Abstract
Consider an evolving population, with genealogy given by a \( \Lambda \)-coalescent that comes down from infinity. We provide rather explicit sampling formulae under this model, for large samples. More precisely, we describe the asymptotic behavior of the site and allele frequency spectrum and the number of segregating sites, as the sample size tends to \( \infty \). A regular variation condition on the driving measure \( \Lambda \) is assumed for some of the almost sure asymptotic results, but most of our results are valid for a general \( \Lambda \)-coalescent that comes down from infinity. The proofs rely in part on the recent analysis of the speed of coming down from infinity for \( \Lambda \)-coalescents, done by the authors in [4]. The second goal of this paper is to investigate a remarkable connection between \( \Lambda \)-coalescents and genealogies of continuous-state branching processes. Our particle representation and the resulting coupling construction offer new perspective on the speed of coming down from infinity, and its consequences, as well as several other results recently obtained in the area.

AMS 2000 Subject Classification. 60J25, 60F99, 92D25

Key words and phrases. \( \Lambda \)-coalescents, exchangeable coalescents, sampling formulae, infinite allele model, genetic variation, particle system representation, continuous-state branching processes, Lévy processes, lookdown construction, Fleming-Viot processes

Running Head: Sampling formulae and particle representations

*Research supported in part by NSERC Discovery Grant, by Alfred P. Sloan Research Fellowship, and by ANR MAEV research grant
1 Introduction and main results

Coalescents with multiple collisions, also known as $\Lambda$-coalescents, are widely studied models for the genealogy of a sample of individuals taken from a given population. This is the class of coalescent processes that has the following properties: (a) time-homogeneous Markovian dynamics, (b) sampling consistency, (c) exchangeable dynamics, (d) well-mixed (i.e., spatially homogeneous, or mean-field) population, and (e) at most one coalescence event can occur at any given time. Relaxing the final assumption leads to the general exchangeable coalescent or $\Xi$-coalescent, discovered by Möhle and Sagitov [32] and Schweinsberg [38].

The following question arises naturally in the context of population genetics: assuming that the genealogy of a sample is suitably described by a $\Lambda$-coalescent, what kind of genetic variation do we expect to see in this sample? A quite complete answer in the special case where only pairwise collisions are possible is classical and well-known as the Ewens sampling formula [20] for the Kingman coalescent [27]. More recently, partial results have been obtained by Berestycki et al. [6] in the particular case of Beta-coalescents (for a general overview of previous results on the subject, we refer the reader to Durrett [18] or Berestycki [7].)

The first main goal of this paper is to address the above question and give rather explicit asymptotic formulae (as the sample size increases to $\infty$) for quantifying the genetic variation. In Theorem [2] a deterministic asymptotic rate of growth for the number of segregating sites and the number of distinct alleles in a sample is obtained in complete generality (i.e.,
for arbitrary \( \Lambda \) such that the corresponding \( \Lambda \)-coalescent comes down from infinity. This asymptotic behavior in probability is proved under both the infinite sites model or the infinite alleles model (precise definitions will be given soon). Furthermore, the above convergence in probability is strengthened to the convergence almost surely, provided that the measure \( \Lambda \) satisfies an additional regular variation condition in the neighborhood of zero. Finally, Theorem 3 derives explicit almost sure asymptotic formulae for the frequency spectrum in the latter regularly varying case. This last result generalities previous work of Berestycki et al. [5, 6] and of Schweinsberg [39], both in the sense that the result is valid for more general measures, and in the sense that the convergence holds almost surely. Our methodology in this first part of the paper is different from that of [6], in particular, it is based on our recent work on the speed of coming down from infinity [4]. The asymptotics in probability for the number of distinct alleles in a sample, under the model where the genealogy is driven by the general (regular) \( \Xi \)-coalescent dynamics, was obtained in parallel by Limic [30] using a different method (see Remark 15).

The second main goal of this work is to explore in detail the link between \( \Lambda \)-coalescents and genealogies of continuous-state branching processes. In particular, we construct explicit couplings between \( \Lambda \)-coalescents and the genealogy of continuous-state branching processes (CSBP) using particle system representations based on the look-down construction. Apart from its interest from a purely theoretical point of view, this coupling gives a new understanding of the asymptotic form of the speed of coming down from infinity, and in turn of the above mentioned sampling formulae (see Theorem 8 and Section 4.2). The reader might also be interested to know the limitations of this technique as explained in Remark 9. Our analysis is based in part on estimating the classical upper and lower (regularity) indices of the corresponding spectrally positive Levy processes. Finally, the explicit coupling leads to a new and transparent argument for identifying a time changed genealogy of the \( \alpha \)-stable CSBP with a particular Beta-coalescent (see Section 4.4), where the original result is due to Birkner et al. [12]. The methodology in this second part of the paper has several points in common with [5, 6], where an analogous link between Beta-coalescents and \( \alpha \)-stable continuous-state branching processes was used.

In the rest of the paper, we denote by \( \Rightarrow \) the convergence in distribution, and by \( d \equiv \) the equivalence in distribution. In addition, we use the standard Bachmann-Landau notation \( \sim, O(\cdot), o(\cdot), \asymp \) for comparing asymptotic behavior of deterministic and stochastic functions and sequences.

### 1.1 Mutation models

We now describe the underlying framework for the sampling results in more detail. Consider a sample of \( n \) individuals, where \( n \) is a fixed number tending to infinity. Assume that the genealogical relationship between these individuals is given by a \( \Lambda \)-coalescent, where \( \Lambda \) is an arbitrary finite measure on \([0,1]\) (precise definitions will be given later). In order to discuss genetic variation, we need to specify a mutation model. The two most widely used models are the \textit{infinite sites model} and the \textit{infinite alleles model}. To familiarize oneself with these models, it is also useful to think in terms of the forward-in-time evolution dynamics for the whole population (and not only in terms of the backward-in-time coalescent dynamics).

In the classical \textit{infinite sites model}, introduced by Kimura [25] in 1969, any individual is affected by neutral mutations at constant rate \( \theta > 0 \). Here it is also assumed that the number of loci (the size of the genome) is large, so that each mutation occurs at a new locus.
In particular, if an individual is affected by a mutation, then all the descendants of this individual carry this mutation (see Figure 1). Conversely, the genetic type of any individual in the sample depends on the entire history of its ancestral lineage. We denote by $S_n$ the number of segregating sites, or the total number of distinct genetic types in the sense just described.

The infinite alleles model is somewhat similar with one essential difference. As above, any individual is affected by a mutation at constant rate $\theta$. However, it is now assumed that every mutation changes the allelic type of the individual into something new, distinct from anything else (already seen or yet unseen) in the population. Thus, the allelic type of an individual in the sample is entirely determined by the most recent mutation affecting the corresponding ancestral lineage. The allelic partition is the partition of the sample (represented by a partition of $\{1, \ldots, n\}$) obtained by identifying the individuals that carry the same allelic type. Denote by $A_n$ the number of blocks in this partition, or equivalently, the total number of allelic types expressed in the sample.

The above random variables can be realized in a natural way on a common probability space as follows. Consider a $\Lambda$-coalescent ($\Pi_t, t \geq 0$) that comes down from infinity, and let $T$ be the associated coalescent tree. Then $T$ is a tree with infinitely many leaves $1, 2, \ldots$, and the root given by the most recent common ancestor between all the individuals. Each branch of $T$ is endowed with a positive number, its length or the size of the interval of time that elapsed between the two defining coalescent events for this branch (the one that started and the one that ended it). Let $P$ be a Poisson process of mutations on the branches of $T$, where the intensity of mutations is constant and equal to $\theta$ per unit length. Restricting $T$ to the first $n$ leaves produces a tree (even if the $\Lambda$-coalescent does not come down from infinity), denoted by $T_n$, that has the law generated by the same $\Lambda$-coalescent started from $n$ particles. The restriction of $P$ to $T_n$ is identified as the mutation process on $T_n$, and it suffices for determining the values of $S_n$ and $A_n$. It is useful to note here that $T$ and $P$ alone determine, simultaneously for all $n$, the values of $A_n$ and $S_n$, as well as various related quantities to be introduced in the sequel. Moreover, thus induced coupling between $T_n$ and $T_m$, for $m < n$, is canonical from the sampling perspective, in that the mutations arriving onto $T_n$ also arrive onto $T_m$. On the asymptotically unlikely event $\{P \cap T_n = \emptyset\}$, we declare $A_n = S_n = 0$.

Figure 1: The genealogical tree $T_n$ for a sample of size $n = 6$. The mutations on the (vertical) branches of $T_n$ are indicated as dots. The dot encircled in black corresponds to the mutation which is not seen under the infinite alleles model. The dot encircled in gray will be referred to in Section 2. Thus we
have \( S_n = 5 \) while \( A_n = 4 \).

### 1.2 Sampling formulae

Let \( \Lambda \) be a finite measure on \([0,1]\). We will assume without further mention that \( \Lambda(1) = 0 \) (for reasons why this can be done without loss of full generality see any of \([4, 33, 37]\)).

**Definition 1.** We say that \( \Lambda \) has (strong) \( \alpha \)-regular variation at zero if \( \Lambda(dx) = f(x)dx \) where \( f(x) \sim Ax^{1-\alpha} \) as \( x \to 0 \) for some \( 1 < \alpha < 2 \) and \( A > 0 \).

For any given finite measure \( \Lambda \) on \([0,1]\), associate a function \( \psi_\Lambda = \psi \) defined by

\[
\psi(q) := \int_{[0,1]} (e^{-qx} - 1 + qx)x^{-2}\Lambda(dx). \tag{1}
\]

In this paper we will usually require that

\[
\int_1^\infty \frac{dq}{\psi(q)} < \infty, \tag{2}
\]

which is known as Grey’s condition. As we will discuss soon (and as was already proved by Bertoin and Le Gall \([11]\)), this is equivalent to the requirement that the \( \Lambda \)-coallescent comes down from infinity. One can check (see e.g. \([21]\) XIII.6) that in the case of strong \( \alpha \)-regular variation,

\[
\psi(q) \sim A\Gamma(2-\alpha)\alpha^{\alpha}(\alpha-1)q^{\alpha}, \text{ as } q \to \infty, \tag{3}
\]

where \( A \) is the constant from Definition 1 so in particular Grey’s condition holds if \( \alpha \in (1,2) \).

Our first result concerns the asymptotic behavior of the number \( S_n \) of segregating sites and the size \( A_n \) of the allelic partition.

**Theorem 2.** Assume that Grey’s condition holds. Let \( X_n \) denote either \( A_n \) or \( S_n \). Then

\[
\frac{X_n}{\int_1^n q\psi(q)^{-1}dq} \to \theta, \tag{4}
\]

in probability as \( n \to \infty \). Moreover, if \( \Lambda \) has (strong) \( \alpha \)-regular variation at zero, then the above convergence holds almost surely, implying

\[
n^{\alpha-2}X_n \to \theta B, \text{ almost surely}, \tag{5}
\]

where \( B = B(A, \alpha) := \alpha(\alpha-1)/[\Gamma(2-\alpha)(2-\alpha)] \).

The above results have corollaries for the asymptotic frequency spectrum both in the infinite sites model and in the infinite alleles model. For each \( n \) consider a sample of size \( n \), and for each \( k \in \{1, \ldots, n\} \), let \( F_{k,n} \) be the number of families of size \( k \) in its allelic partition, and \( M_{n,k} \) be the number of mutations affecting precisely \( k \) of its individuals under the infinite sites model.
Theorem 3. Suppose that $\Lambda$ has (strong) $\alpha$-regular variation at zero. Recall the constant $B \equiv B(\alpha, A)$ from Theorem 2, and define $C = B \cdot (2 - \alpha)/\Gamma(\alpha - 1)$. Let $X_{k,n}$ denote $M_{k,n}$ or $F_{k,n}$, where $1 \leq k \leq n$. As $n \to \infty$,

$$
\frac{X_{k,n}}{n^{2-\alpha}} \to \theta B \frac{(2 - \alpha)\Gamma(k + \alpha - 2)}{k!}, \text{ a.s.}
$$

(6)

Moreover, if $P_1, P_2, \ldots$ are the ordered allele frequencies in the population, then

$$
P_j \sim C_j^{\alpha-2},
$$

(7)

almost surely as $j \to \infty$, and $C = (B/\Gamma(\alpha - 1))^{1/(2-\alpha)}$.

By the properties of the Gamma function, another expression for the constant on the right-hand side of (6) is

$$
\theta B (2 - \alpha) (\alpha - 1) \ldots (\alpha + k - 3) \frac{k!}{k^2}.
$$

As mentioned in the Introduction, the above results are improvements over previously known results, since the convergence (6) was known to hold only in probability in the case of Beta-coalescents (see [6]), while (7) was not known to hold even in this special case.

The following central limit theorem, obtained here with almost no additional work, could be useful for future statistical applications.

Theorem 4. As $n \to \infty$, the infinite vector

$$
\left( \frac{F_{i,n} - \mathbb{E}(F_{i,n})}{n^{2-\alpha} \theta B^{1/2}}, j = 1, 2, \ldots \right)
$$

converges, in the sense of finite-dimensional distributions, to a multivariate Gaussian array with zero mean and covariance matrix

$$
\sigma_{r,s} = \frac{(2 - \alpha)\Gamma(r + s + \alpha - 2)}{r!s!} 2^{2-r-s-\alpha}, r \neq s
$$

$$
\sigma_{r,r} = -\frac{(2 - \alpha)\Gamma(2r + \alpha - 2)}{r!} 2^{2(r-1)-\alpha} + \frac{(2 - \alpha)\Gamma(r + \alpha - 2)}{r!}.
$$

The proofs have two main ingredients. The first one is the following proposition which studies the asymptotics of the law and the mean for the age of a randomly chosen (typical) mutation. More precisely, let the time run in the “coalescent sense”, so that the leaves of the tree are present at time 0, and the number of branches decreases in time. Denote by $M_n$ the time-coordinate (age) of a point chosen at random from $\mathcal{P} \cap T_n$. On the (asymptotically unlikely) event \{\mathcal{P} \cap T_n = \emptyset\} = \{S_n = A_n = 0\}, we set $M_n = 0$ (although this value could be set to anything between 0 and the time of MRCA (the root of $T_n$) and the next result would still be true). Define

$$
g(n) = \begin{cases} 
n^{1-\alpha}, & \text{if } 1 < \alpha < 3/2, \\
n^{-1/2} \log n, & \text{if } \alpha = 3/2, \\
n^{\alpha-2}, & \text{if } 3/2 < \alpha < 2.
\end{cases}
$$

(8)
Proposition 5. Suppose that $\Lambda$ has (strong) $\alpha$-regular variation at zero, for some $\alpha \in (1, 2)$.
(a) We have
\[ \frac{M_n}{n^{1-\alpha}} \Rightarrow \frac{\alpha}{\Gamma(2-\alpha)} \left( U^{-(\alpha-1)/(2-\alpha)} - 1 \right), \quad \text{where } U \overset{d}{=} \text{Unif}[0,1], \] (9)
(b) If in addition $\Lambda[1-\eta, 1] = 0$ for some $\eta > 0$, then there exist $c_1 \equiv c_1(\alpha) \in (0, \infty)$, such that for $g$ given by (8)
\[ \lim_{n \to \infty} \frac{\mathbb{E}(M_n)}{g(n)} = c_1. \] (10)

Remark 6. One can guess that $\alpha = 3/2$ will be “critical” already from (9), the right hand side is integrable random variable if and only if $\alpha < 3/2$. Interestingly, $g(n)$ observed as a function of $\alpha$ decreases on $(1, 3/2)$ and increases on $(3/2, 2)$. It seems difficult to see intuitively why this happens.

Remark 7. We believe that the result (10) should hold without any further restriction on $\Lambda$ than strong regular variation. The techniques used in the proof of (10) can be used to show, with some additional effort, that the sequence $\mathbb{E}(M_n)/g(n)$ is bounded away from 0 and infinity when no assumption is made on the support of $\Lambda$. However, in the interest of brevity we decided to omit these arguments.

The second ingredient of the proof of Theorem 3 is a Tauberian theorem for random partitions of Gnedin, Hansen and Pitman [23]. The assumptions of this theorem were recently extended in an independent but related work of Schweinsberg [39], to deal with convergence in probability (to which the approach of [23] could not apply), allowing him to obtain the convergence in probability of Theorem 3 for the limiting behavior of $F_{k,n}$ (but not for that of $M_{k,n}$).

1.3 Coupling and asymptotic for the number of blocks
A key component in the proofs of Theorems 2 and 3 is a good understanding of the asymptotics for the number of blocks in the underlying $\Lambda$-coalescent at small times. Theorem 1 in [4] states the following:
\[ \lim_{t \to 0} \frac{N^\Lambda(t)}{v(t)} = 1, \quad \text{almost surely}, \] (11)
where $N^\Lambda(t)$ is the number of blocks of $\Pi_t$ and where $v(t) = \inf\{z > 0 : \int_z^\infty \psi(q)^{-1}dq < t\}$, or equivalently, where
\[ \int_{v(t)}^\infty \frac{dq}{\psi(q)} = t. \] (12)
The relation (11) is truly informative only if the $\Lambda$-coalescent comes down from infinity, otherwise, both the numerator and the denominator are infinite at all times. The function $\psi$ may be interpreted as the Lévy exponent of a certain Lévy process. Somewhat remarkably, the function $v$ also has an interpretation in terms of a continuous-state branching process (CSBP) for which $\psi$ is the branching mechanism. As already said in the Introduction, one of our goals is to give a probabilistic explanation of how and why $\psi$ and $v$ arise naturally in the description of the small-time behavior of $\Lambda$-coalescents. In turn, we obtain an intuitive explanation of the form of sampling formulae from Theorems 2 and 3. We note that Bertoin
and Le Gall [9] already speculated about the existence of such a connection, based on analytic computations.

The probabilistic explanation comes from a general coupling between CSBP and \( \Lambda \)-coalescents, that is valid in complete generality (regardless of (2)). It allows us to directly compare the processes \((N_Z(t), t \geq 0)\) and \((N_X(t), t \geq 0)\), where \(N_Z(t)\) is the number of families in the CSBP alive at time \(t\), and \(N_X(t)\) is the number of distinct types in the corresponding generalized Fleming-Viot process. This later class of measure-valued processes was introduced by Bertoin and Le Gall [9], where it was proven that any such process is dual to a particular \( \Lambda \)-coalescent (precise definitions and further details are provided in Section 3.2). As a consequence, we have that \(N_X(t) \overset{d}{=} N_\Lambda(t)\), for each fixed \(t \geq 0\). This together with the coupling construction of Section 4.1 yields the following result.

**Theorem 8.** For each \( \varepsilon \in (0, 1) \),

\[
\mathbb{P} \left( \liminf_{t \to 0} \frac{N_X(t)}{v \left( \frac{1 + \varepsilon}{1 - \varepsilon} t \right)} \geq \frac{1}{1 + \varepsilon}, \quad \limsup_{t \to 0} \frac{N_X(t)}{v \left( \frac{1 - \varepsilon}{1 + \varepsilon} t \right)} \leq \frac{1}{1 - \varepsilon} \right) = 1, \tag{13}
\]

and therefore

\[
\lim_{t \to 0} \mathbb{P} \left[ \frac{1}{(1 + \varepsilon)^2} \cdot v \left( \frac{1 + \varepsilon}{1 - \varepsilon} t \right) \leq N_\Lambda(t) \leq \frac{1}{(1 - \varepsilon)^2} \cdot v \left( \frac{1 - \varepsilon}{1 + \varepsilon} t \right) \right] = 1. \tag{14}
\]

**Remark 9.** Since \(N_X\) and \(N_\Lambda\) have only equal marginal distributions, but they are not equal in distribution as processes (while the first one only decreases by jumps of size 1, at least in the stable case, the second one can decrease by jumps of arbitrary integral length), one cannot obtain more than (14) from (13). This result is clearly weaker than (11), that was obtained in [4] via a sophisticated martingale technique. However, it was the knowledge of this coupling that initiated [4] and suggested the form of the asymptotics in the first place.

Another interesting consequence of the same coupling is that it cast a new light on a result of Birkner et al. [12], connecting Beta-coalescents and \( \alpha \)-stable continuous branching processes by putting it in a more general context. This result was core to the analysis in [5, 6], and we believe that the present approach offers some new insights. Along the way to proving Theorem 8, we will obtain a few related results, some of which are stated next.

**Corollary 10.** The \( \Lambda \)-coalescent comes down from infinity if and only if \(Z\) becomes extinct in finite time. Both are equivalent to (2).

We now briefly introduce some additional concepts related to the regularity of Lévy processes, for rigorous definitions see (77)-(78) in Section 4.3 or Pruitt [35], display (3.3). Given \(X\) a Lévy process with Laplace exponent \( \psi \), denote by \( \delta \) and \( \beta \) its lower and upper index, respectively. Intuitively, \( \beta \) and \( \delta \) indicate the regularity of \(X\) in the sense that if \(M_t := \max_{s \leq t} |X_s - X_0|\), then \(M_t\) oscillates (see, e.g., display (3.4) in [35] or Section 4.3 for a precise statement) between \(t^{1/\delta}\) and \(t^{1/\beta}\) approximately for \(t\) small. We recall that \(0 \leq \delta \leq \beta \leq 2\) (see [35], p. 952 or Section 4.3). As the function \(\psi = \psi_\Lambda\) is the Laplace exponent of a Lévy process, one can associate the lower and upper index \( \delta \) and \( \beta \) directly to each \( \Lambda \), and in turn to each \( \Lambda \)-coalescent.
Corollary 11. If the lower-index \( \delta \) is strictly greater than 1, then for any \( \varepsilon > 0 \),

\[
\frac{N^\Lambda(t)}{t^{-1/(\beta+\varepsilon-1)}} \to \infty, \quad \text{in probability},
\]

and, for any \( \varepsilon \in (0, \delta - 1) \)

\[
\frac{N^\Lambda(t)}{t^{-1/((\delta-\varepsilon)-1)}} \to 0, \quad \text{in probability}.
\]

Remark 12. When Grey’s condition for extinction holds, it is not hard to see (and will be proved in Lemma 31) that \( \beta \geq 1 \). However, there are examples of Lévy measures \( \nu \) (cf. Section 5) such that both \( \beta > 1 \) and Grey’s condition does not hold (that is, the corresponding coalescent does not come down from infinity).

The asymptotics (11) in the sense of convergence in probability can be obtained from Theorem 8 under additional assumptions on \( v \) (that is, on \( \Lambda \)) as the following result shows.

Proposition 13. Assume \( \Lambda\{\{0\}\} = 0 \). Then, the convergence

\[
N^\Lambda(t)/v(t) \to 1 \quad \text{in probability}
\]

holds at least if

\[
\lim_{\varepsilon \to 0} \lim_{t \to 0} \sup_{n} v(t(1 - \varepsilon)) v(t) = 1, \quad \lim_{\varepsilon \to 0} \lim_{t \to 0} \inf_{n} v(t(1 + \varepsilon)) v(t) = 1,
\]

and, in particular, if

\[
\psi(v(t)) = O(v(t)/t), \quad \text{as } t \to 0.
\]

The rest of the paper is organized as follows. Section 2 is devoted to proving the results on the mutation frequency spectrum, announced in Section 1.1. Theorem 8 is proved in Section 4. The necessary particle-system tools are introduced in Section 3. We obtain Corollary 11 in Section 4.3 by studying the regularity of \( \psi \) and \( v \). Section 4.4 revisits the particular case of the Beta coalescent family. The appendix contains an example of measures \( \Lambda \) that are not “well-behaved”, in the sense that the corresponding \( \Lambda \)-coalescent comes down from infinity but the lower and the upper indices are different.

2 Proofs of the sampling formulae

Recall the setting of Section 1.1 and fix some \( \theta > 0 \). For each \( n \in \mathbb{N} \) and \( t \geq 0 \), let \( N^{\Lambda,n}(t) \) denote the number of ancestral lineages of the first \( n \) individuals remaining at time \( t \). In particular, \( (N^{\Lambda,n}(t), t \geq 0) \) is a continuous-time Markov jump process, starting from \( N^{\Lambda,n}(0) = n \). We assume throughout this section that (2) holds, or that equivalently, the \( \Lambda \)-coalescent comes down from infinity:

\[
N^\Lambda(t) := \lim_{n \to \infty} N^{\Lambda,n}(t) = \sup_{n \geq 1} N^{\Lambda,n}(t) < \infty, \quad \forall t > 0.
\]

We will need some further notations. Define for \( k, n \in \mathbb{N} \),

\[
\tau^n_k = \inf\{t \geq 0 : N^{\Lambda,n}(t) \leq k\}, \quad \text{and} \quad \tau^\infty_k = \lim_{n \to \infty} \tau^n_k = \inf\{t \geq 0 : N^\Lambda(t) \leq k\}.
\]
In particular $\tau_1^n = \inf\{t \geq 0 : N^A_{\Lambda n}(t) = 1\}$ is the time of the MRCA for the sample containing the first $n$ individuals, and $L_n = \int_{0}^{\tau_1^n} N^A_{\Lambda n}(t) \, dt$ is the total length of the tree $T_n$.

The genealogical tree $T$ is a path-connected set in $\mathbb{R}^2$. To any point $x$ on the tree one can associate a number $t = t(x)$ called the time-coordinate or the age of $x$, which is defined as the distance from that point to the set of leaves $\text{pf} T$. Let $T_n$ be the subtree of $T$ consisting of all the points in $\mathcal{T}$ having age in $[\tau_n, \tau_1]$. Then $L_n = \int_{\tau_n}^{\tau_1} N^A(u) \, du$ is the length of $T_n$. Finally, define $t_n$ as

$$t_n := \int_n^\infty \frac{dq}{\psi(q)} \equiv v^{-1}(n).$$

The following lemma gathers some asymptotic results which we will use in the rest of the proof. Set

$$\bar{\alpha} = (2 - \alpha)/(\alpha - 1) \in (0, \infty).$$

**Lemma 14.** Assume (3) and define $c = c(A, \alpha) = \alpha/(A\Gamma(2 - \alpha))$ (compare with the constant in (2)). Then, as $n \to \infty$, we have almost surely

1. $\tau_n \sim c n^{1-\alpha}$, and
2. $t_n \sim c n^{-\alpha}$.

Furthermore, there exist $c_1 = c(A) > 0$, and $c_2 = c(A, \alpha) \in (0, \infty)$, such that

3. $\mathbb{P}(\tau_1 > x) \leq e^{-c_1 x}$, for all $x \geq 1$,

4. $\hat{L}_n = \int_{\tau_n}^{\tau_1} N^A(u) \, du \sim \int_1^1 v(u) \, du \sim \frac{\bar{\alpha}}{\alpha} (\tau_n)^{-\bar{\alpha}} \sim \frac{\bar{\alpha}}{\alpha} n^{2-\alpha}$, a.s. as $n \to \infty$,

5. as $x \to 0$

$$\int_x^{\tau_1} u N^A(u) \, du \sim \begin{cases} \frac{c_2 x^{-\bar{\alpha}+1}}{\bar{\alpha}}, & \bar{\alpha} > 1, \\ \frac{c_2 \log(1/x)}{\bar{\alpha}}, & \bar{\alpha} = 1, \text{ a.s. and} \\ \frac{c_2}{\bar{\alpha}}, & \bar{\alpha} < 1, \end{cases}$$

where $Y := \int_0^{\tau_1} u N^A(u) \, du$ is a finite random variable if $\bar{\alpha} < 1$.

**Proof.** Recall (12). Theorem 1 in [4] and (3) yield

$$N^A(t) \sim v(t) \sim c_n^{1/\alpha} t^{-1/\alpha}, \text{ as } t \to 0, \text{ almost surely},$$

(22)

where $c = c(A, \alpha)$ is as specified above. The asymptotic behavior (22) implies that $N^A(\tau_n) = n(1 + o(1))$, almost surely, as $n \to \infty$. Indeed, since $\tau_n \to 0$, we have $N^A(\tau_n) \sim v(\tau_n) = v(\tau_n^{-}) \sim N^A(\tau_n^{-})$, and at the same time, $\mathbb{P}(N^A(\tau_n) \leq n < N^A(\tau_n^{-})) = 1$. Since, again due to (22) $N^A(\tau_n) \sim c_n^{1/\alpha - 1} (\tau_n)^{-\bar{\alpha} - 1}$, we obtain claim 1. Claim 2 is directly seen from (19) and the asymptotic behavior of $v$ in (22).

Claims 4 and 5 are derived similarly. One notes first that, due to $\mathbb{P}(\tau_1 > 0) = 1$ and (22),

$$\int_x^{\tau_1} N^A(u) \, du \sim \int_x^1 v(u) \, du \sim \frac{c^{1/\alpha - 1}}{\alpha} x^{-\bar{\alpha}} \text{ as } x \to 0,$$

and then uses the facts that $\tau_n \to 0$, $t_n \to 0$ as well as claims 1 and 2 to obtain claim 4. For claim 5, we use $u N^A(u) \sim c_n^{1/\alpha} u^{-\bar{\alpha} + 1/\alpha} = c_n^{1/\alpha} u^{-\bar{\alpha}}$, and this uniquely determines $c_2$. If $\bar{\alpha} < 1$, then both $\int_0^1 \psi(u) \, du$ and $\int_0^{\tau_1} u N^A(u) \, du$ are finite.
It remains to verify claim 3. Due to monotonicity of the coalescent and the simple Markov property, we have $\mathbb{P}(\tau_n^\alpha \geq m + 1|\tau_1^\alpha \geq m) \leq \mathbb{P}(\tau_1^\alpha \geq 1)$. In turn, letting $n \to \infty$,

$$\mathbb{P}(\tau_1 \geq m + 1|\tau_1 \geq m) \leq \mathbb{P}(\tau_1 \geq 1), \text{ for each } m \geq 1.$$ 

Hence, by induction, $\mathbb{P}(\tau_1 > m) \leq e^{-cm}$ for all $m \geq 0$, with $c_1 = \log \mathbb{P}(\tau_1 \geq 1)$. $\square$

2.1 Proof of Proposition 5

Let $A_1^n = \{S_n \geq 1\} = \{A_n \geq 1\}$. Since the length $L_n$ of $T_n$ diverges, we have that $\mathbb{P}(A_1^n) \to 1$, as $n \to \infty$. Recall that on $A_1^n$, $M_n$ is the age of a randomly chosen mutation in $T_n$, and that on the complement of $A_1^n$, $M_n$ is set to 0. Due to basic properties of Poisson point processes, on the event $A_1^n$ (of overwhelming probability), the random mutation is positioned as a point $P_n^*$ chosen uniformly at random from $n - 1$. Let $\alpha$ be the subtree $\hat{T}_n$ of $T$. Consider a uniform random point on $\hat{T}_n$ and let $\hat{M}_n$ be its age minus $\tau_n$. Let $\mathcal{E}_n = \{N_{\tau_n}^\Lambda = n\}$ be the event that $\Pi$ ever attains a configuration with exactly $n$ blocks. Due to the consistency property and the Markov property of $(\Pi_t, t \geq 0)$,

$$\mathbb{P}(\hat{M}_n \cdot n^{\alpha - 1} \leq x) = \mathbb{P}(M_n^* \cdot n^{\alpha - 1} \leq x) + O(\mathbb{P}(A_1^n)),$$ for each $x \geq 0$,

therefore (23) is equivalent to

$$\frac{M_n^*}{n^{1 - \alpha}} \Rightarrow c(U^{-(\alpha - 1)/(2 - \alpha)} - 1), \text{ where } U \overset{d}{=} \text{Unif}[0,1], \quad (24)$$

Hence we proceed by studying $M_n^*$.

Now recall the subtree $\hat{T}_n$ of $T$. Consider a uniform random point on $\hat{T}_n$ and let $\hat{M}_n$ be its age minus $\tau_n$. Let $\mathcal{E}_n = \{N_{\tau_n}^\Lambda = n\}$ be the event that $\Pi$ ever attains a configuration with exactly $n$ blocks. Due to the consistency property and the Markov property of $(\Pi_t, t \geq 0)$,

the conditional law of $\hat{T}_n$ given $\mathcal{F}_{\tau_n}$ on the event $\mathcal{E}_n$ equals the law of $T_n$. $\quad (25)$

This clearly induces the equivalence of the conditional law of $\hat{M}_n$ given $\mathcal{F}_{\tau_n}$ on $\mathcal{E}_n$ and the law of $M_n^*$, which will be used below. Recalling (18), we have

$$\mathbb{P}(\hat{M}_n \geq x|\hat{T}_n) = \frac{\int_{\tau_1 + x}^{\tau_1 + x} N_\Lambda(u) \, du}{L_n}, \quad x \in [0, \tau_1 - \tau_n].$$

So, recalling $\bar{\alpha}$ from (20), we have for a fixed $y \in (0,1)$

$$\mathbb{P}((\hat{M}_n/\tau_n + 1)^{-\bar{\alpha}} \leq y) = \mathbb{E}\left[\left.\int_{\tau_n y^{-1/\bar{\alpha}} \wedge \tau_1}^{\tau_1} N_\Lambda(u) \, du\right]/L_n\right]. \quad (26)$$

By the argument used to show claim 4 of Lemma 13 and $\mathbb{P}(\tau_1 > 0) = 1$. we conclude

$$\int_{\tau_n y^{-1/\bar{\alpha}} \wedge \tau_1}^{\tau_1} N_\Lambda(u) \, du \sim c' [\tau_n y^{-1/\bar{\alpha}} \wedge \tau_1]^{-1/(\alpha - 1) + 1} \sim c' [\tau_n y^{-1/\bar{\alpha}}]^{-1/(\alpha - 1) + 1} = c' \tau_n^{-\bar{\alpha}} y, \text{ a.s.}$$

11
as } n \to \infty \text{, where } c' = c^{1/(\alpha-1)}/\bar{\alpha}. \text{ Due to claim 4 of Lemma 14 we now see that the random variable inside the expectation on the RHS of (26) converges to } y, \text{ almost surely. Since }

\[ |\mathbb{E}[\mathbb{P}(\hat{M}_n/\tau_n + 1)^{-\bar{\alpha}} - y | \mathcal{F}_{\tau_n})1_{\mathcal{E}_n}] - y\mathbb{P}(\mathcal{E}_n)| \leq \mathbb{E} \left[ 1_{\mathcal{E}_n} \frac{\int_{\tau_n}^{T_1} \frac{y^{-\alpha} N^{\Lambda}(u)\,du}{L_n}}{y} \right], \]

and since the bounded (by 2) random variable inside the expectation converges to 0 almost surely, the dominated convergence theorem implies that the left-hand side converges to 0. Recalling (25), or its consequence for random points, this quantity can be rewritten as \( |\mathbb{P}((M_n^*/\tau_n + 1)^{-\bar{\alpha}} \leq y) - y\mathbb{P}(\mathcal{E}_n)|. \) Theorem 1.8 in [5] gives \( \mathbb{P}(\mathcal{E}_n) \to \alpha - 1 \) as \( n \to \infty, \) hence \( \mathbb{P}((M_n^*/\tau_n + 1)^{-\bar{\alpha}} \leq y) \to y, \) for all \( y \in [0, 1]. \) This, together with claim 1 of Lemma 14 implies (9).

The proof of (17) is analogous but technically more delicate. Due to \( \mathbb{P}(M_n^* \geq x|\mathcal{T}_n) = \int_{x}^{\tau_n} N^{\Lambda,n}(u)\,du/L_n \) and Fubini’s theorem, we have

\[ \mathbb{E}(M_n^*|\mathcal{T}_n) = \frac{\int_{0}^{\infty} \int_{\tau_n}^{T_1} N^{\Lambda,n}(u)\,du\,dx}{L_n} = \frac{\int_{0}^{T_1} \frac{u N^{\Lambda,n}(u)\,du}{\int_{\tau_n}^{T_1} N^{\Lambda,n}(u)\,du}}. \] (27)

Therefore, \( \mathbb{E}(M_n^*) = \mathbb{E}[Y_n], \) where

\[ Y_n = \frac{\int_{\tau_n}^{T_1} u N^{\Lambda,n}(u)\,du}{\int_{\tau_n}^{T_1} N^{\Lambda,n}(u)\,du}. \] (28)

Due to (25), the variable \( Y_n \) is equal in law to

\[ \hat{Y}_n = \frac{\int_{\tau_n}^{T_1} (u - \tau_n) N^{\Lambda}(u)\,du}{\int_{\tau_n}^{T_1} N^{\Lambda}(u)\,du} = \frac{\int_{\tau_n}^{T_1} u N^{\Lambda}(u)\,du}{\int_{\tau_n}^{T_1} N^{\Lambda}(u)\,du} - \tau_n, \text{ given } \mathcal{F}_{\tau_n}, \text{ on the event } \mathcal{E}_n. \] (29)

Using Lemma 14 one can analyze the asymptotic behavior of \( \hat{Y}_n/g(n) \) in each of the three cases \( \bar{\alpha} > 1, \bar{\alpha} < 1 \) and \( \bar{\alpha} = 1 \) (corresponding respectively to \( \alpha < 3/2, \alpha > 3/2 \) and \( \alpha = 3/2 \)). First observe that \( \tau_n/g(n) \to 0 \) almost surely if \( \alpha \geq 3/2, \) and otherwise \( \tau_n/g(n) \to c \) almost surely, where \( c \) is the constant from claim 1 of Lemma 14. One can apply claims 4 and 5 (plugging in \( \tau_n \) as \( x \)) of Lemma 14 to the first term in (29). More precisely, if we let \( c_3 = c_2\bar{\alpha}/c\bar{\alpha} \) and \( c_4 = \bar{\alpha}/c, \) then

1. if \( \bar{\alpha} > 1 \) then \( \int_{\tau_n}^{T_1} u N^{\Lambda}(u)\,du \sim c_2 \tau_n^{-\bar{\alpha}+1} \) so \( \hat{Y}_n/g(n) \sim c_3 - c. \) almost surely.

2. if \( \bar{\alpha} < 1 \) then \( \int_{\tau_n}^{T_1} u N^{\Lambda}(u)\,du \sim Y \) so \( \hat{Y}_n/g(n) \sim c_4 Y, \) almost surely.

3. if \( \bar{\alpha} = 1 \) then \( \int_{\tau_n}^{T_1} u N^{\Lambda}(u)\,du \sim -c_2 \log \tau_n \sim c_2 (\alpha - 1) \log n \) so \( \hat{Y}_n/g(n) \sim (\alpha - 1) c_2 / c, \) almost surely.

With a slight abuse of notation, let \( Y := \lim_{n \to \infty} \hat{Y}_n/g(n), \) almost surely. Clearly \( \mathbb{P}(Y \geq 0) = 1. \) Denote by \( \mathcal{D}_c(Y) \) the set of points of continuity for the distribution function of \( Y. \) Then for any \( x > 0 \) in \( \mathcal{D}_c(Y) \) and any sequence \( (B_n)_{n \geq 1} \) of events, where \( B_n \in \mathcal{F}_{\tau_n}, \) \( n \geq 1, \) we have

\[ \mathbb{E}(1_{B_n} \mathbb{E}(1_{\{\hat{Y}_n/g(n) \leq x\}} - 1_{\{Y \leq x\}} | \mathcal{F}_{\tau_n})) = o(1), \text{ as } n \to \infty. \] (30)
We claim that
\[ \frac{Y_n}{g(n)} \Rightarrow Y, \text{ as } n \to \infty, \]  
which can be verified as follows. Note that, due to (25), for each fixed \( x > 0 \)
\[ \mathbb{P} \left( \frac{Y_n}{g(n)} \leq x \right) = \mathbb{E} \left[ 1_{\mathcal{E}_n} \mathbb{P}(\hat{Y}_n/g(n) \leq x | \mathcal{F}_{r_n}) \right] / \mathbb{P}(\mathcal{E}_n) \]
Backward martingale convergence and measurability imply \( Y \in \mathcal{F}_0 \) imply \( \lim_n \mathbb{P}(Y \leq x | \mathcal{F}_{r_n}) = \mathbb{P}(Y \leq x) \). Combined with (30) and the fact \( \liminf_n \mathbb{P}(\mathcal{E}_n) > 0 \), this gives \( \lim_n \mathbb{P} \left( \frac{Y_n}{g(n)} \leq x \right) = \mathbb{P}(Y \leq x) \), for each \( n \in \mathcal{D}_c(Y) \), or equivalently, the convergence (31).

To conclude (10) from (31), it thus suffices to show that \( (Y_n/g(n))_{n \geq 1} \) is a uniformly integrable family. In fact we will now show that this family is uniformly bounded in \( L^2 \). Due to (28), we have \( \mathbb{P}(Y_n \leq \tau^n_1) = 1 \), and in particular
\[ Y_n 1_{\{\tau^n_1 \leq g(n)\}} \leq g(n), \] almost surely.

Due to claims 4 and 5 of Lemma 14 we know that \( g(n) \sim c \int_1^n u N^{\Lambda,n}(u) du \int_{t_n}^1 v(u) du \) for some \( c = 1/C \in (0, \infty) \). Therefore
\[
\frac{Y_n}{g(n)} \leq C \frac{\int_{t_n}^{\tau^n_1} u N^{\Lambda,n}(u) du}{\int_{t_n}^1 u v(u) du} \cdot \frac{\int_{t_n}^1 v(u) du}{\int_{t_n}^{\tau^n_1} N^{\Lambda,n}(u) du} 1_{\{\tau^n_1 \leq g(n)\}} + 1. \tag{32}
\]
Denote by \( A_n \) (resp, \( B_n \)) the first (resp, second) ratio on the LHS of (32), so that \( \frac{Y_n}{g(n)} \leq C A_n \cdot B_n 1_{\{\tau^n_1 \leq g(n)\}} + 1 \). We will bound separately the terms \( A_n \) and \( B_n 1_{\{\tau^n_1 \leq g(n)\}} \). Let \( b_n := \min(1 - t_n, \tau^n_1) \leq \tau_1 \) (note that \( b_n \to 1 \wedge \tau_1 \) a.s.), choose some \( k_0 \) such that \( t_{k_0} < 1/2 \), and henceforth assume WLOG that \( n \geq k_0 \). Then \( t_n \leq 1/2 \) so that
\[
A_n \leq \left( \frac{\int_{t_n}^{b_n} u N^{\Lambda,n}(u) du}{\int_{t_n}^1 (u + t_n) v(u + t_n) du} + \frac{\int_{b_n}^{\tau_1} u N^{\Lambda,n}(u) du}{\int_{1/2}^1 v(u) du} \right). \]
For \( h_1, h_2 \) two strictly positive integrable functions over some interval \([a, b]\) we always have that \( \frac{\int_{a}^{b} h_1(u) du}{\int_{a}^{b} h_2(u) du} \leq \sup_{u \in [a, b]} \frac{h_1(u)}{h_2(u)} \). Therefore
\[
A_n \leq \sup_{u \in [0, b_n]} \frac{N^{\Lambda,n}(u)}{v(t_n + u)} + O \left( \int_{b_n}^{\tau_1} u N^{\Lambda,n}(u) du \right) \\
= \sup_{u \in [0, 1]} \frac{N^{\Lambda,n}(u)}{v(t_n + u)} + O \left( N^{\Lambda}(1/2) \tau_1^2 \right), \tag{33}
\]
due to \( b_n \leq 1 \) and \( N^{\Lambda,n}(b_n) = 1_{\{b_n=\tau^n_1-<t_n\}} + N^{\Lambda,n}(1-t_n)1_{\{b_n=1-t_n<\tau^n_1\}} \leq N^{\Lambda}(1/2) \), a.s. Similarly we have
\[
B_n 1_{\{\tau^n_1 \leq g(n)\}} \leq \sup_{u \in [0, \tau_1]} \frac{v(t_n + u)}{N^{\Lambda,n}(u)} \cdot \frac{\int_{t_n}^{\tau_1} v(u + t_n) du}{\int_{t_n}^1 v(u + t_n) du} 1_{\{\tau^n_1 \leq g(n)\}} \leq \sup_{u \in [0, \tau_1]} \frac{v(t_n + u)}{N^{\Lambda,n}(u)} \cdot \frac{\int_{t_n}^{\tau_1} v(u) du}{\int_{t_n}^{t_n+g(n)} v(u) du},
\]
Observe that due to $t_n \sim cn^{1-\alpha}$ (claim 1 of Lemma [14]) we have $t_n = o(g(n))$ if $\alpha \geq 3/2$, and $t_n/g(n) \to c' \in (0, \infty)$ if $\alpha < 3/2$. Due to the asymptotic form (22) for $v$, the sequence $(\int_{t_n}^{t_n + g(n)} v(u)du)_{n \geq k_0}$ is uniformly bounded. We conclude that

$$B_n 1_{\{\tau_n^1 > g(n)\}} = O \left( \sup_{u \in [0, \tau_1]} \frac{v(t_n + u)}{N^{\Lambda,n}(u)} \right),$$

Combining (32)–(34), we obtain

$$Y_n \geq g(n) = O \left[ \left( \sup_{u \in [0, 1]} N^{\Lambda,n}(u) + N^{\Lambda}(1/2)(\tau_1)^2 \right) \left( \sup_{s \in [0, \tau_1]} \frac{v(t_n + u)}{N^{\Lambda,n}(u)} \right) \right] + 1, \forall n \geq k_0.$$

By the Cauchy-Schwarz inequality, it suffices to show that, for all $n$ sufficiently large, each factor in the brackets above is bounded in $L^1$. This follows immediately from:

(i) [4] Eq. (38) applied with $s = 1$ and $n \geq n_0 \vee k_0$ (see above (33) in [4] for the definition of $r(x; s)$, Lemma 19 in [4] for the choice of $n_0$, and note that this step uses the condition $\Lambda[1 - \eta, 1] = 0$ for some $\eta > 0$),

(ii) $N^{\Lambda}(1/2) \in L^p$ for all $p \geq 1$ (a consequence of Theorem 2 in [4]).

(iii) claim 3 of Lemma [14]

This proves the uniform integrability of $(Y_n/g(n))_{n \geq 1}$, and completes the proof of (10).

### 2.2 Proof of Theorem [2]

Recall the construction of Section [14] where the genealogy with mutations is realized for all $n$ simultaneously, with nice monotonicity properties. In this and the next subsection we will often refer to it under the name the full genealogy (construction or coupling).

**Case $X_n = S_n$.** For each $s > 0$ we have, due to Theorem 5 in [4],

$$\lim_{n \to \infty} \frac{\int_0^s N^{\Lambda,n}(t) dt}{\int_0^s v(t_n + t) dt} = 1, \text{ in probability.}$$

(35)

For the Kingman and the regular variation coalescents (see Definition [1]) the above convergence holds almost surely. The total length of $T_n$ is $L_n = \int_0^{\tau_n^1} N^{\Lambda,n}(t) dt$. Observe that $| \int_1^{\tau_n^1} N^{\Lambda,n}(u)du/\int_0^{\tau_n^1} v(t_n + t) dt | \to 0$ almost surely since $| \int_1^{\tau_n^1} N^{\Lambda,n}(u)du | < \infty$ a.s., and $\int_0^{\tau_n^1} v(t_n + t) dt$ diverges in $n$. Applying (35) with $s = 1$, we deduce that

$$L_n \sim \int_0^1 v(t_n + t) dt$$

in probability (i.e., the ratio of the two sides tends to 1 in probability), and almost surely in the regular variation case. Using the facts that $v(t_n) = n$ and $v'(q) = -\psi(v(q))$ for all $q > 0$, and applying a change of variables $q = v(t)$, we obtain

$$\int_0^1 v(t_n + t) dt = \int_{v(1 + t_n)}^n \frac{q}{\psi(q)} dq \sim \int_{v(1)}^n \frac{q}{\psi(q)} dq \sim \int_1^n \frac{q}{\psi(q)} dq, \text{ as } n \to \infty,$$
since \( v(1 + t_n) \to v(1) \in (0, \infty) \), and since the integral of \( q/\psi(q) \) is finite (resp. infinite) over \([a,b]\) (resp. \([a,\infty)\)), for all fixed \( a,b \in (0, \infty)\).

Recall again the fact that \( S_n \) has Poisson \((\theta L_n)\) distribution, given \( T_n \). Now due to \( L_n \to \infty \), almost surely, we obtain

\[
\frac{S_n}{\int_1^n q\psi(q)^{-1} \, dq} \to \theta,
\]

in probability, as claimed. In the regular variation case, this last convergence holds again in the almost sure sense due the fact that in the full genealogy coupling \( S_n \leq S_{n+1} \), for all \( n \), almost surely, and that \( \int_1^n q\psi(q)^{-1} \, dq \) is asymptotic to a multiple of \( n^{2-\alpha} \). To obtain the final claim, we recall (3). Integrating the RHS and recalling (35), we deduce that \( S_n \sim \theta Bn^{2-\alpha} \), almost surely, where \( B \) is as stated in Theorem 2, in consistence with Theorem 1.9 of [5].

For \( X_n = A_n \), our strategy is as follows: we first establish the convergence in probability of \( A_n \) in the general case, and then show the almost sure convergence in the strong regular variation case.

**Case** \( X_n = A_n \), **convergence in probability**. In the full genealogy construction, we have \( A_n \leq S_n \) for each \( n \), almost surely. Therefore, (36) implies that for any \( \varepsilon > 0 \),

\[
P \left( A_n \geq (1 + \varepsilon)\theta \int_1^n q\psi(q)^{-1} \, dq \right) \to 0.
\]

(37)

It remains to prove the matching lower-bound. To do this, for each mutation (or mark) \( x \) on \( T_n \), consider the path \( \gamma = \gamma(x) \subset T_n \) defined as follows. Consider a mutation or mark \( x \in T_n \) with age \( t \). Then \( \gamma(x) \) is defined as the path connecting the mark to the leaf carrying the smallest label possible, while the age of any point \( y \in \gamma \) is at most \( t \). For example, the \( \gamma \) of the mutation encircled in gray on Figure 1 is the path linking it to the leaf labeled by 4.

We say that a mark \( x \) is **unblocked** if \( \gamma(x) \) carries no other mutation than \( x \), and otherwise call it **blocked**. Observe that if \( x \) is unblocked then it is guaranteed to contribute one allelic type to \( A_n \). Intuitively, it is rather likely that \( \gamma(x) \) is unblocked. Indeed, since the age \( M_n \) of a randomly chosen point on \( T_n \) is typically small, then \( e^{-\theta M_n} \approx 1 - \theta M_n \), so the probability that a typical mutation is blocked is of order \( \theta\mathbb{E}(M_n) \to 0 \). This suggests that the proportion of blocked mutations is negligible, which is sufficient to yield the desired result.

More rigorously, given \( T_n \) and \( S_n \), the mutations fall on \( T_n \) as \( S_n \) i.i.d. uniformly chosen random points. For \( 1 \leq i \leq S_n \), let \( K_{i,n} \) be the “good” event that the \( i \)th mutation is unblocked, and define

\[
Y_n := \sum_{i=1}^{S_n} 1_{K_{i,n}},
\]

the total number of unblocked mutations. As already argued, we have \( Y_n \leq A_n \leq S_n \) almost surely, so in view of (36) it suffices to prove

\[
\lim_{n \to \infty} \frac{Y_n}{S_n} = 1, \text{ in probability.}
\]

(38)

Note that, given \( T_n \) and \( S_n \), the events \((K_{i,n})_{i=1,\ldots,S_n} \) are exchangeable. In particular, almost surely,

\[
P(K_{1,n}|T_n, S_n) = P(K_{i,n}|T_n, S_n), \ i = 1, \ldots, S_n.
\]
Note in addition that the age of the mutation corresponding to $K_{1,n}$ is equal in distribution to $M_n$ from Proposition 5. Due to the above discussion, we have

$$P(K_{1,n}|T_n, S_n, M_n) = \left(1 - \frac{M_n}{L_n}\right)^{(S_n-1)_+}$$

(39)

almost surely on the event $\{S_n > 0\}$. We extend the definition of $K_{1,n}$ using the above on the complement $\{S_n = 0\}$, making $K_{1,n}$ certain in this case. Fix $\varepsilon > 0$ and note that by Markov’s inequality, we have

$$P(Y_n \leq (1 - \varepsilon)S_n|T_n, S_n) = P(S_n - Y_n \geq \varepsilon S_n|T_n, S_n) \leq \frac{E(S_n - Y_n|T_n, S_n)}{\varepsilon S_n},$$

with the convention $0/0 = 1$. Therefore, due to the above discussion and (39), we obtain

$$P(Y_n \leq (1 - \varepsilon)S_n) \leq \frac{1}{\varepsilon}E\left[\frac{E(S_n - Y_n|T_n, S_n)}{S_n}\right] = \frac{1}{\varepsilon}E\left[P(K_{1,n}^c|T_n, S_n, M_n)\right]$$

$$\leq \frac{1}{\varepsilon}E\left[1 - \left(1 - \frac{M_n}{L_n}\right)^{(S_n-1)_+}\right].$$

(40)

The random variable (conditional probability) in this last expectation is bounded by 1, almost surely. Therefore, in order to show that it converges to 0 in the mean (in $L^1$), it suffices to show that it converges to 0 in probability. Now note that, since $1 - (1 - x)^n \leq nx$ for $n \in \mathbb{N}$ and $x \geq 0$,

$$1 - \left(1 - \frac{M_n}{L_n}\right)^{(S_n-1)_+} \leq 2\theta M_n 1\{S_n/L_n \leq 2\theta\} + 1\{S_n/L_n > 2\theta\}$$

(41)

So, for a fixed small $\delta > 0$, we have

$$P\left(1 - \left(1 - \frac{M_n}{L_n}\right)^{(S_n-1)_+} > \delta\right) \leq P(2\theta M_n > \delta) + P(S_n > 2\theta L_n).$$

(42)

Due to Proposition 5(a), the first term on the RHS in (42) vanishes as $n \to \infty$, and since $S_n$ has Poisson (rate $\theta L_n$) distribution, given $L_n$, the second term also vanishes. Therefore (40) converges to 0 as $n \to \infty$, implying (38).

Case $X_n = A_n$ with strong $\alpha$-regular variation. Here we use a variation of (41):

$$1 - \left(1 - \frac{M_n}{L_n}\right)^{(S_n-1)_+} \leq \left(M_n \cdot \frac{S_n}{L_n}\right) \wedge 1 \leq 2\theta M_n 1\{S_n/L_n \leq 2\theta\} + \left(M_n \wedge \frac{1}{2\theta}\right) \frac{S_n}{L_n} 1\{S_n/L_n > 2\theta\}.$$

(43)

Therefore, applying the Cauchy-Schwarz inequality,

$$E\left[1 - \left(1 - \frac{M_n}{L_n}\right)^{(S_n-1)_+}\right] \leq 2\theta E(M_n) + \sqrt{E[(M_n \wedge 1/(2\theta))^2]} \cdot E[(S_n/L_n)^2].$$

Due to Proposition 5(b), the first term above is $O(n^{-2\delta})$ for some $\delta > 0$. For the second one, note that $E[(M_n \wedge 1/(2\theta))^2] \leq E(M_n)/(2\theta) = O(n^{-2\delta}/\theta)$ and that $E[(S_n/L_n)^2] = O(\theta^2 \vee \theta)$. Indeed, since $S_n$ is Poisson (rate $\theta L_n$), given $L_n$, we have (assuming $n \geq 3$)

$$E[(S_n/L_n)^2] = E[E(S_n^2|L_n)/L_n^2] = E[\theta^2 + \theta/L_n] \leq \theta^2 + \theta E(1/L_n).$$
where it is simple to verify that $\mathbb{E}(1/L_3) < \infty$. As a consequence, $\mathbb{E}(1 - (1 - M_n/L_n)^{(S_n-1)^+}) = O(n^{-\delta})$. Due to (40), we deduce

$$\mathbb{P}(Y_n \leq (1 - \epsilon)S_n) \leq \frac{cn^{-\delta}}{\epsilon}$$

for some $\delta > 0$, and $c < \infty$ which depends only on $\theta$. Consider the subsequence $n_k = \lfloor k^{2/\delta} \rfloor$, $k \geq 1$. By the Borel-Cantelli lemma and (44), $Y_n/S_n$ tends to 1 along the subsequence $(n_k)$. Moreover, since both $A_n \leq A_{n+1}$ and $S_n \leq S_{n+1}$, for all $n$, almost surely, we have

$$\frac{S_{n_k} A_{n_k}}{S_{n_k+1}} \leq \frac{A_n}{S_n} \leq \frac{A_{n+1}}{S_{n+1}} \frac{S_{n_k+1}}{S_{n_k}}$$

whenever $n \in [n_k, n_{k+1}]$ for some $k \geq 1$. (45)

Since we already verified at the beginning of the argument that $S_{n_k}/S_{n_{k+1}} \sim (n_k/n_{k+1})^{2-\alpha}$, almost surely, and since $(n_k/n_{k+1})^{2-\alpha} \to 1$, as $k \to \infty$, the almost sure convergence along the subsequence $(n_k)_{k \geq 1}$ and (45) imply that $A_n/S_n \to 1$ almost surely. This finishes the proof of Theorem 2. \[\square\]

Remark 15. As already mentioned, the above convergence in probability for $X_n = A_n$ is proved in [30] for a more general class of regular $\Xi$-coalescents, using a compact martingale argument that accounts for all the mutations in a dynamic way (from the point of view of coalescent evolution). However, that approach is not well-suited for obtaining qualitative or quantitative information about a random (typical) mutation. The present approach could be used even in the setting without martingale structure, once given the estimates in the form of Proposition 5. Furthermore, the random mutation analysis enables us to easily identify the asymptotic behavior of $M_{k,n}$ with that of $F_{k,n}$ (see the end of the proof of Theorem 3).

2.3 Proof of Theorem 3

Recall the setting of Theorem 3. We first concentrate on the result (6) in the case of the allelic partition, which we restate here for convenience: if $F_{k,n}$ denotes the number of allelic types in the allelic partition carried by exactly $k$ individuals, then for any fixed $k \geq 1$,

$$\frac{F_{k,n}}{n^{2-\alpha}} \to \theta B(2 - \alpha)(\alpha - 1) \cdots (\alpha + k - 3) / k!, \text{ a.s.}$$

as $n \to \infty$. The key to proving (16) is to apply Corollary 21 in Gnedin, Hansen Pitman [23], which could be thought of as a Tauberian theorem for random exchangeable partitions, that establishes the mutual equivalence between the strong almost sure asymptotics (5), (6), and (7).

We now recall the setting in [23]. Let $\vec{p} = (p_1, p_2, \ldots)$ be a deterministic sequence such that $p_1 \geq p_2 \geq \ldots \geq 0$ and $\sum_i p_i = 1$. Suppose that $\Theta = \Theta_{\vec{p}}$ is an exchangeable random partition on $\mathbb{N}$, obtained by performing the paintbox construction generated by $\vec{p}$ (see, e.g., [1] or Definition 1.2 in [2]). Let $\Theta^n$ denote the restriction of $\Theta$ onto $[n] = \{1, \ldots, n\}$. Let $K_n$ be the number of blocks in $\Theta^n$, and for each $r = 1, \ldots, n$, let $K_{n,r}$ be the number of blocks in $\Theta^n$ containing exactly $r$ elements. The frequency vector $\vec{p}$ is said to be regularly varying with index $\gamma$ if

$$\sum_i 1_{\{p_i \geq x\}} \sim \ell(1/x)x^{-\gamma}$$

as $x \to 0$, where $\ell$ is a slowly varying function.
Lemma 16. (Corollary 21 in [23]) There is equivalence between the following statements.

(a) \( \bar{p} \) is regularly varying with index \( \gamma \).
(b) \( K_n \sim \Gamma(1 - \gamma)n^{\gamma}L(n) \), almost surely as \( n \to \infty \).

If either (a) or (b) holds, then for each fixed \( r \geq 1 \),
\[
K_{n,r} \sim \frac{\gamma \Gamma(r - \gamma)}{r!} n^{\gamma}L(n), \text{ almost surely.}
\]

We refer the reader to Theorem 1.11 in [7] for an overview and a sketch of proof, and to Schweinsberg [39] for a version of this result where the assumptions and conclusions are convergence in probability, rather than almost surely.

Proof of Theorem \( \overline{3} \). We apply the above lemma to the allelic partition \( \Theta \), which is an exchangeable random partition. As the reader is about to see, for this particular application the almost sure convergence in (5) of Theorem 2 is crucial. Since \( \Theta \) is random exchangeable, by Kingman’s representation theorem (Theorem 1.1 in [7]), all the blocks of \( \Theta \) have a well-defined asymptotic frequency. We let \( \bar{P} \) be the sequence of block frequencies in decreasing order. Thus \( \bar{P} \in \nabla_{\leq 1} = \{(p_1, p_2, \ldots, p_r) : p_1 \geq p_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} p_i \leq 1 \} \). Moreover, given \( \bar{P} \), \( \Theta \) has the law of a paintbox partition derived from \( \bar{P} \). Note that \( A_n \) then corresponds to the total number of blocks of \( \Theta^n \), while \( F_{k,n} \) is the number of blocks of size exactly \( k \). In particular, since \( A_n = o(n) \) almost surely, it must be that \( \mathbb{P}(\bar{P} \in \nabla_1) = 1 \), where \( \nabla_1 = \{(p_1, p_2, \ldots) : p_1 \geq p_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} p_i = 1 \} \). That is, \( \Theta \) has no singletons (or no “dust”) almost surely. Moreover, since \( \mathbb{P}(A_n \sim Bn^{2-\alpha}) = 1 \) by (5), then also
\[
\mathbb{P}(A_n \sim Bn^{2-\alpha} | \bar{P}) = 1, \text{ a.s.}
\]
Therefore, Corollary 21 in [23] implies that
\[
\mathbb{P}(\bar{P} \text{ is regularly varying with index } 2 - \alpha) = 1,
\]
and, moreover that, if \( N(x) = \sum_{i \geq 1} 1\{p_i \geq x\} \), then
\[
N(x) \sim \frac{B}{\Gamma(\alpha - 1)} x^{\alpha - 2}, \text{ almost surely.}
\]
Furthermore,
\[
\mathbb{P}\left(F_{k,n} \sim \frac{B \cdot (2 - \alpha)\Gamma(k - 2 + \alpha)}{r!} n^{2-\alpha} \bigg| \bar{P}\right) = 1, \text{ a.s.}
\]
Taking expectations in the last identity yields (46).

It remains us to prove (6) in the case where \( X_{k,n} = M_{k,n} \), the number of genetic types under the infinite sites model. Observe another important property of our full genealogy coupling (cf. Figure 1): if a set of individuals in the sample of size \( n \) belongs to the same allelic type under the infinite alleles model, then they will also carry the same genetic type under the infinite sites model, and the corresponding family in the latter model may only be larger. Thus, for all \( n \geq 1 \), and for all fixed \( k \in \{1, \ldots, n\} \), we have that
\[
\bar{F}_{k,n} \leq \bar{M}_{k,n}, \quad (47)
\]
where \( \bar{F}_{k,n} = \sum_{j=1}^n F_{j,n} \) and \( \bar{M}_{k,n} = \sum_{j=1}^n M_{j,n} \) are the cumulative number of families of size larger or equal to \( k \). Let

\[
c_k = (2 - \alpha)(\alpha - 1) \ldots (\alpha + k - 3)
\]

so that \( F_{k,n} \sim n^{2-\alpha} B c_k \). Observe that \( \bar{F}_{k,n} = A_n \) and thus (since \( \bar{F}_{k+1,n} = \bar{F}_{k,n} \sim F_{k,n} \)) we deduce by induction on \( k \geq 1 \) that \( \bar{F}_{k,n} \sim n^{2-\alpha} B c_k \), where \( \bar{c}_k = \sum_{j=k}^\infty c_j \). Here we use the fact \( \tilde{c}_1 = 1 \) (see [6], Lemma 30 or [34], display (3.38)).

Therefore, Theorem 3 will be proved, provided we show that, for each fixed \( k \geq 1 \),

\[
\bar{M}_{k,n} \leq \bar{F}_{k,n} + o(n^{2-\alpha}), \tag{48}
\]

almost surely as \( n \to \infty \). This can be done by the following adaptation of the argument for Theorem 2.

Fix \( k \geq 1 \). We extend the definition of an unblocked mutation as follows. Recall that any point \( x \in T_n \) corresponds uniquely to a block \( B \) of the coalescing partition \( \Pi_t \), where \( t \) is the age of \( x \). Suppose that \( B = \{i_1 < \ldots < i_m\} \), for some \( m \in \mathbb{N} \), and define \( T(x) \subset T_n \) to be the restriction of the coalescence subtree generated by the paths that lead from \( x \) to the leaves labeled by \( i_1, \ldots, i_m \). For example, for the mutation encircled in black on Figure 1, this subtree has four leaves labeled by \( \{3, 4, 5, 6\} \). Note furthermore that, for \( k = 1 \), \( T(x) \) coincides with the path \( \gamma(x) \) defined in the proof of Theorem 2 and that the total length of \( T(x) \) cannot exceed \( k t \).

Let us say that \( x \in T_n \) is \( k \)-unblocked if \( T(x) \) carries no other mark than \( x \), and otherwise call it \( k \)-blocked. Similarly to the proof of Theorem 2 define \( \bar{K}_{i,n} \) as the event that the \( i \)th mutation (picked at random without replacement) is \( k \)-unblocked, and let \( \bar{Y}_{k,n} := \sum_{i=1}^n 1_{K_{i,n}} \).

Reasoning as for (10), and using the fact that the length of \( T \) which corresponds to the randomly picked mutation is at most \( k M_n \), we obtain, for all fixed \( \varepsilon > 0 \),

\[
\mathbb{P}(\bar{Y}_{k,n} \leq (1 - \varepsilon)S_n) \leq \frac{1}{\varepsilon} \mathbb{E} \left[ \left( 1 - \frac{k M_n}{T_n} \right)^{(S_n-1)_+} \right].
\]

Since \( k \) is fixed, the bound (43) with \( M_n \) replaced by \( k M_n \) will lead to

\[
\mathbb{P}(\bar{Y}_{k,n} \leq (1 - \varepsilon)S_n) \leq \frac{c k n^{-\delta}}{\varepsilon}, \tag{49}
\]

where \( \delta \) is as in (44), and \( c \) depends only on \( \theta \). Therefore, we have as before \( \bar{Y}_{k,n}/S_n \to 1 \) almost surely along the subsequence \( (n_j) \), where \( n_j = \lceil j^{2/\delta} \rceil, j \geq 1 \). In particular, \( S_{n_j} - \bar{Y}_{k,n_j} = o(n_j^{2-\alpha}) \), \( j \geq 1 \).

Denote by \( \bar{M}'_{k,n} \) the number of \( k \)-unblocked mutations that span at least \( k \) leaves. For each such mutation, the corresponding family in the allelic partition is of size at least \( k \), so \( \bar{M}'_{k,n} \leq \bar{F}_{k,n} \). Moreover,

\[
0 \leq \bar{M}_{k,n} - \bar{M}'_{k,n} \leq S_n - \bar{Y}_{k,n}
\]

since \( S_n - \bar{Y}_{k,n} \) accounts for all the \( k \)-blocked mutations, even if they span fewer than \( k \) leaves. Thus, due to the previous observations,

\[
\bar{M}'_{k,n_j} = \bar{M}_{k,n_j} - o(n_j^{2-\alpha}) \leq \bar{F}_{k,n_j} \leq \bar{M}_{k,n_j}, \quad j \geq 1, \tag{50}
\]
implying (48), and in particular \( M_{k,n} \sim F_{k,n} \), along the subsequence \( (n_j)_{j \geq 1} \). Since we already know that \( F_{n,k} \sim \theta Bc_k n^{2-\alpha} \) almost surely, and since both \( (F_{n,k})_{n \geq 1} \) and \( (M_{n,k})_{n \geq 1} \) are non-decreasing, almost surely, reasoning as in (45) gives

\[
M_{k,n} \sim \theta Bc_k n^{2-\alpha},
\]
as \( n \to \infty \), which finishes the proof of Theorem 3.

**Proof of Theorem 4.** The proof of Theorem 4 is a direct application of the discussion below Corollary 21 of Gnedin, Hansen and Pitman [23].

Here is an interesting consequence about the structure of \( T_n \cap \mathcal{P} \). Let \#\( i \)T\( (x) \) be the number of leaves of \( T_n \) contained in \( T(x) \). If \( x \) is a randomly chosen mutation, then denote its \#\( i \)T\( (x) \) simply by \#\( i \)T.

**Corollary 17.** We have \( \lim_{k \to \infty} \lim_{n \to \infty} P(\#\( i \)T \geq k) = 0. \)

This claim is weaker than the statement that \#\( i \)T is stochastically bounded.

**Proof.** Given \( T_n, \mathcal{P} \), the probability that \#\( i \)T \( \geq k \) equals precisely \( M_{k,n}/S_n \). Due to Theorem 3 the almost sure limit of this is \( \bar{c}_k \), defined in the proof above. Since \( M_{k,n}/S_n \in [0,1] \), this convergence is also in \( L^1 \), and the corollary follows due to \( \lim_k \bar{c}_k = 0. \)

### 2.4 Relaxing the condition \( \Lambda[1 - \eta, 1] = 0 \)

Define \( \Lambda_\eta(dx) := \Lambda(dx)1_{[0,1-\eta]}(x) \). Then it is easy to see (or consult, e.g., [4]) that there exists a path-wise full genealogy coupling of the corresponding \( \Lambda \)-coalescent and \( \Lambda_\eta \)-coalescent, so that, almost surely, for each \( n \geq 1 \), \( N_{\Lambda,n}(t) = N_{\Lambda_\eta,n}(t) \), for all \( t \in [0,T_n] \), and \( N_{\Lambda,n}(t) \leq N_{\Lambda_\eta,n}(t) \) for all \( t > T_n \), where \( T_n \) is an exponential (rate \( \int_{[1-\eta,1]} 1/x^2 \Lambda(dx) < \infty \)) random variable. By the “full genealogy” coupling, we also mean that the same realization of the mutation process \( \mathcal{P} \) is used for both the restriction of \( T_n^\Lambda \) onto \( [0,T_n] \), and the restriction of \( T_n^{\Lambda_\eta} \) onto \( [0,T_n] \), simultaneously for all \( n \).

The crucial fact is that then the family of non-negative random variables

\[
\max \left\{ \int_{T_n}^{1+\tau_1^{n:\Lambda}} N_{\Lambda,n}(t) \, dt, \int_{T_n}^{1+\tau_1^{n:\Lambda_\eta}} N_{\Lambda_\eta,n}(t) \, dt \right\}, \quad n \geq 1,
\]
is bounded from above by a finite random variable, almost surely. Therefore, \( L_n^{\Lambda} \sim L_n^{\Lambda_\eta} \) almost surely, as well as, \( S_n^{\Lambda} \sim S_n^{\Lambda_\eta} \) and \( A_n^{\Lambda} \sim A_n^{\Lambda_\eta} \), again almost surely, as \( n \to \infty \). This implies Theorem 2 in the general case. And similarly, in the above coupling, we have \( F_{n,k}^{\Lambda} \sim F_{n,k}^{\Lambda_\eta} \) and \( M_{n,k}^{\Lambda} \sim M_{n,k}^{\Lambda_\eta} \), for each fixed \( k \), almost surely as \( n \to \infty \), yielding Theorems 3 and 4.

### 3 Preliminaries on countable particle systems

In this section we describe a general procedure known as the lookdown construction, enabling one to construct measure valued processes from point processes on \( [0,1] \times \mathbb{R}_+ \). The material discussed in this section is mostly well-known, but we prefer to give a brief account of the theory to set the ground for the construction of the coupling in section 4. Unless stated otherwise, we henceforth assume that \( \Lambda(\{0\}) = 0 = \Lambda(\{1\}) = 0. \)
3.1 Lookdown construction

The lookdown construction was first introduced by Donnelly and Kurtz in 1996 [14]. Their goal was to give a construction of the Fleming-Viot superprocess that provides an explicit description of the genealogy of the individuals in the population (see [22] for a reader-friendly introduction to these notions). Donnelly and Kurtz subsequently modified their construction in [15] to include more general measure-valued processes (such as the Dawson-Watanabe superprocesses). It is this version that we use here, and that we will apply to the generalized Fleming-Viot superprocesses (which are dual to \( \Lambda \)-coalescents) as well as to the ratio processes associated to CSBPs. Our approach here closely follows similar material in [12, 5, 7].

For a given (infinite size) population evolving in continuous time, let genetic types of individuals be encoded as numbers in \([0, 1]\). More precisely, for each \( i \geq 1 \) and \( t \geq 0 \), let \( \xi_i(t) \in [0, 1] \) be the genetic type of the individual \( i \) (or level \( i \)) at time \( t \). As will be seen soon, for our models, the infinite particles system \( ((\xi_i(t), \xi_2(t), \ldots), t \geq 0) \) is such that the limiting empirical measure

\[
\Xi_t(\cdot) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(t)}(\cdot)
\]

exists simultaneously for all \( t \), almost surely. The process \( (\Xi_t(\cdot), t \geq 0) \) is a convenient way to track the evolution of the genetic composition of the population.

We first offer an informal description followed by a formal one in Definition 19. The evolution of \( (\xi_i(t))_{i \geq 1} \) is driven by a point process (i.e. a countable collection of random points) \( \pi = (p_i, t_i)_{i \in \mathbb{N}} \) in \([0, 1] \times \mathbb{R}_+\), and a family of i.i.d. coin tosses. Each atom of \( \pi \) corresponds to a birth (or resampling) event. Changes in \( (\xi_i(t), t \geq 0)_{i \geq 1} \) occur only at birth event times. Let \((p, t) \in \pi\). Then at time \( t \), for each level \( i \geq 1 \), a coin is tossed, where the probability of head equals \( p \), independently over levels. Those levels for which the coin comes up heads (let us denote this set by \( I_{p,t} \)) modify their label to \( \xi_{\min I_{p,t}}(t-) \). In words, each level in \( I_{p,t} \) immediately adopts the type of the smallest level participating in this birth event. For the remaining levels reassign the types so that their relative order immediately prior to this birth event is preserved. More precisely, for each \( i \notin I_{p,t} \), let \( \xi_i(t-) = \xi_{\phi(i)}(t) \) where \( \phi \) is the unique increasing bijection from \( \mathbb{N} \setminus \{\min I_{p,t}\} \) onto \( \mathbb{N} \setminus I_{p,t} \).

Remark 18. This procedure is usually referred to as the modified lookdown construction of Donnelly and Kurtz. In the original construction, the same authors left the types of the levels in the complement of \( I_{p,t} \) unchanged at time \( t \), hence the types \( \xi_i(t-) \), for \( i \in I_{p,t} \setminus \{\min I_{p,t}\} \) got erased from the population at time \( t \). Due to Theorem 1.1 in [15] and Lemma 2.1 in [14], this choice has no effect on the distribution of the resulting empirical measures associated to \((\xi_i(t))_{i \geq 1}\) as long as \((\xi_i(0))_{i \geq 1}\) is an exchangeable family of random variables. Therefore it does not really matter (in the limit) which rule we choose, but for our purposes the modified lookdown rule is more suitable.

The coupling between \( \Lambda \)-coalescents and CSBP that we construct is based on applying this construction with two differently distributed point processes \( \pi \). It is therefore useful to consider a more general framework as follows. Fix \((U_{i,j})_{i,j \geq 1}\), a collection of i.i.d. uniform variables on \([0, 1]\).

Definition 19. Let \( \pi = \{(p_i, t_i) : i \in \mathbb{N}\} \) be a random point process on \([0, 1] \times \mathbb{R}_+\) such that
for any $0 \leq t < \infty$

$$\sum_{i:t_i \leq t} p_i^2 < \infty, \text{ almost surely.} \quad (51)$$

For each $n \geq 1$, construct the label process associated with $\pi$ as follows. Set $\xi^n_0(0) = \xi_i(0)$, $i = 1, \ldots, n$. For each $j \geq 1$ and $i \in \{1, \ldots, n\}$ define

$$A_i(t_j, p_j) \equiv A_j(i) := \{U_{i,j} \leq p_j\} \quad \text{and} \quad i_1(j) := \min\{i \geq 1: A_j(i) \text{ occurs}\}. \quad (52)$$

For $i \leq n$, let

$$m_j(i) := \sum_{l=1}^{i} 1_{A_l(j)}, \quad i \geq 1, \quad (53)$$

be the number of levels smaller or equal to $i$ that participate in the birth event $(p_j, t_j)$. Denote by $J$ the set of atom indices $\{j \geq 1: m_j(n) \geq 2\}$ for which two or more levels in $\{1, \ldots, n\}$ participate in the corresponding birth event. Order the collection of indices in $J$ so that $t_{j_1} < t_{j_2} < \ldots$ (this is almost surely possible due to (51), see Proposition 20 below). Define $(\xi^n_i(t))_{1 \leq i \leq n}$ to be constant over $[t_{j_k}, t_{j_{k+1}})$. Moreover, if $j \in J$, modify the labels at time $t_j$ as follows: for each $1 \leq i \leq n$ declare

$$\xi^n_i(t_j) = \xi^{n-1}_{i-(m_j(i)-1)_+}(t_j-)1_{A_j(i)} + \xi^{n-1}_{i_1(j)}(t_j-)1_{A_j(i)}.$$  \quad (54)

where $m_j(i)$ is defined in (53).

Finally, observe a crucial property of the above construction: if $1 \leq m < n$, then the restriction of $\xi^n$ to the first $m$ levels yields $\xi^m$, in symbols:

$$(\xi^n_1(t), \ldots, \xi^n_m(t)), \quad t \geq 0 \equiv ((\xi^m_1(t), \ldots, \xi^m_m(t)), \ t \geq 0). \quad (55)$$

This fact is a simple consequence of the (lookdown) updating rule (54) that makes the type at level $i$ depend only on the previous types at levels up to (and including) $i$. Therefore, one can unambiguously define the label process $(\xi_i, i = 1, 2, \ldots)$ simultaneously for all $i$, as

$$\xi_i(t) := \xi_i^1(t) := \lim_{n \to \infty} \xi^n_i(t), \quad \forall t \geq 0, \forall i \geq 1. \quad (56)$$

We call $\xi := (\xi_i(t), t \geq 0)_{i \geq 1}$ the label process associated to $\pi$. We may write $\xi^\pi$ for $\xi$ in order to indicate this association.

In the sequel we will often focus on $(N^\pi(t), t \geq 0)$, the number of (distinct) types in the population process, defined by

$$N^\pi(t) := \#\{\xi_1(t), \xi_2(t), \ldots\}, \quad t \geq 0. \quad (57)$$

Note that $N^\pi(t) \in \{1, 2, \ldots\} \cup \{\infty\}$ and $N^\pi(0) = 0$, due to our assumptions on $\xi(0)$.

The next proposition justifies the above definition of $\xi$, and ensures that the corresponding limiting empirical measure exists as a càdlàg Markov process. These facts will be used in the construction of the coupling without further reference in the sequel.

**Proposition 20.** Let $\pi$ be a point process satisfying (51) and let $(\xi_i)_{i \geq 1}$ be its label process. Then the limit $\Xi_t = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(t)}$ exists simultaneously for all $t$ almost surely, and $(\Xi_t, t \geq 0)$ is a càdlàg Markov process.
Definition 21. The process \( (\Xi_t, t \geq 0) \) is the lookdown (measure-valued) process associated to \( \pi \). We may write \( \Xi^n \) instead of \( \Xi_t \) to make explicit the dependence on the point process \( \pi \).

Proof of Proposition 20. The proof is fairly standard, and is included here for completeness. Recall the notation of Definition 19. To show that \( \xi^n \) is well defined, note that, almost surely,

\[
\# \{ j \geq 1 : t_j \in [0, t] \text{ and } m_j(n) \geq 2 \} < \infty, \forall t \geq 0. \tag{58}
\]

Indeed, for each \( j \) the indicator \( 1_{\{m_j(n)\geq2\}} \) has expectation \( 1 - (1 - p_j)^n - np_j(1 - p_j)^{n-1} \leq \binom{n}{j} p_j^2 \), and hypothesis (51) together with Borel Cantelli lemma ensures (58). Thus the dynamic (inductive) update (54) is feasible, and the label process \( \xi \) associated to \( \pi \) is well-defined.

A crucial feature of \( \xi \) is that for each fixed \( t > 0 \), the sequence \((\xi_i(t), i = 1, 2, \ldots)\) is exchangeable. Indeed, \((\xi_i(0), i = 1, 2, \ldots)\) is an exchangeable family, and the transitions preserve the exchangeability. An application of de Finetti’s theorem now yields the existence of the limit

\[
\Xi_t = \lim_{n \to \infty} \Xi_t^n, \quad \text{where } \Xi_t^n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\xi_i(t)}, \tag{59}
\]

for any fixed time \( t \), and hence for all \( t \in \mathbb{Q} \) simultaneously, almost surely. It is elementary to check either directly or by using Dynkin’s criterion \(19\) Theorem 10.13] that \( \Xi \) has Markovian transitions, being a “nice” function of a Markov process \( \xi \). Moreover, as Pitman \(33\) Corollary 7 does for ranked \( \Lambda \)-coalescent process (see also Donnelly and Kurtz \(14\), Theorem 2.4 for Fleming-Viot process with Kingman type resampling and with mutation), one concludes that definition (59) extends to the entire closed half-line \([0, \infty)\), yielding a càdlàg Markov process \( (\Xi_t, t \geq 0) \). The right continuity and left limits are taken with respect to the weak topology on \( \mathcal{M} \), the space of probability measures on \([0, 1]\). In other words, the transition semigroup of \( \Xi \) satisfies the Feller property as a consequence of the “continuity of Kingman’s correspondence” (see, e.g., \(34\), Theorem 11 in Chapter 2), so that \( \Xi \) admits a càdlàg modification (see, for example \(36\), Sections III.6–7). \( \square \)

3.2 Ancestral partitions, Fleming-Viot processes and \( \Lambda \)-coalescents

We next apply Proposition 20 in two different settings, corresponding to the Fleming-Viot process and to the CSBP, respectively. The upshot of this construction is a convenient way to track the respective genealogies. This is achieved through the ancestral partition process, associated to the process \( \xi \) constructed in Proposition 20.

Let \( \pi \) be a point process satisfying (51), and \( \xi^n \) its associated label process. Note that for each \( s > 0 \), the shifted point process \( \pi^{-s} := \{(p, t-s) : (p, t) \in \pi, t \geq s\} \) also satisfies (51), and that, due to the updating rule (54), the label updates of the associated label process \( \{\xi^{n-s}(t), t \geq 0\} \) are the same as those of \( \{\xi^n(t), t \geq s\} \). The difference between the two processes is manifested through their initial states, since for \( i \neq j \) we have \( \xi^{n-s}_{i}(0) \neq \xi^{n-s}_{j}(0) \), almost surely, while it is possible that \( \xi^{n}_{i}(s) = \xi^{n}_{j}(s) \). Now fix some \( T > 0 \).

Definition 22. The ancestral partition process \( (\mathcal{R}^T(t), 0 \leq t \leq T) \) takes values in the space of level partitions (or partitions of \( \mathbb{N} \)). For each \( t \leq T \), \( \mathcal{R}^T(t) \) is defined by the equivalence relation: \( i \sim j \) in \( \mathcal{R}^T(t) \) if and only if \( \xi_i(T) \) and \( \xi_j(T) \) descend from the same level at time \( t \), or equivalently, if \( \xi_i^{n-1}(T-t) = \xi_j^{n-1}(T-t) \) (see also equation (2.3) in \(12\)).
Note that $\mathcal{R}^T(T)$ is the trivial partition $\{\{i\} : i \in \mathbb{N}\}$ and that $\mathcal{R}^T(t_1)$ is a coarser partition than $\mathcal{R}^T(t_2)$, whenever $0 \leq t_1 \leq t_2 \leq T$.

We now briefly recall the definition of generalized Fleming-Viot processes as well as their link to $\Lambda$-coalescents. A generalized (nonspatial, i.e. mean-field) $\Lambda$-Fleming-Viot process $(\rho_t, t \geq 0)$ is a Markov process taking values in the space $\mathcal{M}$ of probability measures on $[0,1]$. Its generator $L$ is defined as follows: given a finite measure $\Lambda$ on $[0,1]$,

$$LF(\mu) = \int_{[0,1]} y^{-2}\Lambda(dy) \int_{[0,1]} \mu(dx) \left( F((1 - y)\mu + y\delta_x) - F(\mu) \right),$$

(60)

where $F : \mathcal{M} \to \mathbb{R}$ is a bounded continuous function. In words, a number $y$ between 0 and 1 is sampled at rate $y^{-2}\Lambda(dy)$. A type $x$ is sampled from $\rho_{t-}$ then $\rho_t$ is obtained from $\rho_{t-}$ by scaling down $\rho_{t-}$ by $(1 - y)$, and adding to the result an atom at $x$ of mass $y$.

**Theorem 23.** Let $\Lambda$ be a finite measure on $[0,1]$. Let $\pi$ be a Poisson point process on $[0,1] \times \mathbb{R}$ with intensity $x^{-2}\Lambda(dx) \otimes dt$. Then the lookdown process $\Xi^{\pi}$ (cf. Definition 21) is a $\Lambda$-generalized Fleming-Viot process with generator (60), started from the uniform measure on $[0,1]$. Furthermore, the ancestral partition process $(\mathcal{R}^T(T-t), 0 \leq t \leq T)$ is the $\Lambda$-coalescent, run for time $T$.

*Proof.* A careful proof of this fact can be found in Lemma 3.6 of [12], that is directly based on the work of Donnelly and Kurtz [15]. We include a simpler proof which relies instead on the duality introduced by Bertoin and Le Gall [9]. We start by the claim that the ancestral partition process $(\mathcal{R}^T(T-t), 0 \leq t \leq T)$ is a $\Lambda$-coalescent. This follows simply from the following observation: let $\pi'$ be the point process obtained from $\pi$ by applying the transformation $(p,t) \mapsto (p,T-t)$. Then $\pi'$ has same law as $\pi$ restricted to $[0,T]$ and is thus a Poisson point process on $[0,1] \times [0,T]$ with intensity $x^{-2}\Lambda(dx) \otimes dt1_{[0,T]}(t)$. Now, at each atom $(x,t)$ of $\pi'$ we flip a coin for each active ancestral lineage with probability of heads equal to $p$ and the lineages that come up heads merge. Hence we see that this is precisely the Poissonian construction of $\Lambda$-coalescents (see, e.g., Theorem 3.2 in [7]).

Let $(\rho_t, t \geq 0)$ be a Fleming-Viot process, and let

$$F_t(x) = \rho_t([0, x]), \quad 0 \leq x \leq 1$$

be the associated *bridge* process. Denote by $F_t^{-1}$ the càdlàg inverse of the map $x \mapsto F_t(x)$. Let $V_1, V_2, \ldots$, be i.i.d. uniform random variables in $[0,1]$, independent of $(\rho_t, t \geq 0)$. By the Glivenko-Cantelli theorem (see, e.g., (7.4) in Chapter 1 of Durrett [17]), noting that $(F_t^{-1}(V_i), i \geq 1)$ are i.i.d. samples from the random measure $\rho_t$, we have for each fixed $t \geq 0$

$$\rho_t = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{F_t^{-1}(V_i)}$$

almost surely,

(61)

where the limit is taken in the sense of the weak topology on probability measures.

Bertoin and Le Gall [10] proved that the $\Lambda$-coalescent $(\Pi_t, t \geq 0)$ is dual to the generalized Fleming-Viot process corresponding to $\Lambda$ in the following sense: if $n \geq 1$ and $f$ is any continuous function on $[0,1]^n$, then

$$\mathbb{E}(f(F_t^{-1}(V_1), \ldots, F_t^{-1}(V_n))) = \mathbb{E}(f(Y(\Pi^n(t), V_1', \ldots, V_n'))),$$

(62)
where \( \Pi^n(t) \) denotes the restriction of \( \Pi_t \) to \([n]\), the random variables \( (V'_1, \ldots, V'_n) \) are i.i.d. uniform on \([0, 1]\), and independent of \((\Pi_t, t \geq 0)\), and where the map \( Y \) is defined as follows:

\[
Y(\pi, x_1, \ldots, x_n) = (y_1, \ldots, y_n) \text{ with } y_j = x_i \text{ for } i = \min\{k : k \sim \pi j\}.
\]

Note that the duality relation (62) has the form of a generalized functional duality in the context of interacting particle systems (see [31]), and should not be confused with the notion of duality between coagulation and fragmentation processes of [33].

We next verify that, for each \( t > 0 \),

\[
(\xi_1(t), \ldots, \xi_n(t)) \overset{d}{=} Y(\Pi^n(t), V'_1, \ldots, V'_n).
\]

This fact is an immediate consequence of our construction. Indeed, at time \( t \) two levels \( i \) and \( j \) have the same type \( \xi_i(t) = \xi_j(t) \) if and only if they descend from the same level at time 0 (since all the \( \xi_i(0) \) are almost surely distinct). Hence \( \xi_i(t) = \xi_j(t) \) if and only if \( i \) and \( j \) belong to the same block of \( \mathcal{R}^t(0) \). Therefore

\[
(\xi_1(t), \ldots, \xi_n(t)) = Y'(\mathcal{R}^t(0), \xi_1(0), \ldots, \xi_n(0)),
\]

where for \( \pi = (B_1, B_2, \ldots) \in \mathcal{P}_n \) and \((x_1, \ldots, x_n) \in [0, 1]^n\) we let

\[
Y'(\pi, x_1, \ldots, x_n) = (y_1, \ldots, y_n) \text{ with } y_j = x_i \text{ for } j \in B_i.
\]

Clearly, as long as the random variables \( \Pi \in \mathbb{P}_n \) and \((X_1, \ldots, X_n) \) (i.i.d. uniform on \([0, 1]\)) are independent one has

\[
Y(\Pi, X_1, \ldots, X_n) \overset{d}{=} Y'(\Pi, X_1, \ldots, X_n),
\]

and since the \( \xi_i(0) \) are i.i.d. uniform on \([0, 1]\) and \( \mathcal{R}^t(0) \overset{d}{=} \Pi(t) \), this proves the claim (63).

Due to (62), one concludes that \((F_{t}^{-1}(V'_1), \ldots, F_{t}^{-1}(V'_n))_{t \geq 0}\) and \((\xi_1(t), \ldots, \xi_n(t))_{t \geq 0}\) have the same one-dimensional marginals. This implies that

\[
\forall t \geq 0, \Xi^t \overset{d}{=} \frac{1}{n} \sum_{i=1}^{n} \delta_{F_{t}^{-1}(V_i)}, \ n \geq 1, \text{ and hence that } \Xi^t \overset{d}{=} \rho_t.
\]

Our argument was carried out under the assumption that the initial state is the uniform law on \([0, 1]\). However, it would equally apply if the \( \xi_i(0) \) were drawn independently from any other law on \([0, 1]\). Since \( \Xi \) and \( \rho \) are both càdlàg Markov processes, they must be equal in distribution. \( \square \)

### 3.3 Lookdown process of a CSBP

Recall \( \psi \) from [11] and consider a CSBP \( (Z(t), t \geq 0) \) with branching mechanism \( \psi \) (see, e.g., [8] or [7] Chapter 4.2 for an elementary introduction). In the sequel, we often refer to any such process as \( \psi \)-CSBP. In this section assume that \( Z \) is started from \( Z(0) = 1 \). Following Bertoin and Le Gall [8], recall existence of a two parameter branching family \((Z_t(x), t \geq 0, x \in [0, 1])\), such that for each fixed \( x \in [0, 1] \), \((Z_t(x), t \geq 0)\) is a \( \psi \)-CSBP started from \( Z_0(x) = x \), independent from the \( \psi \)-CSBP \((Z_t(1) - Z_t(x), t \geq 0)\). In particular \((Z_t(1), t \geq 0) \overset{d}{=} (Z(t), t \geq 0)\). The quantity \( Z_t(x) \) can be interpreted as the population size at time \( t \), descended from
the initial fraction $x$ of the population at time 0. Furthermore, the branching property also implies that, for any $t > 0$, $(Z_t(x), x \in [0,1])$ is a subordinator.

We briefly recall the setting of [12]. For each fixed $t \geq 0$, define $M_t([x_1, x_2]) := Z_t(x_2) - Z_t(x_1)$, for all $0 \leq x_1 \leq x_2 \leq 1$. Then $M_t$ extends to a random measure on $[0,1]$. The process $M = (M_t, t \geq 0)$ is easily seen to be Markovian, with a generator given by (see (1.15) in [12] for the general case) for $t > 0$.

We have $\lim_{t \to 0} N^Z(t) = 1$, almost surely.

Proof. This essentially follows from Theorem 12 in [6] and Corollary 1.4.2 (ii) in Duquesne and Le Gall [10]. (See also Corollary 4.1 in [7] for an elementary sketch of proof). Indeed, when Grey’s condition is satisfied, we may use the construction of [6] for the Donnelly-Kurtz lookdown process, where the labeling process $(\theta_1(t), \theta_2(t), \ldots)$ is directly defined in terms of the excursions of a Continuous Random Tree (CRT) with branching mechanism $\psi$ (see [10] or [6] for the basic terminology and properties of these objects, to which we will refer in this.
proof). Let \((H_s, 0 \leq s \leq T_1)\) be the height process associated with \((Z_s, s \geq 0)\), where \(T_1 := \inf\{u > 0 : L_u^0 > 1\}\) and where \((L_u^0, u \geq 0)\) is the local time process at level 0 of \((H_s, s \geq 0)\).

It follows from the construction in [6] that one can embed the lookdown construction in the CRT so that for any \(t > 0\), \(N^Z(t)\) is exactly the number of excursions of \((H_s, 0 \leq s \leq T_1)\) that reach level \(t\). It follows directly (by excursion theory for \((H_s, 0 \leq s \leq T_1)\)) that \(\{N^Z(t), t > 0\}\) has the law of \((M_{\tilde{v}}(t), t \geq 0)\), where by definition:

\[
\tilde{v}(t) = \mathbf{N}(\sup_{s \geq 0} H_s > t).
\]

Here, \(\mathbf{N}(\cdot)\) denotes the excursion measure of \(H\). By Corollary 1.4.2 (ii) of [16], \(\tilde{v}(t) = v(t) < \infty\), which proves the result.

**Remark 26.** For each fixed \(t > 0\), due to the exchangeability of the sequence \((\xi_i(t), i = 1, 2, \ldots)\), the number of types \(N^Z(t)\) is almost surely equal to the number of atoms of the purely atomic measure \(\Xi_t = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \delta_{\xi_i(t)}\).

**Remark 27.** The property \(\mathbb{P}(N^Z(t) < \infty) = 1\) may seem counter-intuitive in view of the fact that types are not destroyed in any particular application of the updating rule (54). However, an accumulation of many densely placed small lookdown jumps “pushes off” to infinity all but finitely many types in any positive amount of time, whenever Grey’s condition (2) is fulfilled.

## 4 The coupling

### 4.1 Coupling construction

We can now explain the coupling between \(\Lambda\)-coalescents and CSBP. The key idea is to use the following result due to Lamperti, which expresses any CSBP as a time-change of a Lévy process.

Consider a Lévy process \((X_t, t \geq 0)\) with Laplace exponent \(\psi\) given in (1), and assume \(X_0 = x \in (0, 1]\). Define

\[
U^{-1}(t) := \inf \left\{ s > 0 : \int_0^s \frac{du}{X_u} > t \right\}, \quad (67)
\]

and

\[
Z_t = X_{U^{-1}(t)}, \quad t \geq 0. \quad (68)
\]

**Theorem 28.** (Lamperti [28, 29].) The process \((Z_t, t \geq 0)\) is a \(\psi\)-CSBP started from \(Z_0 = x\).

**Construction.** We now describe the coupling between the genealogies of a CSBP and Fleming-Viot processes. Assume that the Lévy process \(X\) and its corresponding CSBP \(Z\) (Lamperti time-changed as above) satisfy \(X_0 = Z_0 = 1\). As before, denote by \(\pi^Z\) the point process of the rescaled jump sizes of \(Z\). Call \(\xi = (\xi_i(t), t \geq 0)_{i \geq 1}\) the label process of \(\pi^Z\) obtained from the lookdown construction applied to \(Z\).

Consider simultaneously the point process \(\pi^X = (\Delta X(t_i), t_i)\) of (unscaled) jump sizes of \(X\), and its associated label process \(\theta = (\theta_i(t), t \geq 0)_{i \geq 1}\), as well as the lookdown measure \(\Theta = (\Theta_t, t \geq 0)\). Then \(\Theta\) is a \(\Lambda\)-Fleming-Viot process, and hence (due to Theorem 23) has a genealogy given by a \(\Lambda\)-coalescent. Indeed, since \(X\) is a Lévy process, due to the Lévy-Itô decomposition, the point process of jumps \(\pi^X = (\Delta X(t_i), t_i)\) is a Poisson point process with intensity \(\nu(dx) \otimes dt\), where \(\nu(dx) = x^{-2} \Lambda(dx)\) is the Lévy measure of \(X\).
Heuristics. For a small $t > 0$, the two point processes $\pi^X$ and $\pi^Z$, restricted to $[0, t]$, are “close to each other”. Indeed, each point $(p, t) \in \pi^X$ also corresponds to a point $(\bar{p}, \bar{t}) \in \pi^Z$, where $\bar{t} = U^{-1}(t)$, and $\bar{p} = p/Z(\bar{t})$. Now, since $(X_t, t \geq 0)$ is almost surely continuous at $t = 0$, the time-change $U^{-1}$ is almost surely differentiable at $t = 0$ with derivative close to 1. Therefore, $U^{-1}(t) \sim t$ as $t \to 0$, and one deduces that for small $t$, $\bar{t} \approx t$. Likewise, invoking the continuity of $Z$ and the fact $Z_0 = 1$, we have $Z(\bar{t}) \approx 1$, hence $(\bar{p}, \bar{t}) \approx (p, t)$.

It is therefore reasonable to believe that for small $t$, $N^X(t) \approx N^Z(t)$, where $N^X(t)$ (resp. $N^Z(t)$) is the number of types in the lookdown process associated to $\pi^X$ (resp. $\pi^Z$) at time $t$. At the same time, by Proposition 22 we also know $N^Z(t) \sim v(t)$ almost surely as $t \to 0$, and all of the above strongly suggests that the same is true for $N^X$ in place of $N^Z$.

Finally, due to Theorem 23, we have

$$N^X(t) \overset{d}{=} N^\Lambda(t), \text{ for each fixed } t \geq 0,$$

where $N^\Lambda(t)$ is (as usual) the number of blocks in the corresponding $\Lambda$-coalescent at time $t$.

The reader can easily check this property by restricting attention to the first $n$ levels, and using the updating rule (54), as well as the fact that $(\pi^X(t), t \in [0, T])$ and $(\pi^X(T-t), t \in [0, T])$ have the same distribution. Therefore, we obtain $N^\Lambda(t) \sim v(t)$ in probability, as $t \to 0$.

We will now turn these heuristic observations into a rigorous argument for Theorem 8, starting with a monotonicity lemma. As indicated in the Introduction, this convergence in fact holds almost surely (see Theorem 1 of [4]).

Definition 29. Given two point processes $\pi$ and $\pi^+$ on $[0, 1] \times \mathbb{R}_+$ on the same probability space, and a random time $T \geq 0$, measurable with respect to the filtration generated by $\pi$ and $\pi^+$, we write $\pi \prec |[0, T] \pi^+$ (or $\pi \prec \pi^+$ on $[0, T]$) if there exists an increasing càdlàg process $r : [0, t] \to \mathbb{R}_+$ such that, almost surely, $r(0) = 0$ and

$$\pi = \{(p_i, t_i) : i \geq 1\} \text{ and } \pi^+ = \{(q_i, r(t_i)) : i \geq 1\},$$

where $p_i \leq q_i$, for each $i \geq 1$ such that $t_i \leq t$.

In words, $\pi \prec \pi^+$ on $[0, T]$, if the atoms of $\pi^+$ are those of $\pi$, time-changed by $r$, and multiplied in size by a (possibly non-constant and random) quantity not smaller than 1. Observe that $r$ preserves the order of the atoms, almost surely. In our main applications, the form of $r$ will be rather simple. Furthermore, the processes $\pi$ and $\pi^+$ of interest will both have infinitely (countably) many atoms in any interval of positive length, almost surely, ensuring that $\{r(T) < \infty\} = \{T < \infty\}$, almost surely.

Consider now $\pi$ and $\pi^+$ such that $\pi \prec \pi^+$ on $[0, T]$ for some finite random time $T$, and both

$$\sum_{i : t_i \leq t} p_i^2 < \infty, \sum_{i : t_i \leq t} q_i^2 < \infty, \forall t \geq 0,$$

almost surely.

One can then construct a coupling of $\Xi^\pi$ (with its label processes $\xi = (\xi_i(t), t \geq 0)_{i \geq 1}$ and $\Xi^\pi^+$ (with its label processes $\xi^+ = (\xi^+_i(t), t \geq 0)_{i \geq 1}$), by using the same collection $\{U_{i, j}\}_{i, j \in \mathbb{N}}$ of i.i.d. uniform random variables to specify the levels participating in the resampling events in Definition 19. Due to $\pi \prec \pi^+$ on $[0, T]$, the following result is obvious by construction:

Lemma 30. If $\pi \prec \pi^+$ on $[0, T]$, then

$$\mathbb{P}(N^{\pi^+}(r(s)) \leq N^\pi(s), \forall s \in [0, T]) = 1.$$
4.2 Application to $N^\Lambda$ asymptotics

Proof of Theorem 3. We start by showing (13) for $\varepsilon$ sufficiently small. The conclusion (14) will then readily follow. Let us assume for the moment that $\text{supp}(\Lambda) \subset [0, \eta]$ where $\eta < 1$, and fix some $\varepsilon \in (0, 1/\eta - 1)$. Consider again the Lévy process $X$ with Laplace exponent $\psi$ such that $X_0 = 1$, and let

$$\pi = \pi^X = \{(\Delta X_t, t) : t > 0\}$$

be the corresponding Poisson point process. Let $\pi_\varepsilon^-$ (resp. $\pi_\varepsilon^+$) be the image of $\pi$ under the map $(t, p) \mapsto (p(1 - \varepsilon), t)$ (resp. $(t, p) \mapsto (p(1 + \varepsilon), t)$). Due to our assumptions on $\text{supp}(\Lambda)$ and the choice of $\varepsilon$, we have that for each atom $(p, t)$ of $\pi$, $p(1 + \varepsilon) < 1$ almost surely. Therefore, both $\pi_\varepsilon^+$ and $\pi_\varepsilon^-$ are Poisson point processes on $(0, 1) \times \mathbb{R}_+$. Let $\nu_\varepsilon^+ \otimes dt$ (resp. $\nu_\varepsilon^- \otimes dt$) be the intensity measure corresponding to $\pi_\varepsilon^+$ (resp. $\pi_\varepsilon^-$). If $f$ is a Borel function on $[0, 1]$, then $\nu_\varepsilon^+$ is obtained by the formula

$$\int_{[0,1]} f(x)\nu_\varepsilon^+(dx) = \int_{[0,1]} f(x(1 + \varepsilon))\nu(dx),$$

and $\nu_\varepsilon^-$ is obtained by an analogous formula with $1 - \varepsilon$ in place of $1 + \varepsilon$. For $\lambda > 0$, let

$$\psi_\varepsilon^+(\lambda) := \int_{(0,1)} (e^{-\lambda x} - 1 + \lambda x)\nu_\varepsilon^+(dx).$$

By the above observation we see that, for each $\lambda > 0$,

$$\psi_\varepsilon^+(\lambda) = \psi(\lambda(1 + \varepsilon)), \quad \psi_\varepsilon^-(\lambda) = \psi(\lambda(1 - \varepsilon)).$$

Therefore, if we let $u_\varepsilon^+(t) := \int_t^\infty d\lambda/\psi_\varepsilon^+(\lambda)$ and $v_\varepsilon^+ = u_\varepsilon^+(t)^{-1}$, we have

$$u_\varepsilon^+(s) = \frac{1}{1 + \varepsilon} u(s(1 + \varepsilon)) \quad \text{and} \quad u_\varepsilon^-(s) = \frac{1}{1 - \varepsilon} u(s(1 - \varepsilon)), $$

hence

$$v_\varepsilon^+(t) = \frac{1}{1 + \varepsilon} v(t(1 + \varepsilon)) \quad \text{and} \quad v_\varepsilon^-(t) = \frac{1}{1 - \varepsilon} v(t(1 - \varepsilon)).$$

Recall that $X_0 = 1$, and define

$$X_t^+ = (1 + \varepsilon)X_t - \varepsilon, \quad \text{and} \quad X_t^- = (1 - \varepsilon)X_t + \varepsilon, \quad t > 0.$$ 

Then it easy to see that both $(X_t^+, t \geq 0)$ and $(X_t^-, t \geq 0)$ are Lévy processes such that $X_0^+ = X_0^- = 1$. Moreover, the Laplace exponent of $X^+$ (resp. $X^-$) is $\psi_\varepsilon^+$ (resp. $\psi_\varepsilon^-$).

Define $T_{\varepsilon}^+ = \inf\{s : |X^+(s) - 1| > \varepsilon\}$ and $T_{\varepsilon}^- = \inf\{s : |X^-(s) - 1| > \varepsilon\}$. Then if $t \leq T_{\varepsilon}^+ \wedge T_{\varepsilon}^-$,

$$\frac{\Delta X^-(t)}{X^-(t)} \leq \frac{\Delta X^-(t)}{1 - \varepsilon} = \Delta X(t) = \frac{\Delta X^+(t)}{1 + \varepsilon} \leq \frac{\Delta X^+(t)}{X^+(t)}. $$

(72)

Using the Lamperti transform, now define two continuous-state branching processes with branching mechanism $\psi_\varepsilon^+$ and $\psi_\varepsilon^-$, respectively, by setting $U_{\varepsilon}^+(t) := \int_0^t \frac{1}{X_u} du$, 

$$U_{\varepsilon}^+(t) := \inf\{s \geq 0 : U_{\varepsilon}(s) > t\} \quad \text{and} \quad Z_{\varepsilon}^+ := X_{U_{\varepsilon}^+(t)}^+, \quad Z_{\varepsilon}^- := X_{U_{\varepsilon}^-(t)}^-, \quad t \geq 0.$$
Finally, (compare with Lemma \[24\]) define \( \pi^Z := \{(\Delta Z^+_{s}/Z^+, s) : s \geq 0 \} \) and \( \pi^Z := \{(\Delta Z^-_{s}/Z^-, s) : s \geq 0 \} \). Due to (72), we have that almost surely

\[
\pi^Z <_{\langle 0, U_-(T^+_\epsilon \wedge T^-_\epsilon) \rangle} \pi \quad \text{(with } r = U^-_\epsilon) \quad \text{and} \quad \pi <_{\langle 0, T^+_\epsilon \wedge T^-_\epsilon \rangle} \pi^Z \quad \text{(with } r = U_+),
\]

where \(<\) is as in Definition \[29\]. Both \( T^+_\epsilon \wedge T^-_\epsilon \) and \( U_-(T^+_\epsilon \wedge T^-_\epsilon) \) are clearly strictly positive and finite, almost surely. Hence, Lemma \[30\] gives that

\[
N^\pi(t) \leq N^\pi^Z(U_-(t)) \quad \text{and} \quad N^\pi(t) \geq N^\pi^Z(U_+(t)), \quad \forall t \leq T^+_\epsilon \wedge T^-_\epsilon, \quad \text{almost surely.}
\]

Proposition \[25\] implies that

\[
\lim_{t \to 0} \frac{N^\pi^Z(t)}{v^-(t)} = \lim_{t \to 0} \frac{N^\pi^Z(t)}{v^+(t)} = 1, \quad \text{almost surely.}
\]

Moreover, it is easy to check that

\[
t/(1 + \epsilon) \leq U_+(t) \leq t/(1 - \epsilon), \quad \forall t \leq T^+_\epsilon \wedge T^-_\epsilon, \quad \text{almost surely.} \tag{73}
\]

This together with \( \mathbb{P}(T^+_\epsilon \wedge T^-_\epsilon > 0) = 1 \) yields

\[
\limsup_{t \to 0} \frac{N^\pi(t)}{v^-(U_-(t))} \leq 1, \quad \text{and} \quad \liminf_{t \to 0} \frac{N^\pi(t)}{v^+(U_+(t))} \geq 1, \tag{74}
\]

almost surely. Due to monotonicity of \( v^\pm \) and (73), we have

\[
\frac{N^\pi(t)}{v^-(U_-(t))} \geq \frac{N^\pi(t)}{v^- (t/(1 + \epsilon))} \quad \text{and} \quad \frac{N^\pi(t)}{v^+(U_+(t))} \leq \frac{N^\pi(t)}{v^+ (t/(1 - \epsilon))}, \quad \forall t \leq T^+_\epsilon \wedge T^-_\epsilon, \quad \text{a.s.} \tag{75}
\]

Combining (74), (74) and (75), and recalling \( \mathbb{P}(T^+_\epsilon \wedge T^-_\epsilon > 0) = 1 \), we can now conclude that

\[
\limsup_{t \to 0} \frac{N^\pi(t)}{v \left( \frac{1 + \epsilon}{1 + \epsilon} \right)} \leq \frac{1}{1 - \epsilon} \quad \text{and} \quad \liminf_{t \to 0} \frac{N^\pi(t)}{v \left( \frac{1 + \epsilon}{1 - \epsilon} \right)} \geq \frac{1}{1 + \epsilon}, \quad \text{almost surely.} \tag{76}
\]

Since \( N^X = N^\pi \) by definition, this gives (13), under the hypothesis that \( \Lambda \) does not give positive mass to a neighborhood of 1. Otherwise, we modify the above argument in the following way. For a fixed \( \eta \in (0, 1) \), since \( x^{-2}\Lambda(dx) \) assigns a finite mass to \((1 - \eta, 1]\), the first time \( T_\eta \) when \( X \) makes a jump of size strictly greater than \( \eta \) is strictly positive, almost surely, being an exponential random variable with finite rate (this resembles the setting of Section \[2.4\]). The analysis (72)–(76) clearly works if \( T^+_\epsilon \wedge T^-_\epsilon \) is everywhere replaced by \( T^+_\epsilon \wedge T^-_\epsilon \wedge T_\eta \), yielding (13).

In particular, almost surely, for all \( t \) sufficiently small,

\[
\frac{1}{1 + \epsilon} \cdot v \left( \frac{1 + \epsilon}{1 - \epsilon} \right) \leq N^X(t) \leq \frac{1}{1 - \epsilon} \cdot v \left( \frac{1 - \epsilon}{1 + \epsilon} \right).
\]

The limit (14) is easily deduced from (69) and this final estimate. \[\square\]

Our coupling argument, based on the particle system construction, does not require Grey's condition (2). Hence, Corollary (10) is an immediate consequence of Theorem (8).
Proof of Proposition 13. The first claim follows by simple calculus manipulations from (14). To see why (16) implies (15), we note that \( \psi : [0, \infty) \to \mathbb{R}^+ \) of (1) is a (strictly) increasing and convex function on \([0, \infty)\). Furthermore, \( v_\psi'(s) = -\psi(v_\psi(s)) \), so that \( v_\psi \) is decreasing with its derivative decreasing in absolute value. Therefore, for \( \varepsilon > 0 \) small enough,

\[
|v(t(1 + \varepsilon)) - v(t)| = \int_t^{t(1+\varepsilon)} |v'(s)|ds \leq |v'(t)|\varepsilon t = \psi(v(t))\varepsilon t.
\]

Similarly,

\[
|v(t(1 - \varepsilon)) - v(t)| = \int_{t(1-\varepsilon)}^t |v'(s)|ds \leq |v'(t(1 - \varepsilon))|t\varepsilon = \psi(v(t(1 - \varepsilon)))t(1 - \varepsilon)\frac{\varepsilon}{1 - \varepsilon}.
\]

Hence (15) will hold provided \( \psi(v(t))t = O(v(t)) \). \( \square \)

4.3 Regularity indices and consequences

Proposition 13 motivates the study of the asymptotic behavior of \( v \) and \( \psi \), which is the topic of this section.

Let \( X = (X_t, t \geq 0) \) be a Lévy process with Laplace exponent \( \psi \) given by (1), where \( \nu(dx) = x^{-2}\Lambda(dx) \), and where we assume again that \( \Lambda(\{0\}) = 0 \). As discussed above, we may also assume that \( \text{supp}(\Lambda) \subset [0,1/2) \). The following definitions and properties of the upper-index \( \beta \) and of the lower-index \( \delta \) of \( X \) can be found in the seminal papers of Blumenthal and Getoor [13] and Pruitt [35]. The upper index is defined by

\[
\beta := \inf \left\{ \alpha > 0 : \int_{|x| \leq 1} |x|^\alpha \nu(dx) < \infty \right\} \in [0, 2].
\]

To define the lower-index, following Pruitt [35], we introduce the function \( h(x) = G(x) + K(x) + M(x) \), where (since in our setting \( \text{supp}(\nu) \subset \mathbb{R}^+ \) and moreover the drift is 0)

\[
G(x) = \nu(y : y > x), \quad K(x) = x^{-2} \int_{y \leq x} y^2 \nu(dy),
\]

and

\[
M(x) = x^{-1} \left| \int_{y \leq x} \frac{y^3}{1 + y^2} \nu(dy) - \int_{y > x} \frac{y}{1 + y^2} \nu(dy) \right|.
\]

The lower index is defined by

\[
\delta := \inf \{ \alpha : \liminf_{x \to 0} x^\alpha h(x) = 0 \}. \tag{78}
\]

The upper index \( \beta \) of (77) is similarly given by

\[
\beta = \inf \{ \alpha : \limsup_{x \to 0} x^\alpha h(x) = 0 \}.
\]

Therefore, it must be

\[
0 \leq \delta \leq \beta \leq 2.
\]
As we already pointed out in the Introduction (see (3.4) in Pruitt [35], and Figure 3), β and δ characterize the asymptotic behavior of X near 0. More precisely, if \( M_t := \sup_{0 \leq s \leq t} |X_s| \), then

\[
\limsup_{t \to 0} M_t / t^\kappa = \begin{cases} 
0 & \text{if } \kappa < 1/\beta \\
\infty & \text{if } \kappa > 1/\beta
\end{cases}, \quad
\liminf_{t \to 0} M_t / t^\kappa = \begin{cases} 
0 & \text{if } \kappa < 1/\delta \\
\infty & \text{if } \kappa > 1/\delta
\end{cases}.
\]

Informally speaking, the following lemma states that as \( t \to 0 \) the function \( q \mapsto \psi(q) \) is of order at most \( q^\beta \) and at least \( q^\delta \).

**Lemma 31.** For each \( \varepsilon > 0 \), there exist finite constants \( c_\varepsilon, \beta \) and \( c_\varepsilon, \delta \) such that for all \( v \) large enough \( c_\varepsilon, \delta v^{\beta - \varepsilon} \leq \psi(v) \leq c_\varepsilon, \beta v^{\beta + \varepsilon} \). Hence if \( \Lambda \) is such that the \( \Lambda \)-coalescent comes down from infinity, then \( \beta \geq 1 \).

**Proof.** Observe that for large \( q \),

\[
\psi(q) \asymp q^2 \int_{[0,1/q]} x^2 \nu(dx) + q \int_{[1/q,1]} x \nu(dx), \quad q \to \infty,
\]

where \( f(q) \asymp g(q) \) means that both \( f = O(g) \) and \( g = O(f) \). Indeed, for \( x \leq 1/q \) one can use Taylor’s approximation to get \( e^{-q x^2} - 1 + q x^2 \in [q^2 x^2 / 6, q^2 x^2 / 2] \) while for \( x \geq 1/q \) an easy computation shows \( e^{-q x^2} - 1 + q x^2 \in [q x / e, q x] \).

By definition (77), we have that \( \int_{[0,1]} x^{\beta + \varepsilon} \nu(dx) < \infty \). Therefore

\[
\sum_{n=0}^{\infty} e^{-n(\beta + \varepsilon)} \nu([e^{-n-1}, e^{-n}]) \leq \sum_{n=0}^{\infty} \int_{e^{-n-1}}^{e^{-n}} x^{\beta + \varepsilon} \nu(dx) = \int_{[0,1]} x^{\beta + \varepsilon} \nu(dx) < \infty.
\]

In particular, there exists a constant \( c > 0 \) such that for all \( n \geq 1 \),

\[
\nu([e^{-n-1}, e^{-n}]) \leq c e^{n(\beta + \varepsilon)}.
\]

As a consequence,

\[
\int_{0}^{1/q} x^2 \nu(dx) \leq \sum_{n=\left\lfloor \log q \right\rfloor}^{\infty} \int_{e^{-n-1}}^{e^{-n}} x^2 \nu(dx) \leq c \sum_{n=\left\lfloor \log q \right\rfloor}^{\infty} e^{n(\beta + \varepsilon)} e^{-2n} \leq c q^{\beta - 2 + \varepsilon},
\]

where the finite positive constant \( c \) may change from one inequality to the next one. Similarly, one estimates

\[
\int_{1/q}^{1} x \nu(dx) \leq c q^{\beta - 1 + \varepsilon}.
\]

Together with (79), this yields the upper bound \( \psi(q) = O(q^{\beta + \varepsilon}) \).

For the lower bound, recall the definition (78) and related notation. Observe that

\[
\int_{y>x} \frac{y}{1+y^2} \nu(dy) \asymp \int_{y>x} y \nu(dy), \quad x \in (0,1).
\]

The first integral in the definition of \( M(x) \) is of order \( \int_{y>x} y \nu(dy) = O(x) \), so it is negligible, in comparison. Also, note that

\[
G(x) = \int_{y>x} \nu(dy) \leq x^{-1} \int_{y>x} y \nu(dy) \asymp M(x).
\]

32
Combining this with the definition of $K(x)$ and (79) one gets
\[ h(x) \simeq \psi(1/x), \text{ as } x \to 0. \] (81)

Due to (78), we have $h(x) \geq Cx^{-\delta + \varepsilon}$ for all $x$ sufficiently small and for some $C > 0$, and the lower bound for $\psi$ now easily follows.

Finally, assume that a given $\Lambda$-coalescent comes down from infinity. Then by Corollary 10, Grey’s condition (2) is satisfied for the corresponding measure $\Lambda$. Since for each $\varepsilon > 0$, $\psi(q) \leq cq^{\beta + \varepsilon}$, we deduce that $\beta \geq 1$.

**Remark 32.** Note that (81) also implies the stated upper bound on $\psi(v)$.

The asymptotic behavior of $\psi(q)$ as $q \to \infty$ induces the asymptotic behavior of $v(t)$ as $t \to 0$:

**Corollary 33.** Assume that the $\Lambda$-coalescent comes down from infinity and recall (12).

(i) If $\beta > 1$, we have $\liminf_{t \to 0} t^{1/(\beta + \varepsilon - 1)} v(t) = +\infty$ for any $\varepsilon \in (0, \beta - 1)$.

(ii) If $\delta > 1$, then $\limsup_{t \to 0} t^{1/(\delta - \varepsilon - 1)} v(t) = 0$ for any $\varepsilon \in (0, \delta - 1)$.

**Proof.** Let $c_{\varepsilon, \beta}$ and $c_{\varepsilon, \delta}$ be the constants from Lemma 31. It follows immediately that
\[
\frac{c_{\varepsilon, \beta}}{\beta + \varepsilon - 1} \cdot v^{-(\beta + \varepsilon) + 1} \leq \int_v^\infty \frac{dq}{\psi(q)} \leq \frac{c_{\varepsilon, \delta}}{\delta - \varepsilon - 1} \cdot v^{-(\delta - \varepsilon) + 1}.
\]

Note that since $\beta \geq 1$ we are able to integrate the lower bound for $1/\psi$, but we use the additional constraint $\delta > 1$ in order to be able to integrate the upper bound for $1/\psi$, and thus derive the right-hand-side of the above inequality. Setting the middle term to $t$, after rearranging, we obtain
\[
\left( \frac{c_{\varepsilon, \beta}}{\beta + \varepsilon - 1} \right)^{1/(\beta - 1 + \varepsilon)} \cdot t^{-1/(\beta - 1 + \varepsilon)} \leq v(t) \leq \left( \frac{c_{\varepsilon, \delta}}{\delta - \varepsilon - 1} \right)^{1/(\delta - 1 - \varepsilon)} \cdot t^{-1/(\delta - 1 - \varepsilon)},
\]
implying both statements. \(\square\)

**Remark 34.** Note that Theorem 8 and Corollary 33 together yield Corollary 11. The above analysis and Theorem 1 of [4] again imply here the stronger almost sure form of convergence.

### 4.4 Application to the Beta-coalescent case

In this section, we use the ideas developed in section 4.1 to give a shorter version of the proof of a result on Beta-coalescents by Birkner et al. [12] (see Theorem 35 below). Let $1 < \alpha < 2$ and let $\Lambda$ be the Beta($2 - \alpha, \alpha$) distribution, that is,
\[ \Lambda(dx) = \frac{1}{\Gamma(2 - \alpha) \Gamma(\alpha)} x^{1-\alpha} (1-x)^{\alpha-1} dx. \]

Consider $(Z_t, t \geq 0)$, a continuous-state branching process with stable branching mechanism $\psi(u) = u^{\alpha}$. This branching mechanism arises when the Lévy measure $\nu$ of the associated Lévy process has the form
\[ \nu(dx) = \frac{\alpha(\alpha - 1)}{\Gamma(2 - \alpha)} x^{-1-\alpha} dx = \frac{1}{\Gamma(-\alpha)} x^{-1-\alpha} dx. \]
Let \((\theta_1(t), \theta_2(t), \ldots, t \geq 0)\) be the Donnelly-Kurtz lookdown label system constructed in Definition \[\text{19}\] associated with the point process \(\pi^Z = \{(\Delta Z_s, s) : s \geq 0\}\) (i.e., with the CSBP \(Z\)). Fix \(T > 0\) and consider the ancestral partition process \((R^T(t), 0 \leq t \leq T)\) of Definition \[\text{22}\] associated with \(\theta\). Consider the following additive functional of \((Z_s, s \geq 0)\):

\[
R_t = \alpha(\alpha - 1)\Gamma(\alpha) \int_0^t Z_s^{1-\alpha} \, ds,
\]

and let \(R_t^{-1}\) be its càdlàg inverse. It is elementary to check that \(R_t^{-1}\) is finite for all \(t\). The next statement was originally obtained in \[\text{12}\], Theorem 2.1.

**Theorem 35.** Fix \(t > 0\) and let \(T = R_t^{-1}\). Then \((R^T(R_t^{-1}), 0 \leq s \leq t)\) has the law of a Beta-coalescent, run on the interval \([0, t]\).

Its original proof in \[\text{12}\] relies on delicate computations. Here we offer an intuitive proof, which is a direct application of our coupling. Theorem \[\text{35}\] was extended by Berestycki et al. (Theorem 1 in \[\text{4}\]) to a partition-valued process defined on the continuous stable random tree. This extension was based on a simple construction of the lookdown process in terms of the excursions of the height process. Thus a straightforward variation of the argument for Theorem \[\text{35}\] given below also applies to proving Theorem 1 of \[\text{6}\].

**Proof of Theorem 35.** Let \(\pi\) be the Poisson point process on \([0, 1] \times [0, \infty)\) with intensity \(x^{-2} \Lambda(dx) \otimes dt\). Let \(\pi'\) be the image of \(\pi^Z\) by the transformation \((p, s) \mapsto (p, R_t^{-1}(s))\). That is, if \((z_i, s_i)_{i \geq 1}\) is an enumeration of the atoms of \(\pi^Z\), and if \(t_i = R_t^{-1}(s_i)\), then

\[
\pi' := \{(z_i, t_i) : i \geq 1\}.
\]

Then observe that for \(0 \leq s \leq t\), the partition \(R^t(t-s)\) constructed relative to \(\pi'\) is identical to the partition \(R^T(R_t^{-1})\) (constructed relative to \(\pi^Z\)) featured in the statement of the theorem. Hence, by Theorem \[\text{23}\] it suffices to show that \(\pi\) and \(\pi'\) have the same law.

To verify this identity in law, let

\[
g(x) = \frac{1}{\Gamma(-\alpha)} x^{-1-\alpha}, \quad x > 0,
\]

so that the Lévy measure \(\nu\) becomes \(\nu(dx) = g(x) \, dx\). By the Lamperti transform, it is easy to see that for each \(s \geq 0\), conditionally on \(\mathcal{F}_s = \sigma(Z_u, 0 \leq u \leq s)\), a jump in the population of size \(x\) occurs at instantaneous rate \(Z_s g(x) \, dx\). That is, for any Borel set \(B \subset [0, \infty)\), if \(N_u(B)\) counts the number of jumps of \(Z\) of size \(x \in B\) during \([0, u]\), then \(\langle N_u(B) - \int_0^u Z_s \nu(B) \, ds, u \geq 0\rangle\) is an \((\mathcal{F}_u)\)-martingale.

After the jump of size \(x\) at time \(s\), a fraction \(p = x/(Z_s + x)\) of the population participates in the birth event in the lookdown process. We have \(x = Z_s p/(1-p)\) and \(dx/dp = Z_s/(1-p)^2\), so an atom of size \(p\) arrives in \(\pi^Z\) at instantaneous rate

\[
\eta_s^Z(dp) = Z_s g\left(\frac{Z_s p}{1-p}\right) : \frac{Z_s}{(1-p)^2} \, dp.
\]

Using (83), we conclude

\[
\eta_s^Z(dp) = \frac{1}{\Gamma(-\alpha)} Z_s^{1-\alpha} p^{-1-\alpha} (1-p)^{\alpha-1} \, dp.
\]
Equivalently, if for $B \subset [0,1]$ a Borel set, $N_u^{\pi^Z}(B)$ denotes the number of points of $\pi^Z$ falling into $B$ during $[0,u]$, then $(N_u^{\pi^Z}(B) - \int_0^u ds \int_B \eta^Z_s(dp), u \geq 0)$ is an $(\mathcal{F}_u)$-martingale.

For each $s \geq 0$, let $\mathcal{G}_s = \mathcal{F}_{R_s^{-1}}$ (it is crucial here that $R_s^{-1}$ is a stopping time relative to the filtration $\mathcal{F}$). Moreover, since $R^{-1}$ is non-decreasing, almost surely, we can assume that $\mathcal{G} = (\mathcal{G}_s, s \geq 0)$ is again a filtration. Due to the previous observations, we conclude that for each fixed Borel set $B \subset [0,1]$, $(N_u^{\pi^Z}(B) - \int_0^{R_s^{-1}} dz \int_B \eta^Z_s(dp), s \geq 0)$ is a $(\mathcal{G}_s)$-martingale.

Now note that, if $N'_s(B)$ denotes the number of points of $\pi'$ falling into $B$ during $[0,s]$, then the above construction of $\pi'$ yields $\mathbb{P}(N_s^{\pi^Z}(B) = N'_s(B), \forall s \geq 0) = 1$. Therefore it suffices to show that, almost surely,

$$
\int_0^{R_s^{-1}} dz \int_B \eta^Z_s(dp) = \int_0^s dz \int_B p^{-2}\Lambda(dp)(dp), \forall s \geq 0,
$$

or equivalently that, almost surely,

$$
\int_0^u ds \int_B \eta^Z_s(dp) = \int_0^{R_u} ds \int_B p^{-2}\Lambda(dp)(dp), \forall u \geq 0,
$$

However, due to \((82)\) and \((84)\), we have

$$
ds \eta^Z_s(dp) = \frac{p^{-1-\alpha}(1-p)^{\alpha-1}}{\Gamma(-\alpha)\alpha(\alpha - 1)\Gamma(\alpha)} dp dR_s \approx \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} p^{-1-\alpha}(1-p)^{\alpha-1} dp dR_s,
$$

using a basic property of the Gamma function. The last identity is equivalent to \((85)\).

\section{Appendix: an instructive example}

In this section we discuss a class of examples that illustrate potential difficulties in analyzing functions $\psi$ and $v$ directly. In particular, we show that for some $\beta \neq \delta$, one can choose the measure $\Lambda$ in such a way that $\psi(q)$ oscillates between $q^\delta$ and $q^\beta$, resulting in analogous oscillations for $v(t)$ between $t^{-1/(\beta-1)}$ and $t^{-1/(\delta-1)}$. In passing, we provide examples of $\Lambda$-coalescents with $\delta = 1$ that come down from infinity. Let $\beta \in (1,2)$ be fixed. Set $a_n = e^{-n}$, $n \geq 0$ and for each $n \geq 0$ define the interval $J_n$ as

$$
J_n = (a_{n+1},a_n).
$$

For a subsequence $(a_{n_k})_{k \geq 0}$ of $(a_n)_{n \geq 0}$ define the measure

$$
\nu(dx; (a_{n_k})_{k \geq 0}) \equiv \nu(dx) = \sum_k 1_{J_{n_k}}(x) \frac{1}{x^{\beta + 1}} dx.
$$

Then it is easy to check that for any choice of such a subsequence, the corresponding measure $\nu$ has the upper index $\beta$. It is moreover easily seen that if $n_k = k$ then (recall definitions \((1)\) and \((12)\)) $\psi(q) \asymp q^\beta$, $u(t) := \int_t^\infty dq/\psi(q) \asymp t^{\beta - 1}$, as $t \to \infty$, and as a result $v(t) \asymp t^{-1/(\beta-1)}$, as $t \to 0$. The remaining calculations however confirm that if one chooses the intervals $J_{n_k}$ sparse enough as $k \to \infty$, the asymptotic behavior of the functions $\psi$, $u$, and $v$ can become quite irregular.
By (79), estimating \( u(t) \) as \( t \to \infty \) (up to constants) amounts to estimating

\[
\int_t^\infty \frac{1}{q^2 \int_{[0,1/q]} x^2 \nu(dx) + q \int_{(1/q,1]} x \nu(dx)} dq.
\]

Define \( k^* = k^*(q) = \max\{k : a_n k \geq 1/q\} \) and set \( \beta_1 := \beta - 1 \) and \( \beta_2 := 2 - \beta \) so that \( \beta_1, \beta_2 > 0 \). First compute

\[
\int_{[0,1/q]} x^2 \nu(dx) = \sum_k \int_{J_n k \cap [0,1/q]} x^{1-\beta} dx
\]

\[
= \frac{1}{\beta_2} \sum_{k:a_n k < 1/q} (a_{n_k - 1}^2 - a_{n_k}^2) + \sum_{k:a_n k \geq 1/q} \int_{a_n k}^{1/q} x^{1-\beta} dx
\]

\[
= \frac{1}{\beta_2} \sum_{k=k^*}^{\infty} e^{-n_k (2-\beta)} (1 - e^{-(2-\beta)}) + \int_{\exp(-n_k, -1) \wedge 1/q}^{1/q} x^{1-\beta} dx
\]

\[
= \frac{1}{\beta_2} \left( 1 - e^{-\beta_k} \right) \sum_{k=k^*}^{\infty} e^{-n_k \beta_2} + \frac{1}{q^{\beta_2}} - \frac{1}{(e^{n_k + 1} \wedge q^{\beta_2})}, \tag{86}
\]

and similarly

\[
\int_{(1/q,1]} x \nu(dx) = \frac{1}{\beta_1} \sum_{l=1}^{k_1} \left( e^{n_l \beta_1} (e^{\beta_1} - 1) + [(q \wedge e^{n_k + 1})^{\beta_1} - e^{n_k \beta_1}] \right). \tag{87}
\]

From now on assume \( q \geq 1 \) and let \( k^* = k^*(q) \) be as defined above. Note that if \( 1/q \in J_{n_{k*}} \) (meaning \( n_{k*} \leq \log(q) < n_{k*} + 1 \)) then

\[
q^2 \cdot \left[ \frac{1}{q^{\beta_2}} - \frac{1}{e^{(n_{k*} + 1)\beta_2}} \right] + q \cdot [q^{\beta_1} - e^{n_{k*} \beta_1}]
\]

\[
q^2 \cdot \left[ \frac{1}{q^{2-\beta}} - \frac{1}{e^{(n_{k*} + 1)(2-\beta)}} \right] + q \cdot [q^{\beta_2} - e^{n_{k*} (\beta_2-1)}] \asymp q^{\beta_3},
\]

where for the last estimate it is best to consider separately the two cases \( \log(q) \in [n_{k*}, n_{k*} + 1/2) \) and \( \log(q) \in [n_{k*} + 1/2, n_{k*} + 1) \). One can check similarly that (still assuming \( 1/q \in J_{n_{k*}} \)) the initial terms, corresponding to the non-negative series from (86) and (87), are of the order at most \( q^{\beta-2} \) and \( q^{\beta-1} \), respectively. Hence,

\[
\psi(q) \asymp q^{\beta}, \quad 1/q \in \bigcup_k J_{n_k}, \tag{88}
\]

which agrees well with the “regular” setting where \( n_k = k \). If on the contrary, \( 1/q \notin \bigcup_k J_{n_k} \), then \( n_{l-1} + 1 \leq \log(q) < n_l \) for some \( l \), and computations (86) and (87) imply

\[
\psi(q) \asymp c_1(\beta) q^2 \sum_{k=l}^{\infty} e^{-n_k (2-\beta)} + c_2(\beta) q \sum_{k=1}^{l-1} e^{n_k (\beta-1)},
\]

where \( c_i(\beta) \in (0, \infty), \quad i = 1, 2 \) are constants depending on \( \beta \) only. Due to the properties of the exponential function we then have

\[
\psi(q) \asymp q^2 e^{-n_l (2-\beta)} + q e^{n_{l-1} (\beta-1)}, \quad 1/q \in (a_{n_l}, a_{n_{l-1}+1}]. \tag{89}
\]
Therefore, we need to estimate up to constants
\[
\sum_k \int_{[t,\infty) \cap [e^{nk},e^{nk+1})} \frac{1}{q^\beta} dq + \sum_l \int_{[t,\infty) \cap [e^{nl-1+1},e^{nl})} \frac{1}{q^2 e^{-nl(2-\beta)} + q e^{nl-1(\beta-1)}} dq.
\]
Assume \(1/t \notin \cup_k J_{nk}\). Then the first series of integrals above can easily be evaluated as being of order
\[
\sum_{n_k \geq \log t} e^{n_k(1-\beta)}.
\]
Using the formula
\[
\int_a^b \frac{dx}{Bx + Cx^2} = \left[ \frac{1}{B} \log \left| \frac{Cx}{C+xB} \right| \right]_a^b,
\]
for each \(l\) such that \(\log t \leq n_l-1+1\), the corresponding summand in the second integral equals
\[
\frac{1}{e^{n_l-1(\beta-1)}} \log \left| \frac{e^{n_l-n_l-1} e^{n_l(\beta-2)} e^{n_l-1+1} + e^{n_l-1(\beta-1)}}{e^{n_l(\beta-1)} + e^{n_l-1(\beta-1)}} \right| \approx \frac{(2-\beta)(n_l-n_l-1)}{e^{n_l-1(\beta-1)}},
\]
since \(\beta < 2\) and \(n_{l-1} \leq n_l\).

From all of the above, we conclude that if \(t = e^{n_{l_0}-1+1}\) for some \(l_0\) we have
\[
u(t) = \int_t^\infty \psi(q) dq \asymp \sum_{n_k \geq \log t} e^{n_k(1-\beta)} + \sum_{l \geq l_0} \frac{n_l - n_{l-1}}{e^{n_l-1(\beta-1)}}.
\]

We construct inductively a particular sequence \(n_k\) by fixing \(\varepsilon > 0\) and imposing
\[n_1 = 1, \ n_{k+1} = n_k + \lfloor \exp(\varepsilon n_k) \rfloor, \ k \geq 1.
\]
The corresponding coalescent comes down from infinity since
\[
\sum_{l \geq 1} \frac{n_l - n_{l-1}}{e^{n_l-1(\beta-1)}} < \infty.
\]

Here \(u(t)\) is not anymore of order \(t^{\beta-1}\) for all large \(t\) since whenever \(t = e^{n_{l_0}-1+1}\) one has \(u(t) \asymp t^{\beta-1+\varepsilon}\). It is left to the reader to check that (78), (81) and (89) together imply that \(1 = \delta < \beta\) in the above examples.

References


of continuous-state branching processes. Probab. Theory Related Fields 117, 249–266
(2000).


41, 307–333.


dition.


Biol. 3, 87–112.

Wiley.

[22] A.M. Etheridge. An introduction to superprocesses. American mathematical society, Uni-
versity Lecture series (Vol. 20).


