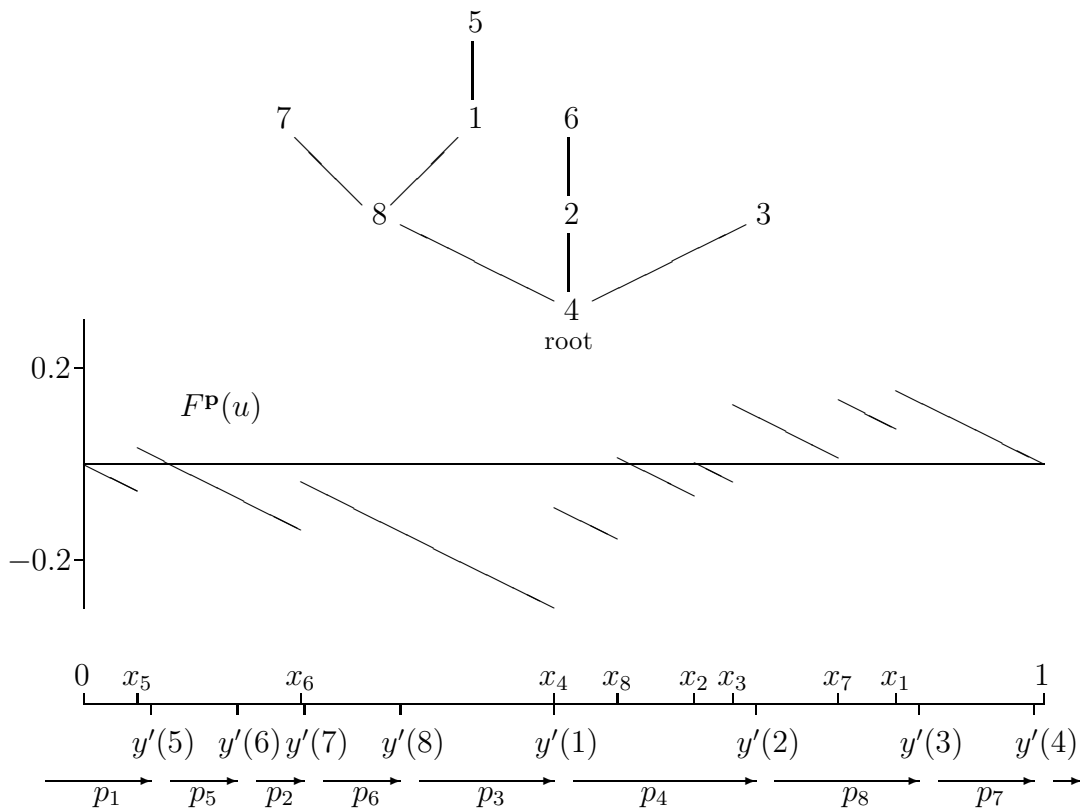


Figure 6: The construction of the tree $\psi_{\mathbf{p}}^{\text{depth}}(x_1, \dots, x_n)$

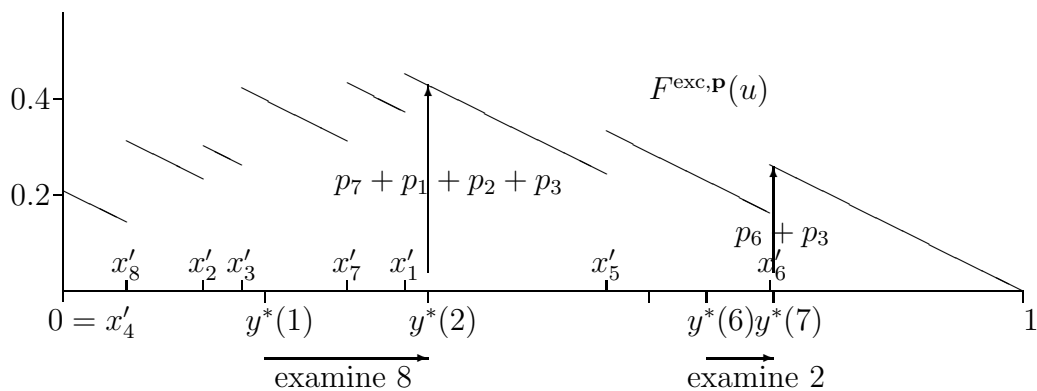


Figure 7: Relation (15) in the depth-first construction

we sketch here. Imagine vertex i is a customer in a line which requires a treatment time p_i . The customer i arrives at time x_i and customers are treated according to the Last In First Out rule. After relocating the the time-origin is at the time when the minimum of the bridge $F^{\mathbf{p}}$ is attained, the first customer in line will also be the last to get out. Then we say that vertex i is a parent of vertex j if customer j arrives in a time-interval when i was being treated. Notice that the tree thus defined is in general different from $\psi_{\mathbf{p}}^{\text{depth}}(x_1, \dots, x_n)$.

It is easy to see, using induction and the same kind of arguments as above, that taking x_1, \dots, x_n to be independent uniform random variables builds a \mathbf{p} -tree (in order that i has k children, k uniform random variables must interrupt the service of i which takes total time p_i , so this has probability p_i^k). It is also easy that the order of customer arrivals (after relocating the time origin) corresponds to the depth-first order on the tree. In particular, the cyclic depth-first random order of vertices in a \mathbf{p} -tree is the uniform cyclic order on the n vertices.

4 Convergence of \mathbf{p} -trees to the ICRT

Here we review known results concerning convergence of \mathbf{p} -trees to the ICRT, and spotlight what new results are required to prove Theorems 1 and 2.

The general notion (1) of exploration process of a continuum random tree can be reinterpreted as follows. Fix $J \geq 1$. Let $(U_j, 1 \leq j \leq J)$ be independent $U(0, 1)$ r.v.s and let $U_{(1)} < U_{(2)} < \dots < U_{(J)}$ be their order statistics. To an excursion-type process $(H_s, 0 \leq s \leq 1)$ associate the random $2J - 1$ -vector

$$\left(H_{U_{(1)}}, \inf_{U_{(1)} \leq s \leq U_{(2)}} H_s, H_{U_{(2)}}, \inf_{U_{(2)} \leq s \leq U_{(3)}} H_s, \dots, H_{U_{(J)}} \right). \quad (16)$$

This specifies a random tree-with-edge-lengths, with J leaves, as follows.

- The path from the root to the i 'th leaf has length $H_{U_{(i)}}$.
- The paths from the root to the i 'th leaf and from the root to the $(i + 1)$ 'st leaf have their branchpoint at distance $\inf_{U_{(i)} \leq s \leq U_{(i+1)}} H_s$.

Now label the i 'th leaf as vertex i' , where $U_{(i)} = U_{i'}$. Write the resulting tree as \mathcal{T}_J^H . Call this the *sampling a function* construction.

On the other hand one can use a continuum random tree \mathcal{T} to define a random tree-with-edge-lengths \mathcal{T}_J as follows.

- Take a realization of \mathcal{T} .
- From the mass measure on that realization, pick independently J points and label them as $\{1, 2, \dots, J\}$.
- Construct the spanning tree on those J points and the root; this is the realization of \mathcal{T}_J .

Call this the *sampling a CRT* construction.

As discussed in detail in [3], the relationship

the exploration process of \mathcal{T} is distributed as $(H_s, 0 \leq s \leq 1)$

is equivalent to

$$\mathcal{T}_J \stackrel{d}{=} \mathcal{T}_J^H, \quad \forall J \geq 1,$$

(the background hypotheses in [3] were rather different, assuming path-continuity for instance, but the ideas go through to our setting.) In our setting, there is an explicit description of the distribution of the spanning tree \mathcal{T}_J^θ derived from the ICRT \mathcal{T}^θ (see [8]), so to prove Theorem 1 it is enough to verify

$$\mathcal{T}_J^\theta \stackrel{d}{=} \mathcal{T}_J^{2Y/\theta_0^2}, \quad \forall J \geq 1 \quad (17)$$

for $Y = Y^\theta$ defined at (7). In principle one might verify (17) directly, but this seems difficult even in the case $J = 1$. Instead we shall rely on weak convergence arguments, starting with the known Proposition 1 below.

Consider a probability distribution $\mathbf{p} = (p_1, \dots, p_n)$ which is *ranked*: $p_1 \geq p_2 \geq \dots \geq p_n > 0$. In the associated \mathbf{p} -tree (9), pick J vertices independently from distribution \mathbf{p} , label them as $[J]$ in order of pick, take the spanning tree on the root and these J vertices, regard each edge as having length 1, and then delete degree-2 vertices to form edges of positive integer length. Call the resulting random tree $\mathcal{S}_J^\mathbf{p}$. Define $\sigma(\mathbf{p}) := \sqrt{\sum_i p_i^2}$. Now consider a sequence $\mathbf{p}_n = (p_{ni})$ of ranked probability distributions which satisfy

$$\lim_n \sigma(\mathbf{p}_n) = 0; \quad \lim_n p_{ni}/\sigma(\mathbf{p}_n) = \theta_i, \quad 1 \leq i \leq I; \quad \lim_n p_{ni}/\sigma(\mathbf{p}_n) = 0, \quad i > I \quad (18)$$

for some limit $\theta = (\theta_0, \dots, \theta_I) \in \Theta_{\text{finite}}$. For a tree \mathbf{t} and a real constant $\sigma > 0$ define $\sigma \otimes \mathbf{t}$ to be the tree obtained from \mathbf{t} by multiplying edge-lengths by σ . The following result summarizes Propositions 2, 3 and 5(b) of [9]. Recall \mathcal{T}_J^θ is obtained by sampling the ICRT \mathcal{T}^θ .

Proposition 1. *For a sequence $\mathbf{p} = \mathbf{p}_n$ satisfying (18), as $n \rightarrow \infty$*

$$\sigma(\mathbf{p}) \otimes \mathcal{S}_J^\mathbf{p} \xrightarrow{d} \mathcal{T}_J^\theta, \quad J \geq 1.$$

The tree $\mathcal{S}_J^\mathbf{p}$ may not be well-defined because two of the J sampled vertices may be the same; but part of Proposition 1 is that this probability tends to zero.

Now consider the “bridge” process $F^\mathbf{p}$ at (10), where from now on the jump times x_1, \dots, x_n are uniformly distributed independent random variables. Standard results going back to Kallenberg [20] show that, under the asymptotic regime (18),

$$(\sigma^{-1}(\mathbf{p})F^\mathbf{p}(s), \quad 0 \leq s \leq 1) \xrightarrow{d} (X_s^{\text{br},\theta}, \quad 0 \leq s \leq 1),$$

where $X^{\text{br},\theta}$ is defined at (4). It follows by an argument that can be found e.g. in [13] (using the continuity of the bridge process at its minimum) that the associated excursion process $F^{\text{exc},\mathbf{p}}$ at (12) satisfies

$$(\sigma^{-1}(\mathbf{p})F^{\text{exc},\mathbf{p}}(s), \quad 0 \leq s \leq 1) \xrightarrow{d} (X_s^\theta, \quad 0 \leq s \leq 1) \quad (19)$$

for X^θ defined at (5).

Recall from section 2 how (Y_s^θ) is constructed as a modification of (X_s^θ) . We next describe a parallel modification of $F^{\text{exc},\mathbf{p}}$ to construct a process $G_J^\mathbf{p}$. Given a realization of the \mathbf{p} -tree obtained via the depth-first construction illustrated in Figure 7, and given $I \geq 0$, let $\mathcal{B}_i \subseteq [n]$ be the set of vertices which are the child of some vertex i in from $\{1, \dots, I\}$. In the setting of

the depth-first construction of the \mathbf{p} -tree from $F^{\text{exc}, \mathbf{p}}$, illustrated in Figure 7, for every vertex $v \in \mathcal{B}_i$, define

$$\begin{aligned} \rho_v(u) &= 0 && 0 \leq u \leq x'_i \\ &= p_v && x'_i < u \leq e(v) - p_v \\ &= e(v) - u && e(v) - p_v \leq u \leq e(v) \\ &= 0 && e(v) \leq u \leq 1. \end{aligned} \tag{20}$$

and then let

$$r_i^{\mathbf{p}}(u) = \sum_{v \in \mathcal{B}_i} \rho_v(u) \tag{21}$$

and

$$G_I^{\mathbf{p}}(u) = F^{\text{exc}, \mathbf{p}}(u) - \sum_{i=1}^I r_i^{\mathbf{p}}(u). \tag{22}$$

We will show in section 6.1 that (19) extends to

Proposition 2. *For a sequence $\mathbf{p} = \mathbf{p}_n$ satisfying (18) with limit $\boldsymbol{\theta} = (\theta_0, \dots, \theta_I)$, as $n \rightarrow \infty$*

$$(\sigma^{-1}(\mathbf{p})G_I^{\mathbf{p}}(u), 0 \leq u \leq 1) \xrightarrow{d} (Y^{\boldsymbol{\theta}}(u), 0 \leq u \leq 1)$$

for $Y^{\boldsymbol{\theta}}$ defined at (7), for the Skorokhod topology.

We finally come to the key issue; we want to show that $G_I^{\mathbf{p}}(\cdot)$ approximates the (discrete) exploration process. In the depth-first construction of the \mathbf{p} -tree \mathcal{T} from $F^{\text{exc}, \mathbf{p}}$, we examine vertex w_i during $(y^*(i-1), y^*(i)]$. Define

$$H^{\mathbf{p}}(u) := \text{height of } w_i \text{ in } \mathcal{T} ; \quad u \in (y^*(i-1), y^*(i)]. \tag{23}$$

Roughly, we show that *realizations* of $\frac{\theta_0^2}{2}\sigma(\mathbf{p})H^{\mathbf{p}}(\cdot)$ and of $\sigma^{-1}(\mathbf{p})G_I^{\mathbf{p}}(\cdot)$ are close. Precisely, we will prove the following in section 6.2

Proposition 3. *Let $\boldsymbol{\theta} \in \Theta_{\text{finite}}$. There exists a sequence $\mathbf{p} = \mathbf{p}_n$ satisfying (18) with limit $\boldsymbol{\theta}$, such that as $n \rightarrow \infty$,*

$$\sup_{u \in [0,1]} \left| \frac{\theta_0^2}{2}\sigma(\mathbf{p})H^{\mathbf{p}}(u) - \sigma(\mathbf{p})^{-1}G_I^{\mathbf{p}}(u) \right| \xrightarrow{p} 0.$$

The next result, Lemma 3, relates the exploration process $H^{\mathbf{p}}$ at (23) to the spanning trees $\mathcal{S}_J^{\mathbf{p}}$. This idea was used in ([10]; proof of Proposition 7) but we say it more carefully here. Given $u^1 \in (0, 1)$ define, as in (23),

$$w^1 = w_i \text{ for } i \text{ specified by } u^1 \in (y^*(i-1), y^*(i)].$$

Given $0 < u^1 < u^2 < 1$, define w^2 similarly, and let vertex b be the branchpoint of the paths from the root to vertices w^1 and w^2 . Distinguish two cases.

Case (i): $w^1 = w^2$ or w^1 is an ancestor of w^2 . In this case $b = w^1$ and so trivially $\text{ht}(b) = \min_{u^1 \leq u \leq u^2} H^{\mathbf{p}}(u)$.

Case (ii): otherwise, b is a strict ancestor of both w^1 and w^2 . In this case we assert

$$\text{ht}(b) = \min_{u^1 \leq u \leq u^2} H^{\mathbf{p}}(u) - 1,$$

because vertex b appears, in the depth-first order, strictly before vertex w^1 . Then consider the set of vertices between w^1 and w^2 (inclusive) in the depth first order. This set contains the child w^* of b which is an ancestor of w^2 or is w^2 itself, and $\text{ht}(w^*) = \text{ht}(b) + 1$. But the set cannot contain any vertex of lesser height.

Now the length of the interval $(y^*(i-1), y^*(i)]$ equals p_{w_i} by construction. So if U^1 has uniform distribution on $(0, 1)$ then the corresponding vertex W^1 at (23) has distribution \mathbf{p} . Combining with the discussion above regarding branchpoint heights gives

Lemma 3. *Fix \mathbf{p} , make the depth-first construction of a \mathbf{p} -tree and define $H^{\mathbf{p}}$ by (23). Fix J . Take U_1, \dots, U_J independent uniform $(0, 1)$ and use them and $H^{\mathbf{p}}$ to define a tree-with-edge-lengths $\mathcal{T}_J^{\mathbf{p}}$ via the “sampling a function” construction below (16). Then this tree agrees, up to perhaps changing heights of branchpoints by 1, with a tree distributed as the tree $\mathcal{S}_J^{\mathbf{p}}$ defined above (18).*

4.1 Proof of Theorem 1

We now show how the ingredients above (of which, Proposition 2 and Proposition 3 remain to be proved later) are enough to prove Theorem 1.

Let $\mathbf{p} = \mathbf{p}_n$ satisfy (18) with limit $\boldsymbol{\theta} \in \Theta_{\text{finite}}$. Fix J and take independent U_1, \dots, U_J with uniform $(0, 1)$ distribution. Proposition 2 implies that as $n \rightarrow \infty$

$$\begin{aligned} & \sigma^{-1}(\mathbf{p}) \left(G_I^{\mathbf{p}}(U_{(1)}), \inf_{U_{(1)} \leq s \leq U_{(2)}} G_I^{\mathbf{p}}(s), G_I^{\mathbf{p}}(U_{(2)}), \inf_{U_{(2)} \leq s \leq U_{(3)}} G_I^{\mathbf{p}}(s), \dots, G_I^{\mathbf{p}}(U_{(J)}) \right) \\ & \xrightarrow{d} \left(Y^{\boldsymbol{\theta}}(U_{(1)}), \inf_{U_{(1)} \leq s \leq U_{(2)}} Y^{\boldsymbol{\theta}}(s), Y^{\boldsymbol{\theta}}(U_{(2)}), \inf_{U_{(2)} \leq s \leq U_{(3)}} Y^{\boldsymbol{\theta}}(s), \dots, Y^{\boldsymbol{\theta}}(U_{(J)}) \right). \end{aligned}$$

By making the particular choice of (\mathbf{p}_n) used in Proposition 3,

$$\begin{aligned} & \frac{1}{2} \theta_0^2 \sigma(\mathbf{p}) \left(H^{\mathbf{p}}(U_{(1)}), \inf_{U_{(1)} \leq s \leq U_{(2)}} H^{\mathbf{p}}(s), H^{\mathbf{p}}(U_{(2)}), \inf_{U_{(2)} \leq s \leq U_{(3)}} H^{\mathbf{p}}(s), \dots, H^{\mathbf{p}}(U_{(J)}) \right) \\ & \xrightarrow{d} \left(Y^{\boldsymbol{\theta}}(U_{(1)}), \inf_{U_{(1)} \leq s \leq U_{(2)}} Y^{\boldsymbol{\theta}}(s), Y^{\boldsymbol{\theta}}(U_{(2)}), \inf_{U_{(2)} \leq s \leq U_{(3)}} Y^{\boldsymbol{\theta}}(s), \dots, Y^{\boldsymbol{\theta}}(U_{(J)}) \right). \end{aligned} \quad (24)$$

Appealing to Lemma 3, this implies

$$\frac{1}{2} \theta_0^2 \sigma(\mathbf{p}) \otimes \mathcal{S}_J^{\mathbf{p}} \xrightarrow{d} \mathcal{T}_J^Y$$

where the right side denotes the tree-with-edge-lengths obtained from sampling the function $Y^{\boldsymbol{\theta}}$, and where convergence is the natural notion of convergence of shapes and edge-lengths ([10] sec. 2.1). Rescaling by a constant factor,

$$\sigma(\mathbf{p}) \otimes \mathcal{S}_J^{\mathbf{p}} \xrightarrow{d} \mathcal{T}_J^{2\theta_0^{-2}Y}.$$

But Proposition 1 showed

$$\sigma(\mathbf{p}) \otimes \mathcal{S}_J^{\mathbf{p}} \xrightarrow{d} \mathcal{T}_J^{\boldsymbol{\theta}}$$

where the right side is the random tree-with-edge-lengths obtained by sampling the ICRT \mathcal{T}^θ . So we have established (17) and thereby proved Theorem 1 in the case $\theta \in \Theta_{\text{finite}}$.

In the case $\sum_i \theta_i < \infty$, write θ^n for the truncated sequence $(\theta_0, \dots, \theta_n, 0, \dots)$, and recall from Lemma 2 that $Y^n = Y^{\theta^n}$ converges uniformly to Y^θ . By previous considerations this entails

$$\mathcal{T}_J^{2\theta_0^{-2}Y^n} \xrightarrow{d} \mathcal{T}_J^\theta$$

for every $J \geq 1$. On the other hand, we have proved that the left-hand term has the same law as $c(\theta^n)^{-1} \otimes \mathcal{T}_J^{c(\theta^n)\theta^n}$ where $c(\theta^n) = (\sum_{0 \leq i \leq n} \theta_i^2)^{-1/2}$ is the renormalization constant so that $c(\theta^n)\theta^n \in \Theta$. It thus remain to show that this converges to \mathcal{T}^θ . Plainly the term $c(\theta^n)$ converges to 1 and is unimportant. The result is then straightforward from the line-breaking construction of the ICRT: \mathcal{T}_J^θ can be build out of the first (at most) $2J$ points (cutpoints and their respective joinpoints) of the superimposition of infinitely many Poisson point processes on the line $(0, \infty)$. It is easily checked that taking only the superimposition of the n first Poisson processes allows us to construct jointly a reduced tree with same law as $c(\theta^n)^{-1} \mathcal{T}_J^{c(\theta^n)\theta^n}$ on the same probability space. So for n large the first $2J$ points of both point processes coincide and we have actually $c(\theta^n)^{-1} \mathcal{T}_J^{c(\theta^n)\theta^n} = \mathcal{T}_J^\theta$ on this probability space. \square

Remark. Theorem 1 essentially consists of an “identify the limit” problem, and that is why we are free to choose the approximating \mathbf{p}_n in Proposition 3. But having proved Theorem 1, we can reverse the proof above to show that (24) holds true for *any* \mathbf{p} satisfying (18) with limiting $\theta \in \Theta_{\text{finite}}$. Indeed, the convergence in (24) is equivalent to that of $\sigma(\mathbf{p}) \otimes \mathcal{S}_J^\mathbf{p}$ to \mathcal{T}_J^θ for every J .

4.2 Skorokhod convergence of the discrete exploration process

Suppose again that the ranked probability \mathbf{p} satisfies (18) with limit $\theta \in \Theta_{\text{finite}}$ with length I . As observed in [10] (Theorem 5 and Proposition 7), the convergence in (24) is equivalent to weak convergence of the rescaled exploration process to Y^θ , but using a certain topology on function space which is weaker than the usual Skorokhod topology. As noted in [10] Example 28, assumption (18) is paradoxically not sufficient to ensure convergence in the usual Skorokhod topology; the obstacle in that example was the presence of exponentially many (in terms of $1/\sigma(\mathbf{p})$) exponentially small p -values. In this section we present some crude sufficient conditions (25,26); Proposition 3 will be a natural consequence of the proof in section 6.2. The hypotheses are as follows.

First, we prevent very small p -values by making the assumption

$$1/p_* = o(\exp(\alpha/\sigma(\mathbf{p}))) \text{ for all } \alpha > 0 \tag{25}$$

where

$$p_* := \min_i p_i.$$

Second, we will assume that most of the small $p(\cdot)$ -weights, as compared with the I first, are of order $\sigma(\mathbf{p})^2$. Write $\bar{\mathbf{p}} = (0, 0, \dots, p_{I+1}, \dots, p_n)$ for the sequence obtained from \mathbf{p} by truncating the first I terms. Let ξ have distribution $\bar{\mathbf{p}}$ on $[n]$, and write $\bar{p}(\xi)$ for the r.v. \bar{p}_ξ .

We assume that there exists some r.v. $0 \leq Q < \infty$ such that the following “moment generating function” convergence holds:

$$\lim_{n \rightarrow \infty} E \left[\exp\left(\frac{\lambda \bar{p}(\xi)}{\sigma(\mathbf{p})^2}\right) \right] = E[\exp \lambda Q] < \infty, \quad (26)$$

for every λ in some neighborhood of 0. This implies that $\bar{p}(\xi)/\sigma(\mathbf{p})^2 \xrightarrow{d} Q$, and also that the moments of all order exist and converge to those of Q .

Then we have

Theorem 3. *Suppose \mathbf{p} satisfies (18) with limit $\theta \in \Theta_{\text{finite}}$. Under extra hypotheses (25,26),*

$$\sigma(\mathbf{p})H^{\mathbf{p}} \xrightarrow{d} \frac{2}{\theta_0^2} Y^\theta \quad (27)$$

in the usual Skorokhod topology.

Remark. The proof (section 6.2) rests upon applying the elementary large deviation inequality $P(S > s) \leq e^{-\lambda s} E \exp(\lambda S)$ to the independent sums involved in (39,41). Hypothesis (26) is designed to make the application very easy; it could surely be replaced by much weaker assumptions, such as plain moment convergence conditions.

We would also guess that the convergence in (27) also holds with $H^{\mathbf{p}}$ replaced by more general exploration processes, and in particular the “classical” one, where each vertex v is visited during an interval of length $1/n$ instead of p_v , or the Harris (or contour) walk on the tree (see e.g. [17, Chapter 2]). We can easily verify the first guess. Consider the \mathbf{p} -tree $\psi_{\mathbf{p}}^{\text{depth}}(X_1, \dots, X_n)$ defined as in section 3.2 out of uniformly distributed independent r.v. Write w_1, \dots, w_n for the vertices in depth-first order, and let $H^n(t)$ be the height of the w_i for which $i/n \leq t < (i+1)/n$ (and with the convention $H^n(1) = H^n(1-)$).

Corollary 3. *Suppose \mathbf{p} satisfies (18) with limit $\theta \in \Theta_{\text{finite}}$. Under extra hypotheses (25,26),*

$$\sigma(\mathbf{p})H^n \xrightarrow{d} \frac{2}{\theta_0^2} Y^\theta \quad (28)$$

in the usual Skorokhod topology.

Proof. By the functional weak law of large numbers for sampling without replacement, we know that if π is a uniform random permutation of the n first integers, the fact that $\max_i p_{n,i} \rightarrow 0$ as $n \rightarrow \infty$ implies that if $(S_n^0(t), 0 \leq t \leq 1)$ is the linear interpolation between points $((i/n, \sum_{1 \leq k \leq i} p_{\pi(i)}), 0 \leq i \leq n)$ then $\sup_{0 \leq t \leq 1} |S_n^0(t) - t| \rightarrow 0$ in probability. Now by the remark at the end of Sect. 3.2, the cyclic order on vertices associated to the depth-first order is uniform, so with the above notation for $i = w_1, \dots, w_n$ the linear interpolation S_n between points $((i/n, \sum_{1 \leq k \leq i} p_{w_k}), 0 \leq i \leq n)$ converges uniformly to the identity in probability, since it is a (random) cyclic permutation of a function distributed as S_n^0 . Noticing that $H^n = H^{\mathbf{p}} \circ S_n^n$, the result follows. \square

The convergence of the Harris walk follows from this proposition by the arguments in [17, Chapter 2.4].

5 Height profile

This section is devoted to the proof of Theorem 2. In this section, we do not assume that $\theta \in \Theta$ has finite length nor that $\theta_0 > 0$.

5.1 Continuity of the cumulative height profile

We first prove the following intermediate lemma. Recall that the *cumulative height process* of the \mathcal{T}^θ is defined as $\bar{W}^\theta(\cdot) = \mu^\theta\{v \in \mathcal{T}^\theta : \text{ht}(v) \leq \cdot\}$, where μ^θ is the mass measure of \mathcal{T}^θ .

Lemma 4. *The cumulative height process \bar{W}^θ is continuous for a.a. realizations of \mathcal{T}^θ . Moreover, it has no flat interval, except its (possibly empty) final constancy interval, equal to $[\sup_{v \in \mathcal{T}^\theta} \text{ht}(v), \infty)$.*

Proof of Lemma 4. Recall the recursive line-breaking construction of \mathcal{T}^θ in the introduction, and the fact from [9] that the tree constructed at stage J is distributed as the reduced tree \mathcal{T}_J^θ of Sect. 3. From this, we see that the leaves labelled $1, 2, \dots$ are a.s. at pairwise different heights, meaning that the measure $d\bar{W}^\theta$ has no atom. Moreover, if \bar{W}^θ had a flat interval (other than the final constancy interval), this would mean that for some $h < \sup_{v \in \mathcal{T}^\theta}$, no leaf picked according to the mass measure can have a height in say $(h - \epsilon, h + \epsilon)$ for some $\epsilon > 0$. But let v be a vertex of \mathcal{T}^θ at height h . By the line-breaking construction, the fact that branches have size going to 0 and the “dense” property of joinpoints, we can find a joinpoint η^* at a distance $< \epsilon/2$ of v and so that the corresponding branch has length $\eta < \epsilon/2$. Since the leaves that are at the right-end of branches of the line-breaking construction are distributed as independent sampled leaves from the mass measure, this contradicts the above statement. \square

5.2 Proof of Theorem 2

The reader can consult [22] for a similar treatment of convergence of the height profile of Galton-Watson trees to a time-changed excursion of a stable Lévy process.

Suppose that $\mathbf{p} = \mathbf{p}^n$ satisfies the asymptotic regime (18). Let $\mathcal{T}^\mathbf{p}$ be the \mathbf{p} -tree, and \mathcal{T}^θ the limiting ICRT. Define \bar{W}^θ as above and recall the notation $u(h)$ in (13). For $h \geq 0$ let

$$W^\mathbf{p}(h) = \sum_{v \in \mathcal{T}^\mathbf{p}, \text{ht}(v) = \lfloor \frac{h}{\sigma(\mathbf{p})} \rfloor} p_v = u\left(\left\lfloor \frac{h}{\sigma(\mathbf{p})} \right\rfloor + 1\right) - u\left(\left\lfloor \frac{h}{\sigma(\mathbf{p})} \right\rfloor\right), \quad h \geq 0$$

and $\bar{W}^\mathbf{p}(h) = u(\lfloor h/\sigma(\mathbf{p}) \rfloor)$. Now let U_1, U_2, \dots be independent uniform(0, 1) random variables. The sequence $((\bar{W}^\mathbf{p})^{-1}(U_j), j \geq 1)$ has the law of the heights of an i.i.d. random sample of vertices of $\mathcal{T}^\mathbf{p}$, chosen according to \mathbf{p} , and the same holds for $((\bar{W}^\theta)^{-1}(U_j), j \geq 1)$ and the tree \mathcal{T}^θ , with the mass measure μ^θ as common law. For $J \geq 1$ let $\bar{W}_J^\mathbf{p}(h)$ be the associated empirical distribution of the first J terms, defined by

$$\bar{W}_J^\mathbf{p}(h) = \frac{1}{J} \sum_{i=1}^J \mathbb{1}_{\{(\bar{W}^\mathbf{p})^{-1}(U_i) \leq h\}} = \frac{1}{J} \sum_{i=1}^J \mathbb{1}_{\{U_i \leq \bar{W}^\mathbf{p}(h)\}},$$

and define $\bar{W}_J^\theta(h)$ in a similar way.

By Proposition 1, we have that the random Stieltjes measure $d\bar{W}_J^\mathbf{P}$ converges in law to $d\bar{W}_J^\theta$ as $n \rightarrow \infty$ for every $J \geq 1$. Moreover, the empirical measure of an i.i.d. J -sample of leaves distributed according to μ^θ converges to μ^θ , implying $d\bar{W}_J^\theta \xrightarrow{d} d\bar{W}^\theta$ as $J \rightarrow \infty$. Thus, for $h \geq 0$ and $J_n \rightarrow \infty$ slowly enough,

$$d\bar{W}_{J_n}^\mathbf{P} \xrightarrow{d} d\bar{W}^\theta.$$

Now let $F_J(x) = J^{-1} \sum_{i=1}^J \mathbb{1}_{\{U_i \leq x\}}$ be the empirical distribution associated to the uniform variables U_1, \dots, U_J . Then $\sup_{h \geq 0} |\bar{W}_J^\mathbf{P}(h) - \bar{W}^\mathbf{P}(h)| \leq \sup_{x \in [0,1]} |F_J(x) - x|$, which by the Glivenko-Cantelli Theorem converges to 0 as $J \rightarrow \infty$, and this convergence is uniform in n . Hence the random measure $d\bar{W}^\mathbf{P}$ converges in distribution to $d\bar{W}^\theta$ for the weak topology on measures. Thanks to Lemma 4 we may improve this to

$$\bar{W}^\mathbf{P}(\cdot) \xrightarrow{d} \bar{W}^\theta(\cdot)$$

where the convergence is weak convergence of processes for the topology of uniform convergence. It is then an elementary consequence of Lemma 4 that $\bar{W}^\mathbf{P}((\bar{W}^\mathbf{P})^{-1}(\cdot))$ converges in law for the uniform convergence topology to the identity function on $[0, 1]$.

Equation (14) can be rewritten as

$$W^\mathbf{P}(h) = F^{\text{exc}, \mathbf{P}}(\bar{W}^\mathbf{P}(h)), \quad h \geq 0, \quad (29)$$

so the convergence in distribution of $\bar{W}^\mathbf{P}$, the fact that its limit is strictly increasing and continuous, and (19) imply that the sequence of random processes $(\sigma(\mathbf{p})^{-1}W^\mathbf{P})$ is tight. Thus, the pair $(\sigma(\mathbf{p})^{-1}W^\mathbf{P}, \bar{W}^\mathbf{P})$ is tight, and up to extraction of a subsequence, we can suppose that $(\sigma(\mathbf{p})^{-1}W^\mathbf{P}, \bar{W}^\mathbf{P}) \xrightarrow{d} (W, \bar{W}')$ for some process W , and where \bar{W}' has the same law as \bar{W}^θ . Suppose further by Skorokhod's embedding theorem that the convergence is almost-sure. By definition

$$\int_0^h \frac{W^\mathbf{P}(u)}{\sigma(\mathbf{p})} du = \bar{W}^\mathbf{P}(h - \sigma(\mathbf{p})) + R(n, h)$$

where $R(n, h) \leq \bar{W}^\mathbf{P}(h) - \bar{W}^\mathbf{P}(h - \sigma(\mathbf{p}))$ goes to 0 uniformly as $n \rightarrow \infty$ by continuity of the limiting \bar{W}' . So necessarily,

$$\int_0^h W(u) du = \bar{W}'_h, \quad h \geq 0$$

for every $h \geq 0$, so that the only possible limit W is the density of $d\bar{W}'$. Therefore, the height profile W^θ of the ICRT exists and $(\sigma(\mathbf{p})^{-1}W^\mathbf{P}, \bar{W}^\mathbf{P}) \xrightarrow{d} (W^\theta, \bar{W}^\theta)$. Looking back at (29) we have

$$\sigma(\mathbf{p})^{-1}W^\mathbf{P}((\bar{W}^\mathbf{P})^{-1}(u)) = \sigma(\mathbf{p})^{-1}F^{\text{exc}, \mathbf{P}}(\bar{W}^\mathbf{P}((\bar{W}^\mathbf{P})^{-1}(u))), \quad 0 \leq u \leq 1,$$

so by the convergence of $\bar{W}^\mathbf{P}((\bar{W}^\mathbf{P})^{-1}(\cdot))$ and (19), we obtain convergence in distribution of the right-hand side to X^θ . By the convergence in law of $W^\mathbf{P}$ this finally implies that $W^\theta((\bar{W}^\theta)^{-1}(\cdot)) \stackrel{d}{=} X^\theta(\cdot)$ and Theorem 2 is proved. \square

Proof of Corollary 2. By the proof of Lemma 4, the only constant interval of the width process of the ICRT is $[\sup_{v \in \mathcal{T}^\theta} \text{ht}(v), \infty)$. Thus the height of the tree, $\sup_{v \in \mathcal{T}^\theta} \text{ht}(v)$, is the first point after which the width process remains constant. By (3), this point has same law as $\int_0^1 ds/X_s^\theta$. \square

6 The exploration process

To shorten notation, for $A \subseteq [n]$ we write $p(A)$ for the quantity $\sum_{j \in A} p_j$.

6.1 Proof of Proposition 2

Let \mathbf{p} satisfy (18) for some limiting $\theta \in \Theta_{\text{finite}}$, with length I . In this subsection we suppose that the \mathbf{p} -tree $\mathcal{T}^\mathbf{p}$ is constructed from the process $F^{\text{exc}, \mathbf{p}}$ by the depth-first search construction of section 3. Moreover, since we have (19) the convergence in law $\sigma(\mathbf{p})^{-1} F^{\text{exc}, \mathbf{p}} \xrightarrow{d} X^\theta$, we suppose by Skorokhod's representation theorem that our probability space is such that the convergence holds almost surely. Recall that in the depth-first search construction of the \mathbf{p} -tree out of the process $F^{\text{exc}, \mathbf{p}}$, the i -th examined vertex $v = w_i$ is examined during an interval $[e(v) - p_v, e(v))$, during which the labels of jumps of $F^{\text{exc}, \mathbf{p}}$ determine the set \mathcal{B}_v of children of v .

We begin with two useful observations. First, if v is a vertex of $\mathcal{T}^\mathbf{p}$ and if $\mathcal{T}_v^\mathbf{p}$ denotes the *fringe subtree* of $\mathcal{T}^\mathbf{p}$ rooted at v , that is, the subtree of descendants of v , then for every vertex w of $\mathcal{T}_v^\mathbf{p}$ one has

$$F^{\text{exc}, \mathbf{p}}(e(w)) \geq F^{\text{exc}, \mathbf{p}}(e(v)) - p(\mathcal{B}_v). \quad (30)$$

To argue this, simply recall formula (15) and notice that $\mathcal{N}(v) \subseteq \mathcal{N}(w) \cup \mathcal{B}_v$.

Second, notice that since $\max_j p_j \rightarrow 0$ and the limiting process X^θ is continuous except for a finite number I of upward jumps, we must necessarily have that a.s. as $n \rightarrow \infty$,

$$\eta_n := \max_{v \in [n]} \left| \inf_{u \in [e(v) - p_v, e(v))} (F^{\text{exc}, \mathbf{p}}(u) - F^{\text{exc}, \mathbf{p}}(e(v) - p_v)) \right| = o(\sigma(\mathbf{p})). \quad (31)$$

Lemma 5. *Almost surely*

$$\max_{j \in [n]} \sigma(\mathbf{p})^{-1} |p_j - p(\mathcal{B}_j \setminus [I])| \rightarrow 0.$$

Proof. As mentioned, for every vertex $v \in [n]$,

$$F^{\text{exc}, \mathbf{p}}(e(v)) - F^{\text{exc}, \mathbf{p}}(e(v) - p_v) = p(\mathcal{B}_v) - p_v.$$

Consider the process $F^{\mathbf{p}\downarrow}$ defined by

$$F^{\mathbf{p}\downarrow}(s) = F^{\text{exc}, \mathbf{p}}(s) - \sum_{1 \leq i \leq I} p_i \mathbb{1}_{\{s \geq x'_i\}}$$

where as above x'_i is the time when $F^{\text{exc}, \mathbf{p}}$ has its jump with size p_i . Easily, $\sigma(\mathbf{p})^{-1}F^{\mathbf{p}\downarrow}$ converges in the Skorokhod space to the process $X^{\theta\downarrow}$ defined by

$$X_s^{\theta\downarrow} = X_s^\theta - \sum_{1 \leq i \leq I} \theta_i \mathbb{1}_{\{s \geq t_i\}}$$

where t_i is the time when X^θ jumps by θ_i . This process is continuous, hence $\max_j p_j \rightarrow 0$ implies

$$\sigma(\mathbf{p})^{-1} \max_v |F^{\mathbf{p}\downarrow}(e(v)) - F^{\mathbf{p}\downarrow}(e(v) - p_v)| \rightarrow 0.$$

Now the quantity $F^{\mathbf{p}\downarrow}(e(v)) - F^{\mathbf{p}\downarrow}(e(v) - p_v)$ equals

$$\begin{aligned} & F^{\text{exc}, \mathbf{p}}(e(v)) - F^{\text{exc}, \mathbf{p}}(e(v) - p_v) - \sum_{1 \leq i \leq I} p_i \mathbb{1}_{\{x'_i \in (e(v) - p_v, e(v))\}} \\ &= p(\mathcal{B}_v) - p_v - p(\mathcal{B}_v \cap [I]) \end{aligned}$$

implying the lemma. \square

Now, for v a non-root vertex of $\mathcal{T}^{\mathbf{p}}$ let $f(v)$ be its parent. For $i \in [I]$ and n large enough, i is not the root (since the limiting X^θ does not begin with a jump), so $f(i)$ exists.

Lemma 6. *Let $i \in I$. Let $\mathcal{M}(i)$ be the set of descendents of $f(i)$ that come strictly before i in depth-first order. Suppose that $f(i) \notin [I]$ for n large enough. Then as $n \rightarrow \infty$, $p(\mathcal{M}(i)) \rightarrow 0$ almost surely.*

Proof. A variation of (30) implies for any $v \in \mathcal{M}(i)$ and n large that

$$F^{\text{exc}, \mathbf{p}}(e(v)) \geq F^{\text{exc}, \mathbf{p}}(e(f(i))) - p(\mathcal{B}_{f(i)} \setminus [I]). \quad (32)$$

Indeed, it is clear that for n large the sets $\mathcal{B}_v \cap [I]$ contain at most one element, otherwise the Skorokhod convergence $\sigma(\mathbf{p})^{-1}F^{\text{exc}, \mathbf{p}} \rightarrow X^\theta$ would fail as two or more upward jumps of non-negligible sizes could occur in an ultimately negligible interval. Moreover, for $v \in \mathcal{M}(i)$, it is clear that $\mathcal{N}(v)$ contains i , hence (32). Thus

$$\inf_{e(f(i)) \leq u \leq e(f(i)) + p(\mathcal{M}(i))} F^{\text{exc}, \mathbf{p}}(u) \geq F^{\text{exc}, \mathbf{p}}(e(f(i))) - p(\mathcal{B}_{f(i)} \setminus [I]) - \eta_n,$$

with η_n defined at (31), since the vertices of $\mathcal{M}(i)$ are visited during the interval $[e(f(i)), e(f(i)) + p(\mathcal{M}(i))]$. Since $\sigma(\mathbf{p})^{-1}F^{\text{exc}, \mathbf{p}}(e(f(i)))$ is easily seen to converge to $X_{t_i}^\theta$, by (31), Lemma 5 and the fact that $f(i) \notin [I]$ for n large, if $p(\mathcal{M}(i))$ did not converge to 0, by extracting along a subsequence we could find an interval $[t_i, t_i + \varepsilon]$ with $\varepsilon > 0$ where $X_u^\theta \geq X_{t_i}^\theta$, and this is a.s. impossible by Lemma 1. \square

The assumption that $f(i) \notin [I]$ may look strange since it is intuitive that the child of some $i \in [I]$ is very unlikely to be in $[I]$ for n large (e.g. by Theorem 2). We actually have:

Lemma 7. *For every $i \in [I]$, almost surely, $\mathcal{B}_i \cap [I] = \emptyset$ for n large, and*

$$\sigma(\mathbf{p})^{-1}p(\mathcal{B}_i) \rightarrow \theta_i.$$

Proof. By Lemma 5 it suffices to prove that a.s. for large n , $\mathcal{B}_i \cap [I] = \emptyset$. Suppose that there exist $i, j \in [I]$ such that j is the child of i in the \mathbf{p} -tree infinitely often. Since $I < \infty$, we may further suppose that $f(i) \notin [I]$ by taking (up to extraction) the least such i in depth-first order. By definition, $F^{\text{exc}, \mathbf{P}}$ has a jump with size i in the interval $[e(f(i)) - p_{f(i)}, e(f(i))]$. Moreover, it follows from the definition of $\mathcal{M}(i)$ that $e(i) - p_i = e(f(i)) + p(\mathcal{M}(i))$. Since the vertex i is examined in the interval $[e(i) - p_i, e(i)]$ and $p(\mathcal{M}(i)) \rightarrow 0$ by the preceding lemma, the fact that $f(j) = i$ implies that the jumps with size p_i and p_j occur within a vanishing interval $[e(f(i)) - p_{f(i)}, e(i)]$. Therefore, the Skorokhod convergence of $\sigma(\mathbf{p})^{-1} F^{\text{exc}, \mathbf{P}}$ to X^θ would fail. \square

Now recall the definition (21) of the processes $r_i^{\mathbf{P}}$ used to build $G_I^{\mathbf{P}}$ in section 4, and that x'_i is the time when $F^{\text{exc}, \mathbf{P}}$ jumps by p_i .

Lemma 8. *For every $i \in [I]$, as $n \rightarrow \infty$, we have*

$$\sigma(\mathbf{p})^{-1} \left| \inf_{x'_i \leq u \leq s} F^{\text{exc}, \mathbf{P}}(u) - F^{\text{exc}, \mathbf{P}}(x'_i-) - r_i^{\mathbf{P}}(s) \right| \rightarrow 0$$

a.s. uniformly in $s \in [x'_i, e(i) + p(\mathcal{T}_i^{\mathbf{P}})]$.

Proof. Let $i \in [I]$, and let $\mathcal{B}_i = \{v_1, v_2, \dots, v_k\}$ (with $k = |\mathcal{B}_i|$) where v_1, v_2, \dots are in depth-first order. For $1 \leq j \leq k$ let also v'_j be the last examined vertex of $\mathcal{T}_{v_j}^{\mathbf{P}}$ in depth-first order, that is, the predecessor of v_{j+1} if $j < k$. Then one has, for every $1 \leq j \leq k$ and $w \in \mathcal{T}_{v_j}^{\mathbf{P}}$

$$F^{\text{exc}, \mathbf{P}}(e(w)) \geq F^{\text{exc}, \mathbf{P}}(e(v_j)) - p(\mathcal{B}_{v_j}),$$

as follows from (30). Rewrite this as

$$F^{\text{exc}, \mathbf{P}}(e(w)) \geq F^{\text{exc}, \mathbf{P}}(e(i)) - \sum_{1 \leq r \leq j-1} p_{v_r}$$

and check that the right hand side equals $F^{\text{exc}, \mathbf{P}}(e(v'_{j-1}))$. In particular, we obtain

$$\left| \inf_{v: e(v) \in [e(i), e(w)]} F^{\text{exc}, \mathbf{P}}(e(v)) - F^{\text{exc}, \mathbf{P}}(e(i)) + \sum_{1 \leq r \leq j-1} p_{v_r} \right| \leq \max_{1 \leq j \leq k} p_{v_j}.$$

Now check that for w a vertex of $\mathcal{T}_{v_j}^{\mathbf{P}}$, one has $r_i^{\mathbf{P}}(e(w)) = \sum_{j \leq r \leq k} p_{v_r}$. For s as in the statement of the lemma deduce, for n large (since $\mathcal{B}_i \cap [I] = \emptyset$ by Lemma 7),

$$\left| \inf_{u \in [x'_i, s]} F^{\text{exc}, \mathbf{P}}(u) - F^{\text{exc}, \mathbf{P}}(e(i)) + p(\mathcal{B}_i) - r_i^{\mathbf{P}}(u) \right| \leq 2 \max_{j \notin [I]} p_j + \eta_n + \eta'_n$$

where

$$\eta'_n = \max_{x'_i \leq u \leq e(i)} |F^{\text{exc}, \mathbf{P}}(u) - F^{\text{exc}, \mathbf{P}}(e(i))|$$

which is $o(\sigma(\mathbf{p}))$ by Lemma 6 and the convergence $\sigma(\mathbf{p})^{-1} F^{\text{exc}, \mathbf{P}} \rightarrow X^\theta$. We conclude, using the fact that $\sigma(\mathbf{p})^{-1} F^{\text{exc}, \mathbf{P}}(x'_i-) \rightarrow X_{t_i-}^\theta$, which is equal to the limit of $\sigma(\mathbf{p})^{-1} (F^{\text{exc}, \mathbf{P}}(e(i)) - p(\mathcal{B}_i))$, as follows from Lemmas 6 and 7. \square

Proof of Proposition 2. We prove that the process $\sigma(\mathbf{p})^{-1}r_i^{\mathbf{P}}$ converges to the $R_i^{\boldsymbol{\theta}}$ of section 2 in the Skorokhod topology, for every i . In view of Lemma 8, and since by definition of r_i one has $r_i(u) = 0$ for $u \geq e(i) + p(\mathcal{T}_i^{\mathbf{P}})$, the only thing to do is to show that $e(v'_k) = e(i) + p(\mathcal{T}_i^{\mathbf{P}})$ converges to the T_i of section 2. Since $e(v'_k) \geq \inf\{s \geq x'_i : r_i^{\mathbf{P}}(s) = 0\}$, we obtain that $\liminf e(v'_k) \geq T_i$. Suppose $\ell = \limsup e(v'_k) > T_i$, and up to extraction suppose that ℓ is actually the limit of $e(v'_k)$. From the fact that $F^{\text{exc}, \mathbf{P}}(e(v'_k)) = F^{\text{exc}, \mathbf{P}}(e(i)) - p(\mathcal{B}_i)$, hence $\sigma(\mathbf{p})^{-1}F^{\text{exc}, \mathbf{P}}(e(v'_k))$ converges to $X_{t_i-}^{\boldsymbol{\theta}}$ by Lemmas 6 and 7, we would find $\ell > T_i$ with $X_{\ell}^{\boldsymbol{\theta}} = X_{t_i-}^{\boldsymbol{\theta}}$ and $X_s^{\boldsymbol{\theta}} \geq X_{t_i-}^{\boldsymbol{\theta}}$ for $s \in [T_i, \ell]$, and this is almost surely impossible by Lemma 1 as $X_{t_i-}^{\boldsymbol{\theta}}$ would be a local minimum of $X^{\boldsymbol{\theta}}$, attained at time T_i .

Without extra argument we cannot conclude that the sum $\sigma(\mathbf{p})^{-1}(F^{\text{exc}, \mathbf{P}} - \sum_{i=1}^I r_i^{\mathbf{P}})$ converges to $X^{\boldsymbol{\theta}} - \sum_{i=1}^I R_i^{\boldsymbol{\theta}}$, but this is nonetheless true for the following reason. The process $R_i^{\boldsymbol{\theta}}$ is continuous except for one jump at t_i , and the process $r_i^{\mathbf{P}}$ has precisely one jump with size $p(\mathcal{B}_i)$ at time x'_i , that is, at the *same* time as the jump of $F^{\text{exc}, \mathbf{P}}$ with size p_i . Together with Lemma 7, we obtain the Skorokhod convergence $\sigma(\mathbf{p})^{-1}G_I^{\mathbf{P}} \rightarrow Y^{\boldsymbol{\theta}}$. \square

6.2 Proof of Theorem 3

As above, we suppose that \mathbf{p} is a ranked probability distribution satisfying (18) for some limiting $\boldsymbol{\theta}$ with length I , and we suppose that the \mathbf{p} -tree $\mathcal{T}^{\mathbf{P}}$ is obtained by the depth-first construction of section 3 out of the process $F^{\text{exc}, \mathbf{P}}$. We are going to show the following result:

Proposition 4. *Under extra hypotheses (25,26) on $(\mathbf{p}^{(n)})$, as $n \rightarrow \infty$*

$$\max_v \left| \frac{\theta_0^2 \sigma(\mathbf{p})}{2} \text{ht}(v) - \sigma(\mathbf{p})^{-1} G_I^{\mathbf{P}}(e(v)) \right| \xrightarrow{p} 0.$$

We first show how Theorem 3 and Proposition 3 are easy consequences of Proposition 4.

Proof of Theorem 3. Since $\sigma(\mathbf{p})^{-1}G_I^{\mathbf{P}}$ converges uniformly in distribution to a continuous process, and since $H^{\mathbf{P}}$ does not vary in the intervals $[e(v) - p_v, e(v))$, the last displayed convergence extends to

$$\max_{u \in [0,1]} \left| \frac{\theta_0^2 \sigma(\mathbf{p})}{2} H^{\mathbf{P}}(u) - \sigma(\mathbf{p})^{-1} G_I^{\mathbf{P}}(u) \right| \xrightarrow{p} 0, \quad (33)$$

and then Proposition 2 implies Theorem 3. \square

Proof of Proposition 3. For Proposition 3, we choose the following approximating sequence $\mathbf{p}^{(n+I)}$ for $\boldsymbol{\theta} \in \Theta_{\text{finite}}$ with length I . Given n , let $z_n = \sqrt{n}/\theta_0$, $s_n = n + z_n \sum_{1 \leq i \leq I} \theta_i$ and

$$\begin{cases} p_i = \frac{z_n \theta_i}{s_n} & \text{if } 1 \leq i \leq I \\ p_i = \frac{1}{s_n} & \text{if } I + 1 \leq i \leq n + I. \end{cases} \quad (34)$$

It is trivial to see that this sequence fulfills hypotheses (25,26). Hence (33) is satisfied, and Proposition 3 is an immediate consequence. \square

We now mention three consequences of hypotheses (25,26) that will be used later. First, notice that $p_* \leq 1/n$ since \mathbf{p} is a probability on $[n]$, so (25) implies

$$n = o(\exp(\alpha/\sigma(\mathbf{p}))) \text{ for all } \alpha > 0. \quad (35)$$

Second, (26) implies convergence of all moments of $\bar{p}(\xi)/\sigma(\mathbf{p})^2$, and in particular

$$\begin{aligned} E\left(\frac{\bar{p}(\xi)}{\sigma(\mathbf{p})^2}\right) &= \sum_{i \notin [I]} p_i^2 / \sigma(\mathbf{p})^2 \\ &= 1 - \sum_{i \in [I]} p_i^2 / \sigma(\mathbf{p})^2 \\ &\xrightarrow{n \rightarrow \infty} \theta_0^2 = E(Q). \end{aligned} \tag{36}$$

Third, for every λ in a neighborhood of 0,

$$\sigma(\mathbf{p})^2 \sum_{i \notin [I]} \left[\exp\left(\frac{\lambda p_i}{\sigma^2}\right) - 1 - \frac{\lambda p_i}{\sigma^2} \right] \xrightarrow{n \rightarrow \infty} E \frac{1}{Q} [\exp(\lambda Q) - 1 - \lambda Q] < \infty. \tag{37}$$

Indeed, the left side can be rewritten as $E\left(\frac{\sigma(\mathbf{p})^2}{\bar{p}(\xi)} \left[\exp\left(\frac{\lambda \bar{p}(\xi)}{\sigma^2}\right) - 1 - \frac{\lambda \bar{p}(\xi)}{\sigma^2} \right]\right)$, where the function $f(x) = (e^{\lambda x} - 1 - \lambda x)/x$ is understood to equal its limit 0 at 0. Since it is bounded in a neighborhood of 0 and dominated by $e^{\lambda x}$ near ∞ , the convergence of this expectation is an easy consequence of (26).

The first step in the proof of Proposition 4 is to relate $H(\cdot)$ to another function $\mathcal{G}(\cdot)$ measuring “sum of small \mathbf{p} -values along path to root”. Let $\mathcal{A}(v)$ be the set of ancestors of v in the \mathbf{p} -tree, and let

$$\mathcal{G}(v) := p(\mathcal{A}(v) \setminus [I]). \tag{38}$$

Lemma 9. *Under extra hypotheses (25,26), as $n \rightarrow \infty$ for fixed $K > 0$*

$$\max_{v: \text{ht}(v) \leq K/\sigma(\mathbf{p})} \left| \sigma(\mathbf{p})\theta_0^2 \text{ht}(v) - \sigma(\mathbf{p})^{-1} \mathcal{G}(v) \right| \xrightarrow{p} 0.$$

Proof. Let V be a \mathbf{p} -distributed random vertex. Fix $\varepsilon > 0$. It is enough to prove that as $n \rightarrow \infty$

$$P\left(|\sigma(\mathbf{p})\theta_0^2 \text{ht}(V) - \sigma(\mathbf{p})^{-1} \mathcal{G}(e(V))| > \varepsilon, \sigma(\mathbf{p})\text{ht}(V) \leq K\right) = o(p_*).$$

Let ξ have distribution \mathbf{p} on $[n]$ and let $(\xi_i, i \geq 1)$ be i.i.d. By the “birthday tree” construction of the \mathbf{p} -tree [15, Corollary 3] we have equality of joint distributions

$$(\text{ht}(V), \mathcal{G}(V)) \stackrel{d}{=} (T - 2, \sum_{i=1}^{T-1} \bar{p}(\xi_i))$$

where

$$T := \min\{j \geq 2 : \xi_j = \xi_i \text{ for some } 1 \leq i < j\}$$

is the first repeat time in the sequence ξ_i . So it is enough to prove

$$P\left(\left|\sigma(\mathbf{p})\theta_0^2(T - 2) - \sigma(\mathbf{p})^{-1} \sum_{i=1}^{T-1} \bar{p}(\xi_i)\right| > \varepsilon, \sigma(\mathbf{p})(T - 2) \leq K\right) = o(p_*).$$

We may replace $T - 2$ by $T - 1$ and θ_0^2 by $E(\frac{\bar{p}(\xi)}{\sigma(\mathbf{p})^2})$ by the above remark. Rewriting in terms of $\tilde{p}(i) := \frac{\bar{p}(i)}{\sigma(\mathbf{p})^2} - E(\frac{\bar{p}(\xi)}{\sigma(\mathbf{p})^2})$, we need to prove

$$P\left(\left|\sum_{i=1}^{T-1} \tilde{p}(\xi_i)\right| > \varepsilon/\sigma(\mathbf{p}), T-1 \leq K/\sigma(\mathbf{p})\right) = o(p_*).$$

Now we are dealing with a mean-zero random walk, and classical fluctuation inequalities (e.g. [18] Exercise 1.8.9) reduce the problem to proving the fixed-time bound

$$P\left(\left|\sum_{1 \leq i \leq K/\sigma(\mathbf{p})} \tilde{p}(\xi_i)\right| \geq \varepsilon/\sigma(\mathbf{p})\right) = o(p_*). \quad (39)$$

We now appeal to assumption (26), which basically says that the sums in question behave as if the summands had distribution $Q - \theta_0^2$ not depending on n . More precisely, the elementary large deviation inequality applied to the probability in (39) but without the absolute values implies that for any small $\lambda > 0$,

$$\log P\left(\sum_{i=1}^{K/\sigma(\mathbf{p})} \tilde{p}(\xi_i) \geq \varepsilon/\sigma(\mathbf{p})\right) \leq -\frac{\lambda\varepsilon}{\sigma(\mathbf{p})} + \frac{K}{\sigma(\mathbf{p})} \log(E(\exp(\lambda\tilde{p}(\xi)))).$$

Assumption (26) and the convergence of the expectation of $\bar{p}(\xi)$ allows us to rewrite the log term on the right as

$$\frac{K}{\sigma(\mathbf{p})} \log E(\exp(\lambda(Q - \theta_0^2))) + \frac{K\eta_\lambda(n)}{\sigma(\mathbf{p})E(\exp(\lambda(Q - \theta_0^2)))},$$

where $\eta_\lambda(n) \rightarrow 0$ as $n \rightarrow \infty$ for any fixed λ . We now choose λ small enough so that $-\lambda\varepsilon + K \log E(\exp(\lambda(Q - \theta_0^2))) = -\delta < 0$ and we let $n \rightarrow \infty$, obtaining the bound $\exp(-\delta'/\sigma(\mathbf{p}))$, for some $\delta' > 0$, for the probability in (39) without absolute values, but the other side of the inequality is similar. Now assumption (25) gives the desired bound (39). \square

The next, rather strange-looking lemma does most of the work in relating the processes $G_I^{\mathbf{P}}(\cdot)$ and $\mathcal{G}(\cdot)$.

Given a probability distribution \mathbf{p} on $[n]$ and given a subset $A \subset [n]$, let \mathbf{q} be the probability distribution obtained by lumping the points A into a single point; that is, $q_1 = p(A)$ and the multiset $\{q_i, i \geq 2\}$ is the multiset $\{p_i, i \notin A\}$. We also let \mathcal{I} be the set of ‘‘large’’ \mathbf{q} -values, except q_1 . Precisely, \mathcal{I} is such that the multisets $\{p_v, v \in [I] \setminus A\} = \{q_v, v \in \mathcal{I}\}$ are equal. Then

Lemma 10. *Suppose $\mathbf{p} = \mathbf{p}^{(n)}$ satisfies the regime (18) and extra hypotheses (25,26). Let $A = A^{(n)} \subset [n]$ and define \mathbf{q} as above. Define a random variable $X = X(\mathbf{q})$ as follows. Take a \mathbf{q} -tree, condition on vertex 1 being the root. Let \mathcal{B}_1 be the set of children of 1, and for each $v \in \mathcal{B}_1$ toss two coins c_1 and c_2 , c_1 a fair coin and $P(c_2 = \text{Heads}) = p(A \setminus [I])/p(A)$, and set*

$$X := \sum \{q_v : v \in \mathcal{B}_1 \setminus \mathcal{I}, \text{coins } c_1 \text{ and } c_2 \text{ land Heads}\}.$$

Suppose $q_1 \leq K\sigma(\mathbf{p})$ and set $\bar{q}_1 = p(A \setminus [I])$. Then for fixed $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, K) > 0$ with

$$P(|X - \frac{1}{2}\bar{q}_1| > \varepsilon\sigma(\mathbf{p})) \leq \exp(-\delta/\sigma(\mathbf{p})) = o(1/n),$$

where the $o(1/n)$ is thus uniform over $q_1 \leq K\sigma(\mathbf{p})$.

Proof. Consider the random variable

$$Y := \sum_{i \notin A \cup [I]} p_i \mathbb{1}_{\{U_i \leq \bar{q}_1/2\}}$$

where the (U_i) are independent uniform(0, 1). The key relation is

$$P(X \in \cdot) \leq \frac{1}{q_1} P(Y \in \cdot). \quad (40)$$

This follows from the *breadth-first* construction of the \mathbf{p} -trees. In that construction of a \mathbf{q} -tree, vertices i are associated with uniform(0, 1) r.v.'s U'_i in such a way that, if vertex 1 happens to be the root, then the children v of 1 are the vertices v for which $U_v := U'_v - U'_1 \bmod 1$ falls within $(0, q_1)$. Thus, writing

$$\begin{aligned} X' &:= \sum \{q_v : v \in \mathcal{B}_1 \setminus \mathcal{I}\} \\ Y' &:= \sum_{i \notin A \cup [I]} p_i \mathbb{1}_{\{U_i \leq q_1\}} \end{aligned}$$

we have

$$X' = Y' \text{ on the event } \{ \text{vertex 1 is root} \}.$$

So

$$P(X' \in \cdot \mid 1 \text{ is root}) \leq \frac{P(Y' \in \cdot)}{P(1 \text{ is root})} = \frac{1}{q_1} P(Y' \in \cdot).$$

The stated inequality (40) follows by applying an independent Bernoulli($\bar{q}_1/(2q_1)$) thinning procedure to both sides.

Now write $c = \bar{q}_1/2$ and let us study the centered version of Y :

$$\tilde{Y} := \sum_{i \notin A \cup [I]} p_i (\mathbb{1}_{\{U_i \leq c\}} - c). \quad (41)$$

The elementary large deviation bound, applied to $\tilde{Y}/\sigma(\mathbf{p})^2$, is: for arbitrary $\lambda > 0$,

$$\log P(\tilde{Y} > \varepsilon\sigma(\mathbf{p})) \leq \frac{-\lambda\varepsilon}{\sigma(\mathbf{p})} + \log E \exp(\lambda\tilde{Y}/\sigma(\mathbf{p})^2).$$

We calculate

$$\begin{aligned} &\log E \exp(\lambda\tilde{Y}/\sigma(\mathbf{p})^2) \\ &= \sum_{i \notin A \cup [I]} \left\{ \frac{-\lambda p_i}{\sigma(\mathbf{p})^2} c + \log \left[1 + c(e^{\lambda p_i/\sigma(\mathbf{p})^2} - 1) \right] \right\} \\ &\leq c \sum_{i \in [n]} \left\{ e^{\lambda p_i/\sigma(\mathbf{p})^2} - 1 - \frac{\lambda p_i}{\sigma(\mathbf{p})^2} \right\}, \end{aligned}$$

since the quantities we are summing are positive, and by (37) the bound is asymptotic to $c\sigma(\mathbf{p})^{-2}\Phi(\lambda)$ for

$$\Phi(\lambda) := E\frac{1}{Q} [\exp(\lambda Q) - 1 - \lambda Q].$$

By hypothesis $c := \bar{q}_1/2 \leq K\sigma(\mathbf{p})$, so $c\sigma(\mathbf{p})^{-2} \leq K\sigma(\mathbf{p})^{-1}$. So there is a constant $C_1 = C_1(K)$ such that

$$\log P(\tilde{Y} > \varepsilon\sigma(\mathbf{p})) \leq \frac{1}{\sigma(\mathbf{p})} (-\lambda\varepsilon + C_1\Phi(\lambda)).$$

But $\Phi'(0) = 0$ and so $\Phi(\lambda) = o(\lambda)$ as $\lambda \downarrow 0$, so the right side is strictly negative for small $\lambda > 0$. So there exists $\delta_1 = \delta_1(\varepsilon, K) > 0$ such that

$$P(\tilde{Y} > \varepsilon\sigma(\mathbf{p})) \leq \exp(-\delta_1/\sigma(\mathbf{p})).$$

Since $Y - \tilde{Y} = c \sum_{i \notin A \cup [I]} p_i \leq \bar{q}_1/2$ we have established the one-sided inequality

$$P(Y - \frac{1}{2}\bar{q}_1 > \varepsilon\sigma(\mathbf{p})) \leq \exp(-\delta_1/\sigma(\mathbf{p})).$$

The other side of the inequality is similar except for this last step: we cannot bound so easily the quantity $\tilde{Y} - Y$. However, by (18),

$$\sum_{i \notin A \cup [I]} p_i = 1 - p(A \cup [I]) \geq 1 - q_1 - \sum_{i=1}^I p_i \geq 1 - C_2\sigma(\mathbf{p})$$

for some $C_2 = C_2(K) < \infty$. Thus $Y - \tilde{Y} \geq c(1 - C_2\sigma(\mathbf{p}))$ and we can conclude as above by the existence of $\delta_2 = \delta_2(\varepsilon, K)$ satisfying

$$P(\frac{1}{2}\bar{q}_1 - Y > \varepsilon\sigma(\mathbf{p})) \leq \exp(-\delta_2/\sigma(\mathbf{p})).$$

So, letting $\delta' = \delta_1 \wedge \delta_2$,

$$P(|Y - \frac{1}{2}\bar{q}_1| > \varepsilon\sigma(\mathbf{p})) \leq 2 \exp(-\delta'/\sigma(\mathbf{p})).$$

Now (40) and hypothesis (25) and its consequence (35) establish Lemma 10 (with any $\delta < \delta'$). \square

For the next lemma, recall the definition of $\mathcal{N}(v)$ around (15) and let $\mathcal{N}^*(v)$ be the subset of vertices of $\mathcal{N}(v)$ which are not in $[I]$ and whose parent is not in $[I]$ either.

Lemma 11. *Fix $j \in [n]$ and a subset $A \subset [n]$ with $j \in A$. Take a random \mathbf{p} -tree and condition on $\mathcal{A}(j) = A$. Let also v_1, \dots, v_k be the children of j that are not in $[I]$ and let $c^*(j) = \sum_{1 \leq l \leq k} b_l p_{v_l}$, where the b_l 's are independent Bernoulli random variables with parameter $1/2$, independent of the \mathbf{p} -tree. Define*

$$X^* := p(\mathcal{N}^*(j)) - c^*(j).$$

Then X^ is distributed as the random variable X in Lemma 10.*

Proof. Order A as v_0, v_1, \dots, j , arbitrarily except for ending with j . Let \mathbf{T}^* be the set of rooted trees on $[n]$ with root v_0 whose path to j is the path v_0, v_1, \dots, j . Let \mathbf{T}^\oplus be the set of rooted trees on $[n] \setminus A \cup \{\oplus\}$ with root \oplus . There is a natural map $\mathbf{T}^* \rightarrow \mathbf{T}^\oplus$: “lump the vertices in A together into a single vertex \oplus ”. It is straightforward to check, from the combinatorial definition (see e.g. [31]) of \mathbf{p} -tree, that this map takes the distribution of \mathbf{p} -tree (conditioned to \mathbf{T}^*) into the distribution of a \mathbf{q} -tree (conditioned on having root \oplus). Also, we have the extra constraint in X^* that the parents of the vertices we are summing on are not in $[I]$, but conditionally on the fact that v has some parent in A , it is easy that the parent is in $[I]$ with probability $p(A \cap [I])/p(A)$. This corresponds to the biased coin-tosses in Lemma 10. And the fair coin-tosses in Lemma 10 reflect the random ordering of branches used in defining the depth-first order, as can be seen from the definition in Section 3 (the set of children of any vertex is put in exchangeable random order). The only exception is on children of j itself, which are all in $\mathcal{N}^*(v)$, so the b_l 's are designated to artificially remove each of them with probability $1/2$. This establishes the lemma. \square

The importance of the lemma is explained by the following formula

$$\max_v |G_I^{\mathbf{p}}(e(v)) - p(\mathcal{N}^*(v)) - c^*(v)| = o(\sigma(\mathbf{p})) \text{ in probability.} \quad (42)$$

Since asymptotically we know that children of $i \in [I]$ are not in $[I]$, and since by Lemmas 5 and 7:

$$\sigma(\mathbf{p})^{-1} \max_{v \notin [I]} p(\mathcal{B}_v \setminus [I]) \xrightarrow{p} 0, \quad (43)$$

so in particular $\max_j \sigma(\mathbf{p})^{-1} c^*(j) \rightarrow 0$ in probability with the notations above, this is a straightforward consequence of Lemma 7 and

Lemma 12. *Suppose that no vertex $i \in [I]$ has a child that is also in $[I]$, then we have for every v*

$$G_I^{\mathbf{p}}(e(v)) = p(\mathcal{N}^*(v)) - \sum_{i \in \mathcal{N}(v) \cap [I]} (p_i - p(\mathcal{B}_i)). \quad (44)$$

Proof. Recall by definition (20) of the processes ρ_k that if k is a child of some $i \in [I]$, $\rho_k(e(v)) = p_k$ whenever v is examined after the parent $f(i)$ of i and strictly before k in depth-first order, and $\rho_k(e(v)) = 0$ otherwise. As a consequence of (15), we thus have

$$G_I^{\mathbf{p}}(e(v)) = p(\mathcal{N}(v)) - \sum_{i \in [I], k \in \mathcal{B}_i} p_k \mathbb{1}_{\{e(f(i)) \leq e(v) < e(k)\}}.$$

A careful examination of this formula shows that a term in the sum on the right is not zero if either v has some ancestor $i \in [I]$, or some ancestor of v has a child $i \in [I]$ that is after v in depth-first order, and these situations are exclusive by the assumption that vertices of $[I]$ do not have children in $[I]$. In the first case, the formula says that we remove all the \mathbf{p} -values of children of i that are after v in depth-first order, in the second case, it says that we remove the \mathbf{p} -values of all the children of i , implying (44). \square

Proof of Proposition 4. Fix $\varepsilon > 0$ and consider arbitrary $v \in [n]$. Recall the definition of $\mathcal{A}(v), \mathcal{G}(v), \mathcal{N}^*(v), c^*(v)$. We assert, from Lemmas 10 and 11, that for any $K > 0$ there exists

$\delta = \delta(\varepsilon, K)$ with

$$P\left(|p(\mathcal{N}^*(v)) - c^*(v) - \frac{1}{2}\mathcal{G}(v)| > \varepsilon\sigma(\mathbf{p}) | \mathcal{A}(v)\right) \leq \exp(-\delta/\sigma(\mathbf{p})) \text{ on } \{\mathcal{G}(v) \leq (K + 2\varepsilon)\sigma(\mathbf{p})\}. \quad (45)$$

To argue (45), note that conditioning on the set $A = \mathcal{A}(v)$ of vertices in the path from the root to v determines the value $\mathcal{G}(v) := p(\mathcal{A}(v) \setminus [I]) = \bar{q}_1$ say. Then Lemmas 10, 11 imply that the conditional distribution of $p(\mathcal{N}^*(v)) - c^*(v)$ has the distribution of X in Lemma 10, The conclusion of Lemma 10 now gives (45).

So for fixed K and arbitrary $v \in [n]$

$$P\left(|p(\mathcal{N}^*(v)) - c^*(v) - \frac{1}{2}\mathcal{G}(v)| > \varepsilon\sigma(\mathbf{p}), \frac{1}{2}\mathcal{G}(v) \leq (K + 2\varepsilon)\sigma(\mathbf{p})\right) \leq \exp(-\delta/\sigma(\mathbf{p})) = o(1/n).$$

Using Boole's inequality gives

$$P\left(\frac{1}{\sigma(\mathbf{p})}|p(\mathcal{N}^*(v)) - c^*(v) - \frac{1}{2}\mathcal{G}(v)| > \varepsilon \text{ for some } v \text{ with } \frac{1}{2\sigma(\mathbf{p})}\mathcal{G}(v) \leq K + 2\varepsilon\right) = o(1).$$

By (42) we may replace $p(\mathcal{N}^*(v)) - c^*(v)$ by $G_I^{\mathbf{p}}(e(v))$ in the previous expression. We now use a slightly fussy truncation procedure. Imposing an extra constraint,

$$\begin{aligned} &P\left(\frac{1}{\sigma(\mathbf{p})} \max_v G_I^{\mathbf{p}}(e(v)) \leq K, \frac{1}{\sigma(\mathbf{p})}|G_I^{\mathbf{p}}(e(v)) - \frac{1}{2}\mathcal{G}(v)| > \varepsilon \text{ for some } v \right. \\ &\quad \left. \text{with } \frac{1}{2\sigma(\mathbf{p})}\mathcal{G}(v) \leq K + 2\varepsilon\right) = o(1). \end{aligned} \quad (46)$$

We claim that we can remove the restriction on v to get

$$P\left(\frac{1}{\sigma(\mathbf{p})} \max_v G_I^{\mathbf{p}}(e(v)) \leq K, \frac{1}{\sigma(\mathbf{p})}|G_I^{\mathbf{p}}(e(v)) - \frac{1}{2}\mathcal{G}(v)| > \varepsilon \text{ for some } v\right) = o(1). \quad (47)$$

Indeed, if v has parent v' then $\mathcal{G}(v) - \mathcal{G}(v') \leq \max_{i \notin [I]} p_i = o(\sigma(\mathbf{p}))$. So if there exists a v with $\frac{1}{2\sigma(\mathbf{p})}\mathcal{G}(v) > K + 2\varepsilon$ then (for large n) there is an ancestor w with $K + \varepsilon < \frac{1}{2\sigma(\mathbf{p})}\mathcal{G}(w) < K + 2\varepsilon$. But if the first event in (46) occurs, one obviously cannot have $\sigma(\mathbf{p})^{-1}|G_I^{\mathbf{p}}(e(w)) - \frac{1}{2}\mathcal{G}(w)| \leq \varepsilon$ by definition of w . Thus the probability in (47) is bounded by twice the probability in (46). This establishes (47). Since Proposition 2 implies $\frac{1}{\sigma(\mathbf{p})} \max_v G_I^{\mathbf{p}}(e(v))$ is tight as $n \rightarrow \infty$, (47) implies

$$\max_v \frac{1}{\sigma(\mathbf{p})}|G_I^{\mathbf{p}}(e(v)) - \frac{1}{2}\mathcal{G}(v)| \xrightarrow{P} 0. \quad (48)$$

Now let us show that the sequence $(\sigma(\mathbf{p}) \max_{v \in [n]} \text{ht}(v), n \geq 1)$ is tight. Fix $\varepsilon > 0$ and let $K > 0$ such that

$$P\left(\frac{1}{\sigma(\mathbf{p})} \max_v G_I^{\mathbf{p}}(e(v)) > K\right) < \varepsilon/2,$$

Then

$$P\left(\sigma(\mathbf{p}) \max_v \text{ht}(v) > K + 1\right) \leq \varepsilon/2 + P\left(\sigma(\mathbf{p}) \max_v \text{ht}(v) > K + 1, \frac{1}{\sigma(\mathbf{p})} \max_v G_I^{\mathbf{p}}(e(v)) < K\right),$$

but by the same kind of argument as above, if $\sigma(\mathbf{p}) \max_v \text{ht}(v) > K$, for n large there must exist some w with $K + 1/2 < \sigma(\mathbf{p}) \text{ht}(w) < K + 1$. By Lemma 9 we then have also $K + 1/2 < \sigma(\mathbf{p})^{-1}\mathcal{G}(v) < K + 1$ with high probability, so (48) implies that the right-hand side in the last expression is $< \varepsilon/2$ for n large. This being proved, Lemma 9 rewrites as $\max_v |\sigma(\mathbf{p})^{-1}\mathcal{G}(v) - \sigma(\mathbf{p})\theta_0^2 \text{ht}(v)| = o(1)$ in probability, which together with (48) establishes the proposition. \square

7 Miscellaneous comments

1. In principle Corollary 2 gives a criterion for boundedness of \mathcal{T}^θ , but one would prefer to have a condition directly in terms of θ . Here are some steps in that direction. From [21, Theorem 1.1], the process $X^{\text{br},\theta}$ may be put in the form $X_s^{\text{br},\theta} = X_s^1 + X_s^2, s \geq 0$, where X^1 is a Lévy process on $[0, \infty)$ and X^2 has exchangeable increments on $[0, 1]$ and in a certain sense behaves less wildly than X^1 . Precisely, X^1 has no drift, its Gaussian part is θ_0 and its Lévy measure is $\Lambda(dx) = \sum_{i \geq 1} \delta_{\theta_i}(dx)$, where $\delta_y(dx)$ is the Dirac mass at y . On the other hand, X^2 can be put in the form

$$X_s^2 = -X_1^1 s + \sum_{i \geq 1} \tau_i (\mathbb{1}_{\{s \leq V_i\}} - s)$$

for some square-summable random family (τ_i) and a sequence V_i of independent r.v.'s with uniform law (notice that X^1 and X^2 are by no means independent). Then, writing $\kappa_{X^{\text{br},\theta}} = \inf\{c > 0 : \sum_{i \geq 1} \theta_i^c < \infty\}$ and $\kappa_{X^2} = \inf\{c > 0 : \sum_{i \geq 1} \tau_i^c < \infty\}$ we have that

$$\kappa_{X^2} \leq \frac{\kappa_{X^{\text{br},\theta}}}{1 + \frac{1}{2}\kappa_{X^{\text{br},\theta}}}, \quad (49)$$

which is what we mean by “behaving less wildly”. It is therefore reasonable that the problem on the finiteness of the integral $\int^1 ds/X_s^\theta$, which is a problem dealing with the behavior at the left of the overall minimum of $X^{\text{br},\theta}$, should be replaced by a problem on the Lévy process X^1 as soon as one can show that the overall minimum of $X^{\text{br},\theta}$ is actually attained at a local minimum of X^1 , and such that locally X^2 is negligible compared to X^1 at this time. Since X^1 has no negative jumps, the time-reversed process has no positive jumps, and such questions are addressed in Bertoin [11] and Millar [29]. Pushing the intuition one step further, by analogy with the standard criterion for non-extinction of continuous-state branching processes and the analogy of ICRT’s and Lévy trees mentioned above, we conjecture that $\int^\infty \Psi^{-1}(\lambda) d\lambda < \infty$ is equivalent to the boundedness of \mathcal{T}^θ , where Ψ is the Laplace exponent of X^1 :

$$\Psi(\lambda) = \theta_0 \lambda^2 / 2 + \sum_{i \geq 1} (\exp(-\lambda \theta_i) - 1 + \lambda \theta_i).$$

2. As we mentioned before, a natural guess would be that the exploration process of \mathcal{T}^θ in the general case $\theta_0 > 0$ is $\frac{2}{\theta_0} Y^\theta$. It is more difficult to get an intuition of what the exploration process of \mathcal{T}^θ should be in the cases when $\theta_0 = 0$, when the Brownian part of X^θ vanishes. By the general theory of continuum random trees, it should be easy to prove that compactness of the tree is enough to obtain the existence of an exploration process for \mathcal{T}^θ , which is the weak limit of $2(\theta_0^n)^{-2} Y^{\theta^n}$ for some $\theta^n \in \Theta \rightarrow \theta$ pointwise with $\theta_0^n > 0$ for every n . But this would not tell much about the look of this process. Another way would be to try to generalize local time methods used in [17], but these do not seem to adapt so easily to bridges with exchangeable increments instead of Lévy processes.

Acknowledgment. Thanks to an anonymous referee for a careful reading of a former version of the paper.

References

- [1] D.J. Aldous: The continuum random tree I. *Ann. Probab.* **19**, 1–28 (1991). MR1085326
- [2] D.J. Aldous: The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pp. 23–70. Cambridge University Press (1991). MR1166406
- [3] D.J. Aldous: The continuum random tree III. *Ann. Probab.* **21**, 248–289 (1993). MR1207226
- [4] D.J. Aldous, G. Miermont and J. Pitman: Brownian Bridge Asymptotics for Random p -Mappings. *Electr. J. Probab.* **9**, 37–56, (2004). Article
- [5] D.J. Aldous, G. Miermont and J. Pitman: Weak Convergence of Random p -Mappings and the Exploration Process of the Inhomogeneous Continuum Random Tree. [arXiv:math.PR/0401115](https://arxiv.org/abs/math.PR/0401115)
- [6] D.J. Aldous and J. Pitman: Brownian bridge asymptotics for random mappings. *Random Structures Algorithms* **5**, 487–512 (1994). MR1293075
- [7] D.J. Aldous and J. Pitman: The standard additive coalescent. *Ann. Probab* **26**, 1703–1726 (1998) MR1675063
- [8] D.J. Aldous and J. Pitman: A family of random trees with random edge-lengths. *Random Structures Algorithms* **15** 176–195 (1999). MR1704343
- [9] D.J. Aldous and J. Pitman: Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Th. Rel. Fields* **118**, 455–482 (2000). MR1808372
- [10] D.J. Aldous and J. Pitman: Invariance principles for non-uniform random mappings and trees. In V. Malyshev and A.M. Vershik, editors, *Asymptotic Combinatorics with Applications in Mathematical Physics*, pp. 113–147. Kluwer Academic Publishers (2002). MR1999358
- [11] J. Bertoin: On the local rate of growth of Lévy processes with no positive jumps. *Stoch. Proc. App.* **55**, 91–100 (1995) MR1312150
- [12] J. Bertoin: A fragmentation process connected to Brownian motion. *Probab. Theory Relat. Fields* **117**, 289–301 (2000). MR1771665
- [13] J. Bertoin: Eternal additive coalescent and certain bridges with exchangeable increments. *Ann. Probab.* **29**, 344–360 (2001). MR1825153
- [14] P. Biane and M. Yor: Sur la loi des temps locaux browniens pris en un temps exponentiel. In *Séminaire de Probabilités XXII*, pp. 454–466. *Lecture Notes in Math.* **1321**, Springer (1988). MR960541

- [15] M. Camarri and J. Pitman: Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.* **5**, Paper 2, 1–18 (2000). MR1741774
- [16] P. Chassaing and G. Louchard: Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures Algorithms* **21**, 76–119 (2002). MR1913079
- [17] T. Duquesne and J.-F. Le Gall: Random trees, Lévy processes and spatial branching processes. *Astérisque* **281** (2002). MR1954248
- [18] R. Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition (1996). MR1609153
- [19] T. Jeulin: Application de la théorie du grossissement à l'étude des temps locaux browniens. In T. Jeulin and M. Yor (eds.): *Grossissements de filtrations: exemples et applications. Lecture notes in Maths* **1118**, Springer, Berlin (1985). MR884713
- [20] O. Kallenberg: Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrsch. Verw. Gebiete* **27**, 23–36 (1973). MR394842
- [21] O. Kallenberg: Path properties for processes with independent and interchangeable increments. *Z. Wahrsch. Verw. Gebiete* **28**, 257–271 (1974). MR402901
- [22] G. Kersting: On the height profile of a conditioned Galton-Watson tree. Preprint (1998).
- [23] F. B. Knight: The uniform law for exchangeable and Lévy process bridges. Hommage à P. A. Meyer et J. Neveu. *Astérisque* **236**, 171–188 (1996). MR1417982
- [24] J.-F. Le Gall: The uniform tree in a Brownian excursion. *Probab. Th. Rel. Fields* **96**, 369–383 (1993). MR1231930
- [25] J.-F. Le Gall: Brownian excursions, trees and measure-valued branching processes. *Ann. Probab.* **19**, 1299–1439 (1991). MR1127710
- [26] J.-F. Le Gall and Y. Le Jan: Branching processes in Lévy processes: The exploration process. *Ann. Probab.* **26**, 213–252 (1998). MR1617047
- [27] J.-F. Marckert and A. Mokkadem: The depth-first processes of Galton-Watson trees converge to the same Brownian excursion. *Ann. Probab.* **31**, 1655–1678 (2003). MR1989446
- [28] G. Miermont: Ordered additive coalescent and fragmentations associated to Lévy processes with no positive jumps. *Electr. J. Probab.* **6**, paper 14 (33 pages) (2001). Article MR1844511
- [29] P.W. Millar: Comparison theorems for sample function growth. *Ann. Probab.* **9**, 330–334 (1981). MR606997
- [30] C.A. O’Cinneide and A.V. Pokrovskii: Nonuniform random transformations. *Ann. Appl. Probab.* **10** (4), 1151–1181 (2000). MR1810869

- [31] J. Pitman: Random mappings, forests and subsets associated with Abel-Cayley-Hurwitz multinomial expansions. *Séminaire Lotharingien de Combinatoire*, Issue 46, 45 pp. (2001). Article MR1877634
- [32] J. Pitman and M. Yor: Arcsine laws and interval partitions derived from a stable subordinator. *Proc. London Math. Soc.* (3) **65**, 326–356 (1992). MR1168191
- [33] D. Revuz and M. Yor: *Continuous martingales and Brownian motion*. Springer, Berlin-Heidelberg, third edition (1999). MR1725357
- [34] W. Vervaat: A relation between Brownian bridge and Brownian excursion. *Ann. Probab.* **7**, 143–149 (1979). MR515820