

## Pathwise construction of tree-valued Fleming-Viot processes

Stephan Gufler\*

### Abstract

In a random complete and separable metric space that we call the lockdown space, we encode the genealogical distances between all individuals ever alive in a lockdown model with simultaneous multiple reproduction events. We construct families of probability measures on the lockdown space and on an extension of it that allows to include the case with dust. From this construction, we read off the tree-valued  $\Xi$ -Fleming-Viot processes and deduce path properties. For instance, these processes usually have a. s. càdlàg paths with jumps at the times of large reproduction events. In the case of coming down from infinity, the construction on the lockdown space also allows to read off a process with values in the space of measure-preserving isometry classes of compact metric measure spaces, endowed with the Gromov-Hausdorff-Prohorov metric. This process has a. s. càdlàg paths with additional jumps at the extinction times of parts of the population.

**Keywords:** lockdown model; tree-valued Fleming-Viot process; evolving coalescent;  $\Xi$ -coalescent; (marked) metric measure space; (marked) Gromov-weak topology; Gromov-Hausdorff-Prohorov topology.

**AMS MSC 2010:** Primary 60K35, Secondary 60J25; 60G09; 92D10.

Submitted to EJP on November 2, 2016, final version accepted on April 6, 2018.

Supersedes arXiv:1404.3682v4.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The lockdown space</b>	<b>5</b>
2.1	Parents and decomposed genealogical distances . . . . .	6
2.2	Extinction of parts of the population . . . . .	8
2.3	The $\Xi$ -lookdown model . . . . .	10

---

\*Goethe-Universität, Institute of Mathematics, 60629 Frankfurt am Main, Germany  
Current affiliation: Technion, Faculty of Industrial Engineering and Management, Haifa 3200003, Israel  
E-mail: [stephan.gufler@gmx.net](mailto:stephan.gufler@gmx.net)

<b>3</b>	<b>Sampling measures and jump times</b>	<b>11</b>
3.1	The case without dust . . . . .	11
3.2	The general case . . . . .	14
<b>4</b>	<b>Stochastic processes</b>	<b>16</b>
4.1	The case without dust . . . . .	16
4.2	The general case . . . . .	21
<b>5</b>	<b>Outline and some definitions for the proof of the central results</b>	<b>24</b>
5.1	Some notation . . . . .	24
5.2	Two-step construction of the point measure of reproduction events . . . .	24
5.3	The general setting . . . . .	25
<b>6</b>	<b>Preservation of exchangeability</b>	<b>26</b>
6.1	Single reproduction events . . . . .	27
6.2	In the lockdown model . . . . .	28
<b>7</b>	<b>Uniform convergence in the lockdown model</b>	<b>32</b>
<b>8</b>	<b>Two families of partitions</b>	<b>33</b>
<b>9</b>	<b>The construction on the lockdown space</b>	<b>39</b>
9.1	The case with dust . . . . .	39
9.2	The case without dust . . . . .	43
<b>10</b>	<b>Proof of Lemmas 7.1 and 7.2</b>	<b>48</b>

## 1 Introduction

Similarly to the measure-valued Fleming-Viot process that is a model for the evolution of the type distribution in a large neutral haploid population, a tree-valued Fleming-Viot process models the evolution of the distribution of the genealogical distances between randomly sampled individuals. The (neutral) tree-valued Fleming-Viot process is introduced in Greven, Pfaffelhuber, and Winter [19] and generalized in [22] to the setting with simultaneous multiple reproduction events. The lockdown model of Donnelly and Kurtz [12, 13] provides a pathwise construction of the measure-valued Fleming-Viot process and more general measure-valued processes. In this article, we give a pathwise construction of the tree-valued Fleming-Viot process from the lockdown model.

Let us sketch the lockdown model that we state in more detail in Section 2. The time axis is  $\mathbb{R}_+$ . In the population model, there are countably infinitely many levels which are labeled by  $\mathbb{N}$ . Each level is occupied by one particle at each time. As time evolves, the particles undergo reproduction events in which particles can increase their levels. We call a particle at a fixed instant in time an individual. We identify each element  $(t, i)$  of  $\mathbb{R}_+ \times \mathbb{N}$  with the individual on level  $i$  at time  $t$ . From the genealogy that is determined by the reproduction events and from given genealogical distances between the individuals at time zero, we define the semi-metric  $\rho$  on  $\mathbb{R}_+ \times \mathbb{N}$  of the genealogical distances between all individuals. We speak of the case with dust if each particle reproduces at finite rate. In the general case, only the rate at which a particle reproduces and has offspring on a given level is finite. In the case without dust, we introduce the lockdown space  $(Z, \rho)$  as the metric completion of  $(\mathbb{R}_+ \times \mathbb{N}, \rho)$ . We allow for simultaneous multiple reproduction events so that we can obtain any  $\Xi$ -coalescent as the genealogy at a fixed time [40, 33, 36, 39, 13, 4].

In Section 3, we state the central results in this article. Theorem 3.1 asserts that a. s.

in the case without dust, the uniform measures  $\mu_t^n = n^{-1} \sum_{i=1}^n \delta_{(t,i)}$  on the individuals on the first  $n$  levels at time  $t$  converge uniformly in compact time intervals to some probability measures  $(\mu_t, t \in \mathbb{R}_+)$  in the Prohorov metric  $d_{\mathbb{P}}^Z$  over the lookdown space,

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} d_{\mathbb{P}}^Z(\mu_t^n, \mu_t) \quad \text{a. s. for all } T \in \mathbb{R}_+. \tag{1.1}$$

We recall that a metric measure space  $(X, r, \mu)$  is a triple that consists of a complete and separable metric space  $(X, r)$  and a probability measure  $\mu$  on the Borel sigma algebra on  $(X, r)$ . The Gromov-Prohorov distance between two metric measure spaces  $(X, r, \mu)$  and  $(X', r', \mu')$  is defined as

$$d_{\text{GP}}((X, r, \mu), (X', r', \mu')) = \inf_{Y, \varphi, \varphi'} d_{\mathbb{P}}^Y(\varphi(\mu), \varphi'(\mu'))$$

where the infimum is over all isometric embeddings  $\varphi : X \rightarrow Y, \varphi' : X' \rightarrow Y$  into complete and separable metric spaces  $Y$ , the Prohorov metric over  $Y$  is denoted by  $d_{\mathbb{P}}^Y$ , and pushforward measures are written as  $\varphi(\mu) = \mu \circ \varphi^{-1}$ . Two metric measure spaces are called isomorphic if their Gromov-Prohorov distance is zero, or equivalently, if there is a measure-preserving isometry between the closed supports of the measures. The Gromov-Prohorov distance is a complete and separable metric on the space  $\mathbb{M}$  of isomorphy classes of metric measure spaces, it induces the Gromov-weak topology in which metric measure spaces converge if and only if the distributions of the matrices of the distances between iid samples (the so-called distance matrix distributions) converge weakly. For the theory of metric measure spaces, we refer to Greven, Pfaffelhuber, and Winter [18] and Gromov [20].

In Section 4, we read off the tree-valued  $\Xi$ -Fleming-Viot processes. While the lookdown model is used in [22] to characterize only versions of the tree-valued  $\Xi$ -Fleming-Viot processes (see Remark 4.4 in [22]), we obtain the whole paths in the present article. Other than in [22], we do not use ultrametricity of the initial state for the techniques in the present article. Therefore, we speak for instance of an  $\mathbb{M}$ -valued  $\Xi$ -Fleming-Viot process when the initial state not necessarily corresponds to an ultrametric tree. Such an  $\mathbb{M}$ -valued  $\Xi$ -Fleming-Viot process is given in the case without dust by  $(\chi_t, t \in \mathbb{R}_+)$ , where  $\chi_t$  is the isomorphy class of the metric measure space  $(Z, \rho, \mu_t)$ . We stress that to construct a tree-valued Fleming-Viot process, we first show a. s. convergence of probability measures in the Prohorov metric on the lookdown space, as in (1.1). Thereafter, we take isomorphy classes to obtain pathwise a tree-valued process. By contrast, in [19], first finite population models are considered and isomorphy classes are taken to obtain tree-valued processes whose convergence in distribution in the Gromov-Prohorov metric is then shown. The limit process is then characterized by a well-posed martingale problem. From our approach, we can also deduce convergence in distribution of tree-valued processes that describe finite population models, see Remark 4.11.

From our approach on the lookdown space, it follows readily that tree-valued  $\Xi$ -Fleming-Viot processes have a. s. càdlàg paths with jumps at the times of large reproduction events (except for some settings in which there is no right-continuity at initial time). In particular, we retrieve the result from [19] that paths are a. s. continuous in the Gromov-weak topology in the case with only binary reproduction events (which is the case associated with the Kingman coalescent). The Gromov-weak topology emphasizes the typical genealogical distances in a sample from the population.

As initially suggested to the author by G. Kersting and A. Wakolbinger, we also consider a process whose state space is endowed with a stronger topology, the Gromov-Hausdorff-Prohorov topology, which highlights also the overall structure of the population. This process has jumps already in the Kingman case, namely at the times when the

shape of the whole genealogical tree changes as all descendants of an ancestor die out (Theorem 3.5 and Proposition 4.2). We call this process a tree-valued evolving  $\Xi$ -coalescent, it can be defined in the case of coming down from infinity which is a subcase of the case without dust.

An  $(\mathbb{R}_+)$ -marked metric measure space is a triple  $(X, r, m)$  that consists of a complete and separable metric space  $(X, r)$  and a probability measure  $m$  on the Borel sigma algebra on the product space  $X \times \mathbb{R}_+$ . The space of isomorphy classes of marked metric measure spaces is introduced in Depperschmidt, Greven, and Pfaffelhuber [9], we recall basic facts in the beginning of Section 4.2. Tree-valued  $\Xi$ -Fleming-Viot processes can be defined as processes with values in the space of isomorphy classes of marked metric measure spaces to include the case with dust, as shown in [22]. To give a pathwise construction, we define in Section 2.1 for each individual  $(t, i)$  a parent  $z(t, i)$ , and we introduce an extended lookdown space  $(\hat{Z}, \rho)$  which also includes the parents of the individuals at time zero. We denote by  $v_t(i)$  the genealogical distance between the individual  $(t, i)$  and its parent, and we consider in Section 3.2 the probability measures

$$m_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{(z(t,i), v_t(i))}$$

on  $\hat{Z} \times \mathbb{R}_+$ . By Theorem 3.10, these measures  $m_t^n$  converge in the Prohorov metric uniformly for  $t$  in compact time intervals to probability measures  $(m_t, t \in \mathbb{R}_+)$ . In Section 4.2, we obtain the tree-valued  $\Xi$ -Fleming-Viot process as the process  $(\hat{\chi}_t, t \in \mathbb{R}_+)$ , where  $\hat{\chi}_t$  is the isomorphy class of the marked metric measure space  $(\hat{Z}, \rho, m_t)$ . In the case without dust,  $z(t, i) = (t, i)$  for all individuals  $(t, i)$ , which yields consistency with the construction in the case without dust.

Besides the space of isomorphy classes of marked metric measure spaces, another possible state space for tree-valued Fleming-Viot processes (in the case with or without dust) is a space of matrix distributions. Here the state  $\xi_t$  at time  $t$  is obtained from the marked metric measure space  $(\hat{Z}, \rho, m_t)$  as follows: We sample an  $m_t$ -iid sequence  $(x(i), v(i))_{i \in \mathbb{N}}$  from  $\hat{Z} \times \mathbb{R}_+$  and consider the infinite matrix  $(\rho'(i, j))_{i, j \in \mathbb{N}}$  given by

$$\rho'(i, j) = (v(i) + \rho(x(i), x(j)) + v(j)) \mathbb{1}\{i \neq j\}. \tag{1.2}$$

Then we define  $\xi_t$  as the conditional distribution of  $\rho'$  given  $(\eta, r_0, v_0)$ . Intuitively, this means we condition on the marked metric measure space  $(\hat{Z}, \rho, m_t)$ . In equation (1.2), we obtain  $\rho'(i, j)$  by sampling parents  $x(i), x(j)$  to whose mutual distance we add their respective distances  $v(i), v(j)$  to their descendants at time  $t$ . We consider the process  $(\xi_t, t \in \mathbb{R}_+)$  in the end of Section 4.2.

In general, we work in one-sided time. In this way, we obtain the path regularity of the processes under consideration for arbitrary initial states, which is applied in [21]. Complementing the results on convergence to equilibrium in [22, Section 9] and [19, Theorem 3], we also show in Section 4.1 that the tree-valued evolving  $\Xi$ -coalescent, started from any initial state, converges to a unique equilibrium, and we define a stationary tree-valued evolving  $\Xi$ -coalescent in two-sided time. We remark that in the Kingman case, the restriction of the lookdown space in two-sided time to the closure of the set of individuals at a fixed time  $t$ , endowed with the probability measure  $\mu_t$ , equals the a. s. compact metric measure space associated with the Kingman coalescent that is studied by Evans [15].

We defer the proofs of the central theorems from Section 3 to the second part of the article whose organization is outlined in Section 5.

Now we discuss more relations to the literature. The lookdown graph of Pfaffelhuber and Wakolbinger [34] can be viewed as a semi-metric space whose completion is a

lookdown space in two-sided time. A lookdown construction of the measure-valued  $\Xi$ -Fleming-Viot process is given by Birkner et al. [4]. Véber and Wakolbinger [41] give a lookdown construction of measure-valued spatial  $\Lambda$ -Fleming-Viot processes with dust using a skeleton structure. To construct the probability measures  $(\mu_t, t \in \mathbb{R}_+)$  on the lookdown space in the case without dust, we use the flow of partitions for which we refer to Foucart [17] and Labbé [29]. A related description of evolving genealogies is the flow of bridges of Bertoin and Le Gall [3].

Pfaffelhuber, Wakolbinger, and Weisshaupt [35] and Dahmer, Knobloch, and Wakolbinger [8] study the compensated total tree length of the evolving Kingman coalescent as a stochastic process with jumps, using also the lookdown model. The times of these jumps correspond to the extinction times of parts of the population. Functionals of evolving coalescents such as the external length have been studied in several works, see for example [26, 7].

For the coming down from infinity property in the setting with simultaneous multiple reproduction events, see e.g. [40, 24, 4, 17]. By methods which differ from those in the present article, it is also shown in [11] that a.s., the states of the tree-valued Fleming-Viot process are non-atomic in the Kingman case. We also mention the work of Athreya, Löhr, and Winter [1] where in particular the Gromov-weak topology and the Gromov-Hausdorff-Prohorov topology are compared. Marked metric measure spaces are applied by Depperschmidt, Greven, and Pfaffelhuber [10, 11] to construct the tree-valued Fleming-Viot process with mutation and selection.

## 2 The lookdown space

We write  $\mathbb{R}_+ = [0, \infty)$ ,  $\mathbb{N} = \{1, 2, \dots\}$ , and we denote the set of partitions of  $\mathbb{N}$  by  $\mathcal{P}$ . For  $n \in \mathbb{N}$ , we write  $[n] = \{1, \dots, n\}$  and we denote the set of partitions of  $[n]$  by  $\mathcal{P}_n$ . We define the restriction  $\gamma_n$  from  $\mathcal{P}$  to  $\mathcal{P}_n$ ,  $\gamma_n(\pi) = \{B \cap [n] : B \in \pi\} \setminus \{\emptyset\}$ . We endow  $\mathcal{P}_n$  with the discrete topology, and  $\mathcal{P}$  with the topology induced by the restriction maps.

Let us first repeat the lookdown model from [22, Section 5.1] which is determined by the genealogy at time 0 and a point measure that encodes the reproduction events.

In the population model, there are countably infinitely many levels which are labeled by  $\mathbb{N}$ . The time axis is  $\mathbb{R}_+$ , and each level is occupied by one particle at each time. To encode the reproduction events that the particles undergo, we use a simple point measure  $\eta$  on  $(0, \infty) \times \mathcal{P}$  with

$$\eta((0, T] \times \mathcal{P}^n) < \infty \quad \text{for all } n \in \mathbb{N} \text{ and } T \in (0, \infty), \quad (2.1)$$

where  $\mathcal{P}^n$  denotes the subset of those partitions of  $\mathbb{N}$  in which not all of the first  $n$  integers are in different blocks, that is,

$$\mathcal{P}^n = \{\pi \in \mathcal{P} : \gamma_n(\pi) \neq \{\{1\}, \dots, \{n\}\}\}. \quad (2.2)$$

For a partition  $\pi \in \mathcal{P}$  and  $i \in \mathbb{N}$ , we denote by  $B_i(\pi)$  the  $i$ -th block of  $\pi$  when the blocks are ordered increasingly according to their smallest elements. Each point  $(t, \pi)$  of  $\eta$  is interpreted as a reproduction event as follows. At time  $t-$ , the particles on levels  $i \in \mathbb{N}$  with  $i > \#\pi$  are removed. Then, for each  $i \in [\#\pi]$ , the particle that was on level  $i$  at time  $t-$  is on level  $\min B_i(\pi)$  at time  $t$  and has offspring on all other levels in  $B_i(\pi)$ . In this way, the level of a particle is non-decreasing as time evolves. For each  $n \in \mathbb{N}$ , only finitely many particles in bounded time intervals are pushed away from one of the first  $n$  levels by condition (2.1).

We consider not only the process that describes the genealogical distances between the individuals at each fixed time as in [22], but we are interested in the genealogical distances between all individuals which we describe by a complete and separable metric

space, the lookdown space. We define an individual as a particle at a fixed instant in time. We identify each element  $(t, i)$  of  $\mathbb{R}_+ \times \mathbb{N}$  with the individual on level  $i$  at time  $t$ . For  $s \in [0, t]$ , we denote by  $A_s(t, i)$  the level of the ancestor of the individual  $(t, i)$  such that the maps  $s \mapsto A_s(t, i)$  and  $t \mapsto A_s(t, i)$  are càdlàg. Let  $\rho_0$  be a semi-metric on  $\mathbb{N}$ . We define the genealogical distance between the individuals on levels  $i$  and  $j$  at time 0 by

$$\rho((0, i), (0, j)) = \rho_0(i, j).$$

More generally, we define the genealogical distance between individuals  $(t, i), (u, j) \in \mathbb{R}_+ \times \mathbb{N}$  by

$$\rho((t, i), (u, j)) = \begin{cases} t + u - 2 \sup\{s \leq t \wedge u : A_s(t, i) = A_s(u, j)\} & \text{if } A_0(t, i) = A_0(u, j) \\ t + u + \rho_0(A_0(t, i), A_0(u, j)) & \text{else.} \end{cases}$$

The genealogical distance  $\rho((t, i), (u, j))$  can be seen as the sum of the distances to the most recent common ancestor of  $(t, i)$  and  $(u, j)$  if these individuals have a common ancestor after time zero. Else it is the genealogical distance of their ancestors at time zero, augmented by the times at which the individuals live. The distinction between these two cases is needed as we work in one-sided time.

The distance  $\rho$  is a semi-metric on  $\mathbb{R}_+ \times \mathbb{N}$  (offspring individuals from the same parent have genealogical distance zero at the time of the reproduction event). We identify individuals with genealogical distance zero, and we take the metric completion. We call the resulting metric space  $(Z, \rho)$  the lookdown space associated with  $\eta$  and  $\rho_0$ . In slight abuse of notation, we refer by  $(t, i) \in \mathbb{R}_+ \times \mathbb{N}$  also to the element of the metric space after the identification of elements with  $\rho$ -distance zero, in this sense we also assume  $\mathbb{R}_+ \times \mathbb{N} \subset Z$ .

For  $t \in \mathbb{R}_+$ , we define a semi-metric  $\rho_t$  on  $\mathbb{N}$  by

$$\rho_t(i, j) = \rho((t, i), (t, j)), \quad i, j \in \mathbb{N}. \tag{2.3}$$

Then  $\rho_t$  describes the genealogical distances between the particles at fixed times, and the process  $(\rho_t, t \in \mathbb{R}_+)$  is the process that is denoted in the same way in Section 5 of [22].

The remainder of this section is organized as follows. In Subsection 2.1, we replace  $\rho_0$  with a decomposed semi-metric and we enlarge the lookdown space by parents of the individuals at time zero. In Subsection 2.2, we consider the two ways in which particles can die and we define extinction times for parts of the population. The construction is randomized in Subsection 2.3 where  $\eta$  becomes a Poisson random measure.

### 2.1 Parents and decomposed genealogical distances

We will use the contents of this section to include the case with dust.

Let  $r_0$  be a semi-metric on  $\mathbb{N}$  and  $v_0 = (v_0(i))_{i \in \mathbb{N}} \in \mathbb{R}_+^{\mathbb{N}}$  such that  $(r_0, v_0)$  satisfies

$$\rho_0(i, j) = (r_0(i, j) + v_0(i) + v_0(j)) \mathbb{1}\{i \neq j\}$$

for all  $i, j \in \mathbb{N}$ . Then  $(r_0, v_0)$  is a decomposition of the semi-metric  $\rho_0$  in the sense of [22, Section 2]. The trivial decomposition  $(r_0, v_0) = (\rho_0, 0)$  always exists.

For each  $(t, i) \in \mathbb{R}_+ \times \mathbb{N}$ , we define the quantity  $v_t(i)$  as in [22, Section 6.1]: For  $j \in \mathbb{N}$ , let  $\mathcal{P}(j) = \{\pi \in \mathcal{P} : \{j\} \notin \pi\}$  be the set of partitions of  $\mathbb{N}$  in which  $j$  does not form a singleton block. If  $\eta(\{s\} \times \mathcal{P}(A_s(t, i))) > 0$  for some  $s \in (0, t]$ , then we set

$$v_t(i) = t - \sup\{s \in (0, t] : \eta(\{s\} \times \mathcal{P}(A_s(t, i))) > 0\},$$

else we set

$$v_t(i) = t + v_0(A_0(t, i)).$$

The quantity  $v_t(i)$  is the time back from the individual  $(t, i)$  until the ancestral lineage is involved in a reproduction event in which it belongs to a non-singleton block, if there is such an event, else  $v_t(i)$  is defined from  $v_0$ .

**Remark 2.1.** For  $t \in \mathbb{R}_+$ , if  $\rho_t$  is a semi-ultrametric (that is,  $\max\{\rho_t(j, k), \rho_t(k, \ell)\} \geq \rho_t(j, \ell)$  for all  $j, k, \ell \in \mathbb{N}$ ) and the condition

$$v_t(i) = \frac{1}{2} \inf_{j \in \mathbb{N} \setminus \{i\}} \rho_t(i, j) \tag{2.4}$$

is satisfied for some  $i \in \mathbb{N}$ , then  $v_t(i)$  is the length of the external branch that ends in the individual  $(t, i)$  in the genealogical tree at time  $t$ . See Remarks 2.4 and 2.5 for more details.

Now we enlarge the set of individuals to the disjoint union  $(\mathbb{R}_+ \times \mathbb{N}) \sqcup \mathbb{N}$ . We call each element  $i$  of  $\mathbb{N} \subset (\mathbb{R}_+ \times \mathbb{N}) \sqcup \mathbb{N}$  the parent of the individual  $(0, i)$ . We extend the semi-metric  $\rho$  to  $(\mathbb{R}_+ \times \mathbb{N}) \sqcup \mathbb{N}$  by

$$\begin{aligned} \rho(i, j) &= r_0(i, j) \quad \text{for } i, j \in \mathbb{N} \\ \text{and } \rho((t, i), j) &= t + v_0(A_0(t, i)) + r_0(A_0(t, i), j) \quad \text{for } (t, i) \in \mathbb{R}_+ \times \mathbb{N}, j \in \mathbb{N}. \end{aligned}$$

That is, the distance between the parents of the individuals  $(0, i)$  and  $(0, j)$  is given by  $r_0(i, j)$ . Furthermore, we define for each individual  $(t, i) \in \mathbb{R}_+ \times \mathbb{N}$  the parent  $z(t, i)$  as the individual  $(t - v_t(i), A_{t-v_t(i)}(t, i))$  if  $v_t(i) < t$ , else we set  $z(t, i) = A_0(t, i)$ . Then  $v_t(i)$  equals the genealogical distance between the individual  $(t, i)$  and its parent.

We identify the elements of  $(\mathbb{R}_+ \times \mathbb{N}) \sqcup \mathbb{N}$  with distance zero and take the metric completion. We call the resulting metric space the extended lookdown space associated with  $\eta$  and  $(r_0, v_0)$ , and we denote it by  $(\hat{Z}, \rho)$ . Here we write again  $\hat{Z} \supset (\mathbb{R}_+ \times \mathbb{N}) \sqcup \mathbb{N}$  in slight abuse of notation. Note that the lookdown space  $(Z, \rho)$  associated with  $\eta$  and  $\rho_0$  is contained in  $(\hat{Z}, \rho)$  as a subspace. Figure 1 below shows an extended lookdown space.

**Remark 2.2.** If  $v_0 = 0$ , then  $\rho_0 = r_0$  and the individuals at time zero in the extended lookdown space  $(\hat{Z}, \rho)$  are identified with their parents as  $\rho((0, i), z(0, i)) = 0$  for all  $i \in \mathbb{N}$ . In this case,  $(\hat{Z}, \rho)$  is equal to the lookdown space  $(Z, \rho)$  associated with  $\eta$  and  $r_0$  from the beginning of Section 2.

**Remark 2.3** (Relation to the decomposed genealogical distances in [22]). We denote the genealogical distances between the parents of individuals  $(t, i), (t, j) \in \mathbb{R}_+ \times \mathbb{N}$  by

$$r_t(i, j) = \rho(z(t, i), z(t, j)).$$

For  $t = 0$ , this is consistent with the definition of  $r_0$  above as  $z(0, i) = i, z(0, j) = j$ , and  $\rho(i, j) = r_0(i, j)$ . For all  $t \in \mathbb{R}_+$  and  $i, j \in \mathbb{N}$ ,

$$\rho_t(i, j) = (v_t(i) + r_t(i, j) + v_t(j)) \mathbb{1}\{i \neq j\}. \tag{2.5}$$

That is, the process  $((r_t, v_t), t \in \mathbb{R}_+)$  of the decomposed genealogical distances between the individuals at fixed times coincides with the process defined from  $\eta$  and  $(r_0, v_0)$  in Section 6.1 of [22]. For  $t = 0$ , equation (2.5) holds by definition of  $\rho_0, r_0$ , and  $v_0$ . That equation (2.5) holds for all  $t \in \mathbb{R}_+$  can be seen from Figure 1. For a formal proof, we distinguish four cases. We always assume  $i \neq j$  in the following.

*Case 1:*  $v_t(i), v_t(j) < t, A_0(t, i) = A_0(t, j)$ . In this case, the definition of  $v_t(i)$  and  $v_t(j)$  implies  $A_s(t, i) \neq A_s(t, j)$  for all  $s \in (t - v_t(i) \vee v_t(j), t]$ . By definition of  $\rho$ , it follows

$$\begin{aligned} \rho(z(t, i), z(t, j)) &= t - v_t(i) - 2 \sup\{s \leq t : A_0(t, i) = A_0(t, j)\} + t - v_j(t) \\ &= \rho((t, i), (t, j)) - v_t(i) - v_t(j), \end{aligned}$$

which is equation (2.5).

**Case 2:**  $v_t(i), v_t(j) < t$ ,  $A_0(t, i) \neq A_0(t, j)$ . In this case, the definition of  $z(t, i)$  and  $z(t, j)$  yields  $A_0(t, i) = A_0(z(t, i))$  and  $A_0(t, j) = A_0(z(t, j))$ . In particular, it follows  $A_0(z(t, i)) \neq A_0(z(t, j))$ . With the definition of  $\rho$ , it follows

$$\begin{aligned} \rho(z(t, i), z(t, j)) &= t - v_t(i) - \rho(A_0(t, i), A_0(t, j)) + t - v_t(j) \\ &= \rho((t, i), (t, j)) - v_t(i) - v_t(j), \end{aligned}$$

which is equation (2.5).

**Case 3:**  $v_t(i) < t, v_t(j) \geq t$ . In this case, it follows that  $A_0(t, i) \neq A_0(t, j)$ . From the definitions, it follows that

$$\rho(z(t, i), z(t, j)) = t - v_t(i) + v_0(A_0(t, i)) + r_0(A_0(t, i), A_0(t, j)).$$

Using  $v_t(j) = t + v_0(A_0(t, j))$  and equation (2.5) for  $t = 0$ , we deduce that

$$v_t(i) + r_t(i, j) + v_t(j) = t + \rho_0(A_0(t, i), A_0(t, j)) + t$$

which is equation (2.5).

**Case 4:**  $v_t(i), v_t(j) \geq t$ . Again by the definitions and by equation (2.5) for  $t = 0$ , we have

$$\begin{aligned} v_t(i) + \rho(z(t, i), z(t, j)) + v_t(j) \\ = t + v_0(A_0(t, i)) + r_0(A_0(t, i), A_0(t, j)) + t + v_0(A_0(t, j)) = 2t + \rho_0(A_0(t, i), A_0(t, j)), \end{aligned}$$

which is equation (2.5).

**Remark 2.4** (Parents and starting vertices of external branches). In this remark, we assume  $v_0(i) = \frac{1}{2} \inf_{j \in \mathbb{N} \setminus \{i\}} \rho_0(i, j)$  for all  $i \in \mathbb{N}$ , and that  $\rho_0$  is a semi-ultrametric. Then, as in Remark 1.1 of [22] and the references therein, we associate with  $\rho_0$  the real tree  $(T_0, d_0)$  that is obtained by identifying the points with distance zero in the semi-metric space  $((-\infty, 0] \times \mathbb{N}, d_0)$ , where  $d_0((s, i), (t, j)) = \max\{\rho_0(i, j) + s + t, |s - t|\}$ . Now we briefly sketch how the extended lookdown space  $(\hat{Z}, \rho)$  can be isometrically embedded into a real tree  $(T, d)$  that contains the genealogical trees of the individuals at all times, and we interpret the parents as starting vertices of external branches.

We define a semi-metric  $d$  on  $\mathbb{R} \times \mathbb{N}$  that coincides on  $\mathbb{R}_+ \times \mathbb{N}$  with the semi-metric  $\rho$  from the beginning of Section 2, that coincides with  $d_0$  on  $(-\infty, 0] \times \mathbb{N}$ , and for  $(s, i) \in \mathbb{R}_+ \times \mathbb{N}$ ,  $(t, j) \in (-\infty, 0] \times \mathbb{N}$ , we set  $d((s, i), (t, j)) = s + d_0(A_0(s, i), (t, j))$ . Then we identify points with  $d$ -distance zero and define  $(T, d)$  as the metric completion. By construction,  $\rho_0(i, j) = d((0, i), (0, j))$  and  $r_0(i, j) = (\rho_0(i, j) - v_0(i) - v_0(j)) \mathbb{1}\{i \neq j\} = d((-v_0(i), i), (-v_0(j), j))$  for all  $i, j \in \mathbb{N}$ . Hence,  $(\hat{Z}, \rho)$  is embedded into  $(T, d)$  by the isometry that maps  $(t, i) \in \mathbb{R}_+ \times \mathbb{N} \subset \hat{Z}$  to  $(t, i) \in T$ , and  $i \in \mathbb{N} \subset \hat{Z}$  to  $(-v_0(i), i) \in T$ .

For each  $t \in \mathbb{R}_+$ , Remark 5.2 in [22] says that  $\rho_t$  is a semi-ultrametric. The associated real tree is given by the subspace  $T_t = (-\infty, t] \times \mathbb{N}$  of  $(T, d)$ . If condition (2.4) is satisfied, then the quantity  $v_t(i)$  and the parent  $z(t, i)$  can be interpreted as the length and the starting vertex, respectively, of the external branch that ends in the leaf  $(t, i)$  of  $T_t$ , see also [22, Remark 2.2].

**Remark 2.5.** In the context of Sections 3 and 4, assumption (2.4) can be checked for  $t = 0$  by Proposition 3.4 in [22], and for  $t > 0$  by Proposition 6.5 in [22].

## 2.2 Extinction of parts of the population

In the beginning of Section 2, we stated the lookdown model as an infinite particle system. The particles undergo reproduction events that are given by the point measure  $\eta$ . In reproduction events encoded by points  $(t, \pi)$  of  $\eta$  where the partition  $\pi$  has finitely

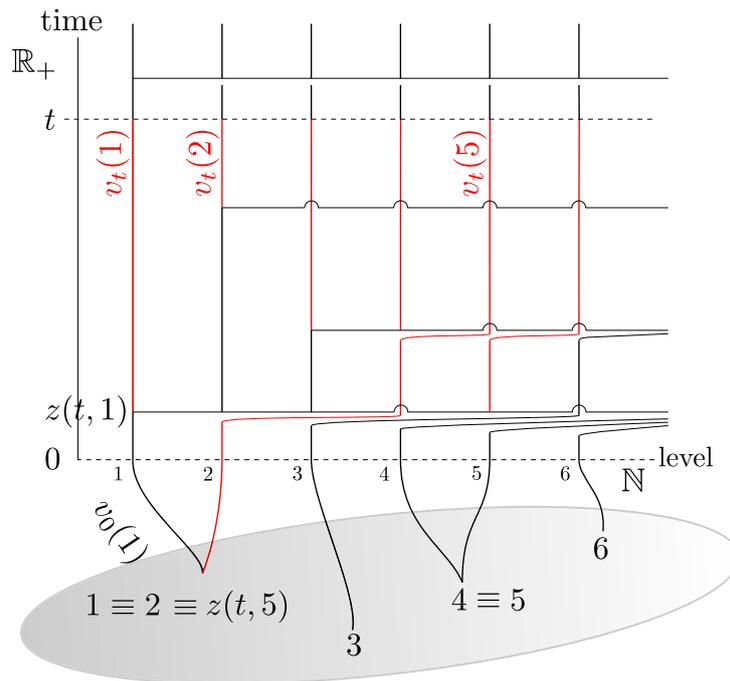


Figure 1: Part of an extended lookdown space. The space  $\mathbb{R}_+ \times \mathbb{N}$  is represented in the upper part of the figure. Time goes upwards and levels go from the left to the right. In the lower part, the metric space obtained from  $\mathbb{N}$ , endowed with the semi-metric obtained from  $r_0$  is symbolized. For each  $i \in \mathbb{N}$ , the junction between the individual  $(0, i)$  and its ancestor  $i$  has length  $v_0(i)$ . Individuals that are in the same block in a reproduction event have genealogical distance zero and are identified. In the figure, they are connected by horizontal lines. In this example, there are no simultaneous multiple reproduction events. The genealogical distances between the individuals  $(t, i)$  and their respective parents  $z(t, i)$  equal  $v_t(i)$ , they are represented by red lines. The genealogical distance between any two individuals is the sum of the lengths of the vertical parts of the path from one individual to the other, plus the distance in the metric space obtained from  $r_0$  if this space has to be traversed.

many blocks, the particles at time  $t$  have only finitely many ancestors among the infinite population at time  $t-$ , hence particles die at time  $t$ . A particle can also die due to an accumulation of reproduction events in which its level is pushed to infinity.

For  $(s, i) \in \mathbb{R}_+ \times \mathbb{N}$  and  $t \in [s, \infty)$ , let  $D_t(s, i)$  be the lowest level that is occupied at time  $t$  by a descendant of the individual  $(s, i)$ , that is,

$$D_t(s, i) = \inf\{j \in \mathbb{N} : A_s(t, j) = i\}$$

with  $D_t(s, i) = \infty$  if and only if there exists no  $j \in \mathbb{N}$  with  $i = A_s(t, j)$ . This quantity corresponds to the forward level process in [34] and to the fixation line in [23]. The map  $t \mapsto D_t(s, i)$  is non-decreasing. Let  $\tau_{s,i}$  be the extinction time of the part of the population that descends from the individual  $(s, i)$ , that is,

$$\tau_{s,i} = \inf\{t \in [s, \infty) : D_t(s, i) = \infty\}.$$

Then the set of times at which such parts of the population become extinct is given by

$$\Theta^{\text{ext}} := \{\tau_{s,i} : s \in \mathbb{R}_+, i \in \mathbb{N}\}. \tag{2.6}$$

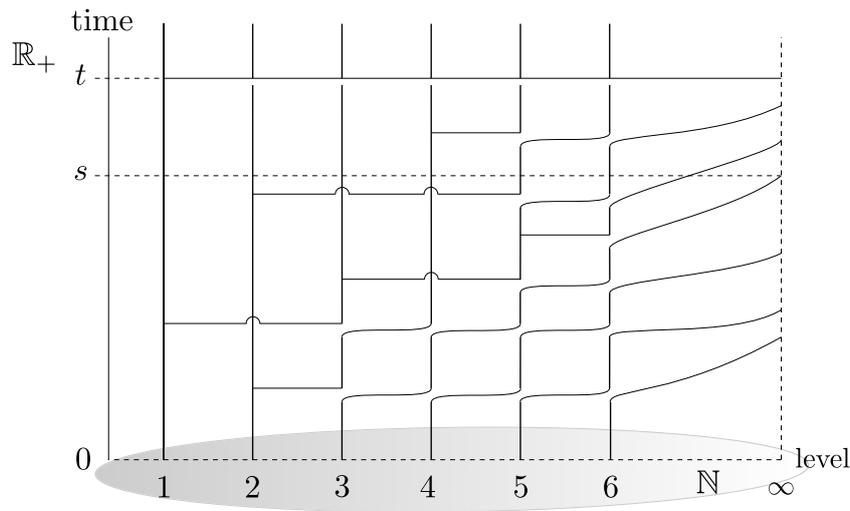


Figure 2: Part of a lookdown space. Only the reproduction events with offspring on the first 6 levels are drawn. At time  $t$ , a reproduction event occurs in which the whole population is replaced by the offspring of the individual on level 1. The limits  $(t-, i) = \lim_{r \uparrow t}(r, i)$ ,  $i \geq 2$  are part of the boundary of the lookdown space. Due to accumulations of jumps, the lines  $t' \mapsto D_{t'}(s', i)$  may hit infinity, here symbolized by a dashed line, similarly to illustrations of the lookdown graph in [34]. This occurs for instance at time  $s$ , the limit  $\lim_{r \uparrow s} D_r(0, 3)$  is part of the boundary. Further elements of the boundary are obtained from Cauchy sequences at fixed times.

**Remark 2.6.** In a reproduction event that is encoded by a point  $(t, \pi)$  of  $\eta$  with  $\#\pi = \infty$ , every individual that sits on a level  $i \in \mathbb{N}$  at time  $t-$  has a descendant at time  $t$ . Hence the two mechanisms mentioned in the beginning of this subsection are the only possibilities for a particle to die.

### 2.3 The $\Xi$ -lookdown model

We recall here the simple point measure  $\eta$  that is used to drive the lookdown model in [22], cf. also the references therein.

Let  $\Delta$  be the simplex

$$\Delta = \{x = (x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, |x|_1 \leq 1\},$$

where  $|x|_p = (\sum_{i \in \mathbb{N}} x_i^p)^{1/p}$ . Let  $\kappa$  be the probability kernel from  $\Delta$  to  $\mathcal{P}$  associated with Kingman's correspondence, that is,  $\kappa(x, \cdot)$  is the distribution of the paintbox partition associated with  $x \in \Delta$ , see e. g. Section 2.3.2 in [2].

Let  $\Xi$  be a finite measure on  $\Delta$ . We decompose

$$\Xi = \Xi_0 + \Xi\{0\}\delta_0.$$

For distinct integers  $i, j \in \mathbb{N}$ , we denote by  $K_{i,j}$  the partition of  $\mathbb{N}$  that contains the block  $\{i, j\}$  and apart from that only singleton blocks. We define a  $\sigma$ -finite measure  $H_\Xi$  on  $\mathcal{P}$  by

$$H_\Xi(d\pi) = \int_{\Delta} \kappa(x, d\pi) |x|_2^{-2} \Xi_0(dx) + \Xi\{0\} \sum_{1 \leq i < j} \delta_{K_{i,j}}(d\pi).$$

In the following sections,  $\eta$  is always a Poisson random measure on  $(0, \infty)$  with intensity  $dt H_\Xi(d\pi)$ . Then  $\eta$  satisfies a. s. condition (2.1), as checked in equation (5.5) of [22]. The lookdown model can therefore be driven by  $\eta$ .

The measure  $\Xi$  is called dust-free if and only if

$$\Xi\{0\} > 0 \quad \text{or} \quad \int |x|_1 |x|_2^{-2} \Xi_0(dx) = \infty.$$

In this case, each particle reproduces with infinite rate, hence  $v_t(i) = 0$  for all  $t \in (0, \infty)$  and  $i \in \mathbb{N}$  a. s. We denote by  $\mathcal{M}_{\text{nd}}$  the subset of finite measures on  $\Delta$  that are dust-free, and by  $\mathcal{M}_{\text{dust}}$  its complement in the set of finite measures on  $\Delta$ .

The extended lookdown space in Figure 1 could be the extended lookdown space associated with a typical realization of an appropriate Poisson random measure  $\eta$  with dust, and a decomposed semimetric  $(r_0, v_0)$ . Figure 2 illustrates the dust-free case.

We speak of a large reproduction event when a particle has offspring on a positive proportion of the levels. We denote by  $\Theta_0$  the set of times at which large reproduction events occur:

$$\Theta_0 = \{t \in (0, \infty) : \text{there exist } \pi \in \mathcal{P} \text{ and } B \in \pi \text{ with } \eta\{(t, \pi)\} > 0 \text{ and } |B| > 0\}. \quad (2.7)$$

Here  $|B| = n^{-1} \lim_{n \rightarrow \infty} B \cap [n]$  denotes the asymptotic frequency of the block  $B$  if it exists. (Note that a. s., all the asymptotic frequencies in the definition of  $\Theta_0$  exist by Kingman's correspondence and as  $\eta$  has countably many points.) The measure  $\Xi_0$  governs the large reproduction events, as opposed to  $\Xi\{0\}$ , which gives the rates of the binary reproduction events (i. e. the reproduction events in which the reproducing particle has only one offspring). A. s. by definition of  $H_\Xi$  and Kingman's correspondence, each reproduction event is either binary or large. In case  $\Xi \in \mathcal{M}_{\text{dust}}$ , the set  $\Theta_0$  equals a. s. the set  $\{t \in (0, \infty) : \eta(\{t\} \times \mathcal{P}) > 0\}$  of reproduction times, as there are a. s. no binary reproduction events.

### 3 Sampling measures and jump times

In this section, we consider mathematical objects that are defined from a realization of the Poisson random measure  $\eta$  and a random (decomposed) distance matrix on an event of probability 1. Stochastic processes are read off from these constructions in Section 4. We defer the proofs of the main statements in Section 3 to the second part of the article which begins in Section 5.

#### 3.1 The case without dust

We construct a family of probability measures on the lookdown space. We consider regularity of this family in the weak topology. In the case of coming down from infinity, we also consider regularity of a family of subsets of the lookdown space with respect to the Hausdorff distance.

Let the Poisson random measure  $\eta$  be defined from the finite measure  $\Xi$  on  $\Delta$  as in Section 2.3. Let  $(X, r, \mu)$  be a metric measure space and  $\rho_0$  be an independent  $\mathbb{R}^{\mathbb{N}^2}$ -valued random variable that has the distance matrix distribution of  $(X, r, \mu)$ . That is, we can and will assume  $\rho_0 = (r(x(i), x(j)))_{i,j \in \mathbb{N}}$  for a  $\mu$ -iid sequence  $x(1), x(2), \dots$  in  $(X, r)$ . We view  $\rho_0$  as a random semi-metric on  $\mathbb{N}$ . Let  $(Z, \rho)$  be the lookdown space associated with  $\eta$  and  $\rho_0$  as defined in Section 2.

For each  $t \in \mathbb{R}_+$  and  $n \in \mathbb{N}$ , let the probability measure  $\mu_t^n$  on  $(Z, \rho)$  be the uniform measure on the first  $n$  individuals at time  $t$ , that is,

$$\mu_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{(t,i)}.$$

Let  $\mu_t$  be the weak limit of  $\mu_t^n$ , provided it exists. For almost all realizations of  $\eta$  and  $\rho_0$ , these weak limits exist simultaneously for all  $t \in \mathbb{R}_+$  by Theorem 3.1 below. The convergence is uniform for the Prohorov metric  $d_P^Z$  on  $(Z, \rho)$  for  $t$  in compact intervals.

**Theorem 3.1.** Assume  $\Xi \in \mathcal{M}_{\text{nd}}$ . Then there exists an event of probability 1 on which the following assertions hold:

(i) For all  $T \in \mathbb{R}_+$ ,

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} d_{\mathbb{P}}^Z(\mu_t^n, \mu_t) = 0.$$

(ii) The map  $t \mapsto \mu_t$  is càdlàg in the weak topology on the space of probability measures on  $(Z, \rho)$ . The set  $\Theta_0$ , defined in (2.7), is the set of jump times.

(iii) For all  $t \in \Theta_0$ , the measure  $\mu_t$  contains atoms, and the left limit  $\mu_{t-}$  is non-atomic.

The proof is given in Section 9.2. Note that the measure  $\mu_t^n$  assigns mass  $1/n$  to each offspring in a reproduction event. At the time  $t$  of a large reproduction event, the measure  $\mu_t$  has an atom on each family of individuals that descend from the same individual at time  $t-$ . Indeed, these individuals have genealogical distance zero and are identified in the lookdown space. In the Kingman case, that is, if  $\Xi = \delta_0$ , there are a.s. no large reproduction events and  $t \mapsto \mu_t$  is a.s. continuous.

For  $t \in \mathbb{R}_+$ , let  $X_t$  be the closure of the set of individuals  $\{t\} \times \mathbb{N}$  at time  $t$ , defined as a subspace of the complete space  $(Z, \rho)$ . Clearly, the probability measures  $\mu_0^n$  do not depend on  $\eta$ . Their weak limit exists a.s. by the following lemma which is essentially Vershik’s proof [42, Theorem 4] of the Gromov reconstruction theorem, see also [22, Proposition 10.5]. We write  $\text{supp } \mu'$  for the closed support of a measure  $\mu'$ .

**Lemma 3.2.** The weak limit  $\mu_0$  of the probability measures  $\mu_0^n$  on  $Z$  exists a.s. The metric measure spaces  $(\text{supp } \mu, r, \mu)$  and  $(X_0, \rho, \mu_0)$  are a.s. measure-preserving isometric.

*Proof.* For our  $\mu$ -iid sequence  $(x(i), i \in \mathbb{N})$ , the empirical measures  $\mu^n := n^{-1} \sum_{i=1}^n \delta_{x(i)}$  converge to  $\mu$  a.s. by the Glivenko-Cantelli theorem. A.s., the isometry  $\{x(i) : i \in \mathbb{N}\} \rightarrow Z, x(i) \mapsto (0, i)$  can be extended to a surjective isometry  $\varphi : (\text{supp } \mu, r) \rightarrow (X_0, \rho)$  with  $\varphi(\mu^n) = \mu_0^n$ , hence also  $\varphi(\mu) = \mu_0$ .  $\square$

In Theorem 3.5 below, we consider measures  $\Xi$  that satisfy the “coming down from infinity”-assumption that there exists an event of probability 1 on which the number  $\#\{A_s(t, j) : j \in \mathbb{N}\}$  of ancestors at time  $s$  of the individuals at time  $t$  is finite for all  $t \in (0, \infty)$  and  $s \in (0, t)$ . Let  $\mathcal{M}_{\text{CDI}}$  denote the subset of those finite measures  $\Xi$  on  $\Delta$  that satisfy this assumption. Then  $\mathcal{M}_{\text{CDI}} \subset \mathcal{M}_{\text{nd}}$ . Indeed, if  $\Xi \in \mathcal{M}_{\text{dust}}$ , then the rate at which a given ancestral lineage merges with any other ancestral lineage is finite (as discussed e.g. in [22, Section 6.2]) which implies  $\#\{A_s(t, j) : j \in \mathbb{N}\} = \infty$  a.s. for all  $0 \leq s < t$ . Furthermore, if  $\Xi \in \mathcal{M}_{\text{CDI}}$ , then the set  $\Theta^{\text{ext}}$  of extinction times, defined in (2.6), is a.s. dense in  $(0, \infty)$ . Indeed, if  $\Theta^{\text{ext}}$  has no points in an interval  $[s, t]$ , then the individuals on all levels at time  $s$  are ancestors of individuals at time  $t$ .

**Remark 3.3.** “Coming down from infinity” usually refers to a property of a  $\Xi$ -coalescent. A  $\Xi$ -coalescent describes the genealogy at a fixed time in our lookdown model. The assumption on  $\Xi$  that for each fixed  $t \in (0, \infty)$ , there exists an event of probability 1 on which  $\#\{A_s(t, j) : j \in \mathbb{N}\} < \infty$  holds for all  $s \in (0, t)$  already implies  $\Xi \in \mathcal{M}_{\text{CDI}}$ . Indeed, from this a priori weaker assumption, it follows that there exists an event of probability 1 on which  $\#\{A_s(t, j) : j \in \mathbb{N}\} < \infty$  for all  $t \in (0, \infty) \cap \mathbb{Q}$  and  $s \in (0, t)$ . For a general  $t \in (0, \infty)$ , there exists a rational  $t' \in (s, t)$ , and monotonicity of the number of ancestors yields

$$\#\{A_s(t, j) : j \in \mathbb{N}\} = \#\{A_s(t', A_{t'}(t, j)) : j \in \mathbb{N}\} \leq \#\{A_s(t', j) : j \in \mathbb{N}\} < \infty.$$

**Remark 3.4.** If  $\Xi \in \mathcal{M}_{\text{CDI}}$ , then there exists an event of probability one on which all subsets  $X_t \subset Z$  with  $t \in (0, \infty)$  are compact. Indeed,  $\#\{A_s(t, j) : j \in \mathbb{N}\} < \infty$  for all

$s \in (0, t)$  implies that the complete subspace  $X_t$  is totally bounded by definition of the metric  $\rho$  on the lookdown space  $Z$ .

If  $(X, r)$  is compact, then  $X_0$  is a.s. compact by Lemma 3.2. This also implies assertion (i) of Theorem 3.5 below for  $t = 0$ .

Recall that the Hausdorff distance between two subsets  $A, B$  of a metric space  $(Y, d)$  is defined as the infimum over those  $\varepsilon > 0$  such that  $d(a, B) < \varepsilon$  for all  $a \in A$  and  $d(A, b) < \varepsilon$  for all  $b \in B$ . The Hausdorff distance is a metric on the set of closed subspaces of  $(Y, d)$ , see e.g. [6].

**Theorem 3.5.** *Assume  $\Xi \in \mathcal{M}_{\text{CDI}}$  and that  $(X, r)$  is compact. Then the following assertions hold on an event of probability 1:*

- (i) *For each  $t \in \mathbb{R}_+$ , the compact set  $X_t$  is the closed support of  $\mu_t$ .*
- (ii) *The map  $t \mapsto X_t$  is càdlàg for the Hausdorff distance on the set of closed subsets of  $(Z, \rho)$ . The set  $\Theta^{\text{ext}}$  is the set of jump times. For each  $t \in \Theta^{\text{ext}}$ , the set  $X_t$  and the left limit  $X_{t-}$  are not isometric.*

The proof is given in Section 9.2.

**Remark 3.6.** By Theorems 3.1 and 3.5, there exists an event of probability 1 on which  $\mu_t = \mu_{t-}$  and  $X_t \subsetneq X_{t-}$  for all  $t \in \Theta^{\text{ext}} \setminus \Theta_0$ . In particular, the closed support of  $\mu_{t-}$  is strictly smaller than  $X_{t-}$  for these  $t$  a.s. The set  $X_{t-} \setminus \text{supp } \mu_{t-} \subset X_{t-}$  is equal to  $X_{t-} \setminus X_t$  a.s., this is the part of the population at time  $t-$  that dies out at time  $t$ .

We conclude this subsection with a side observation (Proposition 3.8) on the intersection of the sets  $\Theta_0$  and  $\Theta^{\text{ext}}$  of jump times in Theorems 3.1 and 3.5. Reproduction events in which the whole population is replaced by finitely many particles and their offspring occur at the times in the set

$$\Theta_f := \{t \in (0, \infty) : \text{there exists } \pi \in \mathcal{P} \text{ with } \eta\{(t, \pi)\} > 0 \text{ and } \#\pi < \infty\}.$$

By the construction in Section 2, all particles with level larger than  $\#\pi$  die in a reproduction event that is encoded by a point  $(t, \pi)$  of  $\eta$ .

**Remark 3.7.** If  $\Xi$  is concentrated on  $\{(x, 0, 0, \dots) : x \in [0, 1]\} \subset \Delta$ , then a.s., no simultaneous multiple reproduction events occur. This case corresponds to the coalescents with multiple collisions ( $\Lambda$ -coalescents). In this case,  $\Theta_f$  is a.s. the set of times at which the whole population is replaced by a single particle and its offspring. If  $\Xi$  is concentrated on  $\{(x, 0, 0, \dots) : x \in [0, 1)\} \subset \Delta$ , then  $\Theta_f = \emptyset$  a.s. More generally,  $\Theta_f = \emptyset$  a.s. if and only if  $\Xi\{x \in \Delta : x_1 + \dots + x_k = 1 \text{ for some } k \in \mathbb{N}\} = 0$ .

**Proposition 3.8.** *A.s.,  $\Theta^{\text{ext}} \cap \Theta_0 = \Theta_f$ .*

For the proof below in this subsection, as well as for later use in Sections 8 and 10, we now express  $\eta$  in terms of a collection of Poisson processes. Recall the set  $\mathcal{P}_n$  of partitions of  $[n] = \{1, \dots, n\}$ , and let  $\pi'_1, \pi'_2, \dots$  be an arbitrary enumeration of the set

$$\bigcup_{n \in \mathbb{N}} (\mathcal{P}_n \setminus \{\{\{1\}, \dots, \{n\}\}\})$$

of finite partitions that consist not only of singleton blocks. For  $k \in \mathbb{N}$ , let

$$\mathcal{P}'_k = \{\pi \in \mathcal{P} : \gamma_n(\pi) = \pi'_k \text{ with } n \text{ such that } \pi'_k \text{ is a partition of } [n]\},$$

and

$$J_{t,k} = \eta((0, t] \times \mathcal{P}'_k)$$

for  $t \in \mathbb{R}_+$ . Then the processes  $(J_{t,k}, t \in \mathbb{R}_+)$  form a collection (indexed by  $k \in \mathbb{N}$ ) of Poisson processes. We endow  $\mathbb{R}_+^{\mathbb{N}}$  with the product topology and consider the  $\mathbb{R}_+^{\mathbb{N}}$ -valued stochastic process

$$J = (J_t, t \in \mathbb{R}_+) = ((J_{t,k}, k \in \mathbb{N}), t \in \mathbb{R}_+) \tag{3.1}$$

Note that  $J$  has independent and stationary increments and a. s. càdlàg paths. Hence,  $J$  is a strong Markov process and Feller continuous (i. e. the elements of its semigroup preserve the set of bounded continuous functions), so that we obtain from e. g. Theorem (5.1) in Chapter I of Blumenthal [5] that  $J$  is quasi-left-continuous. Let  $\mathcal{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$  be the complete filtration induced by  $J$ . Then  $\mathcal{F}_t$  is the sigma field generated by the random measure  $\eta(\cdot \times ((0, t] \times \mathcal{P}))$  and all null events. By condition (2.1),  $J$  stays finite a. s. and the set of jump times of  $J$  equals a. s. the set of reproduction times  $\{t \in (0, \infty) : \eta(\{t\} \times \mathcal{P}) > 0\}$ .

*Proof of Proposition 3.8.* We use notation also from Section 2.2. First we show  $\Theta_f \subset \Theta^{\text{ext}} \cap \Theta_0$  a. s. Let  $t \in \Theta_f$ . Then there exists a point  $(t, \pi)$  of  $\eta$  with  $i - 1 := \#\pi < \infty$ . By condition (2.1), as the level of each particle is non-decreasing in time, and as offspring always has a higher level than the reproducing particle, there exists a. s. a time  $s \in (0, t)$  such that  $D_{s'}(s, i) = i$  for all  $s' \in [s, t)$ . In particular, all descendants of  $(s, i)$  at time  $t-$  have level at least  $i$ . Hence,  $(s, i)$  has no descendants at time  $t$ , that is,  $D_t(s, i) = \infty$ , and it follows  $t \in \Theta^{\text{ext}}$ . Clearly,  $\#\pi < \infty$  implies that  $\pi$  contains blocks of infinite size. The definition of  $H_{\Xi}$  and Kingman’s correspondence imply  $\Theta_f \subset \Theta_0$  a. s. It remains to show that  $\Theta^{\text{ext}} \cap \Theta_0 \subset \Theta_f$  a. s.

On the event of probability 1 on which condition (2.1) holds, particles on any level remain on that level for a positive amount of time. This implies

$$\Theta^{\text{ext}} = \{\tau_{s,i} : s \in \mathbb{Q}_+, i \in \mathbb{N}\} \text{ a. s.}$$

For  $s \in \mathbb{R}_+, i, n \in \mathbb{N}$ , we define the  $\mathcal{F}$ -stopping time

$$\tau_{s,i,n} = \inf\{t \geq s : D_t(s, i) \geq n\}.$$

Then  $\tau_{s,i,n}$  is non-decreasing in  $n$ , and  $\tau_{s,i} = \lim_{n \rightarrow \infty} \tau_{s,i,n}$ . We assume w. l. o. g.  $\Xi(\Delta) > 0$ . Then  $\tau_{s,i,n} \in [s, \infty)$  for all  $s \in \mathbb{R}_+$  and  $i, n \in \mathbb{N}$  a. s. Let  $E_{s,i}$  be the event that  $\tau_{s,i,n} < \tau_{s,i}$  for all  $n \in \mathbb{N}$ .

A. s. by (2.1), on the event  $E_{s,i}^c$  that  $\tau_{s,i} = \tau_{s,i,n}$  for some  $n \in \mathbb{N}$ , a particle on a level below  $n$  at time  $\tau_{s,i,n}-$  dies at time  $\tau_{s,i,n}$  due to a reproduction event that lies in  $\Theta_f$ .

To show the assertion of the proposition, it now suffices to show that  $\tau_{s,i} \notin \Theta_0$  a. s. on  $E_{s,i}$ . We define the  $\mathcal{F}$ -stopping time  $\tilde{\tau}_{s,i}$  by  $\tilde{\tau}_{s,i} = \tau_{s,i} \mathbb{1}_{E_{s,i}} + \infty \mathbb{1}_{E_{s,i}^c}$ . Then the  $\mathcal{F}$ -stopping times

$$\tilde{\tau}_{s,i,n} := \begin{cases} \tau_{s,i,n} & \text{if } \tau_{s,i,n} < \tau_{s,i} \\ \tau_{s,i,n} \vee n & \text{if } \tau_{s,i,n} = \tau_{s,i} \end{cases}$$

form an announcing sequence for  $\tilde{\tau}_{s,i}$ , that is,  $\tilde{\tau}_{s,i,n} < \tilde{\tau}_{s,i}$  a. s. and  $\tilde{\tau}_{s,i} = \lim_{n \rightarrow \infty} \tilde{\tau}_{s,i,n}$  a. s. Quasi-left-continuity of  $J$  implies  $J_{\tilde{\tau}_{s,i}-} = J_{\tilde{\tau}_{s,i}}$  a. s. Hence, a. s. on  $E_{s,i}$ , no reproduction event occurs at time  $\tilde{\tau}_{s,i} = \tau_{s,i}$ , and we have  $\tau_{s,i} \notin \Theta_0$ .  $\square$

**Remark 3.9.** In the Kingman case, the set  $\Theta^{\text{ext}}$  of extinction times is described by Poisson processes by Dahmer, Knobloch, and Wakolbinger [8, Proposition 1], see also the references therein.

### 3.2 The general case

We construct a family of probability measures on the Cartesian product of the extended lockdown space and the mark space  $\mathbb{R}_+$ . Let  $(r_0, v_0)$  be an independent  $\mathbb{R}^{\mathbb{N}^2} \times \mathbb{R}^{\mathbb{N}}$ -valued random variable that has the marked distance matrix distribution of an

$(\mathbb{R}_+)$ -marked metric measure space  $(X, r, m)$ . That is,  $(X, r)$  is a complete and separable metric space,  $m$  is a probability measure on the Borel sigma algebra on the product space  $X \times \mathbb{R}_+$ , and we may assume that  $(x(i), v(i))_{i \in \mathbb{N}}$  is an  $m$ -iid sequence in  $X \times \mathbb{R}_+$  and set  $r_0(i, j) = r(x(i), x(j))$  for  $i, j \in \mathbb{N}$ . Then we can view  $r_0$  as a random semi-metric on  $\mathbb{N}$ . Let  $(\hat{Z}, \rho)$  be the extended lookdown space associated with  $\eta$  and  $(r_0, v_0)$ , as defined in Section 2.1. We endow  $\hat{Z} \times \mathbb{R}_+$  with the product metric  $d^{\hat{Z} \times \mathbb{R}_+}((z, v), (z', v')) = \rho(z, z') \vee |v - v'|$ . We denote the Prohorov metric on the space of probability measures on  $\hat{Z} \times \mathbb{R}_+$  by  $d_P^{\hat{Z} \times \mathbb{R}_+}$ . Recall from Section 2.1 also the parent  $z(t, i)$  and the genealogical distance  $v_t(i)$  between the individual  $(t, i)$  and its parent. For each  $t \in \mathbb{R}_+$  and  $n \in \mathbb{N}$ , we define a probability measure  $m_t^n$  on  $\hat{Z} \times \mathbb{R}_+$  by

$$m_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{(z(t,i), v_t(i))} \tag{3.2}$$

The first component  $m_t^n(\cdot \times \mathbb{R}_+)$  lays mass on the parents of the first  $n$  individuals at time  $t$ . The second component  $m_t^n(\hat{Z} \times \cdot)$  records the genealogical distances to these parents. Let  $m_t$  denote the weak limit of  $m_t^n$  provided it exists. This existence is addressed in Theorem 3.10 below in the case with dust, and in Lemma 3.11, Remark 3.13, and Corollary 3.14 below in the case without dust.

**Theorem 3.10.** *Assume  $\Xi \in \mathcal{M}_{\text{dust}}$ . Then the following assertions hold on an event of probability 1:*

(i) *For all  $T \in \mathbb{R}_+$ ,*

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} d_P^{\hat{Z} \times \mathbb{R}_+}(m_t^n, m_t) = 0.$$

(ii) *The map  $t \mapsto m_t$  is càdlàg in the weak topology on the space of probability measures on  $\hat{Z} \times \mathbb{R}_+$ . The set  $\Theta_0$ , defined in (2.7), is the set of jump times.*

(iii) *For each  $t \in (0, \infty)$ , the left limit  $m_{t-}$  satisfies  $m_{t-}(\hat{Z} \times \{0\}) = 0$ . For each  $t \in \Theta_0$ , it holds  $m_t(\hat{Z} \times \{0\}) > 0$ .*

(iv) *If  $m$  is purely atomic, then  $m_t$  and  $m_{t-}$  are purely atomic for all  $t \in (0, \infty)$ .*

The proof is given in Section 9.1. In Proposition 9.2, the measures  $m_t$  are stated explicitly. At the times of large reproduction events,  $v_t(i) = 0$  for all individuals  $i$  with levels in a non-singleton block. This yields the positive mass of  $m_t(\hat{Z} \times \cdot)$  in zero asserted in Theorem 3.10(iii).

**Lemma 3.11.** *The weak limit  $m_0$  of the probability measures  $m_0^n$  on  $\hat{Z} \times \mathbb{R}_+$  exists a. s.*

*Proof.* This follows analogously to Lemma 3.2. □

**Remark 3.12.** We recall that in case  $\Xi \in \mathcal{M}_{\text{nd}}$ , the rate at which each particle reproduces is infinite. In this case, there exists an event of probability 1 on which  $v_t(i) = 0$  and  $(t, i) = z(t, i)$  for all  $t \in (0, \infty)$  and  $i \in \mathbb{N}$ , cf. [22, Section 6.2].

As in Subsection 3.1, we define the probability measures

$$\mu_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{(t,i)} \tag{3.3}$$

on  $\hat{Z}$ , and we denote their weak limits by  $\mu_t$ .

**Remark 3.13.** If  $m(X \times \{0\}) = 1$ , then  $m = \mu \otimes \delta_0$  for a probability measure  $\mu$  on  $X$ , hence  $v_0 = 0$  a.s. and  $r_0$  has the distance matrix distribution of the metric measure space  $(X, r, \mu)$ . By Remark 2.2, the extended lookdown space  $(\hat{Z}, \rho)$  coincides in this case a.s. with the lookdown space associated with  $\eta$  and  $r_0$ . With this identification, the assertions of Theorem 3.1 also hold in the context of the present subsection if  $\Xi \in \mathcal{M}_{\text{nd}}$  and  $m(\hat{Z} \times \{0\}) = 1$ .

Moreover, in the case without dust, the following corollary to Theorem 3.1 also holds for the extended lookdown space and the more general initial configuration  $(r_0, v_0)$  in the present subsection.

**Corollary 3.14.** Assume  $\Xi \in \mathcal{M}_{\text{nd}}$ . Then a.s., the probability measures  $\mu_t$  exist for all  $t \in (0, \infty)$ . The map  $t \mapsto \mu_t$  is a.s. càdlàg on  $(0, \infty)$  in the weak topology on the space of probability measures on  $(\hat{Z}, \rho)$ , and  $\Theta_0$  is a.s. the set of jump times. Moreover, the family of probability measures  $(\mu_t, t \in (0, \infty))$  satisfies a.s. assertion (iii) of Theorem 3.1. A.s., also the probability measures  $m_t$  exist for all  $t \in (0, \infty)$  and satisfy  $m_t = \mu_t \otimes \delta_0$ .

*Proof.* Let  $(Z', \rho')$  be the lookdown space associated with  $\eta$  and  $(0)_{i,j \in \mathbb{N}}$ . By Theorem 3.1, the probability measures  $\mu'_t$  defined on  $Z'$  analogously to  $\mu_t$  satisfy the assertion.

Let  $\varepsilon > 0$ , let  $Z_\varepsilon$  be the closure of  $[\varepsilon, \infty) \times \mathbb{N}$  in  $(\hat{Z}, \rho)$ , and  $Z'_\varepsilon$  the closure of  $[\varepsilon, \infty) \times \mathbb{N}$  in  $(Z', \rho')$ . The construction in the beginning of Section 2 yields  $\rho'((t, i), (u, j)) \wedge \varepsilon = \rho((t, i), (u, j)) \wedge \varepsilon$  for all  $(t, i), (u, j) \in [\varepsilon, \infty) \times \mathbb{N}$ . Hence, the map from  $[\varepsilon, \infty) \times \mathbb{N} \subset Z'_\varepsilon$  to  $Z_\varepsilon$ , given by  $(t, i) \mapsto (t, i)$ , can be extended to a homeomorphism  $h : Z'_\varepsilon \rightarrow Z_\varepsilon$ . Hence a.s., the weak limits

$$\mu_t = \text{w-}\lim_{n \rightarrow \infty} \mu_t^n = \text{w-}\lim_{n \rightarrow \infty} h(\mu_t'^n) = h(\mu_t')$$

exist for all  $t \in [\varepsilon, \infty)$  and the assertion on  $(\mu_t, t \in (0, \infty))$  follows. The assertion on  $(m_t, t \in (0, \infty))$  now follows from Remark 3.12, the definitions (3.2) and (3.3) of  $m_t^n$  and  $\mu_t^n$ , and the definitions of  $m_t$  and  $\mu_t$  as weak limits.  $\square$

## 4 Stochastic processes

### 4.1 The case without dust

From the construction on the lookdown space in Section 3.1, we now read off stochastic processes with values in the space  $\mathbb{M}$  of isomorphy classes of metric measure spaces and in the space  $\mathbb{M}$  of strong isomorphy classes of compact metric measure spaces. First we recall these state spaces from the literature [18, 31, 16].

As stated in the introduction, we call two metric measure spaces  $(X', r', \mu')$ ,  $(X'', r'', \mu'')$  isomorphic if there exists an isometry  $\varphi$  from the closed support  $\text{supp } \mu' \subset X'$  to  $\text{supp } \mu'' \subset X''$  with  $\mu'' = \varphi(\mu')$ . We denote the isomorphy class by  $\llbracket X', r', \mu' \rrbracket$ . We endow the space  $\mathbb{M}$  of isomorphy classes of metric measure spaces with the Gromov-Prohorov metric  $d_{\text{GP}}$  which is complete and separable and induces the Gromov-weak topology, as shown in [18].

Moreover, we call two metric measure spaces  $(X', r', \mu')$ ,  $(X'', r'', \mu'')$  strongly isomorphic if they are measure-preserving isometric, that is, if there exists a surjective isometry  $\varphi : X \rightarrow X'$  with  $\mu'' = \varphi(\mu')$ . We denote the strong isomorphy class by  $[X', r', \mu']$ . We endow the space  $\mathbb{M}$  of strong isomorphy classes of compact metric measure spaces with the Gromov-Hausdorff-Prohorov metric  $d_{\text{GHP}}$ , given by

$$d_{\text{GHP}}(\llbracket X', r', \mu' \rrbracket, \llbracket X'', r'', \mu'' \rrbracket) = \inf_{Y, \varphi', \varphi''} \{d_{\text{P}}^Y(\varphi'(\mu'), \varphi''(\mu'')) \vee d_{\text{H}}^Y(\varphi'(X'), \varphi''(X''))\}$$

where the infimum is over all isometric embeddings  $\varphi' : X' \rightarrow Y$ ,  $\varphi'' : X'' \rightarrow Y$  into complete and separable metric spaces  $Y$ . Here we denote by  $d_{\text{P}}^Y$  and by  $d_{\text{H}}^Y$  the Prohorov and the Hausdorff distance, respectively, over a metric space  $Y$ . Then  $(\mathbb{M}, d_{\text{GHP}})$  is

a complete and separable metric space, see [31, 16], and  $d_{\text{GHP}}$  induces the Gromov-Hausdorff-Prohorov topology on  $\mathbf{M}$ . The Hausdorff distance in the definition of  $d_{\text{GHP}}$  compares the metric spaces also where the probability measures charges them with negligible mass.

We work with the lookdown space  $(Z, \rho)$ , the families of sampling measures  $(\mu_t, t \in \mathbb{R}_+)$ , and the subspaces  $X_t$  of the lookdown space from Section 3.1. Recall that the randomness comes from a Poisson random measure  $\eta$  that is characterized by  $\Xi \in \mathcal{M}_{\text{nd}}$ , and from an independent random variable  $\rho_0$  with the distance matrix distribution of a metric measure space  $(X, r, \mu)$ .

We say that a Markov process  $(Y_t, t \in \mathbb{R}_+)$  with values in a (not necessarily locally compact) metric space is Feller continuous if for each  $t \in \mathbb{R}_+$ , the law of  $Y_t$  depends continuously with respect to the weak topology on the initial state.

**Proposition 4.1.** *Assume  $\Xi \in \mathcal{M}_{\text{nd}}$ . Then a Feller-continuous strong Markov process with values in  $\mathbf{M}$  is given a. s. by  $(\llbracket Z, \rho, \mu_t \rrbracket, t \in \mathbb{R}_+)$ .*

**Proposition 4.2.** *Assume  $\Xi \in \mathcal{M}_{\text{CDI}}$  and that  $(X, r)$  is compact. Then a Feller-continuous strong Markov process with values in  $\mathbf{M}$  is given a. s. by  $(\llbracket X_t, \rho, \mu_t \rrbracket, t \in \mathbb{R}_+)$ .*

By Remark 3.4, the assumption  $\Xi \in \mathcal{M}_{\text{CDI}}$  ensures that a. s., the spaces  $X_t$  are compact for all  $t \in \mathbb{R}_+$ . The proofs of Propositions 4.1 and 4.2 are given further below in this subsection.

We call the process in Proposition 4.1 an  $\mathbf{M}$ -valued  $\Xi$ -Fleming-Viot process with initial state  $\llbracket X, r, \mu \rrbracket$ . By Remark 4.4 below, this process is the  $\mathbb{U}$ -valued  $\Xi$ -Fleming-Viot process from [22, Section 7.1] if  $\rho_0$  is a semi-ultrametric. We call the process in Proposition 4.2 an  $\mathbf{M}$ -valued evolving  $\Xi$ -coalescent starting from  $\llbracket X, r, \mu \rrbracket$ . Note that  $\llbracket Z, \rho, \mu_t \rrbracket$  in Proposition 4.1 above can be replaced by  $\llbracket X_t, \rho, \mu_t \rrbracket$  as  $\mu_t$  is supported by  $X_t$  for all  $t \in \mathbb{R}_+$  a. s.

**Remark 4.3** (Isomorphism classes and strong isomorphism classes). In this remark, we assume  $\Xi \in \mathcal{M}_{\text{CDI}}$  and that  $(X, r)$  is compact. By construction, the  $\mathbf{M}$ -valued  $\Xi$ -Fleming-Viot process in Proposition 4.2 depends on  $(X, r, \mu)$  only through the isomorphism class  $\llbracket X, r, \mu \rrbracket$ . In particular, it does not depend on  $X \setminus \text{supp } \mu$ . As  $\text{supp } \mu_0 = X_0$  a. s. by Lemma 3.2, the strong isomorphism class  $\llbracket X, r, \mu \rrbracket$  is not necessarily the initial state.

Moreover, let  $\mathbf{M}_c$  be the space of isomorphism classes of compact metric measure spaces, and let  $f : \mathbf{M}_c \rightarrow \mathbf{M}$ ,  $\llbracket X', r', \mu' \rrbracket \mapsto [\text{supp } \mu', r', \mu']$  be the function that maps an isomorphism class to the strong isomorphism class of a representative where the measure has full support. Using Theorem 3.5(i), we then obtain that  $(f(\chi_t), t \in \mathbb{R}_+)$  is an  $\mathbf{M}$ -valued evolving  $\Xi$ -coalescent if  $(\chi_t, t \in \mathbb{R}_+)$  is an  $\mathbf{M}$ -valued  $\Xi$ -Fleming-Viot process with initial state  $\chi \in \mathbf{M}_c$ . Conversely, let  $g : \mathbf{M} \rightarrow \mathbf{M}$ ,  $\llbracket X', r', \mu' \rrbracket \mapsto \llbracket X', r', \mu' \rrbracket$ . Then  $(g(\mathcal{X}_t), t \in \mathbb{R}_+)$  is an  $\mathbf{M}$ -valued  $\Xi$ -Fleming-Viot process if  $(\mathcal{X}_t, t \in \mathbb{R}_+)$  is an  $\mathbf{M}$ -valued evolving  $\Xi$ -coalescent. Also note that  $g$  is continuous. By Remark 4.6 below, the function  $f$  is measurable.

We denote by  $\mathfrak{D}$  the space of semi-metrics on  $\mathbb{N}$ . We do not distinguish between a semi-metric  $\rho' \in \mathfrak{D}$  and the distance matrix  $(\rho'(i, j))_{i, j \in \mathbb{N}}$  and we consider  $\mathfrak{D}$  as a subspace of  $\mathbb{R}^{\mathbb{N}^2}$  which we endow with the product topology. Let the  $\mathfrak{D}$ -valued Markov process  $(\rho_t, t \in \mathbb{R}_+)$  be defined from  $\eta$  and  $\rho_0$  as in equation (2.3). For each  $t \in \mathbb{R}_+$ , the random variable  $\rho_t$  is exchangeable by [22, Proposition 5.8], that is,  $\rho_t$  and  $p(\rho_t)$  are equal in distribution for all bijections  $p : \mathbb{N} \rightarrow \mathbb{N}$ . Here the action of the group of bijections  $\mathbb{N} \rightarrow \mathbb{N}$  on  $\mathfrak{D}$  is defined by  $p(\rho') = (\rho'(p(i), p(j)))_{i, j \in \mathbb{N}}$  for  $\rho' \in \mathfrak{D}$  and a bijection  $p : \mathbb{N} \rightarrow \mathbb{N}$ .

**Remark 4.4.** Recall the measurable map  $\psi : \mathfrak{D} \rightarrow \mathbf{M}$  from [22, Section 3.3]. By construction and Theorem 3.1(i), we have  $\llbracket Z, \rho, \mu_t \rrbracket = \psi(\rho_t)$  for all  $t \in \mathbb{R}_+$  a. s.

*Proof of Proposition 4.1 (beginning).* By Remark 4.4,  $\llbracket Z, \rho, \mu_t \rrbracket$  is a random variable. The Markov property follows, precisely as in [22, Theorem 4.1], from an application of Theorem 2 of Rogers and Pitman [37] to the Markov process  $(\rho_t, t \in \mathbb{R}_+)$ , the measurable map  $\psi : \mathfrak{D} \rightarrow \mathbb{M}$ , and the probability kernel from  $\mathbb{M}$  to  $\mathfrak{D}$  given by  $(\chi, B) \mapsto \nu^\chi(B)$ . Here we use the exchangeability of  $\rho_t$ .

Feller continuity can be shown as in Corollary 8.2 of [22]. □

The proof of Proposition 4.2 is analogous. Instead of the map  $\psi$ , we need another map  $v$  which we now define. Let  $\mathfrak{D}_c \subset \mathfrak{D}$  be the space of totally bounded semi-metrics on  $\mathbb{N}$ ,

$$\mathfrak{D}_c = \left\{ \rho' \in \mathfrak{D} : \lim_{n \rightarrow \infty} \sup_{j > n} \inf_{i \leq n} \rho'(i, j) = 0 \right\}.$$

Let  $v : \mathfrak{D}_c \rightarrow \mathbb{M}$  be the function that maps  $\rho' \in \mathfrak{D}_c$  to the strong isomorphy class  $[X', \rho', \mu']$  of the compact metric measure space  $(X', \rho', \mu')$  defined as follows:  $(X', \rho')$  is the completion of the metric space obtained by identifying the elements with  $\rho'$ -distance zero in  $(\mathbb{N}, \rho')$ . The probability measure  $\mu'$  on  $(X', \rho')$  is the weak limit  $w\text{-}\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \delta_i$  if it exists, else we set  $\mu' = \delta_1$ . The following lemma is analogous to [22, Proposition 3.7].

**Lemma 4.5.** *The function  $v : \mathfrak{D}_c \rightarrow \mathbb{M}$  is measurable.*

Again, we refer by an element of a semi-metric space also to the corresponding element of the completion of the metric space that is obtained by identifying points with distance zero.

*Proof.* For  $n \in \mathbb{N}$ , let  $\mathfrak{D}_n \subset \mathbb{R}^{n^2}$  be the space of semi-metrics on  $[n]$ , again we do not distinguish between semi-metrics and distance matrices. Let  $v_n : \mathfrak{D}_n \rightarrow \mathbb{M}$  be the function that maps  $\rho' \in \mathfrak{D}_n$  to the strong isomorphy class of the metric measure space  $(X', \rho', n^{-1} \sum_{i=1}^n \delta_i)$ , where  $(X', \rho')$  is the metric space obtained by identifying the elements of  $[n]$  with  $\rho'$ -distance zero. Clearly, the map  $v_n$  is continuous. To show this formally, we define analogously a metric measure space  $(X'', \rho'', n^{-1} \sum_{i=1}^n \delta_i)$  from another  $\rho'' \in \mathfrak{D}_n$ . From [31, Proposition 6], it follows

$$d_{\text{GHP}} \left( \left( X', \rho', n^{-1} \sum_{i=1}^n \delta_i \right), \left( X'', \rho'', n^{-1} \sum_{i=1}^n \delta_i \right) \right) \leq \frac{1}{2} \max_{i, j \leq n} |\rho'(i, j) - \rho''(i, j)|,$$

we use the coupling  $\nu = n^{-1} \sum_{i=1}^n \delta_{(i, i)}$  on  $X' \times X''$  and the correspondence  $\mathfrak{R} = \{(i, i) : i \in [n]\} \subset X' \times X''$ .

Now let  $\rho' \in \mathfrak{D}_c$ , and let  $(X', \rho', \mu')$  be defined as in the definition of  $v(\rho')$  above. Using the definition of the Gromov-Hausdorff-Prohorov metric, we obtain that

$$d_{\text{GHP}}(v(\rho'), v_n((\rho'(i, j))_{i, j \leq n})) \leq d_{\text{P}}^{X'} \left( \mu', n^{-1} \sum_{i=1}^n \delta_i \right) \vee d_{\text{H}}^{X'}(X', [n]) \rightarrow 0 \quad (n \rightarrow \infty)$$

if the weak limit  $\mu'$  of the measures  $n^{-1} \sum_{i=1}^n \delta_i$  on  $X'$  exists. This yields the assertion. □

*Proof of Proposition 4.2 (beginning).* By Remark 3.4, there exists an event of probability 1 on which  $\rho_t \in \mathfrak{D}_c$  for all  $t \in \mathbb{R}_+$ . By construction and Theorems 3.1(i) and 3.5(i), we have  $\llbracket X_t, \rho, \mu_t \rrbracket = v(\rho_t)$  for all  $t \in \mathbb{R}_+$  a. s. Hence, Lemma 4.5 yields that  $\llbracket X_t, \rho, \mu_t \rrbracket$  is a random variable. The Markov property follows as in [22, Theorem 4.1] from an application of [37, Theorem 2] to the Markov process  $(\rho_t, t \in \mathbb{R}_+)$ , the measurable map  $v : \mathfrak{D}_c \rightarrow \mathbb{M}$ , and the probability kernel from  $\mathbb{M}$  to  $\mathfrak{D}_c$  given by  $(\chi, B) \mapsto \nu^\chi(B)$ . Here we use exchangeability of  $\rho_t$ . To check that Condition (a) in [37, Theorem 2] is satisfied, we note that  $v(\rho') = v(\rho_t)$  a. s. for  $t \in \mathbb{R}_+$  and a random variable  $\rho'$  with conditional distribution  $\nu^{v(\rho_t)}$  given  $v(\rho_t)$ . This a. s. equality follows as in the proof of [22, Proposition 10.5], we also use Theorem 3.5(i). □

**Remark 4.6.** As a by-product of Lemma 4.5, let us deduce measurability of the canonical map  $f : \mathbb{M}_c \rightarrow \mathbb{M}$  from Remark 4.3 (this answers a question posed to the author by H. Sulzbach). We consider  $\chi \in \mathbb{M}_c$  and a random variable  $\rho$  with the distance matrix distribution  $\nu^\chi$ . As in the proof of [22, Proposition 10.5], it follows that  $f(\chi) = v(\rho)$  a. s. Hence, for a Borel subset  $A \subset \mathbb{M}$ , we obtain the equivalence

$$\mathbb{1}\{f(\chi) \in A\} = 1 \iff \int \nu^\chi(d\rho') \mathbb{1}\{v(\rho') \in A\} = 1 \iff \nu^\chi(v^{-1}(A)) = 1.$$

The function that maps a metric measure space  $\chi \in \mathbb{M}_c$  to its distance matrix distribution  $\nu^\chi$  is continuous by definition of the Gromov-Prohorov topology (see [18]). Lemma 4.5 now implies that  $\{\chi \in \mathbb{M}_c : f(\chi) \in A\}$  is a measurable subset of  $\mathbb{M}_c$ . We remark that measurability of  $f$  can also be obtained as a consequence of [1, Corollary 5.6] and e. g. [25, Theorem 15.1].

Path regularity of the  $\mathbb{M}$ -valued  $\Xi$ -Fleming-Viot process and the  $\mathbb{M}$ -valued evolving  $\Xi$ -coalescent follows from Section 3.1:

**Proposition 4.7.** *Assume  $\Xi \in \mathcal{M}_{\text{nd}}$ . Then a. s., the process  $(\llbracket Z, \rho, \mu_t \rrbracket, t \in \mathbb{R}_+)$  has càdlàg paths in the Gromov-weak topology and  $\Theta_0$  is the set of jump times.*

**Proposition 4.8.** *Assume  $\Xi \in \mathcal{M}_{\text{CDI}}$ . Then a. s., the process  $(\llbracket X_t, \rho, \mu_t \rrbracket, t \in \mathbb{R}_+)$  has càdlàg paths in the Gromov-Hausdorff-Prohorov topology and  $\Theta_0 \cup \Theta^{\text{ext}}$  is the set of jump times.*

*Proof of Proposition 4.7.* By Theorem 3.1(ii) and the definition of the Gromov-Prohorov metric, it follows that a. s., the map  $t \mapsto \llbracket Z, \rho, \mu_t \rrbracket$  is càdlàg and the set of jump times is not larger than  $\Theta_0$ . By Theorem 3.1(iii) and as the atomicity properties only depend on the isomorphy classes, it follows that a. s., the set of jump times is not smaller than  $\Theta_0$ . □

*Proof of Proposition 4.8.* This is analogous to Proposition 4.7. We use Theorems 3.1(ii) and 3.5(ii), the definition of the Gromov-Hausdorff-Prohorov metric, and the atomicity properties from Theorem 3.1(iii) which are determined by the strong isomorphy classes. We also use that  $X_{t-}$  and  $X_t$  are isometric if  $\llbracket X_{t-}, \rho, \mu_{t-} \rrbracket = \llbracket X_t, \rho, \mu_t \rrbracket$ . □

*Proof of Proposition 4.1 (end).* The strong Markov property can be deduced by standard arguments (cf. e. g. the proof of Theorem 4.2.7 in [14]) from Feller continuity and a. s. right continuity of the sample paths (Proposition 4.7). □

Now we study Feller continuity of the  $\mathbb{M}$ -valued evolving  $\Xi$ -coalescent.

**Lemma 4.9.** *Let  $(X^n, r^n, \mu^n)$  be a sequence of compact metric measure spaces such that  $\llbracket X^n, r^n, \mu^n \rrbracket$  converges to  $\llbracket X, r, \mu \rrbracket$  in the Gromov-weak topology. Assume  $\Xi \in \mathcal{M}_{\text{CDI}}$  and let  $(\mathcal{X}_t^n, t \in \mathbb{R}_+)$  be an  $\mathbb{M}$ -valued evolving  $\Xi$ -coalescent starting from  $\llbracket X^n, r^n, \mu^n \rrbracket$ . Then for each  $t \in (0, \infty)$ , the random variable  $\mathcal{X}_t^n$  converges in distribution to  $\llbracket X_t, \rho, \mu_t \rrbracket$  in  $\mathbb{M}$ , endowed with the Gromov-Hausdorff-Prohorov topology.*

*Proof.* Let  $t, \varepsilon > 0$  and  $n \in \mathbb{N}$ . Let  $\rho_0^n$  be a random variable with distribution  $\nu^{\llbracket X^n, r^n, \mu^n \rrbracket}$  that is independent of  $\eta$ . Recall the definition of  $\rho_0$  from Section 3.1. Let  $(Z', \rho')$  be the lookdown space associated with  $\eta$  and  $\rho_0^n$ . Let  $X'_t \subset Z'$  be the closure of the set  $\{t\} \times \mathbb{N}$  of individuals at time  $t$  therein, and define a probability measure  $\mu'_t$  on  $Z'$  analogously to  $\mu_t$ . Then a. s., the map  $X_t \rightarrow X'_t, (t, i) \mapsto (t, i)$  can be extended to a measure-preserving homeomorphism  $h$  between  $(X_t, \rho, \mu_t)$  and  $(X'_t, \rho', \mu'_t)$ . The correspondence  $\mathfrak{R} = \{(x, h(x)) : x \in X_t\} \subset X_t \times X'_t$  has distortion  $\max\{|\rho_0^n(i, j) - \rho_0(i, j)| : i, j \in A_0(t, \mathbb{N})\}$ , where we write  $A_0(t, \mathbb{N}) = \{A_0(t, \ell) : \ell \in \mathbb{N}\}$ . With the coupling  $\nu(dx dx') = \mu_t(dx) \delta_{h(x)}(dx')$  of  $\mu_t$  and  $\mu'_t$ , Proposition 6 in [31] implies

$$\begin{aligned} & \mathbb{P}(d_{\text{GHP}}([X'_t, \rho', \mu'_t], [X_t, \rho, \mu_t]) \geq \varepsilon) \\ & \leq \mathbb{P}(\max\{|\rho_0^n(i, j) - \rho_0(i, j)| : i, j \in [k]\} \geq 2\varepsilon) + \mathbb{P}(\#A_0(t, \mathbb{N}) > k) \end{aligned}$$

for all  $k \in \mathbb{N}$ . W.l.o.g., we may assume  $\mathcal{X}_t^n = [X'_t, \rho', \mu'_t]$  for all  $t \in \mathbb{R}_+$  a.s., and that the distance matrices  $\rho_0^n$  converge in probability. We let  $n$  and then  $k$  tend to infinity.  $\square$

*Proof of Proposition 4.2 (end).* As the map  $g : \mathbb{M} \rightarrow \mathbb{M}$  in Remark 4.3 is continuous, we can make in Lemma 4.9 also the stronger assumption that  $[X^n, r^n, \mu^n]$  converges to  $[X, r, \mu]$  in the Gromov-Hausdorff-Prohorov topology. This yields Feller continuity of  $([X_t, \rho, \mu_t], t \in \mathbb{R}_+)$ . The strong Markov property can now be deduced using a.s. right continuity of the sample paths (Proposition 4.8).  $\square$

The tree-valued  $\Xi$ -Fleming-Viot process converges to equilibrium, as shown in [22, Proposition 9.1]. Now we show a similar result for the  $\mathbb{M}$ -valued evolving  $\Xi$ -coalescent.

Assume  $\Xi \in \mathcal{M}_{\text{CDI}}$  and (analogously to Section 2.3), let  $\bar{\eta}$  be a Poisson random measure on  $\mathbb{R} \times \mathcal{P}$  with intensity  $dt H_{\Xi}(d\pi)$ . From  $\bar{\eta}$ , we define a lookdown space  $(\bar{Z}, \bar{\rho})$  in two-sided time as the completion of the space of individuals  $\mathbb{R} \times \mathbb{N}$  with respect to the semi-metric  $\bar{\rho}$ , given by

$$\bar{\rho}((t, i), (u, j)) = t + u - 2 \sup\{s \in (-\infty, t \wedge u] : \bar{A}_s(t, i) = \bar{A}_s(u, j)\}, \quad (4.1)$$

where  $\bar{A}_s(t, i)$  denotes the level of the ancestor of the individual  $(t, i)$  when particles and reproduction events are defined precisely as in Section 2.

Analogously to Theorem 3.1, on an event of probability 1, the probability measures  $\bar{\mu}_t^n = n^{-1} \sum_{i=1}^n \delta_{(t, i)}$  on  $(\bar{Z}, \bar{\rho})$  weakly converge as  $n \rightarrow \infty$  for all  $t \in \mathbb{R}$ , we denote the limits by  $\bar{\mu}_t$ . For  $t \in \mathbb{R}$ , we denote by  $\bar{X}_t$  the closure of  $\{t\} \times \mathbb{N}$  in  $(\bar{Z}, \bar{\rho})$ . A stationary  $\mathbb{M}$ -valued evolving  $\Xi$ -coalescent is given by  $([\bar{X}_t, \bar{\rho}, \bar{\mu}_t], t \in \mathbb{R})$ . We call a random variable that is distributed as  $[\bar{X}_0, \bar{\rho}, \bar{\mu}_0]$  an  $\mathbb{M}$ -valued  $\Xi$ -coalescent measure tree, in analogy to the coalescent measure trees in [22, 19]. As  $[\bar{X}_0, \bar{\rho}, \bar{\mu}_0]$  is a.s. an ultrametric measure space, this random variable can be seen as a random tree.

In the next proposition, we show that the  $\mathbb{M}$ -valued  $\Xi$ -Fleming-Viot process  $([X_t, \rho, \mu_t], t \in \mathbb{R}_+)$  that is defined from  $\eta$  and  $[X, r, \mu]$  in this section converges to equilibrium.

**Proposition 4.10.** *The  $\mathbb{M}$ -valued random variable  $[X_t, \rho, \mu_t], t \in \mathbb{R}_+$  converges in distribution in  $(\mathbb{M}, d_{\text{GHP}})$  to an  $\mathbb{M}$ -valued  $\Xi$ -coalescent measure tree as  $t \rightarrow \infty$ .*

As in [22, 12], we use a coupling argument in the proof. In the present context, the topology is stronger than in [22], but as we restrict to  $\Xi \in \mathcal{M}_{\text{CDI}}$ , there exists a coupling of the tree-valued evolving  $\Xi$ -coalescents with arbitrary initial state and of the stationary process such that these processes coincide after an a.s. finite random time.

*Proof of Proposition 4.10.* Assume that the Poisson random measure  $\eta$  is the restriction of  $\bar{\eta}$  to  $(0, \infty) \times \mathcal{P}$ . Then  $[\bar{X}_t, \bar{\rho}, \bar{\mu}_t] = [X_t, \rho, \mu_t]$  on the event  $\{\text{diam } X_t < 2t\}$ . By the properties of the Poisson random measure  $\eta$  and as the event  $\{\text{diam } X_1 < 2\}$  is independent of  $\rho_0$ , the events  $\{\text{diam } X_t < 2t\}, t \in \mathbb{N}$  are independent and have the same positive probability. Hence, the random time  $\tau = \inf\{t \in \mathbb{R}_+ : \text{diam } X_t < 2t\}$  is geometrically bounded. The assertion follows as  $\text{diam } X_t < 2t$  for all  $t > \tau$ , and as  $[\bar{X}_t, \bar{\rho}, \bar{\mu}_t]$  is an  $\mathbb{M}$ -valued  $\Xi$ -coalescent measure tree.  $\square$

**Remark 4.11** (Convergence of  $\mathbb{M}$ -valued  $\Xi$ -Cannings processes). As an immediate consequence of the uniform convergence in Theorem 3.1(i), we obtain

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} d_{\text{GP}}(\llbracket Z, \rho, \mu_t^n \rrbracket, \llbracket Z, \rho, \mu_t \rrbracket) = 0 \quad \text{a.s.} \quad (4.2)$$

for each  $T \in \mathbb{R}_+$ . The process  $(\llbracket Z, \rho, \mu_t^n \rrbracket, t \in \mathbb{R}_+)$  may be called an  $\mathbb{M}$ -valued  $\Xi$ -Cannings process. In the case without simultaneous multiple reproduction events, it coincides with the tree-valued  $\Lambda$ -Cannings process discussed in [27, Section 4.2], and in the case without multiple reproduction events with the tree-valued Moran process from [19, Definition 2.19]. This can be seen by an application of [37, Theorem 2] similarly to the proof of Lemma 6.7 below, see also Section 2 of [8]. Then the convergence (4.2) implies the assertion of Theorem 2 in [19] for a special choice of the approximating sequence of the initial state.

From the uniform convergence in Theorem 3.10(i), similar statements can be deduced for processes from the next subsection. In [21], convergence of tree-valued Cannings chains is studied by different methods.

#### 4.2 The general case

We include the case with dust by using marked metric measure spaces. Let us first recall some facts from [9]. An  $(\mathbb{R}_+)$ -marked metric measure space  $(X, r, m)$  is a triple that consists of a complete and separable metric space  $(X, r)$  and a probability measure  $m$  on the Borel sigma algebra on the product space  $X \times \mathbb{R}_+$ . Two marked metric measure spaces  $(X, r, m), (X', r', m')$  are called isomorphic if there exists an isometry  $\varphi$  between the closed supports  $\text{supp } m(\cdot \times \mathbb{R}_+) \subset X$  and  $\text{supp } m'(\cdot \times \mathbb{R}_+) \subset X'$  such that the measurable map  $\hat{\varphi} : \text{supp } m \rightarrow \text{supp } m'$ , given by  $\hat{\varphi}(x, v) = (\varphi(x), v)$ , satisfies  $m' = \hat{\varphi}(m)$ . We endow the space  $\hat{\mathbb{M}}$  of isomorphy classes of marked metric measure spaces with the marked Gromov-Prohorov metric, which is defined by

$$d_{\text{mGP}}((X, r, m), (X', r', m')) = \inf_{Y, \hat{\varphi}, \hat{\varphi}'} d_{\text{P}}^Y(\hat{\varphi}(m), \hat{\varphi}'(m'))$$

where the infimum is over all isometric embeddings  $\varphi : X \rightarrow Y, \varphi' : X' \rightarrow Y$  into complete and separable metric spaces  $(Y, d^Y)$ . The maps  $\hat{\varphi} : X \times \mathbb{R}_+ \rightarrow Y \times \mathbb{R}_+$  and  $\hat{\varphi}' : X' \times \mathbb{R}_+ \rightarrow Y \times \mathbb{R}_+$  are defined by  $\hat{\varphi}(x, v) = (\varphi(x), v)$  and  $\hat{\varphi}'(x', v) = (\varphi'(x'), v)$ . The space  $Y \times \mathbb{R}_+$  is endowed with the product metric  $d^{Y \times \mathbb{R}_+}((y, v), (y', v')) = d^Y(y, y') \vee |v - v'|$ . Then  $(\hat{\mathbb{M}}, d_{\text{mGP}})$  is a complete and separable metric space. The marked distance matrix distribution  $\nu^{(X, r, m)}$  of (an isomorphy class of) a marked metric measure space  $(X, r, m)$  is defined as the distribution of the random variable  $(r(x(i), x(j)))_{i, j \in \mathbb{N}}, (v(i))_{i \in \mathbb{N}}$  where  $(x(i), v(i))_{i \in \mathbb{N}}$  is an  $m$ -iid sequence in  $X \times \mathbb{R}_+$ . The metric  $d_{\text{mGP}}$  induces the marked Gromov-weak topology in which a sequence of marked metric measure spaces converges if and only if their marked distance matrix distributions converge weakly.

From the construction on the extended lookdown space in Section 3.2, we now read off a stochastic processes with values in  $(\hat{\mathbb{M}}, d_{\text{mGP}})$ .

We work with the extended lookdown space  $(\hat{Z}, \rho)$  as defined in Section 3.2. This random metric space is constructed from the Poisson random measure  $\eta$  which is characterized by a finite measure  $\Xi$  on the simplex  $\Delta$ , and from an independent random variable  $(r_0, v_0)$  with the marked distance matrix distribution of a marked metric measure space  $(X, r, \mu)$ . The following proposition, which is proved below in this subsection, is the analog of Proposition 4.1.

**Proposition 4.12.** *Assume that  $\Xi$  is a finite measure on  $\Delta$ . Then a Feller-continuous strong Markov process with values in  $\hat{\mathbb{M}}$  is given a. s. by  $(\llbracket \hat{Z}, \rho, m_t \rrbracket, t \in \mathbb{R}_+)$ .*

We call  $(\llbracket \hat{Z}, \rho, m_t \rrbracket, t \in \mathbb{R}_+)$  an  $\hat{\mathbb{M}}$ -valued  $\Xi$ -Fleming-Viot process with initial state  $\llbracket X, r, m \rrbracket$ . By Remark 4.13 below, this process is the tree-valued  $\Xi$ -Fleming-Viot process from [22, Section 7.2] if the restriction of  $\rho$  to  $\{0\} \times \mathbb{N}$  is ultrametric.

Recall the space  $\mathcal{D}$  of semi-metrics on  $\mathbb{N}$ . We define the space  $\hat{\mathcal{D}} = \mathcal{D} \times \mathbb{R}_+^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}^2} \times \mathbb{R}^{\mathbb{N}}$ , where  $\mathbb{R}^{\mathbb{N}^2} \times \mathbb{R}^{\mathbb{N}}$  is endowed with the product topology. The elements of  $\hat{\mathcal{D}}$  are called marked distance matrices or decomposed semi-metrics on  $\mathbb{N}$ . We define a  $\hat{\mathcal{D}}$ -valued

stochastic process  $((r_t, v_t), t \in \mathbb{R}_+)$  from the Poisson random measure  $\eta$  and  $(r_0, v_0)$  as in Remark 2.3. By Proposition 6.3 of [22], this process is Markovian. In Lemma 6.4, we will show that for each  $t \in \mathbb{R}_+$ , the  $\hat{\mathfrak{D}}$ -valued random variable  $(r_t, v_t)$  is exchangeable. That is, its distribution is invariant under the action of the group of bijections  $\mathbb{N} \rightarrow \mathbb{N}$ , defined by  $p(r, v) = ((r(p(i)), p(j)))_{i, j \in \mathbb{N}}, (v(p(i)))_{i \in \mathbb{N}}$  for  $(r, v) \in \hat{\mathfrak{D}}$  and any bijection  $p : \mathbb{N} \rightarrow \mathbb{N}$ .

**Remark 4.13.** Recall the measurable map  $\hat{\psi} : \hat{\mathfrak{D}} \rightarrow \hat{\mathfrak{M}}$  from [22, Section 3.3]. Theorem 3.10 yields  $\llbracket \hat{Z}, \rho, m_t \rrbracket = \hat{\psi}(r_t, v_t)$  for all  $t \in \mathbb{R}_+$  a. s.

*Proof of Proposition 4.12 (beginning).* The Markov property follows as in [22, Theorem 4.1], we apply [37, Theorem 2] to the process  $((r_t, v_t), t \in \mathbb{R}_+)$ , the measurable map  $\hat{\psi} : \hat{\mathfrak{D}} \rightarrow \hat{\mathfrak{M}}$ , and the probability kernel from  $\hat{\mathfrak{M}}$  to  $\hat{\mathfrak{D}}$ , given by  $(\chi, B) \mapsto \nu^\chi(B)$ . Here we use the exchangeability of  $(r_t, v_t)$ .

Feller continuity can be shown precisely as in Corollary 8.2 of [22]. □

Let us deduce path regularity:

**Proposition 4.14.** *Assume that one of the following conditions hold: (i)  $\Xi \in \mathcal{M}_{\text{dust}}$ , or (ii)  $\Xi \in \mathcal{M}_{\text{nd}}$  and  $m(X \times \{0\}) = 1$ . Then the process  $(\llbracket \hat{Z}, \rho, m_t \rrbracket, t \in \mathbb{R}_+)$  has a. s. càdlàg paths in the marked Gromov-weak topology and  $\Theta_0$  is the set of jump times.*

*Proof.* We argue as in the proof of Proposition 4.7. In case  $\Xi \in \mathcal{M}_{\text{dust}}$ , we use Theorem 3.10. If  $\Xi \in \mathcal{M}_{\text{nd}}$  and  $m(\hat{Z} \times \{0\}) = 1$ , we use Remark 3.13 and Theorem 3.1. □

**Proposition 4.15.** *Assume  $\Xi \in \mathcal{M}_{\text{nd}}$  and  $m(X \times \{0\}) < 1$ . Then the process  $(\llbracket \hat{Z}, \rho, m_t \rrbracket, t \in (0, \infty))$  has a. s. càdlàg paths in the marked Gromov-weak topology and  $\Theta_0$  is the set of jump times. The process  $(\llbracket \hat{Z}, \rho, m_t \rrbracket, t \in \mathbb{R}_+)$  is a. s. not right-continuous at time 0.*

*Proof.* The first assertion follows from Corollary 3.14. The definitions of  $m_t$  and of the marked distance matrix distribution, and Remark 3.12 yield that  $\nu^{\llbracket \hat{Z}, \rho, m_t \rrbracket}(\mathfrak{D} \times \{0\}) = m_t(\hat{Z} \times \{0\}) = 1$  for all  $t \in (0, \infty)$  a. s. As the marked metric measure spaces  $(Z, \rho, m_0)$  and  $(X, r, m)$  are a. s. isomorphic (see e. g. [22, Proposition 10.5]), the assumptions also yield  $\nu^{\llbracket \hat{Z}, \rho, m_0 \rrbracket}(\mathfrak{D} \times \{0\}) = m_0(\hat{Z} \times \{0\}) = m(X \times \{0\}) < 1$  a. s. Hence, the probability measures  $\nu^{\llbracket \hat{Z}, \rho, m_t \rrbracket}$  do a. s. not weakly converge to  $\nu^{\llbracket \hat{Z}, \rho, m_0 \rrbracket}$ . By definition of the marked Gromov-weak topology, it follows that  $t \mapsto \llbracket \hat{Z}, \rho, m_t \rrbracket$  is a. s. not right-continuous at 0. □

*Proof of Proposition 4.12 (end).* The strong Markov property can be deduced from Feller continuity and the a. s. right continuity of the sample paths that is given by Propositions 4.14 and 4.15. □

From the construction in Section 3.2, we also read off the  $\Xi$ -Fleming-Viot process with values in a space of matrix distributions from [22, Section 7.3]. As in [22, Section 2], let

$$\alpha : \hat{\mathfrak{D}} \rightarrow \mathfrak{D}, \quad (r, v) \mapsto ((r(i, j) + v(i) + v(j)) \mathbb{1}\{i \neq j\})_{i, j \in \mathbb{N}} \tag{4.3}$$

be the continuous function that maps a decomposed semi-metric to the corresponding semi-metric. The distance matrix distribution of a marked metric measure space  $\chi$  is defined as  $\alpha(\nu^\chi)$ , where  $\nu^\chi$  denotes the marked distance matrix distribution of  $\chi$ . It depends only on the isomorphy class of the marked metric measure space. Let  $\rho_0 = \alpha(r_0, v_0)$ , then  $\rho_0$  is the restriction of  $\rho$  to  $\{0\} \times \mathbb{N}$ . As we will apply [22, Proposition 3.4], we assume in the remainder of this subsection that  $\rho_0$  is a semi-ultrametric (in which

case the restriction of  $\rho$  to  $\{0\} \times \mathbb{N}$  is ultrametric). Then the stochastic process of the distance matrix distributions of the  $\hat{U}$ -valued  $\Xi$ -Fleming-Viot process, given by  $(\xi_t, t \in \mathbb{R}_+) = (\alpha(\nu^{\llbracket \hat{Z}, \rho, m_t \rrbracket}), t \in \mathbb{R}_+)$  a. s., is the  $\mathcal{U}^{\text{erg}}$ -valued  $\Xi$ -Fleming-Viot process from [22, Section 7.3] where  $\mathcal{U}^{\text{erg}}$  refers to the space

$$\mathcal{U}^{\text{erg}} = \left\{ \alpha(\nu^\chi) : \chi \in \hat{\mathbb{M}}, \alpha(\nu^\chi)\text{-a. a. } \rho \in \mathfrak{D} \text{ are semi-ultrametrics} \right\}$$

of ultrametric distance matrix distributions of marked metric measure spaces. (This notation is used in [22] as  $\mathcal{U}^{\text{erg}}$  is shown there to be the set of distributions of semi-ultrametrics that are invariant and ergodic under the action of the group of finite permutations.) We endow  $\mathcal{U}^{\text{erg}}$  with the Prohorov metric.

In spite of Proposition 4.15, the  $\mathcal{U}^{\text{erg}}$ -valued  $\Xi$ -Fleming-Viot process is always a. s. right-continuous at time 0 by the following proposition which is applied in [21].

**Proposition 4.16.** *Assume that  $\Xi$  is a finite measure on  $\Delta$ . Then  $(\xi_t, t \in \mathbb{R}_+)$  has a. s. càdlàg paths and  $\Theta_0$  is a. s. the set of jump times.*

*Proof.* Propositions 4.14 and 4.15, the definition of the marked Gromov-weak topology, and continuity of  $\alpha$  yield that on an event of probability 1, the path  $(0, \infty) \rightarrow \mathcal{U}^{\text{erg}}, t \mapsto \xi_t$  is a. s. càdlàg in  $d_P$  with a. s. no jumps outside  $\Theta_0$ , and  $\llbracket \hat{Z}, \rho, m_t \rrbracket \neq \llbracket \hat{Z}, \rho, m_{t-} \rrbracket$  for all  $t \in \Theta_0$ . By Propositions 3.4 and 10.5 in [22], it follows that  $\xi_t \neq \xi_{t-}$  for all  $t \in \Theta_0$  a. s.

Right continuity at time 0 follows from Proposition 4.14 under the assumptions therein. For the general case, we use that  $(\xi_t, t \in \mathbb{R}_+)$  solves the martingale problem  $(C, \mathcal{C})$  defined in [22, Section 7.3]. We briefly recall the definition of the domain  $\mathcal{C}$ . For  $n \in \mathbb{N}$ , let  $\gamma_n$  be the restriction from  $\mathbb{R}^{\mathbb{N}^2}$  to  $\mathbb{R}^{n^2}$ ,  $\gamma_n(\rho') = (\rho'(i, j))_{i, j \in [n]}$ . Let  $\mathcal{C}_n$  be the set of functions  $\phi \circ \gamma_n : \mathbb{R}^{\mathbb{N}^2} \rightarrow \mathbb{R}$ , where  $\phi$  is a bounded differentiable function  $\mathbb{R}^{n^2} \rightarrow \mathbb{R}$  with bounded uniformly continuous derivative. Then we set  $\mathcal{C} = \bigcup_{n \in \mathbb{N}} \mathcal{C}_n$  and  $\mathcal{C} = \{\mathcal{U}^{\text{erg}} \rightarrow \mathbb{R}, \xi \mapsto \xi\phi : \phi \in \mathcal{C}\}$ , where we use the notation  $\xi\phi = \int \xi(d\rho')\phi(\rho')$ . By definition of  $C$  in [22, Section 7.3],  $C\Psi$  is bounded for each  $\Psi \in \mathcal{C}$ .

The set  $\mathcal{C}$  is convergence determining in  $\mathcal{U}^{\text{erg}}$ , see [22, Remark 4.6]. There also exists a countable subset of  $\mathcal{C}$  that generates the weak topology on  $\mathcal{U}^{\text{erg}}$ . Indeed, by smoothing indicator functions of rational intervals, one finds a countable subset  $\mathcal{C}' \subset \mathcal{C}$  that generates the product topology on  $\mathbb{R}^{\mathbb{N}^2}$ . Let  $\mathcal{C}''$  be the set of finite products of functions in  $\mathcal{C}'$ . Then the algebra  $\mathcal{C}''$  is convergence determining in  $\mathbb{R}^{\mathbb{N}^2}$  by e. g. [30, Theorem 2.7]. That is,  $\mathcal{C}' := \{\xi \mapsto \xi\phi : \phi \in \mathcal{C}''\} \subset \mathcal{C}$  generates the weak topology on  $\mathcal{U}^{\text{erg}}$ .

For  $\Psi \in \mathcal{C}$ , we consider the process  $(M_t, t \in \mathbb{R}_+)$  that is given by

$$M_t = \Psi(\xi_t) - \int_0^t C\Psi(\xi_s) ds$$

and which is a bounded martingale. From

$$\mathbb{E} \left[ \Psi(\xi_t) - \Psi(\xi_0) - \int_0^t C\Psi(\xi_s) ds \right] = 0,$$

continuity of  $\Psi$ , and as  $C\Psi$  is bounded, we obtain that  $\xi_t$  converges in distribution to  $\xi_0$  as  $t \rightarrow 0$ . As in the proof of Proposition 4.15, the marked metric measure spaces  $(X, r, m)$  and  $(\hat{Z}, \rho, m_0)$  are a. s. isomorphic. Hence,  $\xi_0 = \alpha(\nu^{(X, r, m)})$  a. s. As  $(X, r, m)$  is deterministic, it follows that  $\xi_t$  converges to  $\xi_0$  also in probability. By martingale convergence and as  $(M_t, t \in (0, \infty))$  has a. s. càdlàg paths, it follows that the limit  $\lim_{t \downarrow 0} M_t$  exists a. s., see e. g. [38, Theorem II.69.4]. By the convergence in probability we already know, the definition of  $M_t$ , and as  $C\Psi$  is bounded, the limit must be  $\Psi(\xi_0)$  a. s., hence  $\Psi(\xi_t)$  converges to  $\Psi(\xi_0)$  a. s. As  $\mathcal{C}$  contains a countable subset that generates the weak topology on  $\mathcal{U}^{\text{erg}}$ , it follows that  $\xi_t$  converges to  $\xi_0$  a. s.  $\square$

## 5 Outline and some definitions for the proof of the central results

The aim of the remaining sections is to prove in Section 9 Theorems 3.1, 3.5, and 3.10. In Section 6, we prove exchangeability properties for the (decomposed) genealogical distances between the individuals in the lookdown model at various stopping times. In Section 7, we state convergence results for processes of certain asymptotic frequencies that depend on these (decomposed) genealogical distances. We apply these convergence results in particular to families of partitions in Section 8: In the case without dust, we consider the flow of partitions. For the case with dust, we introduce a family of partitions that fits to the decomposition of genealogical distances. Using these families of partitions, we construct the probability measures on the (extended) lookdown space in Section 9. We prove the convergence results from Section 7 in Section 10 using the exchangeability properties from Section 6.

Now we collect some definitions that we will use in the remaining sections.

### 5.1 Some notation

For  $n \in \mathbb{N}$ , we continue using the notation  $[n] = \{1, \dots, n\}$ . We also write  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$  and  $[0] = \emptyset$ . Recall the space  $\mathfrak{D} \subset \mathbb{R}^{\mathbb{N}^2}$  of distance matrices, and that we do not distinguish distance matrices from semi-metrics on  $\mathbb{N}$ . Recall also the space  $\hat{\mathfrak{D}} = \mathfrak{D} \times \mathbb{R}_+^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}^2} \times \mathbb{R}^{\mathbb{N}}$  of marked distance matrices or decomposed semi-metrics on  $\mathbb{N}$ . Here  $\mathbb{R}^{\mathbb{N}^2}$  and  $\mathbb{R}^{\mathbb{N}^2} \times \mathbb{R}^{\mathbb{N}}$  are endowed with the product topology. We denote by  $\mathfrak{D}_n \subset \mathbb{R}^{n^2}$  the space of semi-metrics on  $[n]$  which we do not distinguish from distance matrices. We denote by  $\hat{\mathfrak{D}}_n = \mathfrak{D}_n \times \mathbb{R}_+^n \subset \mathbb{R}^{n^2} \times \mathbb{R}^n$  the space of decomposed semi-metrics on  $[n]$  or marked distance matrices. Recall also the space  $\mathcal{P}$  of partitions of  $\mathbb{N}$  which is endowed with the topology induced by the restriction maps. We denote by  $\mathcal{P}_n$  of set of partitions of  $[n]$ . We denote by  $\gamma_n$  the restriction maps  $\gamma_n : \mathfrak{D} \rightarrow \mathfrak{D}_n, \rho \mapsto (\rho(i, j))_{i, j \in [n]}$ ,  $\gamma_n : \hat{\mathfrak{D}} \rightarrow \hat{\mathfrak{D}}_n, (r, v) \mapsto ((r(i, j))_{i, j \in [n]}, (v(i))_{i \in [n]})$ , and  $\gamma_n : \mathcal{P} \rightarrow \mathcal{P}_n$ . For a partition  $\pi \in \mathcal{P}$  and  $i \in \mathbb{N}$ , we denote the block of  $\pi$  that contains  $i$  by  $B(\pi, i)$ .

We denote the set of the minimal elements of the blocks of  $\pi$  by  $M(\pi) = \{\min B : B \in \pi\}$ . For a subset  $B \subset \mathbb{N}$ , we denote the relative frequency by  $|B|_n = n^{-1} \#(B \cap [n])$  for  $n \in \mathbb{N}$ , and the asymptotic frequency by  $|B| = \lim_{n \rightarrow \infty} |B|_n$  (if it exists).

### 5.2 Two-step construction of the point measure of reproduction events

We will use the following definitions from Section 6.2 onwards. As in Section 2.3, let  $\Xi$  be a finite measure on the simplex  $\Delta$ . We decompose  $\Xi = \Xi_0 + \Xi\{0\}\delta_0$ . Let  $\eta$  be a Poisson random measure on  $(0, \infty) \times \mathcal{P}$  with intensity  $dt H_{\Xi}(d\pi)$  as in Section 2.3. We assume w. l. o. g. that  $\eta$  is constructed as the sum  $\eta = \eta_K + \eta_0$  of two independent Poisson random measures, defined as follows. We define  $\eta_K$  as a Poisson random measure on  $(0, \infty) \times \mathcal{P}$  with intensity

$$dt \Xi\{0\} \sum_{1 \leq i < j} \delta_{K_{i,j}}(d\pi),$$

where  $K_{i,j}$  denotes the partition in  $\mathcal{P}$  that contains the block  $\{i, j\}$  and apart from that only singletons. The point measure  $\eta_K$  encodes the Kingman part, that is, the binary reproduction events. If  $\Xi_0(\Delta) = 0$ , we set  $\eta_0 = 0$ . Let  $\zeta_0$  be a Poisson random measure on  $(0, \infty) \times \Delta$  with intensity  $dt |x|_2^{-2} \Xi_0(dx)$ . In case  $\Xi_0(\Delta) > 0$ , let  $((t^k, y^k), k \in \mathbb{N})$  be a collection of  $\zeta_0$ -measurable random variables with values in  $(0, \infty) \times \Delta$  such that

$$\zeta_0 = \sum_{k \in \mathbb{N}} \delta_{(t^k, y^k)} \quad \text{a. s.}$$

Let  $(\pi^k, k \in \mathbb{N})$  be a collection of  $\mathcal{P}$ -valued random variables that are conditionally independent given  $((t^k, y^k), i \in \mathbb{N})$  such that  $\pi^k$  has conditional distribution  $\kappa(y^k, \cdot)$

given  $((t^k, y^k), k \in \mathbb{N})$ . Thereby,  $\kappa$  is the probability kernel from  $\Delta$  to  $\mathcal{P}$  associated with Kingman's correspondence as in Section 2.3. We define the Poisson random measure  $\eta_0$  on  $(0, \infty) \times \mathcal{P}$  by

$$\eta_0 = \sum_{k \in \mathbb{N}} \delta_{(t^k, \pi^k)}. \tag{5.1}$$

For each point  $(t, y)$  of  $\zeta_0$  and the associated point  $(t, \pi)$  of  $\eta_0$ , the vector  $y$  gives the asymptotic frequencies (in decreasing order) of the blocks of the partition  $\pi$ , that is, the relative family sizes in the large reproduction event at time  $t$ .

Recall that  $\Xi \in \mathcal{M}_{\text{dust}}$  implies  $\Xi\{0\} = 0$ . Also note that the set  $\Theta_0$  of times of large reproduction events, defined in (2.7), satisfies

$$\Theta_0 = \{t \in (0, \infty) : \zeta_0(\{t\} \times \Delta) > 0\} \quad \text{a. s.}$$

### 5.3 The general setting

Let  $(r_0, v_0)$  be a  $\hat{\mathcal{D}}$ -valued random variable that is independent of  $\eta$ . Using the map  $\alpha : \hat{\mathcal{D}} \rightarrow \mathcal{D}$  from (4.3), we define a  $\mathcal{D}$ -valued random variable by  $\rho_0 = \alpha(r_0, v_0)$ . Then  $\rho_0$  can also be considered as an arbitrary  $\mathcal{D}$ -valued random variable that is independent of  $\eta$ . In this way, we unify the settings of Sections 3.1 and 3.2.

Let  $(\hat{Z}, \rho)$  be the extended lookdown space associated with  $\eta$  and  $(r_0, v_0)$  as in Section 2.1. Recall that  $(\hat{Z}, \rho)$  contains the lookdown space  $(Z, \rho)$  associated with  $\eta$  and  $\rho_0$  as a subspace. We endow  $\hat{Z} \times \mathbb{R}_+$  with the product metric  $d^{\hat{Z} \times \mathbb{R}_+}$  defined in the beginning of Section 3.2. We define the  $\mathcal{D}$ -valued process  $(\rho_t, t \in \mathbb{R}_+)$  of the genealogical distances between the individuals at fixed times as in equation (2.3). Then for each  $n \in \mathbb{N}$ , the process  $(\gamma_n(\rho_t), t \in \mathbb{R}_+)$  of the restrictions to the first  $n$  levels jumps only at the reproduction events that are encoded by a partition in  $\mathcal{P}^n$  as in the beginning of Section 2. On the event of probability 1 on which condition (2.1) holds,  $(\gamma_n(\rho_t), t \in \mathbb{R}_+)$  thus jumps only finitely often in bounded time intervals. Between these jumps, the genealogical distances  $\rho_t(i, j)$  with  $i \neq j, i, j \in [n]$  grow linearly with slope 2. Using also the definition of the metric  $\rho$  from the beginning of Section 2 (there in particular that the map  $t \mapsto A_s(t, i)$  is càdlàg), we deduce that the process  $(\rho_t, t \in \mathbb{R}_+)$  has a. s. càdlàg paths. We denote the left limits by  $\rho_{t-}$ . The process  $(\rho_t, t \in \mathbb{R}_+)$  is also Markovian by [22, Proposition 5.4].

We define the  $\hat{\mathcal{D}}$ -valued process  $((r_t, v_t), t \in \mathbb{R}_+)$  of the decomposed genealogical distances between the individuals at fixed times as in Section 2.1 (including Remark 2.3). In case  $\Xi \in \mathcal{M}_{\text{dust}}$ , the following condition is a. s. satisfied (as in [22], where this is condition (6.2) which is checked in Section 6.2 therein):

$$\eta\left((0, t] \times \hat{\mathcal{P}}^n\right) < \infty \quad \text{for all } t \in (0, \infty) \text{ and } n \in \mathbb{N}. \tag{5.2}$$

Here  $\hat{\mathcal{P}}^n = \{\pi \in \mathcal{P} : \{\{1\}, \dots, \{n\}\} \not\subset \pi\}$  is the set of those partitions of  $\mathbb{N}$  in which the first  $n$  integers are not all in singleton blocks. By definition of the population model in the beginning of Section 2, condition (5.2) implies that the particles on the first  $n \in \mathbb{N}$  levels reproduce at only finitely many times in bounded time intervals. Then for  $i \in [n]$ , the maps  $t \mapsto (z(t, i), v_t(i))$  are càdlàg with jumps only at such reproduction times. This follows from the definition of  $v_t(i)$  and  $z(t, i)$  in Section 2.1. Between such jumps, the parent  $z(t, i)$  remains constant and the quantity  $v_t(i)$  grows linearly with slope 1. As a consequence, the process  $((r_t, v_t), t \in \mathbb{R}_+)$  is a. s. càdlàg if  $\Xi \in \mathcal{M}_{\text{dust}}$ . Also recall from Remark 3.12 that  $\Xi \in \mathcal{M}_{\text{nd}}$  implies  $v_t = 0$  and  $\rho_t = r_t$  for all  $t \in (0, \infty)$  a. s. Hence,  $((r_t, v_t), t \in (0, \infty))$  is a. s. càdlàg for all finite measures  $\Xi$  on  $\Delta$ . We denote the left limits by  $(r_{t-}, v_{t-})$ . Note that the left limits  $\rho_{t-}$  and  $(r_{t-}, v_{t-})$  can be defined like  $\rho_t$  and  $(r_t, v_t)$ , respectively, by ignoring a possible reproduction event at time  $t$ . By [22, Proposition 6.3], the process  $((r_t, v_t), t \in \mathbb{R}_+)$  is also Markovian.

## 6 Preservation of exchangeability

In this section, we extend the exchangeability results on the genealogical distances in the lookdown model from Sections 5.3 and 11.1 of [22]: We consider invariance under permutations that leave the first  $b \in \mathbb{N}$  levels unchanged, and we show exchangeability properties of the decomposed genealogical distances at various stopping times. In Subsection 6.1, we show that exchangeability properties are preserved in single reproduction events. In Subsection 6.2, we concatenate these reproduction events to show exchangeability properties in the lookdown model. Exchangeability in the lookdown model is also studied in [12, 13, 4, 29, 28].

Let us first repeat from [22, Sections 5.1 and 6.1] the effect of a reproduction event on the (decomposed) genealogical distances. For  $n \in \mathbb{N}$ ,  $\pi \in \mathcal{P}_n$  and  $i \in [n]$ , let  $\pi(i)$  be the integer  $k$  such that  $i$  is in the  $k$ -th block of  $\pi$  when blocks are ordered increasingly according to their respective smallest element. With each element  $\pi$  of  $\mathcal{P}_n$ , we associate a transformation  $\mathcal{D}_n \rightarrow \mathcal{D}_n$ , which we also denote by  $\pi$ , by

$$\pi(\rho) = (\rho(\pi(i), \pi(j)))_{i,j \in \mathbb{N}}.$$

By comparison with the construction in the beginning of Section 2, we see that a reproduction event that is encoded by a point  $(t, \pi)$  of  $\eta$  results in a jump of the genealogical distances that is given by

$$\gamma_n(\rho_t) = \gamma_n(\pi)(\gamma_n(\rho_{t-})), \tag{6.1}$$

which is equation (5.3) in [22]. Recall the set  $\mathcal{P}^n$  from equation (2.2). Clearly,  $\gamma_n(\pi)(\rho) = \rho$  for each  $\pi \in \mathcal{P} \setminus \mathcal{P}^n$  and  $\rho \in \mathcal{D}_n$ .

To account for the decomposed genealogical distances, we use the set  $\mathcal{S}_n$  of semi-partitions of  $[n]$ . A semi-partition  $\sigma$  of  $[n]$  is a system of nonempty disjoint subsets of  $[n]$ , which we call blocks. The union  $\cup \sigma$  of the blocks needs not comprise all elements of  $[n]$ . For each semi-partition  $\sigma \in \mathcal{S}_n$ , there exists a unique partition  $\pi \in \mathcal{P}_n$  that has the same non-singleton blocks as  $\sigma$ , that is,  $\{B \in \pi : \#B \geq 2\} = \{B \in \sigma : \#B \geq 2\}$ ; we define  $\sigma(i) = \pi(i)$  for  $i \in [n]$ . With each element  $\sigma$  of  $\mathcal{S}_n$ , we associate a transformation  $\hat{\mathcal{D}}_n \rightarrow \hat{\mathcal{D}}_n$ , which we also denote by  $\sigma$ , by  $\sigma(r, v) = (r', v')$ , where

$$v'(i) = v(\sigma(i)) \mathbb{1}\{i \notin \cup \sigma\}$$

and

$$r'(i, j) = (v(\sigma(i)) \mathbb{1}\{i \in \cup \sigma\} + r(\sigma(i), \sigma(j)) + v(\sigma(j)) \mathbb{1}\{j \in \cup \sigma\}) \mathbb{1}\{i \neq j\}$$

for  $i, j \in [n]$ . Furthermore, we define the map

$$\varsigma_n : \mathcal{P} \rightarrow \mathcal{S}_n, \quad \pi \mapsto \{B \cap [n] : B \in \pi, \#B \geq 2\} \setminus \{\emptyset\} \tag{6.2}$$

which removes all singleton blocks from  $\pi$  and restricts the obtained semi-partition to  $[n]$ . In particular,  $\hat{\mathcal{P}}^n = \{\pi \in \mathcal{P} : \varsigma_n(\pi) \neq \emptyset\}$  for all  $n \in \mathbb{N}$ . By construction, each point  $(t, \pi)$  of  $\eta$  results in a jump of the decomposed genealogical distances at time  $t$  given by

$$\gamma_n(r_t, v_t) = \varsigma_n(\pi)(\gamma_n(r_{t-}, v_{t-})), \tag{6.3}$$

this is equation (6.1) in [22]. For each point  $\pi \in \mathcal{P} \setminus \hat{\mathcal{P}}^n$  and  $(r, v) \in \hat{\mathcal{D}}_n$ , we have  $\varsigma_n(\pi)(r, v) = (r, v)$ .

For  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$ , we denote by  $\mathcal{S}_{n,b}$  the set of semi-partitions  $\sigma \in \mathcal{S}_n$  that satisfy  $\sigma(i) = i$  for all  $i \in [b]$ . These are the  $\sigma \in \mathcal{S}_n$  such that no element of  $[b]$  is in a non-singleton block, that is, for each  $i \in [b]$ , either  $i \notin \cup \sigma$  or  $\{i\} \in \sigma$ . Therefore,

$$\mathcal{S}_{n,b} = \{\varsigma_n(\pi) : \pi \in \mathcal{P} \setminus \mathcal{P}^b\}. \tag{6.4}$$

Similarly, we define  $\mathcal{P}_{n,b}$  as the set of partitions of  $[n]$  such that none of the first  $b$  integers are in non-singleton blocks. This is the set of partitions  $\pi \in \mathcal{P}_n$  with  $\pi(i) = i$  for all  $i \in [b]$ . We have

$$\mathcal{P}_{n,b} = \{\gamma_n(\pi) : \pi \in \mathcal{P} \setminus \mathcal{P}^b\}. \tag{6.5}$$

Let  $S_n$  be the group of permutations of  $[n]$ . We define the action of  $S_n$  on  $\mathfrak{D}_n$  by  $p(\rho) = (\rho(p(i), p(j)))_{i,j \in [n]}$  for  $p \in S_n$  and  $\rho \in \mathfrak{D}_n$ . Analogously, we set  $p(r, v) = ((r(p(i), p(j)))_{i,j \in [n]}, (v(p(i)))_{i \in [n]})$  for  $(r, v) \in \hat{\mathfrak{D}}_n$ , and  $p(\sigma) = \{p(B) : B \in \sigma\}$  for  $\sigma \in \mathcal{S}_n \supset \mathcal{P}_n$ . For  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$ , we define the group

$$S_{n,b} = \{p \in S_n : p(i) = i \text{ for all } i \in [b]\} \tag{6.6}$$

of permutations of  $[n]$  that leave the first  $b$  integers unchanged. We say that a random (marked) distance matrix in  $\mathfrak{D}_n$  or  $\hat{\mathfrak{D}}_n$ , or a random (semi-)partition of  $[n]$  is  $(n, b)$ -exchangeable if its distribution is invariant under the action of  $S_{n,b}$ . For  $n \in \mathbb{N}$ , the usual exchangeability is recovered as  $(n, 0)$ -exchangeability. For  $b \in \mathbb{N}$ , no restriction is meant by  $(b, b)$ -exchangeability.

### 6.1 Single reproduction events

In Lemma 6.1 below, we show that  $(n, b)$ -exchangeability of distance matrices is preserved under the transformations associated with independent  $(n, b)$ -exchangeable partitions which will later encode reproduction events in the lookdown model. This extends Lemma 11.1 of [22] to  $(n, b)$ -exchangeability. In the subsequent Lemma 6.2, we consider marked distance matrices and semi-partitions.

**Lemma 6.1.** *Let  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$ . Let  $\tilde{\rho}$  be an  $(n, b)$ -exchangeable random variable with values in  $\mathfrak{D}_n$  and let  $\pi$  be an independent  $(n, b)$ -exchangeable random variable with values in  $\mathcal{P}_{n,b}$ . Then  $\pi(\tilde{\rho})$  is  $(n, b)$ -exchangeable.*

*Proof.* Let  $p \in S_{n,b}$ . As in the proof of Lemma 11.1 in [22], see equations (6.2) and (6.3) therein, there exists a map  $f : \mathcal{P}_n \rightarrow S_n$  that satisfies

$$\pi'(i) = f(\pi')(p(\pi')(p(i))) \tag{6.7}$$

for all  $\pi' \in \mathcal{P}_n$  and  $i \in [n]$ , and

$$\pi'(\rho') = p(p(\pi')(f(\pi')(\rho'))) \tag{6.8}$$

for all  $\pi' \in \mathcal{P}_n$  and  $\rho' \in \mathfrak{D}_n$ .

For each  $\pi' \in \mathcal{P}_{n,b}$ , the definition of  $p$  implies  $p(\pi') \in \mathcal{P}_{n,b}$  and  $\pi'(i) = i = p(\pi')(p(i))$  for all  $i \in [b]$ . Hence,  $f(\pi') \in S_{n,b}$  for each  $\pi' \in \mathcal{P}_{n,b}$ .

This allows to conclude analogously to the proof of Lemma 11.1 in [22]. □

**Lemma 6.2.** *Let  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$ . Let  $(\tilde{r}, \tilde{v})$  be a  $(n, b)$ -exchangeable random variable with values in  $\hat{\mathfrak{D}}_n$  and let  $\sigma$  be an independent  $(n, b)$ -exchangeable random variable with values in  $\mathcal{S}_{n,b}$ . Then  $\sigma(\tilde{r}, \tilde{v})$  is  $(n, b)$ -exchangeable.*

*Proof.* Recall from (4.3) the map  $\alpha : \hat{\mathfrak{D}}_n \rightarrow \mathfrak{D}_n$ . Let  $(r', v') \in \hat{\mathfrak{D}}_n$ ,  $\rho' = \alpha(r', v')$ ,  $\sigma' \in \mathcal{S}_n$ , and let  $\pi'$  be the partition in  $\mathcal{P}_n$  with the same non-singleton blocks as  $\sigma'$ . Then,

$$\alpha(\sigma'(r', v')) = \pi'(\rho') \tag{6.9}$$

by definition of the transformations on  $\mathfrak{D}_n$  and  $\hat{\mathfrak{D}}_n$  associated with each element of  $\mathcal{P}_n$  and  $\mathcal{S}_n$ , respectively. Writing  $\sigma'(r', v') = (r'', v'')$ , we obtain from equation (6.9) and the definition of the map  $\alpha$  that

$$r'' = (((\pi'(\rho'))_{i,j} - v''(i) - v''(j)) \mathbb{1}\{i \neq j\})_{i,j \in [n]}. \tag{6.10}$$

Let  $p \in S_{n,b}$  and the map  $f : \mathcal{P}_n \rightarrow S_n$  be defined as in the proof of Lemma 6.1. For  $i \in [n]$ , it holds  $i \in \cup \sigma'$  if and only if  $p(i) \in \cup p(\sigma')$ . From equation (6.7), we obtain

$$v''(i) = v'(\pi'(i)) \mathbb{1}\{i \notin \cup \sigma'\} = v'(f(\pi')(p(\pi')(p(i)))) \mathbb{1}\{p(i) \notin \cup p(\sigma')\}.$$

Using also equations (6.8) and (6.10), we deduce

$$\sigma'(r', v') = p(p(\sigma')(f(\pi')(r', v'))).$$

We conclude analogously to the proof of Lemma 11.1 in [22]. □

Recall the partitions of the form  $K_{i,j}$  which contain only  $\{i, j\}$  as a non-singleton block. These partitions encode binary reproduction events. In the next lemma, we consider the exchangeability after a transformation associated with a partition of the form  $K_{i,j}$  is applied to an  $(n, b)$ -exchangeable distance matrix.

**Lemma 6.3.** *Let  $n \in \mathbb{N}$ ,  $b \in [n - 1] \cup \{0\}$ ,  $i, j \in [b + 1]$  with  $i < j$ , and  $\pi = K_{i,j}$ . Let  $\tilde{\rho}$  be an  $(n, b)$ -exchangeable random variable with values in  $\mathcal{D}_n$ . Then  $\pi(\tilde{\rho})$  is  $(n, b + 1)$ -exchangeable.*

*Proof.* For all  $k \in [n]$ ,

$$\pi(k) = \begin{cases} k & \text{if } k < j \\ i & \text{if } k = j \\ k - 1 & \text{if } k > j. \end{cases}$$

Let  $p \in S_{n,b+1}$ . Then,

$$\pi(p^{-1}(k)) = \begin{cases} \pi(k) & \text{if } k \leq b + 1 \\ p^{-1}(k) - 1 & \text{if } k > b + 1. \end{cases}$$

Let  $p'$  be the permutation in  $S_{n,b}$  such that  $p'(k-1) = p^{-1}(k)-1$  for all  $k \in [n]$  with  $k > b+1$ . The permutation  $p'$  indeed exists and is unique as  $p' \in S_{n,b}$  implies  $p'(k-1) = k-1$  for all  $k \in [b+1]$  with  $k \geq 2$ , as  $p \in S_{n,b+1}$  implies  $b < p^{-1}(k) - 1 \leq n$  for all  $k \in [n]$  with  $k > b+1$ , and and only one possibility remains for  $p'(n)$  as  $p^{-1}$  is injective.

It follows

$$\pi(p^{-1}(k)) = p'(\pi(k))$$

for all  $k \in [n]$ . To see this, we use that  $\pi(k) \leq b$  for all  $k \in [b+1]$ , and that  $\pi(k) = k-1$  for all  $k \in [n]$  with  $k > b+1$ . From  $\pi(k) = p'(\pi(p(k)))$  for all  $k \in [n]$ , and by definition of the transformations on  $\mathcal{D}_n$  associated with the elements of  $\mathcal{P}_n$ , it follows  $\pi(\tilde{\rho}) = p(\pi(p'(\tilde{\rho})))$ . The assertion follows as  $p'(\tilde{\rho})$  and  $\tilde{\rho}$  are equal in distribution. □

### 6.2 In the lookdown model

We formulate most results in this subsection for the process of marked distance matrices  $((r_t, v_t), t \in \mathbb{R}_+)$ . To apply these results to the distance matrices  $(\rho_t, t \in \mathbb{R}_+)$ , note that the construction in Section 2.1 implies that  $\rho_t = \alpha(r_t, v_t)$  for all  $t \in \mathbb{R}_+$ , with  $\alpha$  defined in (4.3). Also recall from Remark 3.12 that  $\Xi \in \mathcal{M}_{\text{nd}}$  implies that  $v_t = 0$  and  $\rho_t = r_t$  for all  $t \in (0, \infty)$  a. s.

For  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$ , we say a random variable is conditionally  $(n, b)$ -exchangeable (given a sigma-algebra or a random variable) if its conditional distribution is a. s. invariant under the action of the group  $S_{n,b}$ .

The following lemma generalizes Proposition 5.8 in [22] and shows that conditioned on the event that until time  $t$  no reproduction events affect the genealogical distances between the first  $b$  individuals, the marked distance matrix  $\gamma_n(r_t, v_t)$  is  $(n, b)$ -exchangeable if this holds for  $\gamma_n(r_0, v_0)$ . This assertion also holds conditionally given the point measure  $\zeta_0$ .

**Lemma 6.4.** *Let  $n \in \mathbb{N}$ ,  $b \in [n] \cup \{0\}$ , and  $t \in \mathbb{R}_+$ . Assume that  $\gamma_n(r_0, v_0)$  is  $(n, b)$ -exchangeable. Then conditionally given  $\zeta_0$ , the marked distance matrix*

$$\mathbb{1}\{\eta((0, t] \times \mathcal{P}^b) = 0\} \gamma_n(r_t, v_t)$$

is  $(n, b)$ -exchangeable.

The proof is analogous to [22, Proposition 5.8]. It relies on representations (as in equation (6.11)) of the (decomposed) genealogical distances in terms of reproduction events and the growth of the (decomposed) genealogical distances. We use that these constituents preserve exchangeability and that reproduction events that affect the (decomposed) genealogical distances between the first  $n$  levels do a. s. not accumulate. The latter property is ensured in case  $\Xi \in \mathcal{M}_{\text{nd}}$  by condition (2.1), and in case  $\Xi \in \mathcal{M}_{\text{dust}}$  by condition (5.2). To account for the growth of the (decomposed) genealogical distances according to the description in Section 5.3, we define for  $t \in \mathbb{R}_+$  and  $n \in \mathbb{N}$  the maps

$$\lambda_t : \mathfrak{D}_n \rightarrow \mathfrak{D}_n, \quad \rho \mapsto \rho + \underline{2}_n t$$

and

$$\hat{\lambda}_t : \hat{\mathfrak{D}}_n \rightarrow \hat{\mathfrak{D}}_n, \quad (r, v) \mapsto (r, v + \underline{1}_n t),$$

where we write  $\underline{1}_n = (1)_{i \in [n]}$  and  $\underline{2}_n = 2(\mathbb{1}\{i \neq j\})_{i, j \in [n]}$ . The jumps of the (decomposed) genealogical distances are described in equations (6.1) and (6.3). Recall also the restriction  $\gamma_n : \mathcal{P} \rightarrow \mathcal{P}_n$  and the map  $\varsigma_n$  defined in (6.2).

*Proof of Lemma 6.4.* In this proof, we always condition on the event  $\{\eta((0, t] \times \mathcal{P}^b) = 0\}$ . This does not affect the distribution of the Poisson random measure  $\eta(\cdot \cap ((0, \infty) \times (\mathcal{P} \setminus \mathcal{P}^b)))$ .

On an event of probability 1, let  $(t_1, \pi_1), (t_2, \pi_2), \dots$  with  $0 < t_1 < t_2 < \dots$  be the points of  $\eta$  in  $(0, \infty) \times (\mathcal{P}^n \setminus \mathcal{P}^b)$ . Let  $L = \eta((0, t] \times (\mathcal{P}^n \setminus \mathcal{P}^b))$ . Conditionally given  $\zeta_0$  and  $(t_1, t_2, \dots)$ , the random partitions  $\pi_1, \pi_2, \dots$  are independent and  $\gamma_n(\pi_k)$  is  $(n, b)$ -exchangeable for each  $k \in \mathbb{N}$ . Moreover, equation (6.5) yields  $\gamma_n(\pi_k) \in \mathcal{P}_{n,b}$  for all  $k \in [L]$  a. s. By accounting for the growth of the genealogical distances and their jumps, and as we condition on  $\{\eta((0, t] \times \mathcal{P}^b) = 0\}$ , we obtain

$$\gamma_n(\rho_t) = \lambda_{t-t_L} \circ \gamma_n(\pi_L) \circ \lambda_{t_L-t_{L-1}} \circ \dots \circ \gamma_n(\pi_1) \circ \lambda_{t_1}(\gamma_n(\rho_0)) \quad \text{a. s.} \quad (6.11)$$

on  $\{L \geq 1\}$ , and  $\gamma_n(\rho_t) = \lambda_t(\gamma_n(\rho_0))$  a. s. on  $\{L = 0\}$ . Lemma 6.1 implies that the distance matrix  $\gamma_n(\rho_t)$  is  $(n, b)$ -exchangeable conditionally given  $\zeta_0$ . In case  $\Xi \in \mathcal{M}_{\text{nd}}$ , this also holds for the marked distance matrix  $(r_t, v_t)$  as  $v_t = 0$  and  $r_t = \rho_t$  a. s. by Remark 3.12.

An analogous argument applies in case  $\Xi \in \mathcal{M}_{\text{dust}}$ . As (5.2) holds a. s. in this case, we can now define  $(t_1, \pi_1), (t_2, \pi_2), \dots$  with  $0 < t_1 < t_2 < \dots$  to be the points of  $\eta$  in  $(0, \infty) \times (\hat{\mathcal{P}}^n \setminus \mathcal{P}^b)$  a. s. Now let  $L = \eta((0, t] \times (\hat{\mathcal{P}}^n \setminus \mathcal{P}^b))$ . Conditionally given  $\zeta_0$  and  $(t_1, t_2, \dots)$ , the random partitions  $\pi_1, \pi_2, \dots$  are independent and the random semi-partition  $\varsigma_n(\pi_k)$  is  $(n, b)$ -exchangeable for all  $k \in \mathbb{N}$ . Moreover, equation (6.4) yields  $\varsigma_n(\pi_k) \in \mathcal{S}_{n,b}$  for all  $k \in [L]$  a. s. By accounting for the growth of the decomposed genealogical distances and their jumps, we obtain

$$\gamma_n(r_t, v_t) = \hat{\lambda}_{t-t_L} \circ \varsigma_n(\pi_L) \circ \hat{\lambda}_{t_L-t_{L-1}} \circ \dots \circ \varsigma_n(\pi_1) \circ \hat{\lambda}_{t_1}(\gamma_n(r_0, v_0)) \quad \text{a. s.}$$

on  $\{L \geq 1\}$ , and  $\gamma_n(r_t, v_t) = \hat{\lambda}_t(\gamma_n(r_0, v_0))$  a. s. on  $\{L = 0\}$ . The assertion follows from Lemma 6.2.  $\square$

In the next corollary, we set  $(r_{0-}, v_{0-}) = (r_0, v_0)$ .

**Corollary 6.5.** Assume that  $(r_0, v_0)$  is exchangeable. Let  $\tau$  be a  $\zeta_0$ -measurable and a.s. finite random time. Then  $(r_\tau, v_\tau)$  and  $(r_{\tau-}, v_{\tau-})$  are exchangeable.

*Proof.* For  $k \in \mathbb{N}$ , we define a  $\zeta_0$ -measurable random time  $\tau^k$  that assumes countably many values by  $\tau^k = (j + 1)/k$  on the event  $\{\tau \in [j/k, (j + 1)/k)\}$  for  $j \in \mathbb{N}_0$ . For  $n \in \mathbb{N}$ ,  $p \in S_n$ , and bounded continuous  $\phi$ , Lemma 6.4 with  $b = 0$  yields

$$\begin{aligned} \mathbb{E}[\phi(\gamma_n(r_{\tau^k}, v_{\tau^k}))] &= \sum_{j \in \mathbb{N}_0} \mathbb{E}[\phi(\gamma_n(r_{j/k}, v_{j/k})); \tau^k = j/k] \\ &= \sum_{j \in \mathbb{N}_0} \mathbb{E}[\phi(p(\gamma_n(r_{j/k}, v_{j/k}))); \tau^k = j/k] = \mathbb{E}[\phi(p(\gamma_n(r_{\tau^k}, v_{\tau^k})))] \end{aligned}$$

We let  $k$  tend to infinity. The assertion follows as  $t \mapsto \gamma_n(r_t, v_t)$  is càdlàg a.s. To prove the assertion for  $(r_{\tau-}, v_{\tau-})$ , we replace  $\tau^k$  with  $\tilde{\tau}^k = \lfloor \tau^k \rfloor / k$ .  $\square$

In the next lemma, we see that at the time of the first reproduction event that affects the genealogical distances between the first  $b$  individuals, conditioned on this reproduction event being binary, the matrix of the genealogical distances between the first  $n$  individuals is  $(n, b + 1)$ -exchangeable if it is  $(n, b)$ -exchangeable at time zero.

**Lemma 6.6.** Let  $n \in \mathbb{N}$  and  $b \in [n - 1]$  with  $n, b \geq 2$ . Assume  $\Xi\{0\} > 0$  and that  $\gamma_n(\rho_0)$  is  $(n, b)$ -exchangeable. Let

$$\tau = \inf\{t > 0 : \eta((0, t] \times \mathcal{P}^b) > 0\}.$$

Then,  $\tau < \infty$  a.s. and the distance matrix

$$\mathbb{1}\{\zeta_0(\{\tau\} \times \Delta) = 0\} \gamma_n(\rho_\tau)$$

is  $(n, b + 1)$ -exchangeable.

*Proof.* The assumptions  $b \geq 2$  and  $\Xi\{0\} > 0$  imply  $\tau < \infty$  a.s. and  $\mathbb{P}(\zeta_0(\{\tau\} \times \Delta) = 0) > 0$ . In this proof, we always condition on the event  $\{\zeta_0(\{\tau\} \times \Delta) = 0\}$ .

On an event of probability 1, let  $(t_1, \pi_1), (t_2, \pi_2) \dots$  with  $0 < t_1 < t_2 < \dots$  be the points of  $\eta$  in  $(0, \infty) \times \mathcal{P}^n$ , and let  $L = \eta((0, \tau] \times \mathcal{P}^n)$ . Then,  $\tau = t_L$  a.s. The partitions  $\gamma_n(\pi_1), \gamma_n(\pi_2), \dots$  are conditionally independent given  $\zeta_0, L$ , and  $(t_1, \dots, t_L)$ . Conditionally given  $\zeta_0, L$ , and  $(t_1, \dots, t_L)$ , the partitions  $\gamma_n(\pi_k)$  for  $k \in [L - 1]$  are also  $(n, b)$ -exchangeable and by equation (6.5) in  $\mathcal{P}_{n,b}$ . On the event  $\{\zeta_0(\{\tau\} \times \Delta) = 0\}$ , conditionally given  $\zeta_0, L$  and  $(t_1, \dots, t_L)$ , the partition  $\gamma_n(\pi_L)$  a.s. contains one block with two elements in  $[b]$  and apart from that only singleton blocks. This follows from the construction of  $\eta$  in Section 5.2. By accounting for the growth and the jumps of the genealogical distances, we obtain

$$\gamma_n(\rho_\tau) = \gamma_n(\pi_L) \circ \lambda_{t_L - t_{L-1}} \circ \dots \circ \gamma_n(\pi_1) \circ \lambda_{t_1}(\gamma_n(\rho)) \quad \text{a.s.}$$

The assertion follows from Lemmas 6.1 and 6.3.  $\square$

We also consider exchangeability properties at certain stopping times. For  $n \in \mathbb{N}$ , we denote by  $S_\infty^n$  the group of bijections  $p : \mathbb{N} \rightarrow \mathbb{N}$  with  $p(i) = i$  for all  $i > n$ . We denote the space of probability measures on  $\hat{\mathcal{D}}$  by  $\mathcal{M}_1(\hat{\mathcal{D}})$ , and we define the measurable function

$$\beta_n : \hat{\mathcal{D}} \rightarrow \mathcal{M}_1(\hat{\mathcal{D}}), \quad (r, v) \mapsto \frac{1}{n!} \sum_{p \in S_\infty^n} \delta_{p(r,v)}.$$

Intuitively, the image of  $(r, v)$  under  $\beta_n$  corresponds to the orbit of  $(r, v)$  under the action of the group  $S_\infty^n$ .

For  $t \in \mathbb{R}_+$ , we denote by  $\mathcal{F}_t^n$  the sigma-algebra generated by  $(\beta_n(r_s, v_s), s \in [0, t])$ . A filtration is defined by  $\mathcal{F}^n = (\mathcal{F}_t^n, t \in \mathbb{R}_+)$ . We will use the following lemma in Section 8 to study asymptotic frequencies that are invariant under permutation of the first  $n$  levels.

**Lemma 6.7.** *Let  $n \in \mathbb{N}$  and assume that  $(r_0, v_0)$  is exchangeable. Let  $\tau$  be a finite  $\mathcal{F}^n$ -stopping time. Then  $\gamma_n(r_\tau, v_\tau)$  is exchangeable.*

*Proof.* We show that for each  $t \in \mathbb{R}_+$ , the marked distance matrix  $\gamma_n(r_t, v_t)$  is exchangeable conditionally given  $\mathcal{F}_t^n$ . The assertion then follows for stopping times that assume countably many values, and by an approximation argument as in the proof of Corollary 6.5 also for all finite stopping times.

Let  $K$  be the probability kernel from  $\mathcal{M}_1(\hat{\mathcal{D}})$  to  $\hat{\mathcal{D}}$  given by  $K(\xi, \cdot) = \xi$  for  $\xi \in \mathcal{M}_1(\hat{\mathcal{D}})$ . By Lemma 6.4, the marked distance matrix  $(r_t, v_t)$  is exchangeable, hence

$$\begin{aligned} & \mathbb{P}((r_t, v_t) \in B, \beta_n(r_t, v_t) \in B') \\ &= \mathbb{P}(p(r_t, v_t) \in B, \beta_n(p(r_t, v_t)) \in B') = \mathbb{P}(p(r_t, v_t) \in B, \beta_n(r_t, v_t) \in B') \end{aligned}$$

for all measurable  $B, B'$  and all  $p \in S_\infty^n$ . This implies that  $K$  is a regular conditional distribution of  $(r_0, v_0)$  given  $\beta_n(r_0, v_0)$ , and of  $(r_t, v_t)$  given  $\beta_n(r_t, v_t)$ . Now we apply Theorem 2 of [37] to the Markov process  $((r_t, v_t), t \in \mathbb{R}_+)$ , the measurable function  $\beta_n$ , and the probability kernel  $K$  to obtain that for each  $t \in \mathbb{R}_+$ , the marked distance matrix  $(r_t, v_t)$  has the same conditional distribution given  $\beta_n(r_t, v_t)$  as given  $\mathcal{F}_t^n$ . This follows from equation (1) in [37] and implies the assertion.  $\square$

Furthermore, recall the groups of permutations  $S_{n,b}$  for  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$  from (6.6). We write  $\mathcal{M}$  for the space of probability measures on  $\hat{\mathcal{D}}_n$ , and we define the function

$$\beta_{n,b} : \hat{\mathcal{D}}_n \rightarrow \mathcal{M}, \quad (r, v) \mapsto \frac{1}{(n-b)!} \sum_{p \in S_{n,b}} \delta_{p(r,v)}.$$

For  $t \in \mathbb{R}_+$ , we denote by  $\mathcal{F}_t^{n,b}$  the sigma-algebra generated by  $(\beta_{n,b}(r_s, v_s), s \in [0, t])$ . A filtration is defined by  $\mathcal{F}^{n,b} = (\mathcal{F}_t^{n,b}, t \in \mathbb{R}_+)$ . We will use the following lemma in Section 10 where we consider some relative frequencies in the lookdown model that do not change under permutations that leave the first  $b$  levels fixed.

**Lemma 6.8.** *Let  $n \in \mathbb{N}$  and  $b \in [n] \cup \{0\}$ . Assume that  $\gamma_n(r_0, v_0)$  is  $(n, b)$ -exchangeable. Let  $\tau$  be a finite  $\mathcal{F}^{n,b}$ -stopping time. Then the marked distance matrix*

$$\mathbb{1}\{\eta((0, \tau] \times \mathcal{P}^b) = 0\} \gamma_n(r_\tau, v_\tau)$$

*is  $(n, b)$ -exchangeable.*

*Proof.* We show that for each  $t \in \mathbb{R}_+$ , the marked distance matrix

$$\mathbb{1}\{\eta((0, t] \times \mathcal{P}^b) = 0\} \gamma_n(r_t, v_t)$$

*is  $(n, b)$ -exchangeable conditionally given  $\mathcal{F}_t^{n,b}$ .*

We enlarge the spaces  $\hat{\mathcal{D}}_n$  and  $\mathcal{M}$  by a coffin state  $\partial$ . Let  $K$  be the probability kernel from  $\mathcal{M}$  to  $\hat{\mathcal{D}}_n$  such that  $K(\partial, \{\partial\}) = 1$ , and such that  $K(\xi, \cdot) = \xi$  for all  $\xi \in \mathcal{M} \setminus \{\partial\}$ .

Let  $\tau' = \inf\{t > 0 : \eta((0, t] \times \mathcal{P}^b) > 0\}$  and set  $R_t = \gamma_n(r_t, v_t)$  for  $t < \tau'$ , and  $R_t = \partial$  for  $t \geq \tau'$ . Analogously to the proof of Lemma 6.7,  $K$  is a regular conditional distribution of  $\gamma_n(r_0, v_0)$  given  $(\beta_{n,b}(\gamma_n(r_0, v_0)))$  as a consequence of the assumed exchangeability. For all  $t \in \mathbb{R}_+$ , Lemma 6.4 implies that  $K$  is a regular conditional distribution of  $R_t$  given  $\beta_{n,b}(R_t)$ , where we set  $\beta_{n,b}(\partial) = \partial$ . We apply Theorem 2 of Rogers and Pitman [37] to the Markov process  $(R_t, t \in \mathbb{R}_+)$ , the measurable function  $\beta_{n,b}$ , and the probability kernel  $K$  to obtain that for each  $t \in \mathbb{R}_+$ , the random variable  $R_t$  has the same conditional distribution given  $\mathcal{F}_t^{n,b}$  as given  $\beta_{n,b}(R_t)$ . This implies the assertion as in the proof of Lemma 6.7.  $\square$

### 7 Uniform convergence in the lockdown model

Donnelly and Kurtz [13] prove that the measure-valued processes whose states are the uniform measures of the types on the first  $n$  levels in the lockdown model converge a. s. as  $n$  tends to infinity. Lemmas 7.1 and 7.2 below give the existence of and uniform convergence to asymptotic frequencies of subsets of individuals that are characterized by the (marked) genealogical distances.

For  $\ell \in \mathbb{N}$ , let

$$b_\ell : \mathcal{P} \rightarrow \mathbb{N}_0, \quad b_\ell(\pi) = \ell - \#(\gamma_\ell(\pi))$$

and

$$\hat{b}_\ell : \mathcal{P} \rightarrow \mathbb{N}_0, \quad \hat{b}_\ell(\pi) = \#\{i \in [\ell] : \{i\} \notin \pi\}.$$

Moreover, let

$$N^\ell(I) = \int_{I \times \mathcal{P}} b_\ell(\pi) \eta(ds \, d\pi)$$

and

$$\hat{N}^\ell(I) = \int_{I \times \mathcal{P}} \hat{b}_\ell(\pi) \eta(ds \, d\pi)$$

for each interval  $I \subset (0, \infty)$ . The random variable  $N^\ell(I)$  is the number of newborn particles on the first  $\ell$  levels in the time interval  $I$ . The random variable  $\hat{N}^\ell(I)$  counts in the reproducing particles in each reproduction event, in particular, it also takes account of reproduction events in which only the reproducing particles occupy levels in  $[\ell]$ .

**Lemma 7.1.** *Assume that  $\rho_0$  is exchangeable. Let  $b \in \mathbb{N}$  and let  $f$  be a measurable function from  $\mathbb{R}_+ \times \mathbb{R}_+^b$  to  $\{0, 1\}$ . For  $t \in \mathbb{R}_+$  and  $i, n \in \mathbb{N}$ , define*

$$Y_i^b(t) = f(t, (\rho_t(j, b + 1 + i))_{j \in [b]})$$

and

$$X_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i^b(t).$$

Impose the assumption on  $f$  that

$$|X_n(t) - X_n(s)| \mathbb{1}\{\eta((s, t] \times \mathcal{P}^b) = 0\} \leq \frac{1}{n} N^{b+1+n}(s, t) \tag{7.1}$$

for all  $n \in \mathbb{N}$  and  $0 \leq s < t$ . Then there exists a process  $(X(t), t \in \mathbb{R}_+)$  with  $\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} |X_n(t) - X(t)| = 0$  a. s. for all  $T \in \mathbb{R}_+$ .

**Lemma 7.2.** *Assume  $\Xi \in \mathcal{M}_{\text{dust}}$  and that  $(r_0, v_0)$  is exchangeable. Let  $b \in \mathbb{N}$  and let  $f$  be a measurable function from  $\mathbb{R}_+ \times \mathbb{R}_+^{2b} \times \mathbb{R}_+$  to  $\{0, 1\}$ . For  $t \in \mathbb{R}_+$  and  $i, n \in \mathbb{N}$ , define*

$$Y_i^b(t) = f(t, (r_t(j, b + i), v_t(j))_{j \in [b]}, v_t(b + i))$$

and

$$X_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i^b(t).$$

Impose the assumption on  $f$  that

$$|X_n(t) - X_n(s)| \mathbb{1}\{\eta((s, t] \times \hat{\mathcal{P}}^b) = 0\} \leq \frac{1}{n} \hat{N}^{b+n}(s, t) \tag{7.2}$$

for all  $n \in \mathbb{N}$  and  $0 \leq s < t$ . Then there exists a process  $(X(t), t \in \mathbb{R}_+)$  with  $\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} |X_n(t) - X(t)| = 0$  a. s. for all  $T \in \mathbb{R}_+$ .

We defer the proofs of these lemmas to Section 10.

### 8 Two families of partitions

From the process  $(\rho_t, t \in \mathbb{R}_+)$ , we now read off the flow of partitions  $(\Pi_{s,t}, 0 \leq s \leq t)$ . This process corresponds to the dual flow of partitions in Foucart [17] and to the flow of partitions in Labbé [29]. We define the random partition  $\Pi_{s,t}$  of  $\mathbb{N}$  by

$$i \text{ and } j \text{ are in the same block of } \Pi_{s,t} \iff \rho_t(i, j) < 2(t - s)$$

for all  $i, j \in \mathbb{N}$  with  $i \neq j$ . That is,  $i$  and  $j$  are in the same block of  $\Pi_{s,t}$  if and only if  $A_s(t, j) = A_s(t, i)$  which means that the individuals  $(t, i)$  and  $(t, j)$  have a common ancestor at time  $s$ . For each  $s \in \mathbb{R}_+$  and  $n \in \mathbb{N}$ , the process  $t \mapsto \gamma_n(\Pi_{s,s+t})$  jumps only at the times of reproduction events that are encoded by a partition in  $\mathcal{P}^n$ . These times do not accumulate on the event of probability 1 on which condition (2.1) is satisfied. For all  $0 \leq s \leq t$ , the random partition  $\Pi_{s,t}$  is exchangeable. This follows from Lemma 6.4 (where we may assume w. l. o. g., as  $\Pi_{s,t}$  is  $\eta$ -measurable, that  $\rho_0$  is exchangeable). We will apply the flow of partitions in the dust-free case.

For application in the case with dust, we define for each  $a \in \mathbb{N}$  and  $\varepsilon > 0$  a family  $(\Pi_t^{a,\varepsilon}, t \in \mathbb{R}_+)$  of partitions of  $\mathbb{N}$ . Similarly to the partition induced by  $\sim^\varepsilon$  in [22, Section 10.4], our intention is that individuals at time  $t$  whose levels are in the same block of  $\Pi_t^{a,\varepsilon}$  should have parents that are close to each other in the extended lookdown space (the definition of a parent is given Section 2.1). We will define  $\Pi_t^{a,\varepsilon}$  accordingly except for at most one block.

First, we define for each  $t \in \mathbb{R}_+$  and  $I \subset [a]$  the subset of  $\mathbb{N}$

$$C_t^{a,\varepsilon,I} = \{i \in \mathbb{N} : v_t(i) \geq t\} \cap \bigcap_{k \in I: v_t(k) \geq t} \{i \in \mathbb{N} : r_t(i, k) \vee |v_t(i) - v_t(k)| < \varepsilon\} \\ \cap \bigcap_{\ell \in [a] \setminus I: v_t(\ell) \geq t} \{r_t(i, \ell) \vee |v_t(i) - v_t(\ell)| \geq \varepsilon\}.$$

**Remark 8.1.** Clearly,  $(C_t^{a,\varepsilon,I}, I \subset [a])$  is a family of subsets whose union is  $\{i \in \mathbb{N} : v_t(i) \geq t\}$ . Any two such subsets are either disjoint or equal. For  $I \subset [a]$  and  $i, j \in C_t^{a,\varepsilon,I}$ , the construction in Section 2.1 implies that the parents of the individuals  $(t, i)$  and  $(t, j)$  are also the parents of the individuals  $(0, A_0(t, i))$  and  $(0, A_0(t, j))$ , respectively. If moreover  $I \neq \emptyset$ , then the genealogical distance  $r_t(i, j)$  between these parents is less than  $2\varepsilon$ , and  $|v_t(i) - v_t(j)| < 2\varepsilon$ .

Now we let  $i, j \in \mathbb{N}$  be in the same block of  $\Pi_t^{a,\varepsilon}$  if and only if one of the following two conditions is satisfied:

- (i)  $v_t(i) = v_t(j) < t$  and  $r_t(i, j) = 0$
- (ii) There exists  $I \subset [a]$  such that  $i, j \in C_t^{a,\varepsilon,I}$ .

Condition (i) means that the individuals  $(t, i)$  and  $(t, j)$  have the same parent in the (extended) lookdown space, and that this parent lives after time zero. That is,  $(t, i)$  and  $(t, j)$  have a common ancestor at time zero, and the individuals on each of the ancestral lineages of the individuals  $(t, i)$  and  $(t, j)$  are in singleton blocks in each reproduction event until these ancestral lineages merge. The individual in which these ancestral lineages merge is the parent of both the individuals  $(t, i)$  and  $(t, j)$ , when we identify individuals with genealogical distance zero. In this sense, the individuals  $(t, i)$  and  $(t, j)$  may be called siblings.

Using the definitions of  $v_t(i)$  and  $r_t(i, j)$ , it can be seen that for each  $n \in \mathbb{N}$ , on the event of probability 1 on which condition (5.2) is satisfied, the process  $t \mapsto \gamma_n(\Pi_t^{a,\varepsilon})$  jumps only at the times of reproduction events that are encoded by a partition in  $\hat{\mathcal{P}}^n$ , and that these times do not accumulate.

In the next two lemmas, we show that the asymptotic frequencies in  $\Pi_{s,t}$  and  $\Pi_t^{a,\varepsilon}$ , respectively, exist simultaneously for uncountably many  $t$  on an event of probability 1, and that the relative frequencies converge uniformly for  $t$  in compact intervals. On an event of probability 1, the left limits

$$\Pi_{s,t-} := \lim_{s' \uparrow t} \Pi_{s,s'} = \{ \{j \in \mathbb{N} : \rho_{t-}(i, j) < 2(t-s)\} : i \in \mathbb{N} \}$$

and

$$\Pi_t^{a,\varepsilon} := \lim_{s \uparrow t} \Pi_s^{a,\varepsilon} = \{ \{j \in \mathbb{N} : v_{t-}(i) = v_{t-}(j) < t, r_{t-}(i, j) = 0\} : i \in \mathbb{N} \}$$

exist for all  $t \in (0, \infty)$  and  $s \in [0, t)$ . The partitions  $\Pi_{s,t-}$  and  $\Pi_t^{a,\varepsilon}$  are left limits with respect to the topology on  $\mathcal{P}$  that is generated by the restriction maps  $\gamma_n, n \in \mathbb{N}$ . They can also be defined like  $\Pi_{s,t}$  and  $\Pi_t^{a,\varepsilon}$ , respectively, except that a possible reproduction event at time  $t$  is ignored. In the next two lemmas, we also show regularity properties in  $t$ , and that taking (left) limits in  $t$  commutes with taking asymptotic frequencies.

**Lemma 8.2.** *Let  $s, T \in \mathbb{R}_+$  and  $b \in \mathbb{N}$ . Then,*

$$\lim_{n \rightarrow \infty} \sup_{t \in [s, s+T]} | |B(\Pi_{s,t}, b)|_n - |B(\Pi_{s,t-}, b)| | = 0 \quad \text{a. s.} \tag{8.1}$$

*The paths  $[s, \infty) \rightarrow [0, 1], t \mapsto |B(\Pi_{s,t}, b)|$  are càdlàg a. s. Furthermore,  $\lim_{\varepsilon \downarrow 0} |B(\Pi_{s,t-\varepsilon}, b)| = |B(\Pi_{s,t-}, b)|$  for all  $t \in (s, \infty)$  a. s.*

**Lemma 8.3.** *Let  $T \in \mathbb{R}_+, a, k \in \mathbb{N}$ , and  $\varepsilon > 0$ . Assume  $\Xi \in \mathcal{M}_{\text{dust}}$  and that  $(r_0, v_0)$  is exchangeable. Then,*

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} | |B(\Pi_t^{a,\varepsilon}, k)|_n - |B(\Pi_t^{a,\varepsilon}, k)| | = 0 \quad \text{a. s.} \tag{8.2}$$

*The paths  $t \mapsto |B(\Pi_t^{a,\varepsilon}, k)|$  are càdlàg a. s. Furthermore,  $\lim_{s \uparrow t} |B(\Pi_s^{a,\varepsilon}, k)| = |B(\Pi_t^{a,\varepsilon}, k)|$  for all  $t \in (0, \infty)$  a. s.*

A result similar to Lemma 8.2 is Proposition 2.13 in Labbé [29] which is applied there to study relations between the lookdown model and flows of bridges.

*Proof of Lemma 8.2.* By  $\eta$ -measurability of the random variables in the assertion, we can assume w. l. o. g. that  $\rho_0$  is exchangeable. By time homogeneity, it suffices to consider the case  $s = 0$ . We choose  $f$  in Lemma 7.1 such that

$$f(t, (\rho_t(j, b+1+i))_{j \in [b]}) = \mathbb{1} \{ \rho_t(b, b+1+i) < 2(t-s) \}$$

for all  $t \in \mathbb{R}_+$  and  $i \in \mathbb{N}$ . On the right-hand side, we have the indicator variable of the event that  $b$  and  $b+1+i$  are in the same block of  $\Pi_{s,t}$ , which can be written as  $b+1+i \in B(\Pi_{s,t}, b)$ . Hence,

$$X_n(t) \leq \frac{b+1+n}{n} |B(\Pi_{s,t}, b)|_{b+1+n} \leq X_n(t) + \frac{b+1}{n} \tag{8.3}$$

for all  $t \in \mathbb{R}_+$  and  $n \in \mathbb{N}$ . Here  $X_n(t)$  is defined as in Lemma 7.1 so that  $nX_n(t)$  counts the individuals at time  $t$  on the levels  $b+1+1, \dots, b+1+n$  that have a common ancestor with the individual  $(t, b)$  after time  $s$ .

By the construction in the beginning of Section 2, the number  $nX_n(t)$  can only jump at times  $t$  with  $\eta(\{t\} \times \mathcal{P}^{b+1+n}) > 0$ . If there are no newborn particles on the first  $b$  levels at such a jump time (i. e. if the reproduction event at time  $t$  is encoded by a partition in  $\mathcal{P}^{b+1+n} \setminus \mathcal{P}^b$ ), then  $|nX_n(t) - nX_n(t-)|$  is bounded from above by the number of newborn particles on the first  $b+1+n$  levels. Indeed, we then have  $nX_n(t) - nX_n(t-) = a - a' + a''$ ,

where  $a$  is the number of newborn particles on levels  $b + 1 + 1, \dots, b + 1 + n$  that descend from an individual at time  $t-$  on a level in  $B(\Pi_{s,t-}, b)$ . By  $a'$ , we denote here the number of particles at time  $t-$  that are moved from a level in  $B(\Pi_{s,t-}, b) \cap \{b + 1 + 1, \dots, b + 1 + n\}$  to a level higher than  $b + 1 + n$  at time  $t$ . We set  $a'' = 1$  if the particle on level  $b + 1$  at time  $t-$  belongs to the same block of  $\Pi_{s,t-}$  as  $b$  and is pushed at time  $t$  to a level in  $b + 1 + 1, \dots, b + 1 + n$  to make way for a newborn on level  $b + 1$ . Else we set  $a'' = 0$ . Note that neither  $a + a''$  nor  $a'$  exceed the number of newborns on the first  $b + 1 + n$  levels.

Hence, if there are no newborn particles on the first  $b$  levels between time  $t'$  and  $t$  for some  $s \leq t' \leq t$ , then the numbers of levels  $b + 1 + 1, \dots, b + 1 + n$  that are in the same block as  $b$  of the partitions  $\Pi_{s,t'}$  and  $\Pi_{s,t}$  can differ by at most the number of newborn particles on the first  $b + 1 + n$  levels between time  $t'$  and  $t$ . This is condition (7.1), hence by Lemma 7.1,  $X_n(t)$  converges a. s. uniformly in compact intervals to  $X(t) := |B(\Pi_{0,t}, b)|$ . The estimates (8.3) yield the convergence (8.1).

On the event of probability 1 on which condition (2.1) holds, the paths  $t \mapsto |B(\Pi_{0,t}, b)|_n$  are càdlàg and  $\lim_{s' \uparrow t} |B(\Pi_{0,s'}, b)|_n = |B(\Pi_{0,t-}, b)|_n$  for all  $t \in (0, \infty)$  and  $n \in \mathbb{N}$ . This implies that the paths  $t \mapsto |B(\Pi_{0,t}, b)|$  are càdlàg a. s., and that  $\lim_{s' \uparrow t} |B(\Pi_{0,s'}, b)| = c_t$  for some  $c_t \in [0, 1]$  for each  $t \in (0, \infty)$  a. s. To show the assertion on the left limits, let  $\varepsilon > 0$ , and choose on an event of probability 1 a sufficiently large integer  $n_0$  such that  $||B(\Pi_{0,t}, b)|_n - |B(\Pi_{0,t}, b)|| < \varepsilon$  for all  $t \in [0, T]$  and  $n \geq n_0$ . Then,  $\limsup_{s' \uparrow t} |B(\Pi_{0,s'}, b)|_n \leq c_t + \varepsilon$  and  $\liminf_{s' \uparrow t} |B(\Pi_{0,s'}, b)|_n \geq c_t - \varepsilon$  for all  $n \geq n_0$  and  $t \in [0, T]$ . It follows  $c_t = |B(\Pi_{0,t-}, b)|$  for all  $t \in (0, \infty)$  a. s.  $\square$

*Proof of Lemma 8.3.* We choose  $b = a \vee k$  and  $f$  in Lemma 7.2 such that

$$f(t, (r_t(j, b + i), v_t(j))_{j \in [b]}, v_t(b + i)) = \mathbb{1} \left( \{r_t(k, b + i) = 0, v_t(b + i) < t\} \cup \bigcup_{I \subset [a]} \{k, b + i \in C_t^{a, \varepsilon, I}\} \right)$$

for all  $t \in \mathbb{R}_+$  and  $i \in \mathbb{N}$ . Here, the event that  $k$  and  $b + i$  are in the same block of  $\Pi_t^{a, \varepsilon}$  stands in the the indicator variable. Hence,

$$X_n(t) \leq \frac{b + n}{n} |B(\Pi_t^{a, \varepsilon}, k)|_{b+n} \leq X_n(t) + \frac{b}{n} \tag{8.4}$$

for all  $n \in \mathbb{N}$  and  $t \in \mathbb{R}_+$  a. s., with  $X_n(t)$  as defined in Lemma 7.2.

By the construction in the beginning of Section 2 and in Section 2.1, the number  $nX_n(t)$  can only jump at times  $t$  with  $\eta(\{t\} \times \hat{\mathcal{P}}^{b+n}) > 0$ . If there are no particles on the first  $b$  levels that are newborn or progenitor in the reproduction event at such a time  $t$  (i. e. if the reproduction event at time  $t$  is encoded by a partition in  $\hat{\mathcal{P}}^{b+n} \setminus \hat{\mathcal{P}}^b$ ), then  $nX_n(t)$  cannot increase and can decrease by at most the number of particles on levels  $b + 1, \dots, b + n$  that are newborn or progenitor in the reproduction event at time  $t$ . Indeed, such a newborn or progenitor particle becomes its own parent so that the its new level cannot belong to the block  $B(\Pi_t^{a, \varepsilon}, k)$ . As there are no newborns on the first  $b$  levels, each progenitor particle on a level in  $b + 1, \dots, b + n$  at time  $t$  had at time  $t-$  a level in  $b + 1, \dots, b + n$  which may have been in  $B(\Pi_t^{a, \varepsilon}, k)$ . The number of newborn particles on the first  $b + n$  levels equals the number of other particles that are pushed from one of these levels to a level above  $b + n$ . The particles at time  $t-$  on a level in  $B(\Pi_t^{a, \varepsilon}, k) \cap \{b + 1, \dots, b + n\}$  that are not progenitors and not pushed above level  $n + b$  have a level in  $B(\Pi_t^{a, \varepsilon}, k) \cap \{b + 1, \dots, b + n\}$  at time  $t$ . This yields condition (7.2), hence by Lemma 7.2 and the estimates (8.4),  $X_n(t)$  converges a. s. uniformly in compact intervals to  $X(t) := |B(\Pi_t^{a, \varepsilon}, k)|$ , which yields the convergence (8.2).

On the event of probability 1 on which condition (5.2) holds, the processes  $t \mapsto |B(\Pi_t^{a,\varepsilon}, k)|_n$  are càdlàg for all  $n \in \mathbb{N}$  a. s., hence the other assertions can be deduced from (8.2).  $\square$

To construct probability measures in the next section, we will need families of partitions with proper frequencies. A partition  $\pi$  is said to have proper frequencies if  $\sum_{B \in \pi} |B| = 1$ , that is, the asymptotic frequencies of its blocks exist and sum up to 1. In case  $\Xi \in \mathcal{M}_{\text{nd}}$ , the partition  $\Pi_{s,t}$  has proper frequencies a. s. for each  $0 \leq s < t$ . This follows from [40, Proposition 30]. Indeed, that the  $\mathcal{P}$ -valued process  $(\Pi_{t,(t-s)-}, s \in [0, t])$  is a  $\Xi$ -coalescent up to time  $t$  can be seen, for instance, by comparing the construction from the point measure  $\eta$  in Section 2 and the Poisson construction of Schweinsberg [40, Section 3]. The next two lemmas show that uncountably many partitions have proper frequencies on an event of probability 1. They give the existence of lower bounds on the number of blocks whose asymptotic frequencies add up to  $1 - \varepsilon$  that are uniform for  $t$  in compact intervals.

**Lemma 8.4.** *Assume  $\Xi \in \mathcal{M}_{\text{nd}}$ . Let  $s \in \mathbb{R}_+$ ,  $T \in (0, \infty)$ , and  $\varepsilon \in (0, T)$ . Then, on an event of probability 1, there exists an integer  $k$  such that*

$$\sum_{i \in M(\Pi_{s,t}) \cap [k]} |B(\Pi_{s,t}, i)| > 1 - \varepsilon$$

for all  $t \in [s + \varepsilon, s + T]$ . In particular, the partition  $\Pi_{s,t}$  has proper frequencies for all  $t \in (s, \infty)$  a. s.

**Lemma 8.5.** *Assume  $\Xi \in \mathcal{M}_{\text{dust}}$  and that  $(r_0, v_0)$  is exchangeable. Let  $a \in \mathbb{N}$ ,  $\varepsilon, \tilde{\varepsilon} > 0$ , and  $T \in \mathbb{R}_+$ . Then, on an event of probability 1, there exists an integer  $k$  such that*

$$\sum_{i \in M(\Pi_t^{a,\varepsilon}) \cap [k]} |B(\Pi_t^{a,\varepsilon}, i)| > 1 - \tilde{\varepsilon}$$

for all  $t \in [0, T]$ . In particular, the partition  $\Pi_t^{a,\varepsilon}$  has proper frequencies for all  $t \in \mathbb{R}_+$  a. s.

A property like the assertion of Lemma 8.4 is also considered in Section 6.1 of Labbé [29].

*Proof of Lemma 8.4.* Again we assume w. l. o. g. that  $\rho_0$  is exchangeable, and it suffices to consider the case  $s = 0$ .

**Step 1.** There exists an event of probability 1 on which for all  $t \in (\varepsilon, \infty)$  with  $\#\Pi_{0,t} = \infty$ , the partitions  $\Pi_{0,t}$  and  $\Pi_{0,t-}$  do not contain any singleton blocks. Indeed, the partition  $\Pi_{0,\varepsilon}$  contains a. s. no singletons. For  $t \in (\varepsilon, \infty)$ , an implication of  $\#\Pi_{0,t} = \infty$  ( $\#\Pi_{0,t-} = \infty$ ) is that  $\#\Pi_{\varepsilon,t} = \infty$  ( $\#\Pi_{\varepsilon,t-} = \infty$ , respectively). This also implies that all individuals at time  $\varepsilon$  have a descendant at time  $t$  (at time  $t-$ ) as the trajectories of the particles in the population model do not cross, see [22, Remark 5.1]. Hence, the assertion of Step 1 holds with the event of probability 1 that  $\Pi_{0,\varepsilon}$  contains no singletons.

**Step 2.** In this step, we assume  $\Xi_0(\Delta) > 0$  and we show that  $\Pi_{0,t^{k-}}$  has proper frequencies for all  $k \in \mathbb{N}$  a. s. By Corollary 6.5, the random partition  $\Pi_{0,t^{k-}}$  is exchangeable. On the event that the partition  $\#\Pi_{0,t^{k-}}$  has finitely many blocks, it has proper frequencies a. s. If  $\mathbb{P}(\#\Pi_{0,t^{k-}} = \infty) > 0$ , then also  $\Pi_{0,t^{k-}}$ , conditioned on the event  $\{\#\Pi_{0,t^{k-}} = \infty\}$ , is exchangeable. The assertion of step 2 now follows from step 1 and Kingman’s correspondence.

**Step 3.** Recall the filtration  $\mathcal{F}^\ell$  from Section 6.2. For each  $\ell \in \mathbb{N}$ , the process

$$\left( \sum_{i \in M(\Pi_{0,t}) \cap [\ell]} |B(\Pi_{0,t}, i)|, t \in \mathbb{R}_+ \right)$$

is adapted with respect to the usual augmentation of  $\mathcal{F}^\ell$  and has a. s. càdlàg paths by Lemma 8.2. Hence,

$$\vartheta_{\varepsilon,\ell} := \inf \left\{ t \geq \varepsilon : \sum_{i \in M(\Pi_{0,t}) \cap [\ell]} |B(\Pi_{0,t}, i)| < 1 - \varepsilon \right\}$$

is a stopping time with respect to the usual augmentation of  $\mathcal{F}^n$  for all  $n \in [\ell]$ . As  $\vartheta_{\varepsilon,j} \geq \vartheta_{\varepsilon,\ell}$  for integers  $j \geq \ell$ , it follows that

$$\vartheta_\varepsilon := \sup_{\ell \in \mathbb{N}} \vartheta_{\varepsilon,\ell} = \lim_{\ell \rightarrow \infty} \vartheta_{\varepsilon,\ell}$$

is a stopping time with respect to the usual augmentation of  $\mathcal{F}^n$  for all  $n \in \mathbb{N}$ .

By Lemma 6.7, the distance matrix  $\rho_{\vartheta_\varepsilon \wedge T}$  is exchangeable, hence the partition  $\Pi_{0,\vartheta_\varepsilon \wedge T}$  is exchangeable. On the event that the partition  $\Pi_{0,\vartheta_\varepsilon \wedge T}$  has finitely many blocks, it has proper frequencies a. s. If  $P(\#\Pi_{0,\vartheta_\varepsilon \wedge T} = \infty) > 0$ , then  $\Pi_{0,\vartheta_\varepsilon \wedge T}$ , conditioned on the event  $\{\#\Pi_{0,\vartheta_\varepsilon \wedge T} = \infty\}$ , remains exchangeable. It follows from step 1 and Kingman’s correspondence that  $\Pi_{0,\vartheta_\varepsilon \wedge T}$  has a. s. proper frequencies.

*Step 4.* A. s.,  $||B_i(\Pi_{0,t})|_n - |B_i(\Pi_{0,t-})|_n| \leq 1/n$  for all  $n, i \in \mathbb{N}$  and all  $t \in (0, \infty) \setminus \{t^k : k \in \mathbb{N}\}$  as only binary reproduction events can occur at these times. Lemma 8.2 now implies  $|B_i(\Pi_{0,t})| = |B_i(\Pi_{0,t-})|$  for all  $i \in \mathbb{N}$  and  $t \in (0, \infty) \setminus \{t^k : k \in \mathbb{N}\}$  a. s. Hence,  $|\Pi_{0,t}|_1 = |\Pi_{0,t-}|_1$  for all  $t \in (0, \infty) \setminus \{t^k : k \in \mathbb{N}\}$  a. s., where  $|\pi|_1$  denotes the sum of the asymptotic frequencies of the blocks of a partition  $\pi \in \mathcal{P}$ .

It follows that the partitions  $\Pi_{0,(\vartheta_\varepsilon \wedge T)-}$  and  $\Pi_{0,\vartheta_\varepsilon \wedge T}$  have proper frequencies a. s. Hence, there exists a. s.  $\ell \in \mathbb{N}$  such that

$$\sum_{i \in M(\Pi_{0,\vartheta_\varepsilon \wedge T}) \cap [\ell]} |B(\Pi_{0,\vartheta_\varepsilon \wedge T}, i)| > 1 - \varepsilon$$

and

$$\sum_{i \in M(\Pi_{0,(\vartheta_\varepsilon \wedge T)-}) \cap [\ell]} |B(\Pi_{0,(\vartheta_\varepsilon \wedge T)-}, i)| > 1 - \varepsilon.$$

By Lemma 8.2, there exists a. s.  $\delta > 0$  such that

$$\sum_{i \in M(\Pi_{0,t}) \cap [\ell]} |B(\Pi_{0,t}, i)| > 1 - \varepsilon$$

for all  $t \in (\vartheta_\varepsilon \wedge T - \delta, \vartheta_\varepsilon \wedge T + \delta)$ . This implies  $\vartheta_{\varepsilon,j} \notin (\vartheta_\varepsilon - \delta, \vartheta_\varepsilon + \delta)$  for all  $j \geq \ell$  a. s. on the event  $\{\vartheta_\varepsilon < T\}$ , hence  $\{\vartheta_\varepsilon < T\}$  is a null event.

The assertion follows as  $T \in (0, \infty)$  and  $\varepsilon \in (0, T)$  can be chosen arbitrarily.  $\square$

In the proof of Lemma 8.5 which is given below, we will use the following lemma which strengthens Lemma 11.2 of [22].

**Lemma 8.6.** Assume  $\Xi \in \mathcal{M}_{\text{dust}}$  and let  $t \in (0, \infty)$ ,  $i \in \mathbb{N}$ . Then a. s. on the event  $\{v_t(i) < t\}$ , there exists an integer  $j \in \mathbb{N} \setminus \{i\}$  with  $v_t(i) = v_t(j)$  and  $r_t(i, j) = 0$ .

*Proof.* We use the points  $(t^k, \pi^k)$  of  $\eta_0$  from Section 5.2. W.l. o. g., we assume  $\Xi_0(\Delta) > 0$  and that the times  $t^k \wedge t$  are stopping times with respect to the filtration  $(\mathcal{F}_s, s \in \mathbb{R}_+)$  that is defined after equation (3.1). The latter property can be achieved e. g. by setting  $\Delta_\ell = \{x \in \Delta : 1/(\ell + 1) < x_1 \leq 1/\ell\}$  and  $t^{\ell,0} = 0$  for each  $\ell \in \mathbb{N}$ , and  $t^{\ell,n} = \inf\{t > t^{\ell,n-1} : \zeta_0((t^{\ell,n-1}, t] \times \Delta_\ell) > 0\} \wedge t$  for each  $n \in \mathbb{N}$ , and associating each  $k \in \mathbb{N}$  with a pair  $(\ell, n)$ .

Lemma 6.4 yields for each  $k \in \mathbb{N}$  that the sequence  $(\mathbb{1}\{t - v_t(j) \leq t^k\}, j \in \mathbb{N})$  is exchangeable. Here we also use that this sequence is  $\eta$ -measurable so that we can

assume w.l.o.g. that  $(r_0, v_0)$  is exchangeable. The de Finetti theorem implies that a.s. either no or infinitely many elements of this sequence equal 1. Note that the random subset  $A_k := \{A_{t^k}(t, j) : j \in \mathbb{N}, t - v_t(j) \leq t^k\} \subset \mathbb{N}$  is measurable with respect to  $t^k$  and the process  $(J_{t^k+s} - J_{t^k}, s \in \mathbb{R}_+)$ , where  $J$  is defined in (3.1). Now the strong Markov property of  $J$ , our assumption on  $t^k$ , and the definition of  $\pi^k$  yield that  $A_k$  and  $\pi^k$  are independent. As  $\Xi\{0\} = 0$ , by Kingman's correspondence, and as  $A_k$  and  $\pi^k$  are independent, all non-singleton blocks of  $\pi^k$  have an infinite intersection with  $A_k$  a.s. on the event  $\{\#A_k = \infty\}$ . Furthermore, the definition of  $v_{t^k}$  yields that  $A_{t^k}(t, j) \neq A_{t^k}(t, j')$  for all distinct  $j, j' \in \mathbb{N}$  with  $t - v_t(j) \leq t^k, t - v_t(j') \leq t^k$ . Hence,  $\#A_k = \#\{j \in \mathbb{N} : t - v_t(j) \leq t^k\}$ .

By definition of  $v_t(i)$  and condition (5.2), there exists a.s. on  $\{v_t(i) < t\}$  an integer  $k \in \mathbb{N}$  such that  $t - v_t(i) = t^k$  and  $A_{t^k}(t, i)$  is in a non-singleton block of  $\pi^k$ . The above implies that there exists a.s. on  $\{v_t(i) < t\}$  an integer  $j \in \mathbb{N} \setminus \{i\}$  such that  $t - v_t(j) \leq t^k$  and  $A_{t^k}(t, j)$  is in the same block of  $\pi^k$  as  $A_{t^k}(t, i)$ . The definition of  $v_t(j)$  now yields  $v_t(j) = v_t(i)$  and  $r_t(i, j) = 0$ .  $\square$

We will apply Lemma 8.6 through the following corollary.

**Corollary 8.7.** Assume  $\Xi \in \mathcal{M}_{\text{dust}}$  and that  $(r_0, v_0)$  is exchangeable. Let  $a \in \mathbb{N}$  and  $\varepsilon > 0$ . Then on an event of probability 1, none of the partitions  $\Pi_t^{a, \varepsilon}, t \in \mathbb{R}_+$  contains singleton blocks.

*Proof.* Let  $q \in \mathbb{R}_+$ . A.s. by Lemma 8.6, only integers  $i \in \mathbb{N}$  with  $v_q(i) \geq q$  can form singleton blocks of  $\Pi_q^{a, \varepsilon}$ . As those integers belong to finitely many blocks by definition of  $\Pi_q^{a, \varepsilon}$ , it follows that  $\Pi_q^{a, \varepsilon}$  contains a.s. at most finitely many singleton blocks. As the partition  $\Pi_q^{a, \varepsilon}$  is exchangeable by Lemma 6.4, Kingman's correspondence implies that  $\Pi_q^{a, \varepsilon}$  contains a.s. no singleton blocks.

On the event of probability 1 on which condition (5.2) holds, there exists for each  $t \in \mathbb{R}_+$  and  $i \in \mathbb{N}$  a time  $q(t, i) \in (t, \infty) \cap \mathbb{Q}$  with  $\eta((t, q(t, i)] \times \hat{\mathcal{P}}^i) = 0$ , as (5.2) implies that the points of  $\eta(\cdot \times \hat{\mathcal{P}}^i)$  do not accumulate. By construction, if  $i$  forms a singleton block in  $\Pi_t^{a, \varepsilon}$ , then  $i$  forms a singleton block also in  $\Pi_{q(t, i)}^{a, \varepsilon}$ , as the particle remains on level  $i$  and does not reproduce. This implies the assertion.  $\square$

*Proof of Lemma 8.5.* We proceed similarly to the proof of Lemma 8.4. We assume w.l.o.g.  $\Xi_0(\Delta) > 0$ .

*Step 1.* For all  $k \in \mathbb{N}$ , the partition  $\Pi_{t^k}^{a, \varepsilon}$  is exchangeable by Corollary 6.5. By Corollary 8.7 and Kingman's correspondence, it follows that  $\Pi_{t^k}^{a, \varepsilon}$  has proper frequencies a.s.

*Step 2.* We set for  $\ell \in \mathbb{N}$

$$\vartheta_{\varepsilon, \ell} = \inf \left\{ t \geq 0 : \sum_{i \in M(\Pi_t^{a, \varepsilon}) \cap [\ell]} |B(\Pi_t^{a, \varepsilon}, i)| < 1 - \varepsilon \right\},$$

We deduce from Lemma 8.3 that  $\vartheta_{\varepsilon, \ell}$  is a stopping time with respect to the usual augmentation of  $\mathcal{F}^n$  for all  $n \in [\ell]$ . Then we define  $\vartheta_\varepsilon = \sup_{\ell \in \mathbb{N}} \vartheta_{\varepsilon, \ell}$  which is for all  $n \in \mathbb{N}$  a stopping time with respect to the usual augmentation of  $\mathcal{F}^n$ . Let  $T \in \mathbb{R}_+$ . The partition  $\Pi_{\vartheta_\varepsilon \wedge T}^{a, \varepsilon}$  is exchangeable by Lemma 6.7. We deduce as in step 1 that it has proper frequencies a.s.

*Step 3.* We conclude as in the proof of Lemma 8.4, using Lemma 8.3.  $\square$

Recall the set of measures  $\mathcal{M}_{\text{CDI}}$  from Section 3.1. The following lemma will be used in the proof of Theorem 3.5(i).

**Lemma 8.8.** Assume  $\Xi \in \mathcal{M}_{\text{CDI}}$ . Let  $s \in \mathbb{R}_+$ . Then a.s.,  $|B(\Pi_{s, t}, i)| > 0$  for all  $t \in (s, \infty)$  and  $i \in \mathbb{N}$ .

*Proof.* Again we assume w.l. o. g. that  $\rho_0$  is exchangeable, and it suffices to consider the case  $s = 0$ . Let  $\varepsilon > 0$ . For each  $k \in \mathbb{N}$ , the process

$$\left( \min_{i \in [k]} |B(\Pi_{0,t}, i)|, t \in \mathbb{R}_+ \right)$$

is adapted with respect to the usual augmentation of the filtration  $\mathcal{F}^k$  (defined in Section 6.2) and has a. s. càdlàg paths by Lemma 8.2. Consequently,

$$\vartheta_k := \inf \left\{ t \geq \varepsilon : \min_{i \in [k]} |B(\Pi_{0,t}, i)| = 0 \right\}$$

is a stopping time with respect to the usual augmentation of  $\mathcal{F}^n$  for all  $n \in [k]$ . It follows that

$$\vartheta := \inf_{k \in \mathbb{N}} \vartheta_k = \lim_{k \rightarrow \infty} \vartheta_k$$

is a stopping time with respect to the usual augmentation of  $\mathcal{F}^n$  for all  $n \in \mathbb{N}$ . By Lemma 6.7, the distance matrix  $\gamma_n(\rho_{\vartheta \wedge T})$  is exchangeable for each  $T \in [\varepsilon, \infty)$  and  $n \in \mathbb{N}$ . Hence, the partition  $\Pi_{0, \vartheta \wedge T}$  is exchangeable. The assumption  $\Xi \in \mathcal{M}_{\text{CDI}}$  implies  $\#\Pi_{0, \vartheta \wedge T} < \infty$  a. s. Kingman's correspondence now implies that each block of  $\Pi_{0, \vartheta \wedge T}$  has a positive asymptotic frequency a. s. Hence, by Lemma 8.2, there exists an event of probability 1 on which all blocks of the partitions  $\Pi_{0,t}$  with  $t$  in a right neighborhood of  $\vartheta \wedge T$  have positive asymptotic frequencies. By definition of  $\vartheta$ , it follows  $\mathbb{P}(\vartheta < T) = 0$ . The assertion follows as  $T$  and  $\varepsilon$  can be chosen arbitrarily.  $\square$

## 9 The construction on the lookdown space

Now we apply the results from the last section to prove the assertions from Section 3. We use the coupling characterization of the Prohorov distance, namely (see e. g. [14, Theorem 3.1.2]) that in a separable metric space  $(Y, d)$ , the Prohorov distance between probability measures  $\mu$  and  $\mu'$  on the Borel sigma algebra is given by

$$d_{\text{P}}^Y(\mu, \mu') = \inf_{\nu} \inf \{ \varepsilon > 0 : \nu \{ (y, y') \in Y^2 : d(y, y') > \varepsilon \} < \varepsilon \},$$

where the first infimum is over all couplings  $\nu$  of the probability measures  $\mu$  and  $\mu'$ .

### 9.1 The case with dust

In this subsection, we always consider the case  $\Xi \in \mathcal{M}_{\text{dust}}$ . Let  $(X, r, m)$  be a marked metric measure space, and let  $X \times \mathbb{R}_+$  be endowed with the product metric  $d((x, v), (x', v')) = r(x, x') \vee |v - v'|$ . Let  $(x(i), v(i))_{i \in \mathbb{N}}$  be an  $m$ -iid sequence in  $X \times \mathbb{R}_+$  that is independent of  $\eta$ . We set  $(r_0, v_0) = ((r(x(i), x(j)))_{i, j \in \mathbb{N}}, v)$ . Then  $(r_0, v_0)$  is distributed according to the marked distance matrix distribution of  $(X, r, m)$ . With the extended lookdown space  $(\hat{Z}, \rho)$  associated with  $\eta$  and  $(r_0, v_0)$ , we are in the setting of Subsection 3.2.

*Proof of Theorem 3.10 (beginning).* We begin with the proof of item (i). For  $\varepsilon > 0$  and  $a, n \in \mathbb{N}$ , we define the probability measures

$$m_t^{a, \varepsilon, n} = \sum_{i \in M(\Pi_t^{a, \varepsilon})} |B(\Pi_t^{a, \varepsilon}, i)|_n \delta_{(z(t, i), v_t(i))}.$$

on  $\hat{Z} \times \mathbb{R}_+$ . Clearly,  $|B(\Pi_t^{a, \varepsilon}, i)|_n = 0$  for all  $i \in M(\Pi_t^{a, \varepsilon})$  with  $i > n$ .

Let  $T \in \mathbb{R}_+$ . Using that the Prohorov distance is bounded from above by the total variation distance, we obtain

$$\begin{aligned} & \lim_{n, \ell \rightarrow \infty} \sup_{t \in [0, T]} d_{\mathbb{P}}^{\hat{Z} \times \mathbb{R}_+} \left( m_t^{a, \varepsilon, n}, m_t^{a, \varepsilon, \ell} \right) \\ & \leq \lim_{n, \ell \rightarrow \infty} \sup_{t \in [0, T]} \sum_{i \in M(\Pi_t^{a, \varepsilon})} \left| |B(\Pi_t^{a, \varepsilon}, i)|_n - |B(\Pi_t^{a, \varepsilon}, i)|_\ell \right| \\ & \leq \lim_{k \rightarrow \infty} \lim_{n, \ell \rightarrow \infty} \sup_{t \in [0, T]} \sum_{i \in M(\Pi_t^{a, \varepsilon}) \cap [k]} \left| |B(\Pi_t^{a, \varepsilon}, i)|_n - |B(\Pi_t^{a, \varepsilon}, i)|_\ell \right| \\ & \quad + 2 \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{t \in [0, T]} \sum_{\substack{i \in M(\Pi_t^{a, \varepsilon}) \\ i > k}} |B(\Pi_t^{a, \varepsilon}, i)|_n = 0 \quad \text{a. s.} \end{aligned}$$

The first summand on the right-hand side equals zero a. s. by Lemma 8.3. A. s., the second summand equals zero as for each  $\tilde{\varepsilon} > 0$ , there exist integers  $k$  and  $n_0$  such that

$$\inf_{t \in [0, T]} \sum_{i \in M(\Pi_t^{a, \varepsilon}) \cap [k]} |B(\Pi_t^{a, \varepsilon}, i)|_n \geq 1 - \tilde{\varepsilon}$$

for all  $n \geq n_0$  by Lemmas 8.5 and 8.3.

Now we compare the probability measures  $m_t^n$  and  $m_t^{a, \varepsilon, n}$ . A coupling  $\nu$  of these probability measures is given by

$$\nu = \sum_{i=1}^n \frac{1}{n} \delta_{((z(t, i), v_t(i)), (z(t, j), v_t(j)))},$$

where  $j = \min B(\Pi_t^{a, \varepsilon}, i)$  in each summand. By definition of  $\Pi_t^{a, \varepsilon}$  and Remark 8.1,

$$d^{\hat{Z} \times \mathbb{R}_+}((z(t, i), v_t(i)), (z(t, j), v_t(j))) = r_t(i, j) \vee |v_t(i) - v_t(j)| < 2\varepsilon$$

for all  $i, j \in \mathbb{N}$  that are in the same block of  $\Pi_t^{a, \varepsilon}$  and not in  $C_t^{a, \varepsilon, \emptyset}$ . The coupling characterization of the Prohorov metric implies

$$\begin{aligned} & d_{\mathbb{P}}^{\hat{Z} \times \mathbb{R}_+} (m_t^n, m_t^{a, \varepsilon, n}) \\ & \leq \nu \left\{ (y, y') \in (\hat{Z} \times \mathbb{R}_+)^2 : d^{\hat{Z} \times \mathbb{R}_+}(y, y') \geq 2\varepsilon \right\} + 2\varepsilon \\ & \leq |C_t^{a, \varepsilon, \emptyset}|_n + 2\varepsilon. \end{aligned}$$

By construction,  $|C_t^{a, \varepsilon, \emptyset}|_n \leq |C_0^{a, \varepsilon, \emptyset}|_n$ . This follows from the definition of  $C_t^{a, \varepsilon, \emptyset}$  in Section 8, from the definition of  $(r_t, v_t)$ , as a particle at time  $s$  on a level  $i$  loses the property that  $v_s(i) \geq s$  if it reproduces, and as it can only increase its level in a reproduction event. By exchangeability (or Lemma 8.3), we have  $\lim_{n \rightarrow \infty} |C_0^{a, \varepsilon, \emptyset}|_n = |C_0^{a, \varepsilon, \emptyset}|$  a. s.

The triangle inequality yields

$$\lim_{n, \ell \rightarrow \infty} \sup_{t \in [0, T]} d_{\mathbb{P}}^{\hat{Z} \times \mathbb{R}_+} (m_t^n, m_t^\ell) \leq 2|C_0^{a, \varepsilon, \emptyset}| + 4\varepsilon \quad \text{a. s.} \tag{9.1}$$

Letting first  $a \rightarrow \infty$  and then  $\varepsilon \downarrow 0$ , we obtain from Lemma 9.1 below that the left-hand side of (9.1) equals zero a. s. By definition of  $m_t$  and as  $\hat{Z} \times \mathbb{R}_+$  is complete, this implies assertion (i).

As recalled in Section 5.3, the map  $t \mapsto (z(t, i), v_t(i))$  is a. s. càdlàg with respect to  $d^{\hat{Z} \times \mathbb{R}_+}$ , hence the map  $t \mapsto m_t^n$  is a. s. càdlàg in the weak topology on  $\hat{Z} \times \mathbb{R}_+$ . Jump times can only lie in the set  $\Theta_0$ , which equals a. s. the set of reproduction times. The

uniformity of the convergence in assertion (i) implies that also  $t \mapsto m_t$  is a. s. càdlàg in the weak topology on  $\hat{Z} \times \mathbb{R}_+$  with no jump times outside  $\Theta_0$ .

W.l. o. g., we assume  $\Xi(\Delta) > 0$ . As  $\Xi \in \mathcal{M}_{\text{dust}}$ , this implies  $\Xi_0(\Delta) > 0$  and  $\Theta_0 = \{t^k : k \in \mathbb{N}\}$  a. s. Now we deduce that  $m_{t^k}(\hat{Z} \times \{0\}) > 0$  a. s. all for  $k \in \mathbb{N}$ , which is part of assertion (iii). By Corollary 6.5, the sequence  $(v_{t^k}(i), i \in \mathbb{N})$  is exchangeable. From the definition (2.7) of  $\Theta_0$ , the definition of  $v_t(i)$  in Section 2.1, and as condition (5.2) is a. s. satisfied, it follows that  $|\{i \in \mathbb{N} : v_{t^k}(i) = 0\}| > 0$  a. s. Hence, the de Finetti theorem and the definition of  $m_t$  yield that a. s., the empirical measures  $n^{-1} \sum_{i=1}^n \delta_{v_{t^k}(i)}$  converge weakly to the directing measure  $m_{t^k}(\hat{Z} \times \cdot)$  on  $\mathbb{R}_+$  which satisfies  $m_{t^k}(\hat{Z} \times \{0\}) > 0$ .

On the event of probability 1 on which condition (5.2) is satisfied,  $v_{t^k-}(i) > 0$  for all  $i \in \mathbb{N}$  a. s. Analogously to the above, this yields  $m_{t^k-}(\hat{Z} \times \{0\}) = 0$  a. s. This is another part of assertion (iii). As a consequence, the set of jump times is a. s. not smaller than  $\Theta_0$ , which yields assertion (ii). After completing the proof of assertion (i) by proving Lemma 9.1 below, we give in Proposition 9.2 a representation of the probability measures  $m_t$  from which assertion (iv) will follow by Remark 9.4. A. s.,  $v_t(i) > 0$  (by condition (5.2)) and  $m_t = m_{t-}$  (by item (ii)) for all  $t \in (0, \infty) \setminus \Theta_0$  and  $i \in \mathbb{N}$ . Hence, the remainder of assertion (iii) also follows from Proposition 9.2 below.  $\square$

**Lemma 9.1.** *Let  $\varepsilon > 0$ , and let  $C_0^{a,\varepsilon,\emptyset}$  be defined as in this subsection from a random variable  $(r_0, v_0)$  that has the marked distance matrix distribution of a marked metric measure space  $(X, r, m)$ . Then,  $\lim_{a \rightarrow \infty} |C_0^{a,\varepsilon,\emptyset}| = 0$  a. s.*

*Proof.* From the definitions of  $C_0^{a,\varepsilon,\emptyset}$  and of the extended lookdown space, we obtain

$$C_0^{a,\varepsilon,\emptyset} = \left\{ j \in \mathbb{N} : (x(j), v(j)) \in (X \times \mathbb{R}_+) \setminus \bigcup_{i=1}^a \mathcal{U}_\varepsilon^{X \times \mathbb{R}_+}(x(i), v(i)) \right\}$$

for all  $a \in \mathbb{N}$ , where  $\varepsilon$ -balls are defined by  $\mathcal{U}_\varepsilon^{X \times \mathbb{R}_+}(x', v') = \{(x'', v'') \in X \times \mathbb{R}_+ : r(x', x'') \vee |v' - v''| < \varepsilon\}$  for  $(x', v') \in X \times \mathbb{R}_+$ .

As  $(x(a+j), v(a+j))_{j \in \mathbb{N}}$  is an  $m$ -iid sequence in  $X \times \mathbb{R}_+$  that is independent of  $(x(i), v(i))_{i \in [a]}$ , the law of large numbers yields

$$|C_0^{a,\varepsilon,\emptyset}| = m \left( (X \times \mathbb{R}_+) \setminus \bigcup_{i=1}^a \mathcal{U}_\varepsilon^{X \times \mathbb{R}_+}(x(i), v(i)) \right) \quad \text{a. s.}$$

By separability,  $X \times \mathbb{R}_+$  can be covered by countably many balls of diameter  $\varepsilon/2$ , and each ball with positive mass contains elements of the sequence  $(x(i), v(i))_{i \in \mathbb{N}}$  a. s. Using also continuity of  $m$  from above, this implies

$$\lim_{a \rightarrow \infty} m \left( (X \times \mathbb{R}_+) \setminus \bigcup_{i=1}^a \mathcal{U}_\varepsilon^{X \times \mathbb{R}_+}(x(i), v(i)) \right) = 0 \quad \text{a. s.}$$

This yields the assertion.  $\square$

Now we give an explicit representation of the probability measures  $m_t$ . For  $t \in \mathbb{R}_+$ , we denote by  $\Pi_t$  the partition of  $\mathbb{N}$  in which integers  $i, j$  are in the same block if and only if  $v_t(i) = v_t(j) < t$  and  $r_t(i, j) = 0$ , which is condition (i) on p. 33. We may call  $\Pi_t$  the partition of siblings.

The individuals at time  $t$  whose parents are also parents of individuals at time 0 are on the levels in the set

$$C_t = \{i \in \mathbb{N} : v_t(i) \geq t\}.$$

All elements of  $C_t$  form singleton blocks in  $\Pi_t$ . For  $a \in \mathbb{N}, \varepsilon > 0$ , the non-singleton blocks of  $\Pi_t$  are also blocks of  $\Pi_t^{a,\varepsilon}$ . Hence, Corollary 8.7 implies that a. s.,  $C_t$  equals the union of the singleton blocks of  $\Pi_t$  for each  $t \in \mathbb{R}_+$ . We also define the map

$$\theta_t : \hat{Z} \times \mathbb{R}_+ \rightarrow \hat{Z} \times \mathbb{R}_+, \quad (z, s) \mapsto (z, s + t)$$

with is continuous for  $d^{\hat{Z} \times \mathbb{R}_+}$ .

**Proposition 9.2.** Assume  $\Xi \in \mathcal{M}_{\text{dust}}$ . Then on an event of probability 1,

$$m_t = \sum_{i \in M(\Pi_t)} |B(\Pi_t, i)| \delta_{(z(t,i), v_t(i))} + |C_t| \theta_t(m_0) \tag{9.2}$$

for all  $t \in \mathbb{R}_+$ .

**Remark 9.3.** Proposition 9.2 allows to describe the probability measure  $m_t(\cdot \times \mathbb{R}_+)$  on the extended lookdown space  $\hat{Z}$  as follows. With probability given by the asymptotic frequency of the individuals at time  $t$  whose ancestral lineages do not coalesce with other ancestral lineages within the time interval  $(0, t]$ , we sample according to  $m_0(\cdot \times \mathbb{R}_+)$ . For each block in a reproduction event at a time  $\tau'$  in  $(0, t]$ , we draw the individual on, say, the lowest level in this block (which is identified with the individuals on all other levels in this block, as they have genealogical distance zero) with probability given by the asymptotic frequency of the individuals at time  $t$  that descend from this block and whose ancestral lineages do not coalesce with any other ancestral lineages in the time interval  $(\tau', t]$ .

**Remark 9.4.** The marked metric measure spaces  $(X, r, m)$  and  $(\hat{Z}, \rho, m_0)$  are isomorphic (as defined in Section 4.2). This follows from the definition of  $(\hat{Z}, \rho)$  and the Gromov reconstruction theorem (cf. e. g. [22, Proposition 10.5]). Hence,  $m_0$  is purely atomic if and only if  $m$  is purely atomic, and the assertion on the atomicity of  $m_t$  in Theorem 3.10(iv) follows from Proposition 9.2.

Proposition 9.2 and Lemmas 8.3 and 8.5 imply that on an event of probability 1, each weak limit  $m_{t-}$  with  $t \in (0, \infty)$  is also the sum of countably many atoms and a multiple of  $\theta_t(m_0)$ . Here we also use that the map  $\mathbb{R}_+ \times \hat{Z} \times \mathbb{R}_+ \rightarrow \hat{Z} \times \mathbb{R}_+, (s, z', v') \mapsto \theta_s(z', v')$  is continuous also in  $s$ . This yields the assertion on the atomicity of the left limits  $m_{t-}$  in Theorem 3.10(iv).

*Proof of Proposition 9.2.* For all  $a \in \mathbb{N}, \varepsilon > 0$ , and  $t \in \mathbb{R}_+$ , the definition of  $C_t^{a,\varepsilon,I}$  implies

$$C_t = \bigcup_{I \subset [a]} C_t^{a,\varepsilon,I}.$$

Lemma 8.3 implies the existence of the asymptotic frequencies

$$|C_t| = \sum_{I \subset [a]} |C_t^{a,\varepsilon,I}|$$

for all  $t \in \mathbb{R}_+$  a. s. Let us denote the right-hand side of equation (9.2) by  $m'_t$ . On an event of probability one,  $m'_t$  is a well-defined probability measure for all  $t \in \mathbb{R}_+$ , and  $t \mapsto m'_t$  is càdlàg with respect to  $d_P^{\hat{Z} \times \mathbb{R}_+}$ . This follows from Lemmas 8.3 and 8.5, and as a. s.,  $C_t$  equals the union of the singleton blocks of  $\Pi_t$  for each  $t$ . As also  $t \mapsto m_t$  is a. s. càdlàg, it suffices to show that (9.2) holds a. s. for a fixed  $t \in \mathbb{R}_+$ .

For  $i, j \in \mathbb{N}$  that are in the same block of  $\Pi_t$ , we have  $v_t(i) = v_t(j)$  and  $z(t, i) = z(t, j)$  in  $\hat{Z}$  by definition. Hence,

$$m_t^n = \sum_{i \in M(\Pi_t)} |B(\Pi_t, i)|_n \delta_{(z(t,i), v_t(i))}.$$

Let  $A_t^n = \{A_0(t, i) : i \in C_t \cap [n]\}$  be the set of the ancestral levels at time 0 of the individuals on the levels in  $C_t \cap [n]$  at time  $t$ . For finite sets  $A \subset \mathbb{N}$ , let

$$m_0^A = \frac{1}{\#A \vee 1} \sum_{i \in A} \delta_{(z(0,i), v_0(i))}.$$

Using the definitions of  $z(t, i)$  and  $v_t(i)$ , we can write

$$m_t^n = \sum_{i \in M(\Pi_t) \setminus C_t} |B(\Pi_t, i)|_n \delta_{(z(t,i), v_t(i))} + |C_t|_n \theta_t \left( m_0^{A_t^n} \right). \tag{9.3}$$

Note that  $A_t^n$  is the set of those levels at time zero that are occupied by particles that do not reproduce until time  $t$  and that are not above level  $n$  at time  $t$ . This implies that  $C_t \cap [n]$  and  $A_t^n$  are bijective. In particular,  $\bigcup_{n \in \mathbb{N}} A_t^n$  is infinite on the event  $\{|C_t| > 0\}$ .

Now we work with arguments that we encountered already in Lemmas 3.2 and 3.11. Recall that  $(r_0, v_0) = ((r(x(i), x(j)))_{i,j \in \mathbb{N}}, (v(i))_{i \in \mathbb{N}})$  where  $(x(i), v(i))_{i \in \mathbb{N}}$  is an  $m$ -iid sequence in  $X \times \mathbb{R}_+$  that is independent of  $\eta$ . For  $A \subset \mathbb{N}$ , we define the empirical measure

$$m^A = \frac{1}{\#A \vee 1} \sum_{i \in A} \delta_{(x(i), v(i))}$$

on  $X \times \mathbb{R}_+$ . As the sets  $A_t^n$  and  $C_t$  are  $\eta$ -measurable, the Glivenko-Cantelli theorem implies that the weak convergence

$$m = w\text{-}\lim_{n \rightarrow \infty} m^{A_t^n}$$

holds a. s. on  $\{|C_t| > 0\}$ . By construction of the lookdown space  $\hat{Z}$ , the map  $\{(x(i), v(i)) : i \in \mathbb{N}\} \rightarrow \hat{Z} \times \mathbb{R}_+$ ,  $(x(i), v(i)) \mapsto (i, v(i))$  can be extended a. s. to an isometry  $\varphi$  from the closed support of  $m$  in  $X \times \mathbb{R}_+$  to  $\hat{Z} \times \mathbb{R}_+$  which satisfies  $m_0^{A_t^n} = \varphi(m^{A_t^n})$  and  $m_0 = \varphi(m)$ . Thus the weak convergence

$$m_0 = w\text{-}\lim_{n \rightarrow \infty} m_0^{A_t^n}$$

holds a. s. on  $\{|C_t| > 0\}$ .

By exchangeability (or Lemma 8.3), also the relative frequencies in expression (9.3) converge a. s. to the corresponding asymptotic frequencies. Hence,  $m_t^n$  converges weakly to  $m_t'$  on an event of probability 1. This yields the assertion.  $\square$

### 9.2 The case without dust

In this subsection, we always consider the case  $\Xi \in \mathcal{M}_{\text{nd}}$ . Let  $(x(i), i \in \mathbb{N})$  be an iid sequence in a metric measure space  $(X, r, \mu)$  that is independent of  $\eta$ . We assume  $\rho_0 = (r(x(i), x(j)))_{i,j \in \mathbb{N}}$ . With the lookdown space  $(Z, \rho)$  associated with  $\eta$  and  $\rho_0$ , we are in the setting of Section 3.1.

*Proof of Theorem 3.1 (beginning).* We work on an event of probability 1 on which in particular the assertions of Lemmas 8.2 and 8.4 hold simultaneously for all  $s \in \mathbb{Q}_+$ , and we mostly omit ‘a. s.’ We define for each  $t \in (0, \infty)$ ,  $s \in (0, t) \cap \mathbb{Q}$ , and  $n \in \mathbb{N}$  a probability measure  $\mu_t^{(n,s)}$  on  $Z$  by

$$\mu_t^{(n,s)} = \sum_{i \in M(\Pi_{s,t})} |B(\Pi_{s,t}, i)|_n \delta_{(t,i)}.$$

There exists a coupling  $\nu$  of the probability measures  $\mu_t^{(n,s)}$  and  $\mu_t^n$  given by

$$\nu = \sum_{i=1}^n \frac{1}{n} \delta_{((t, \min B(\Pi_{s,t}, i)), (t, i))}.$$

As  $\rho((t, \min B(\Pi_{s,t}, i)), (t, i)) \leq 2(t - s)$ , the coupling characterization of the Prohorov metric implies

$$d_{\mathbb{P}}^Z \left( \mu_t^{(n,s)}, \mu_t^n \right) \leq 2(t - s). \tag{9.4}$$

Let  $\varepsilon \in (0, \infty) \cap \mathbb{Q}$ ,  $T \in (\varepsilon, \infty)$ , and  $\tau \in (0, \varepsilon) \cap \mathbb{Q}$ . First we consider  $t \in [\tau, T]$ . Let  $s_0 = 0$  and  $s_j = \tau + (j - 1)\varepsilon$  for  $j \in \mathbb{N}$ . By Lemma 8.4, there exists for each  $j \in \mathbb{N}_0$  an integer  $\ell_j$  such that

$$\sum_{i \in M(\Pi_{s_j,t}) \cap [\ell_j]} |B(\Pi_{s_j,t}, i)| > 1 - \varepsilon$$

for all  $t \in [s_{j+1}, s_{j+2}]$ . We set  $\ell = \max\{\ell_j : j \in \mathbb{N}_0, s_{j+1} \leq T\}$ . By Lemma 8.2, there exists an integer  $n'$  such that

$$\sum_{i \in M(\Pi_{s_j,t}) \cap [\ell]} |B(\Pi_{s_j,t}, i)|_n > 1 - \varepsilon \tag{9.5}$$

for all  $n \geq n'$ ,  $j \in \mathbb{N}_0$  with  $s_{j+1} \leq T$ , and  $t \in [s_{j+1}, s_{j+2}]$ . For all  $t \in [\tau, T]$ , all  $k, n \geq n'$ , and  $j \in \mathbb{N}_0$  such that  $t \in [s_{j+1}, s_{j+2}]$ , the bound (9.5) yields

$$d_{\mathbb{P}}^Z \left( \mu_t^{(n,s_j)}, \mu_t^{(k,s_j)} \right) \leq \sum_{\substack{i \in M(\Pi_{s_j,t}): \\ i \leq \ell}} \left| |B(\Pi_{s_j,t}, i)|_n - |B(\Pi_{s_j,t}, i)|_k \right| + \varepsilon$$

as the Prohorov distance is bounded from above by the total variation distance. By Lemma 8.2, this expression converges to  $\varepsilon$  uniformly in  $t \in [\tau, T]$  as  $n, k \rightarrow \infty$ . Furthermore, for all  $n, k \in \mathbb{N}$ ,  $t \in [\tau, T]$ , and  $j$  with  $t \in [s_{j+1}, s_{j+2}]$ , the bound (9.4) yields

$$\begin{aligned} d_{\mathbb{P}}^Z \left( \mu_t^n, \mu_t^k \right) &\leq d_{\mathbb{P}}^Z \left( \mu_t^n, \mu_t^{(n,s_j)} \right) + d_{\mathbb{P}}^Z \left( \mu_t^{(n,s_j)}, \mu_t^{(k,s_j)} \right) + d_{\mathbb{P}}^Z \left( \mu_t^{(k,s_j)}, \mu_t^k \right) \\ &\leq 8\varepsilon + d_{\mathbb{P}}^Z \left( \mu_t^{(n,s_j)}, \mu_t^{(k,s_j)} \right). \end{aligned}$$

Hence,  $\lim_{n \rightarrow \infty} \sup_{t \in [\tau, T]} d_{\mathbb{P}}^Z(\mu_t^n, \mu_t) \leq 8\varepsilon$  for all  $\tau \in (0, \varepsilon) \cap \mathbb{Q}$  a. s. □

In the next part of the proof, we choose a random  $\tau$  and show uniform convergence of  $\mu_t^n$  in  $[0, \tau]$ . To this aim, we define for each  $b \in \mathbb{N}$ ,  $\varepsilon > 0$ , and  $t \in \mathbb{R}_+$  a collection  $(A_t^{b,\varepsilon,I}, I \subset [b])$  of disjoint subsets of  $\mathbb{N}$  (whose union is  $\mathbb{N}$ ) by

$$A_t^{b,\varepsilon,I} = \bigcap_{j \in I} \{i \in \mathbb{N} : \rho_t(j, i) < \varepsilon + 2t\} \cap \bigcap_{k \in [b] \setminus I} \{i \in \mathbb{N} : \rho_t(k, i) \geq \varepsilon + 2t\}.$$

This partitions the set of individuals at time  $t$  into blocks such that, if  $t$  is small, then two individuals are close if they are in a common block  $A_t^{b,\varepsilon,I}$  with  $I \neq \emptyset$ .

**Lemma 9.5.** *Let  $\tilde{T}, \varepsilon > 0$ ,  $b \in \mathbb{N}$ , and  $I \subset [b]$ . Then*

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, \tilde{T}]} \left| |A_t^{b,\varepsilon,I}|_n - |A_t^{b,\varepsilon,I}| \right| = 0 \quad \text{a. s.}$$

*Proof.* In Lemma 7.1, choose  $f$  such that

$$f(t, (\rho_t(j, b+1+i))_{j \in [b]}) = \prod_{j \in I} \mathbb{1}\{\rho_t(j, b+1+i) < \varepsilon + 2t\} \prod_{k \in [b] \setminus I} \mathbb{1}\{\rho_t(k, b+1+i) \geq \varepsilon + 2t\}$$

for all  $t \in \mathbb{R}_+$  and  $i \in \mathbb{N}$ . The product on the right-hand side is the indicator variable of  $\{b+1+i \in A_t^{b,\varepsilon,I}\}$ . By the same argument as in the proof of Lemma 8.2, condition (7.1) is satisfied. The assertion now follows from Lemma 7.1 as

$$X_n(t) \leq \frac{b+1+n}{n} |A_t^{b,\varepsilon,I}|_{b+1+n} \leq X_n(t) + \frac{b+1}{n}$$

and  $X(t) = |A_t^{b,\varepsilon,I}|$  for all  $t \in \mathbb{R}_+$  a. s., where  $X_n(t)$  and  $X(t)$  are defined in Lemma 7.1. □

In a metric space  $(Y, d)$ , we denote by  $\mathcal{U}_\varepsilon^Y(z) = \{y' \in Y : d(y', y) < \varepsilon\}$  for  $y \in Y$  some  $\varepsilon$ -balls.

**Lemma 9.6.** *Let  $b \in \mathbb{N}$  and  $\varepsilon > 0$ . Then*

$$|A_0^{b,\varepsilon,\emptyset}| = 1 - \mu \left( \bigcup_{i=1}^b \mathcal{U}_\varepsilon^X(x(i)) \right) \quad \text{a. s.}$$

*Proof.* By the definitions of  $A_0^{b,\varepsilon,\emptyset}$  and  $Z$ ,

$$A_0^{b,\varepsilon,\emptyset} = \left\{ j \in \mathbb{N} : x(j) \notin \bigcup_{i=1}^b \mathcal{U}_\varepsilon^X(x(i)) \right\}.$$

Clearly,  $(x(b+i), i \in \mathbb{N})$  is a  $\mu$ -iid sequence in  $X$  that is independent of  $(x(i), i \in [b])$ . The assertion follows from the law of large numbers.  $\square$

*Proof of Theorem 3.1 (end).* By separability of  $(X, r)$  and continuity of  $\mu$  from above (analogously to the proof of Lemma 9.1), we can (and do) choose a (random) integer  $b$  such that

$$\mu \left( \bigcup_{i=1}^b \mathcal{U}_\varepsilon^X(x(i)) \right) > 1 - \varepsilon.$$

By condition (2.1) and Lemmas 9.6 and 9.5, we can (and do) choose (random)  $\tau \in (0, \varepsilon) \cap \mathbb{Q}$  and  $n_0 \in \mathbb{N}$  such that  $\eta((0, \tau] \times \mathcal{P}^b) = 0$  and  $|A_t^{b,\varepsilon,\emptyset}|_n < \varepsilon$  for all  $t \in [0, \tau]$  and  $n \geq n_0$ .

By construction,  $\rho((t, i), (t, j)) < 2\varepsilon + 4t$  for all  $t \in \mathbb{R}_+$ ,  $I \neq \emptyset$ , and  $i, j \in A_t^{b,\varepsilon,I}$ . Using a coupling  $\nu$  of  $\mu_t^n$  and  $\mu_t^k$  such that

$$\nu \left( \left( \{t\} \times A_t^{b,\varepsilon,I} \right) \times \left( \{t\} \times A_t^{b,\varepsilon,I} \right) \right) = \mu_t^n \left( \{t\} \times A_t^{b,\varepsilon,I} \right) \wedge \mu_t^k \left( \{t\} \times A_t^{b,\varepsilon,I} \right)$$

for all  $I \subset [b]$ , we obtain for  $n, k \geq n_0$  and  $t \in [0, \tau]$  from the coupling characterization of the Prohorov metric that

$$\begin{aligned} & d_{\mathbb{P}}^Z(\mu_t^n, \mu_t^k) \\ & \leq \nu \{ (z, z') \in Z^2 : \rho(z, z') > 2\varepsilon + 4t \} + 2\varepsilon + 4t \\ & \leq \nu \left( Z \times \left( \{t\} \times A_t^{b,\varepsilon,\emptyset} \right) \right) + \sum_{\substack{I \subset [b]: \\ I \neq \emptyset}} \nu \left( \left( Z \setminus \left( \{t\} \times A_t^{b,\varepsilon,I} \right) \right) \times \left( \{t\} \times A_t^{b,\varepsilon,I} \right) \right) + 2\varepsilon + 4t \\ & \leq \mu_t^k \left( \{t\} \times A_t^{b,\varepsilon,\emptyset} \right) + \sum_{\substack{I \subset [b]: \\ I \neq \emptyset}} \left| \mu_t^n \left( \{t\} \times A_t^{b,\varepsilon,I} \right) - \mu_t^k \left( \{t\} \times A_t^{b,\varepsilon,I} \right) \right| + 2\varepsilon + 4t \\ & \leq 7\varepsilon + \sum_{\substack{I \subset [b]: \\ I \neq \emptyset}} \left| |A_t^{b,\varepsilon,I}|_n - |A_t^{b,\varepsilon,I}|_k \right|. \end{aligned}$$

Lemma 9.5 implies  $\lim_{n,k \rightarrow \infty} \sup_{t \in [0, \tau]} d_{\mathbb{P}}^Z(\mu_t^n, \mu_t^k) = 7\varepsilon$ . Altogether,

$$\lim_{n,k \rightarrow \infty} \sup_{t \in [0, T]} d_{\mathbb{P}}^Z(\mu_t^n, \mu_t^k) = 8\varepsilon.$$

Assertion (i) follows as  $\varepsilon$  can be chosen arbitrarily small, and as  $(Z, \rho)$  is complete.

Now we come to assertion (ii). By condition (2.1), the map  $\mathbb{R}_+ \rightarrow Z$ ,  $t \mapsto (t, i)$  is càdlàg for each  $i \in \mathbb{N}$ . Hence, the map  $t \mapsto \mu_t^n$  is càdlàg for each  $n \in \mathbb{N}$ , and by item (i)

also the map  $t \mapsto \mu_t$  is càdlàg. For  $n \in \mathbb{N}$  and  $t \in (0, \infty)$ , we define the probability measure

$$\mu_{t-}^n = \frac{1}{n} \sum_{i=1}^n \delta_{(t-, i)}$$

on  $Z$ , with the left limit  $(t-, i) = \lim_{s \uparrow t} (s, i)$ . Then  $\mu_{t-}^n = \text{w-}\lim_{s \uparrow t} \mu_s^n$ . Let  $\mu_{t-} = \text{w-}\lim_{s \uparrow t} \mu_s$ . From the uniform convergence in item (i), it follows that  $\mu_{t-}^n$  converges weakly to  $\mu_{t-}$  as  $n \rightarrow \infty$ . Note that  $d_{\mathbb{P}}^Z(\mu_{t-}^n, \mu_t^n) \leq 1/n$  for all  $t \in (0, \infty) \setminus \Theta_0$  and all  $n \in \mathbb{N}$  a. s., as only a binary reproduction event can occur at such a time  $t$ . It follows that  $\mu_{t-} = \mu_t$  for all  $t \in (0, \infty) \setminus \Theta_0$  a. s. That the set of jump times is not smaller than  $\Theta_0$  follows from assertion (iii) which we now prove.

W.l.o.g., we assume  $\Xi_0(\Delta) > 0$ , then we have  $\Theta_0 = \{t^k : k \in \mathbb{N}\}$  a. s. (For  $\Xi_0(\Delta) = 0$ , nothing remains to prove as  $\Theta_0 = \emptyset$  a. s. in this case.) Condition (2.1) implies that  $\rho_{t^k-}(i, j) > 0$  for all  $k, i, j \in \mathbb{N}$  a. s. For each  $k \in \mathbb{N}$ , the random variable  $\rho_{t^k-}$  is exchangeable by Corollary 6.5. On an event of probability 1, let  $\chi$  be the isomorphy class of the metric measure space  $(Z, \rho, \mu_{t^k-})$ . As in Remark 4.4, we have  $\chi = \psi(\rho_{t^k-})$ . Let  $\rho'$  be a random variable whose conditional distribution given  $\chi$  is the distance matrix distribution of  $\chi$ . Then by [22, Remark 10.4], the random variables  $\rho'$  and  $\rho_{t^k-}$  are (unconditionally) equal in distribution. Hence,  $\rho'(i, j) > 0$  a. s. for all  $i, j \in \mathbb{N}$  which implies that  $\mu_{t^k-}$  is a. s. non-atomic.

By the definition of the population model in Section 2 and the definition (2.7) of  $\Theta_0$ , there exists for each  $k \in \mathbb{N}$  an  $i \in \mathbb{N}$  such that  $|\{j \in \mathbb{N} : \rho_{t^k}(i, j) = 0\}| = |B(\pi^k, i)| > 0$ . It can now be shown as above that  $\mu_{t^k}$  contains an atom. More simply, the Portmanteau theorem and item (i) imply

$$\mu_{t^k} \{(t, i)\} \geq \limsup_{n \rightarrow \infty} \mu_{t^k}^n \{(t, i)\} = |\{j \in \mathbb{N} : \rho_{t^k}(i, j) = 0\}| > 0. \quad \square$$

*Proof of Theorem 3.5.* Recall that  $\text{supp } \mu_t$  denotes the closed support of  $\mu_t$ . For  $z \in Z$  and  $\varepsilon > 0$ , we denote  $\varepsilon$ -balls in  $Z$  by  $\mathcal{U}_\varepsilon^Z = \{z' \in Z : \rho(z', z) < \varepsilon\}$ . and  $\mathcal{B}_\varepsilon^Z = \{z' \in Z : \rho(z', z) \leq \varepsilon\}$ . Up to null events,

$$\begin{aligned} & \{\text{supp } \mu_t \neq X_t \text{ for some } t \in (0, \infty)\} \\ & \subset \bigcup_{i \in \mathbb{N}} \left\{ \mu_t \left( \mathcal{U}_{2(t-s)}^Z(t, i) \right) = 0 \text{ for some } t \in (0, \infty), s \in (0, t) \cap \mathbb{Q} \right\} \\ & = \bigcup_{i \in \mathbb{N}} \left\{ |B(\Pi_{s,t}, i)| = 0 \text{ for some } t \in (0, \infty), s \in (0, t) \cap \mathbb{Q} \right\} \end{aligned}$$

and we have a null event in the last line by Lemma 8.8. By Lemma 3.2,  $X_0$  is the closed support of  $\mu_0$  a. s. This shows assertion (i).

For  $t \in \mathbb{R}_+$  and  $n \in \mathbb{N}$ , we denote the subset  $\{t\} \times [n]$  of  $Z$  by  $X_t^n$ , and more generally, for  $M \subset \mathbb{N}$ , we denote the subset  $\{t\} \times M$  of  $Z$  by  $X_t^M$ . We mostly omit ‘a. s.’ in the following.

In the proof of assertion (ii), we begin with right continuity. Let  $t \in \mathbb{R}_+$  and  $\varepsilon > 0$ . By construction,  $X_u \subset \mathcal{B}_{u-t}^Z(X_t)$  for all  $u \geq t$ . As  $X_t$  is compact and as  $\{t\} \times \mathbb{N}$  is dense in  $X_t$ , there exists  $n \in \mathbb{N}$  such that

$$X_t \subset \mathcal{B}_\varepsilon^Z(X_t^n).$$

By condition (2.1), we may choose  $\delta \in (0, \varepsilon)$  sufficiently small such that  $\eta((t, t + \delta) \times \mathcal{P}^n) = 0$ . Then,

$$\mathcal{B}_\varepsilon^Z(X_t^n) \subset \mathcal{B}_{2\varepsilon}^Z(X_u^n) \subset \mathcal{B}_{2\varepsilon}^Z(X_u)$$

for all  $u \in [t, t + \delta)$ . Thus, we obtain the bound

$$d_{\mathbb{H}}^Z(X_t, X_u) \leq 2\varepsilon$$

for the Hausdorff distance  $d_{\mathbb{H}}^Z$  over  $Z$ . This proves right continuity of the map  $t \mapsto X_t$  in  $d_{\mathbb{H}}^Z$ .

We now turn to the left limits. We write  $A_s(t, \mathbb{N}) = \{A_s(t, i) : i \in \mathbb{N}\}$ . Recall from Section 2.2 the level  $D_t(s, i)$  of the descendant of an individual  $(s, i)$  at time  $t \geq s$ . Let  $t \in (0, \infty)$  and  $\varepsilon \in (0, t)$ . For  $s \in [0, t)$ , we denote the closure of  $(s, t) \times \mathbb{N}$  in  $Z$  by  $X_{s,t}$ . We define the closed subset

$$X_{t-} = \bigcap_{s \in (0,t)} X_{s,t}$$

of  $Z$ . By construction,  $X_{t-} \supset X_t$ . We claim that  $d_{\mathbb{H}}^Z(X_s, X_{t-}) \rightarrow 0$  as  $s \uparrow t$ . From the definition of  $X_{t-}$ , we have  $X_{t-} \subset \mathcal{B}_{t-s}^Z(X_s)$  for all  $s \in (0, t)$ . Let  $M = \bigcap_{s \in [t-\varepsilon, t)} A_{t-\varepsilon}(s, \mathbb{N})$ . As  $\Xi \in \mathcal{M}_{\text{CDL}}$ , it holds  $\#M < \infty$ . That is, at time  $t-$ , the number of families of individuals that descend from the same ancestor at time  $t - \varepsilon$  is finite. As the map  $(t - \varepsilon/2, \infty) \rightarrow 2^{\mathbb{N}}, s \mapsto A_{t-\varepsilon}(s, \mathbb{N})$  is non-increasing and by condition (5.2) piecewise constant, there exists  $\delta \in (0, \varepsilon)$  such that  $A_{t-\varepsilon}(s, \mathbb{N}) = M$  for all  $s \in (t - \delta, t)$ . For all  $i \in M$  and  $s, s' \in (t - \varepsilon, t)$ , the definitions of  $M$  and  $D_s(t - \varepsilon, i)$  yield  $D_s(t - \varepsilon, i) < \infty$  and  $\rho((s, D_s(t - \varepsilon, i)), (s', D_{s'}(t - \varepsilon, i))) = |s - s'|$ , hence the limit  $x := \lim_{s \uparrow t} (s, D_s(t - \varepsilon, i))$  exists in the complete subspace  $X_{s',t}$ . Also note that  $\rho(x, (t - \varepsilon, i)) = \varepsilon$  and  $x \in X_{t-}$ . Thus,  $X_{t-\varepsilon}^M \subset \mathcal{B}_{\varepsilon}^Z(X_{t-})$ ,

$$X_s \subset \mathcal{B}_{\varepsilon}^Z(X_{t-\varepsilon}^M) \subset \mathcal{B}_{2\varepsilon}^Z(X_{t-}),$$

and  $d_{\mathbb{H}}^Z(X_s, X_{t-}) \leq 2\varepsilon$  for all  $s \in (t - \delta, t)$ .

Now we show  $X_t = X_{t-}$  for  $t \in (0, \infty) \setminus \Theta^{\text{ext}}$ . Let  $x \in X_{t-}$ . Then there exists a sequence  $((s_k, i_k) : k \in \mathbb{N})$  in  $(0, t) \times \mathbb{N} \subset Z$  with  $0 < s_1 < s_2 < \dots$  that converges to  $x$ . For each  $k \in \mathbb{N}$ , there exists  $\ell \in \mathbb{N}$  such that  $\rho((s_n, i_n), (s_\ell, i_\ell)) < 2(s_\ell - s_k)$  for all  $n \geq \ell$ . This implies  $A_{s_k}(s_n, i_n) = A_{s_k}(s_\ell, i_\ell)$  for all  $n \geq \ell$ . We set  $j_k = A_{s_k}(s_\ell, i_\ell)$ . Then  $j_k = A_{s_k}(s_n, j_n)$  for all  $n \geq \ell \in \mathbb{N}$ , hence  $D_s(s_k, j_k) < \infty$  for all  $s \in [s_k, t)$ . By our assumption on  $t$ , the definition (2.6) of  $\Theta^{\text{ext}}$ , and condition (5.2), this implies  $j'_k := D_t(s_k, j_k) < \infty$  for all  $k \in \mathbb{N}$ . The sequence  $((s_k, j_k) : k \in \mathbb{N})$  converges to  $x$  as

$$\rho((s_k, j_k), x) = \lim_{n \rightarrow \infty} \rho((s_k, j_k), (s_n, i_n)) \leq \lim_{n \rightarrow \infty} \rho((s_k, i_k), (s_n, i_n)) = \rho((s_k, i_k), x)$$

for all  $k \in \mathbb{N}$ . Also the sequence  $((t, j'_k) : k \in \mathbb{N})$  converges to  $x$  as

$$\rho((t, j'_k), (s_k, j_k)) = t - s_k$$

for all  $k \in \mathbb{N}$ . This implies  $x \in X_t$ .

Now let  $t \in \Theta^{\text{ext}}$ . Then by (2.6), there exists  $\varepsilon \in (0, t)$  and  $i \in \mathbb{N}$  with  $D_t(t - \varepsilon, i) = \infty$  and  $D_s(t - \varepsilon, i) < \infty$  for all  $s \in [t - \varepsilon, t)$ . That is, the descendants of some ancestor at time  $t - \varepsilon$  die out at time  $t$ . The space  $(X_t, \rho \wedge t)$  is ultrametric, cf. [22, Remark 5.2] and equation (2.3). Hence, a semi-metric  $\rho^{(\varepsilon)}$  on  $X_t$  is given by  $\rho^{(\varepsilon)} = (\rho \wedge t - \varepsilon) \vee 0$ . We denote by  $X_t^{(\varepsilon)}$  the space obtained from  $X_t$  by identifying elements with  $\rho^{(\varepsilon)}$ -distance 0. As  $X_t$  is compact, the space  $X_t^{(\varepsilon)}$  is finite. Also the space  $(X_{t-}, \rho \wedge t)$  is ultrametric. Indeed, for each  $x, y, z \in X_{t-}$  and each  $s \in (0, t)$ , there exist  $x', y', z' \in X_s$  with  $\rho(x, x') \leq t - s$ ,  $\rho(y, y') \leq t - s$ , and  $\rho(z, z') \leq t - s$ . This implies

$$\begin{aligned} (\rho \wedge t)(x, z) &\leq (\rho \wedge s)(x', z') + 2(t - s) \\ &\leq \max\{(\rho \wedge s)(x', y'), (\rho \wedge s)(y', z')\} + 2(t - s) \\ &\leq \max\{(\rho \wedge t)(x, y), (\rho \wedge t)(y, z)\} + 4(t - s). \end{aligned}$$

Hence, a semi-metric  $\rho^{(\varepsilon)}$  on  $X_{t-}$  is given by  $\rho^{(\varepsilon)} = (\rho \wedge t - \varepsilon) \vee 0$ . We denote by  $X_{t-}^{(\varepsilon)}$  the space obtained from  $X_{t-}$  by identifying elements with  $\rho^{(\varepsilon)}$ -distance 0. With  $i$  as above, the limit  $x := \lim_{s \uparrow t} (s, D_s(t - \varepsilon, i))$  exists in  $X_{t-}$ . As  $A_{t-\varepsilon}(t, j) \neq i$  for all  $j \in \mathbb{N}$ , it follows

that  $\rho((s, A_s(t, j)), (s, D_s(t - \varepsilon, i))) \geq 2(s - t + \varepsilon)$  for all  $s \in (t - \varepsilon, t)$ . Taking the limit  $s \uparrow t$ , we obtain that  $\rho((t, j), x) \geq 2\varepsilon$ . As  $\{t\} \times \mathbb{N}$  is dense in  $X_t$ , it also follows that  $\rho(x', x) \geq 2\varepsilon$  for all  $x' \in X_t$ . As  $X_t \subset X_{t-}$ , this implies that the cardinality of  $X_{t-}^{(\varepsilon)}$  is greater than the cardinality of  $X_t^{(\varepsilon)}$ . Hence,  $X_{t-}^{(\varepsilon)}$  and  $X_t^{(\varepsilon)}$  are not isometric. It follows that  $X_{t-}$  and  $X_t$  are not isometric. As a consequence,  $d_{\mathbb{H}}^Z(X_{t-}, X_t) > 0$ , that is, the map  $t \mapsto X_t$  is discontinuous in  $\Theta^{\text{ext}}$ .  $\square$

### 10 Proof of Lemmas 7.1 and 7.2

We work in the context of Section 7, using also the definitions from Section 5. The proofs in the present section are adaptations of the proofs of Lemmas 3.4 and 3.5 of Donnelly and Kurtz [13] and of Lemma 3.2 of Birkner et al. [4]. We also mention Lemma 6.2 of Labbé [29].

We define stochastic processes  $(U(t), t \in \mathbb{R}_+)$  and  $(\hat{U}(t), t \in \mathbb{R}_+)$  by

$$U(t) = \int_{(0,t] \times \Delta} |x|_2^2 \zeta_0(ds dx)$$

and

$$\hat{U}(t) = \int_{(0,t] \times \Delta} |x|_1 \zeta_0(ds dx).$$

For  $t \in \mathbb{R}_+$ , the random variable  $U(t)$  equals the sum of the squared asymptotic frequencies of the blocks that encode the reproduction events up to time  $t$ . The random variable  $\hat{U}(t)$  equals the sum of the asymptotic frequencies of these blocks. By the properties of the Poisson random measure  $\zeta_0$ ,

$$\mathbb{E}[U(t)] = t \int_{\Delta} |x|_2^2 |x|_2^{-2} \Xi_0(dx) < \infty,$$

hence the random variable  $U(t)$  is a. s. finite. In case  $\Xi \in \mathcal{M}_{\text{dust}}$ , we also have that

$$\mathbb{E}[\hat{U}(t)] = t \int_{\Delta} |x|_1 |x|_2^{-2} \Xi_0(dx) < \infty,$$

and  $\hat{U}(t)$  is a. s. finite.

*Proof of Lemma 7.1.* We assume w. l. o. g.  $\Xi(\Delta) > 0$ . Let  $T, \varepsilon, c > 0$ . For  $\ell \in \mathbb{N}$ , we set  $\alpha_0^\ell = 0$  and inductively for  $k \in \mathbb{N}_0$

$$\alpha_{k+1}^\ell = \inf \{t > \alpha_k^\ell : U(t) > U(\alpha_k^\ell) + \ell^{-4}\} \wedge (\alpha_k^\ell + \ell^{-4}).$$

We also set  $k_\ell = 2\lceil(c + T)\ell^4\rceil$ , then we have that  $\mathbb{P}(\alpha_{k_\ell}^\ell \leq T, U(T) \leq c) = 0$ .

For an arbitrary integer  $n_\ell \geq \ell$ , we define for  $k \in \mathbb{N}_0$

$$\beta_k^\ell = \inf \{t > \alpha_k^\ell : \eta((\alpha_k^\ell, t] \times \mathcal{P}^b) > 0\},$$

$$\tilde{\alpha}_k^\ell = \inf \{t > \alpha_k^\ell : |X_{n_\ell}(t) - X_{n_\ell}(\alpha_k^\ell)| \geq 4\varepsilon\} \wedge (\alpha_k^\ell + 1),$$

and

$$\tilde{\beta}_k^\ell = \inf \{t > \beta_k^\ell : |X_{n_\ell}(t) - X_{n_\ell}(\beta_k^\ell)| \geq 4\varepsilon\} \wedge (\beta_k^\ell + 1).$$

Recall from Section 5.2 the point measure  $\eta_0$ . Each point  $(t, \pi)$  of  $\eta_0$  stands for a large (i. e. non-Kingman) reproduction event at time  $t$  that is governed by the partition  $\pi$ . If  $\Xi_0(\Delta) = 0$ , we have  $\eta_0 = 0$  a. s.

We define the event  $E^\ell$  as the intersection of the two events  $E_1^\ell$  and  $E_2^\ell$  that are defined as follows: We set

$$E_1^\ell = \bigcap_{k \in \mathbb{N}_0} \{ \eta((\alpha_k^\ell \wedge (T+1), \alpha_{k+1}^\ell \wedge (T+1)) \times \mathcal{P}^b) \leq 1 \},$$

where we use the notation  $(c, c') = \emptyset$  for  $c' \leq c$ . We define

$$E_2^\ell = \left\{ |\pi|_2^2 > \ell^{-4} \text{ for all } \pi \in \mathcal{P}^b \text{ with } \eta_0((0, T+1] \times \{\pi\}) > 0 \right\},$$

where  $|\pi|_2^2 = \sum_{B \in \pi} |B|^2$ . Note that a.s., the asymptotic frequencies of the blocks  $B$  of the partitions  $\pi$  that form part of the points of  $\eta_0$  exist. This follows by Kingman's correspondence from the definition (5.1) of  $\eta_0$  and as  $\eta_0$  has a.s. at most countably many points.

We claim that

$$\mathbb{P} \left( \bigcup_{\ell_0 \in \mathbb{N}} \bigcap_{\ell \geq \ell_0} E^\ell \right) = 1. \tag{10.1}$$

This claim means that the complement  $(E^\ell)^c$  occurs for a.s. at most finitely many  $\ell$ . As  $\alpha_{k+1}^\ell \leq \alpha_k^\ell + \ell^{-4}$ , the event  $E_1^\ell$  occurs if  $\ell^{-4}$  goes below the minimal distance between points of  $\eta(\cdot \times \mathcal{P}^b)$  in  $(0, T+1]$ . This minimal distance is positive on the event of probability 1 on which  $\eta((0, T+1] \times \mathcal{P}^b) < \infty$  holds by condition (2.1). Hence,  $E_1^\ell$  occurs for all but a.s. at most finitely many  $\ell$ . Moreover, all points  $(t, \pi)$  of  $\eta_0$  satisfy  $|\pi|_2^2 > 0$  a.s. by the definition of  $\eta_0$  from  $\zeta_0$  in Section 5.2 and as  $\Xi_0\{0\} = 0$  implies that for all points  $(t, y)$  of  $\zeta_0$ , the first component  $y_1$  of  $y$  is positive a.s. Hence, also  $E_2^\ell$  occurs for all but a.s. at most finitely many  $\ell$ , which yields the claim (10.1).

Recall the strong Markov property of the process  $J$  from (3.1). For each  $k \in \mathbb{N}_0$ , the sequence  $(Y_i^b(\alpha_k^\ell), i \in \mathbb{N})$  is exchangeable by Corollary 6.5 as  $\alpha_k^\ell$  is  $\zeta_0$ -measurable. The distance matrix

$$\mathbb{1} \{ \eta((\alpha_k^\ell, \tilde{\alpha}_k^\ell] \times \mathcal{P}^b) = 0 \} \gamma_{b+1+n_\ell} \left( \rho_{\tilde{\alpha}_k^\ell} \right)$$

is  $(b+1+n_\ell, b)$ -exchangeable by Lemma 6.8 and the strong Markov property of  $J$  at  $\alpha_k^\ell$ . Hence, the vector

$$\left( \mathbb{1} \{ \eta((\alpha_k^\ell, \tilde{\alpha}_k^\ell] \times \mathcal{P}^b) = 0 \} Y_i^b(\tilde{\alpha}_k^\ell), i \in [n_\ell] \right)$$

is exchangeable.

If  $b \geq 2$  and  $\Xi\{0\} > 0$ , then the distance matrix

$$\mathbb{1} \{ \zeta_0(\{\beta_k^\ell\} \times \Delta) = 0 \} \gamma_{b+1+n_\ell} \left( \rho_{\beta_k^\ell} \right)$$

is  $(b+1+n_\ell, b+1)$ -exchangeable by Lemma 6.6 and the strong Markov property of  $J$  at  $\alpha_k^\ell$ . Hence, the vector

$$\left( \mathbb{1} \{ \zeta_0(\{\beta_k^\ell\} \times \Delta) = 0 \} Y_i^b(\beta_k^\ell), i \in [n_\ell] \right)$$

is exchangeable. The distance matrix

$$\mathbb{1} \left\{ \zeta_0(\{\beta_k^\ell\} \times \Delta) = 0, \eta\left(\left(\beta_k^\ell, \tilde{\beta}_k^\ell\right] \times \mathcal{P}^b\right) = 0 \right\} \gamma_{b+1+n_\ell} \left( \rho_{\tilde{\beta}_k^\ell} \right)$$

is  $(b+1+n_\ell, b+1)$ -exchangeable by Lemma 6.8 and the strong Markov property of  $J$  at  $\beta_k^\ell$ . Hence, the vector

$$\left( \mathbb{1} \left\{ \zeta_0(\{\beta_k^\ell\} \times \Delta) = 0, \eta\left(\left(\beta_k^\ell, \tilde{\beta}_k^\ell\right] \times \mathcal{P}^b\right) = 0 \right\} Y_i^b(\tilde{\beta}_k^\ell), i \in [n_\ell] \right)$$

is exchangeable. If  $b < 2$  or  $\Xi\{0\} = 0$ , then it suffices to work with the stopping times  $\alpha_k^\ell$  and  $\tilde{\alpha}_k^\ell$ .

By Lemma A.2 in [13], there exists a number  $\eta_\varepsilon$  that depends only on  $\varepsilon$  (not on  $n_\ell$ ) such that

$$\begin{aligned} \mathbb{P}(|X_{n_\ell}(\alpha_k^\ell) - X_\ell(\alpha_k^\ell)| \geq \varepsilon) &\leq 2e^{-\eta_\varepsilon \ell}, \\ \mathbb{P}(|X_{n_\ell}(\tilde{\alpha}_k^\ell) - X_\ell(\tilde{\alpha}_k^\ell)| \geq \varepsilon, \eta((\alpha_k^\ell, \tilde{\alpha}_k^\ell] \times \mathcal{P}^b) = 0) &\leq 2e^{-\eta_\varepsilon \ell}, \\ \mathbb{P}(|X_{n_\ell}(\beta_k^\ell) - X_\ell(\beta_k^\ell)| \geq \varepsilon, \zeta_0(\{\beta_k^\ell\} \times \Delta) = 0) &\leq 2e^{-\eta_\varepsilon \ell}, \end{aligned}$$

and  $\mathbb{P}(|X_{n_\ell}(\tilde{\beta}_k^\ell) - X_\ell(\tilde{\beta}_k^\ell)| \geq \varepsilon, \zeta_0(\{\beta_k^\ell\} \times \Delta) = 0, \eta((\beta_k^\ell, \tilde{\beta}_k^\ell] \times \mathcal{P}^b) = 0) \leq 2e^{-\eta_\varepsilon \ell}.$

Let

$$\begin{aligned} H_k = &|X_{n_\ell}(\alpha_k^\ell) - X_\ell(\alpha_k^\ell)| \vee |X_{n_\ell}(\tilde{\alpha}_k^\ell \wedge \beta_k^\ell \wedge \alpha_{k+1}^\ell) - X_\ell(\tilde{\alpha}_k^\ell \wedge \beta_k^\ell \wedge \alpha_{k+1}^\ell)| \\ &\vee |X_{n_\ell}(\beta_k^\ell \wedge \alpha_{k+1}^\ell) - X_\ell(\beta_k^\ell \wedge \alpha_{k+1}^\ell)| \vee |X_{n_\ell}(\tilde{\beta}_k^\ell \wedge \alpha_{k+1}^\ell) - X_\ell(\tilde{\beta}_k^\ell \wedge \alpha_{k+1}^\ell)|. \end{aligned}$$

As

$$\begin{aligned} \{\beta_k^\ell < \alpha_{k+1}^\ell\} \cap \{\alpha_k^\ell \leq T\} \cap E^\ell &\subset \{\zeta_0(\{\beta_k^\ell\} \times \Delta) = 0\}, \\ \{\tilde{\beta}_k^\ell < \alpha_{k+1}^\ell\} \cap \{\alpha_k^\ell \leq T\} \cap E^\ell &\subset \left\{ \eta\left(\left(\beta_k^\ell, \tilde{\beta}_k^\ell\right] \times \mathcal{P}^b\right) = 0 \right\}, \end{aligned}$$

and

$$\{\tilde{\alpha}_k^\ell < \beta_k^\ell\} \subset \{\eta((\alpha_k^\ell, \tilde{\alpha}_k^\ell] \times \mathcal{P}^b) = 0\}$$

up to null events for all  $k \in \mathbb{N}_0$ , the above implies

$$\mathbb{P}\left(\max_{k < k_\ell: \alpha_k^\ell \leq T} H_k \geq \varepsilon, E^\ell\right) \leq 16\lceil(c+T)\ell^4\rceil e^{-\eta_\varepsilon \ell}.$$

For  $k \in \mathbb{N}_0$ , we have that  $\tilde{\alpha}_k^\ell \geq \beta_k^\ell \wedge \alpha_{k+1}^\ell$  a. s. on the event

$$\{H_k < \varepsilon\} \cap \left\{ \sup_{t \in [\alpha_k^\ell, \beta_k^\ell \wedge \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| < \varepsilon \right\}.$$

Indeed, the intersection of this event with  $\{\tilde{\alpha}_k^\ell < \beta_k^\ell \wedge \alpha_{k+1}^\ell\}$  is a null event as it holds on this event that

$$|X_{n_\ell}(\tilde{\alpha}_k^\ell) - X_{n_\ell}(\alpha_k^\ell)| < 3\varepsilon,$$

whereas we have

$$|X_{n_\ell}(\tilde{\alpha}_k^\ell) - X_{n_\ell}(\alpha_k^\ell)| \geq 4\varepsilon \quad \text{a. s.}$$

by definition of  $\tilde{\alpha}_k^\ell$  and right continuity.

Similarly,  $\tilde{\beta}_k^\ell \geq \alpha_{k+1}^\ell$  a. s. on the event

$$\{H_k < \varepsilon\} \cap \left\{ \sup_{t \in [\beta_k^\ell, \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\beta_k^\ell)| < \varepsilon \right\}.$$

By Lemma 10.1 below and the Markov inequality,

$$\begin{aligned} &\sum_{\ell \in \mathbb{N}} k_\ell \mathbb{P}(N^{b+1+\ell}(0, \alpha_1^\ell) > \ell\varepsilon) \\ &\leq \sum_{\ell \geq b+1} 2\lceil(c+T)\ell^4\rceil \frac{1}{(\ell\varepsilon)^4} \mathbb{E}\left[(N^{2\ell}(0, \alpha_1^\ell))^4\right] + \sum_{\ell=1}^b k_\ell < \infty. \end{aligned} \tag{10.2}$$

Using the strong Markov property of  $J$  at  $\alpha_k^\ell$  and the assumption (7.1) on  $f$ , we obtain

$$\begin{aligned} & \mathbb{P} \left( \max_{k < k_\ell} \sup_{t \in [\alpha_k^\ell, \beta_k^\ell \wedge \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| > \varepsilon \right) \\ & \leq \sum_{k < k_\ell} \mathbb{P} \left( \sup_{t \in [\alpha_k^\ell, \beta_k^\ell \wedge \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| > \varepsilon \right) \\ & = k_\ell \mathbb{P} \left( \sup_{t \in [0, \beta_0^\ell \wedge \alpha_1^\ell]} |X_\ell(t) - X_\ell(\alpha_1^\ell)| > \varepsilon \right) \\ & \leq k_\ell \mathbb{P} (N^{b+1+\ell} (0, \beta_0^\ell \wedge \alpha_1^\ell) > \ell\varepsilon) \\ & \leq k_\ell \mathbb{P} (N^{b+1+\ell} (0, \alpha_1^\ell) > \ell\varepsilon) \end{aligned} \tag{10.3}$$

for all  $\ell \in \mathbb{N}$ . By (10.2) and (10.3),

$$\sum_{\ell \in \mathbb{N}} \mathbb{P} \left( \max_{k < k_\ell} \sup_{t \in [\alpha_k^\ell, \beta_k^\ell \wedge \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| > \varepsilon \right) < \infty.$$

After replacing  $[\alpha_k^\ell, \beta_k^\ell \wedge \alpha_{k+1}^\ell]$  with  $[\beta_k^\ell \wedge (T+1), \alpha_{k+1}^\ell \wedge (T+1)]$  and intersecting with the event  $E^\ell$  in (10.3), the calculation from (10.2) and (10.3) also yields

$$\begin{aligned} & \mathbb{P} \left( \max_{k < k_\ell} \sup_{t \in [\beta_k^\ell \wedge (T+1), \alpha_{k+1}^\ell \wedge (T+1)]} |X_\ell(t) - X_\ell(\beta_k^\ell)| > \varepsilon, E^\ell \right) \\ & \leq k_\ell \mathbb{P} (N^{b+1+\ell} (0, \alpha_1^\ell) > \ell\varepsilon). \end{aligned}$$

Indeed, we can use assumption (7.1) in the same way as in the third line of (10.3) because  $\eta((\beta_0^\ell \wedge (T+1), \alpha_1^\ell \wedge (T+1)) \times \mathcal{P}^b) = 0$  on the event  $E^\ell$ . Again we deduce

$$\sum_{\ell \in \mathbb{N}} \mathbb{P} \left( \max_{k < k_\ell} \sup_{t \in [\beta_k^\ell \wedge (T+1), \alpha_{k+1}^\ell \wedge (T+1)]} |X_\ell(t) - X_\ell(\beta_k^\ell)| > \varepsilon, E^\ell \right) < \infty.$$

Altogether, it follows that there exist  $\delta_\ell$  which do not depend on  $n_\ell$  such that  $\sum_{\ell=1}^\infty \delta_\ell < \infty$  and

$$\mathbb{P} \left( \sup_{t \in [0, T]} |X_{n_\ell}(t) - X_\ell(t)| > 4\varepsilon, U(T) \leq c, E^\ell \right) < \delta_\ell$$

for all  $\ell \in \mathbb{N}$ .

By Corollary 6.5 and the de Finetti Theorem, there exists an event of probability 1 on which the limits  $X(t) = \lim_{n \rightarrow \infty} X_n(t)$  exist for all  $t \in \mathbb{Q}_+$ . Hence,

$$\begin{aligned} & \mathbb{P} \left( \sup_{t \in [0, T] \cap \mathbb{Q}} |X(t) - X_\ell(t)| > 4\varepsilon, U(T) \leq c, E^\ell \right) \\ & = \mathbb{P} \left( \sup_{t \in [0, T] \cap \mathbb{Q}} \liminf_{n \rightarrow \infty} |X_n(t) - X_\ell(t)| > 4\varepsilon, U(T) \leq c, E^\ell \right) \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P} \left( \inf_{j \geq n} \sup_{t \in [0, T] \cap \mathbb{Q}} |X_j(t) - X_\ell(t)| > 4\varepsilon, U(T) \leq c, E^\ell \right) \leq \delta_\ell. \end{aligned}$$

The Borel-Cantelli lemma allows to deduce that

$$\mathbb{P} \left( \bigcup_{\ell_0 \in \mathbb{N}} \bigcap_{\ell \geq \ell_0} \left( \left\{ \sup_{t \in [0, T] \cap \mathbb{Q}} |X(t) - X_\ell(t)| \leq 4\varepsilon \right\} \cup \{U(T) > c\} \cup (E^\ell)^c \right) \right) = 1. \quad (10.4)$$

By (10.1), the event  $(E^\ell)^c$  occurs for a. s. at most finitely many  $\ell$ . Hence, (10.4) implies that there exists a random integer  $L < \infty$  such that a. s. on the event  $\{U(T) \leq c\}$ , we have

$$\sup_{t \in [0, T] \cap \mathbb{Q}} |X(t) - X_\ell(t)| \leq 4\varepsilon$$

for all  $\ell \geq L$ . As a consequence, we obtain a. s. on the event  $\{U(T) \leq c\}$  that

$$|X_n(t) - X_\ell(t)| \leq |X_n(t) - X(t)| + |X_\ell(t) - X(t)| \leq 8\varepsilon$$

for all  $t \in [0, T]$  and  $n \geq \ell \geq L$ . Here we used right continuity of  $X_n$  and  $X_\ell$ . It follows that

$$\mathbb{P} \left( \lim_{n, \ell \rightarrow \infty} \sup_{t \in [0, T]} |X_n(t) - X_\ell(t)| = 0, U(T) \leq c \right) = 1.$$

The assertion follows by letting  $c$  tend to infinity. □

**Lemma 10.1.** *For  $\ell \in \mathbb{N}$ , let  $\alpha_1^\ell$  be defined as in the proof of Lemma 7.1. Then there exists a constant  $C$  such that  $\mathbb{E}[(N^{2\ell}(0, \alpha_1^\ell))^4] \leq C\ell^{-2}$  for all  $\ell \in \mathbb{N}$ .*

The proof extends the argument presented on p. 44 in [4] where additional assumptions on  $\Xi$  are required to ensure that the process used there instead of  $U(t)$  is finite.

*Proof.* First, let  $x \in \Delta$ . For  $\ell \in \mathbb{N}$ , let  $(X_1, X_2, \dots)$  have infinite multinomial distribution with parameters  $(\ell, x_1, x_2, \dots)$ , that is, we may consider iid random variables  $U_1, U_2, \dots$  with uniform distribution on  $[0, 1]$  and set

$$X_i = \# \left\{ j \in [\ell] : \sum_{k=1}^{i-1} x_k < U_j < \sum_{k=1}^i x_k \right\}$$

for  $i \in \mathbb{N}$ . The infinite multinomial distribution appears in the context of  $\Xi$ -coalescents e. g. in [32]. For a random partition  $\pi$  in  $\mathcal{P}$  with distribution  $\kappa(x, \cdot)$ , Kingman's correspondence implies that  $b_\ell(\pi)$  and  $\lim_{n \rightarrow \infty} \sum_{i=1}^n (X_i - 1)_+$  are equal in distribution. Here we write  $(x - 1)_+ = \max\{x - 1, 0\}$ . We use the inequalities  $[(x - 1)_+]^2 \leq x^{(2)}$  and  $[(x - 1)_+]^4 \leq 3x^{(4)} + 3x^{(2)}$  for  $x \in \mathbb{N}_0$ , where  $x^{(k)} = x!/(x - k)!$ . Inserting also mixed factorial moments of multinomial distributions, we obtain a constant  $C'$  such that

$$\mathbb{E} \left[ \left( \sum_{i=1}^n (X_i - 1)_+ \right)^2 \right] \leq \sum_{\substack{i, j=1 \\ i \neq j}}^n \mathbb{E} [X_i^{(2)} X_j^{(2)}] + \sum_{i=1}^n \mathbb{E} [X_i^{(2)}] \leq \ell^4 |x|_2^4 + \ell^2 |x|_2^2$$

and

$$\mathbb{E} \left[ \left( \sum_{i=1}^n (X_i - 1)_+ \right)^4 \right] \leq C' (\ell^8 |x|_2^8 + \ell^6 |x|_2^6 + \ell^6 |x|_4^4 |x|_2^2 + \ell^4 |x|_2^4 + \ell^4 |x|_4^4 + \ell^2 |x|_2^2)$$

for all  $n \in \mathbb{N}$ . Taking the limit  $n \rightarrow \infty$  on the left-hand side, we obtain upper bounds for  $\kappa(x, \cdot) b_\ell^2$  and  $\kappa(x, \cdot) b_\ell^4$ .

Let

$$N_0^\ell(I) = \int_{I \times \mathcal{P}} b_\ell(\pi) \eta_0(ds d\pi)$$

for each interval  $I \subset \mathbb{R}_+$ . The random variable  $N_0^\ell(I)$  is the number of newborn particles on the first  $\ell$  levels in the large reproduction events in the time interval  $I$ . The random variable  $N^{2\ell}(0, \alpha_1^\ell) - N_0^{2\ell}(0, \alpha_1^\ell) = \eta_{\mathbb{K}}((0, \alpha_1^\ell) \times \mathcal{P}^{2\ell})$  is stochastically bounded from above by a Poisson random variable with mean  $\Xi\{0\}\ell^{-4} \binom{2\ell}{2}$ .

Here we only show  $\mathbb{E}[(N_0^{2\ell}(0, \alpha_1^\ell)]^4] \leq C'''\ell^{-2}$  for an appropriate constant  $C'''$  and all  $\ell \in \mathbb{N}$ . W.l. o. g. we assume  $\Xi_0(\Delta) > 0$ . Recall  $((t^i, y^i, \pi^i), i \in \mathbb{N})$  from Section 5. We have

$$\begin{aligned} & \mathbb{E} [(N_0^{2\ell}(0, \alpha_1^\ell))^4 \mid \zeta_0] \\ = & \mathbb{E} \left[ \sum_{\substack{i_1, \dots, i_4 \in \mathbb{N}: \\ t^{i_1}, \dots, t^{i_4} < \alpha_1^\ell}} b_{2\ell}(\pi^{i_1}) \cdots b_{2\ell}(\pi^{i_4}) \mid ((t^i, y^i), i \in \mathbb{N}) \right] \\ \leq & C''' \left[ \left( \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} \kappa(y^i, \cdot) b_{2\ell} \right)^4 \right. \\ & + \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} \kappa(y^i, \cdot) b_{2\ell}^2 \left( \sum_{j \in \mathbb{N}: t^j < \alpha_1^\ell} \kappa(y^j, \cdot) b_{2\ell} \right)^2 \\ & + \left( \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} \kappa(y^i, \cdot) b_{2\ell} \right)^2 \\ & + \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} \kappa(y^i, \cdot) b_{2\ell}^3 \sum_{j \in \mathbb{N}: t^j < \alpha_1^\ell} \kappa(y^j, \cdot) b_{2\ell} \\ & \left. + \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} \kappa(y^i, \cdot) b_{2\ell}^4 \right] \text{ a. s.} \end{aligned}$$

for a combinatorial constant  $C'''$ . Now we estimate  $b_{2\ell} \leq b_{2\ell}^2$  and  $b_{2\ell}^3 \leq b_{2\ell}^4$ , and we insert the bounds for  $\kappa(y^i, \cdot) b_{2\ell}^2$  and  $\kappa(y^i, \cdot) b_{2\ell}^4$ . From the definitions of  $(U(t), t \in \mathbb{R}_+)$  and  $\alpha_1^\ell$ , we have

$$\sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} |y^i|_2^2 \leq \ell^{-4}$$

which yields the assertion. □

*Proof of Lemma 7.2.* The proof is analogous to the proof of Lemma 7.1. Again we fix  $T, c, \varepsilon > 0$  and assume  $\Xi(\Delta) > 0$ . We work only with the stopping times  $\alpha_k^\ell$  and  $\tilde{\alpha}_k^\ell$  which we define for  $\ell \in \mathbb{N}$ ,  $k \in \mathbb{N}_0$ , and some arbitrarily large integers  $n_\ell \geq \ell$  as follows:

$$\alpha_0^\ell = 0,$$

$$\tilde{\alpha}_k^\ell = \inf \{ t > \alpha_k^\ell : |X_{n_\ell}(t) - X_{n_\ell}(\alpha_k^\ell)| \geq 4\varepsilon \} \wedge (\alpha_k^\ell + 1),$$

and

$$\alpha_{k+1}^\ell = \inf \{ t > \alpha_k^\ell : \hat{U}(t) > \hat{U}(\alpha_k^\ell) + \ell^{-1} \}.$$

We set  $k_\ell = 2\lceil c\ell \rceil$ . Then we have that  $\mathbb{P}(\alpha_{k_\ell} \leq T, \hat{U}(T) \leq c) = 0$ . Let

$$E^\ell = \left\{ |\pi|_2^2 > \ell^{-1} \text{ for all } \pi \in \hat{\mathcal{P}}^b \text{ with } \eta_0((0, T + 1] \times \{\pi\}) > 0 \right\}.$$

A. s., there exists a random integer  $L < \infty$  such that  $E^\ell$  occurs for all  $\ell \geq L$ . This follows as  $\eta((0, T + 1] \times \hat{\mathcal{P}}^b) < \infty$  by (5.2) and as  $|\pi|_2^2 > 0$  for all points  $(t, \pi)$  of  $\eta_0$ . A. s. on  $E^\ell$ , no reproduction events that are characterized by a partition in  $\hat{\mathcal{P}}^b$  occur in the open time

intervals  $(\alpha_k^\ell, \alpha_{k+1}^\ell)$  as long as  $\alpha_{k+1}^\ell < T + 1$ , that is,  $\eta((\alpha_k^\ell, \alpha_{k+1}^\ell) \times \hat{\mathcal{P}}^b) = 0$  for such  $k$ . Here we use that  $\eta = \eta_0$  a.s. by our assumption that  $\Xi \in \mathcal{M}_{\text{dust}}$ .

We replace the distance matrices  $(\rho_t, t \in \mathbb{R}_+)$  from the proof of Lemma 7.1 by the marked distance matrices  $((r_t, v_t), t \in \mathbb{R}_+)$  to obtain exchangeability for the sequence  $(Y_i^b(\alpha_k^\ell), i \in \mathbb{N})$  and the vector

$$\left( \mathbb{1} \left\{ \eta \left( (\alpha_k^\ell, \tilde{\alpha}_k^\ell] \times \hat{\mathcal{P}}^b \right) = 0 \right\} Y_i^b \left( \tilde{\alpha}_k^\ell \right), i \in [n_\ell] \right).$$

We set

$$H_k = |X_{n_\ell}(\alpha_k^\ell) - X_\ell(\alpha_k^\ell)| \vee |X_{n_\ell}(\tilde{\alpha}_k^\ell \wedge \alpha_{k+1}^\ell) - X_\ell(\tilde{\alpha}_k^\ell \wedge \alpha_{k+1}^\ell)|$$

for  $k \in \mathbb{N}_0$ . Applying Lemma A.2 of [13] as in the proof of Lemma 7.1, and using that

$$\{\tilde{\alpha}_k^\ell < \alpha_{k+1}^\ell\} \cap \{\alpha_k^\ell \leq T\} \cap E^\ell \subset \left\{ \eta \left( (\alpha_k^\ell, \tilde{\alpha}_k^\ell] \times \hat{\mathcal{P}}^b \right) = 0 \right\}$$

for all  $k \in \mathbb{N}_0$ , we obtain

$$\mathbb{P} \left( \max_{k < k_\ell: \alpha_k^\ell \leq T} H_k \geq \varepsilon, E^\ell \right) \leq 8 \lceil c\ell \rceil e^{-\eta\varepsilon\ell}.$$

We claim that

$$\sum_{\ell \in \mathbb{N}} k_\ell \mathbb{P} \left( \hat{N}^{b+\ell}(0, \alpha_1^\ell) > \ell\varepsilon \right) < \infty \tag{10.5}$$

which we prove further below. From the strong Markov property of  $J$  at  $\alpha_k^\ell$  and by the assumption (7.2) on  $f$ , we obtain

$$\begin{aligned} & \mathbb{P} \left( \max_{k < k_\ell} \sup_{t \in [\alpha_k^\ell, \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| > \varepsilon, E^\ell \right) \\ & \leq k_\ell \mathbb{P} \left( \hat{N}^{b+\ell}(0, \alpha_1^\ell) > \ell\varepsilon, E^\ell \right) \end{aligned}$$

for all  $\ell \in \mathbb{N}$ , and the claim (10.5) implies

$$\sum_{\ell \in \mathbb{N}} \mathbb{P} \left( \max_{k < k_\ell} \sup_{t \in [\alpha_k^\ell, \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| > \varepsilon, E^\ell \right) < \infty. \tag{10.6}$$

As in the proof of Lemma 7.1, on the event

$$\{H_k < \varepsilon\} \cap \left\{ \sup_{t \in [\alpha_k^\ell, \alpha_{k+1}^\ell]} |X_\ell(t) - X_\ell(\alpha_k^\ell)| < \varepsilon \right\},$$

we have  $\tilde{\alpha}_k^\ell \geq \alpha_{k+1}^\ell$  a.s., hence (10.6) yields

$$\mathbb{P} \left( \sup_{t \in [0, T]} |X_{n_\ell}(t) - X_\ell(t)| > 4\varepsilon, U(T) \leq c, E^\ell \right) < \delta_\ell$$

for a summable sequence  $(\delta_\ell)$  that does not depend on the  $n_\ell$ , and the assertion follows as in the proof of Lemma 7.1.

It remains to prove the claim (10.5). By the Markov inequality, we have

$$\mathbb{P} \left( \hat{N}^{b+\ell}(0, \alpha_1^\ell) > \ell\varepsilon \right) \leq e^{-\ell\varepsilon} \mathbb{E} \left[ \mathbb{E} \left[ \exp \left( \hat{N}^{b+\ell}(0, \alpha_1^\ell) \right) \mid \zeta_0 \right] \right].$$

We show (similarly to the proof of Lemma 10.1) that  $\sup_{\ell \in \mathbb{N}} \mathbb{E}[\exp(\hat{N}^{2\ell}(0, \alpha_1^\ell)) \mid \zeta_0]$  is bounded, this will imply the claim.

For  $x \in \Delta$  and a random partition  $\pi$  with distribution  $\kappa(x, \cdot)$ , the random variable  $\hat{b}_{2\ell}(\pi)$  is binomially distributed with parameters  $2\ell$  and  $|x|_1$ . Using monotone convergence, conditional independence, and inserting moment generating functions of binomial distributions, we obtain

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( \hat{N}^{2\ell} (0, \alpha_1^\ell) \right) \mid \zeta_0 \right] \\ &= \mathbb{E} \left[ \exp \left( \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} \hat{b}_{2\ell}(\pi^i) \right) \mid ((t^i, y^i), i \in \mathbb{N}) \right] \\ &= \lim_{n \rightarrow \infty} \prod_{i \in [n]: t^i < \alpha_1^\ell} \mathbb{E} \left[ \exp \left( \hat{b}_{2\ell}(\pi^i) \right) \mid ((t^i, y^i), i \in \mathbb{N}) \right] \\ &= \lim_{n \rightarrow \infty} \prod_{i \in [n]: t^i < \alpha_1^\ell} (1 - |y^i|_1 + |y^i|_1 e)^{2\ell} \\ &= \exp \left( \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} 2\ell \log (1 + |y^i|_1 (e - 1)) \right) \\ &\leq \exp \left( 2\ell \sum_{i \in \mathbb{N}: t^i < \alpha_1^\ell} |y^i|_1 (e - 1) \right) \leq \exp(2(e - 1)) \quad \text{a. s.} \end{aligned}$$

The last inequality follows from the definitions of  $(\hat{U}(t), t \in \mathbb{R}_+)$  and  $\alpha_1^\ell$ . □

### List of notation

Here we collect notation that is used globally in the article.

#### Miscellaneous

$\mathbb{R}_+ = [0, \infty)$ ,  $\mathbb{N} = \{1, 2, 3, \dots\}$ ,  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ ,  $[n] = \{1, \dots, n\}$  for  $n \in \mathbb{N}$ ,  $[0] = \emptyset$

$d_{\mathbb{P}}^X, d_{\mathbb{H}}^X$ : Prohorov metric and Hausdorff distance over  $X$

supp  $\mu$ : closed support of the measure  $\mu$

$\varphi(\mu) = \mu \circ \varphi^{-1}$ : pushforward measure under a measurable function  $\varphi$

$\gamma_n$ : restriction map in various contexts (p. 5/l. 25, p. 24/l. 24)

$\mathcal{U}_\varepsilon^X(x) = \{y \in X : d(x, y) < \varepsilon\}$ ,  $\mathcal{B}_\varepsilon^X(x) = \{y \in X : d(x, y) \leq \varepsilon\}$ : balls in a metric space  $(X, d)$

#### (Marked) distance matrices

$\mathfrak{D}$ : space of semi-metrics on  $\mathbb{N}$  (p. 17/l. -10)

$\mathfrak{D}_n$ : space of semi-metrics on  $[n]$  (p. 24/l. 19)

$\hat{\mathfrak{D}}$ : space of decomposed semi-metrics on  $\mathbb{N}$  (p. 21/l. -3)

$\hat{\mathfrak{D}}_n$ : space of decomposed semi-metrics on  $[n]$  (p. 24/l. 21)

#### Partitions and semi-partitions

$\mathcal{P}$ : Set of partitions of  $\mathbb{N}$  (p. 5/l. 23)

$K_{i,j}$ : partition of  $\mathbb{N}$  that contains only  $\{i, j\}$  and singleton blocks (p. 10/l. -6)

$\mathcal{P}_n$ : Set of partitions of  $[n]$ , associated transformations (p. 26/l. 14)

$\mathcal{P}^n$ : Set of partitions of  $\mathbb{N}$  in which the first  $n$  integers are not all in different blocks, equation (2.2)

$\hat{\mathcal{P}}^n$ : Set of partitions of  $\mathbb{N}$  in which the first  $n$  integers are not all in singleton blocks, (p. 25/l. -13)

$B(\pi, i)$ : block of the partition  $\pi \in \mathcal{P}$  that contains  $i \in \mathbb{N}$   
 $\pi(i) = k$  such that  $i$  is in the  $k$ -th block of  $\pi$  (p. 26/l. 11)  
 $M(\pi)$ : set of the minimal elements of the blocks of  $\pi \in \mathcal{P}$  (p. 24/l. 27)  
 $|B|_n = n^{-1}(\#B \cap [n])$ ,  $B = \lim_{n \rightarrow \infty} |B|_n$  for  $B \subset \mathbb{N}$ ,  $n \in \mathbb{N}$ : relative and asymptotic frequency  
 $\#\pi$ : number of blocks of a partition  $\pi$   
 $\mathcal{S}_n$  set of semi-partitions of  $[n]$ , associated transformations (p. 26/l. 22)  
 $\Delta = \{x = (x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, |x|_1 \leq 1\}$   
 $|x|_p = (\sum_{i \in \mathbb{N}} x_i^p)^{1/p}$  for  $x \in \Delta$   
 $\kappa(x, \cdot)$ : paintbox distribution associated with  $\Delta$  (p. 10/l. -11)

### Genealogy in the lockdown model

$\eta$ : point measure on  $(0, \infty) \times \mathcal{P}$  that encodes the reproduction events (p. 5/l. -16, p. 10/l. -15, p. 24/l. -12)  
 $\eta_0, \eta_K$ : restrictions of  $\eta$  to large and binary reproduction events, respectively (p. 24/l. -12)  
 $\zeta_0$  point measure on  $(0, \infty) \times \Delta$  that encodes the family sizes in the large reproduction events (p. 24/l. -6)  
 $(t, i)$ : individual on level  $i$  at time  $t$  (p. 6/l. 2)  
 $A_s(t, i)$ : level of the ancestor at time  $s$  of  $(t, i)$  (p. 6/l. 3)  
 $(Z, \rho)$ : lockdown space with genealogical distance (p. 6/l. 18)  
 $\rho_t(i, j) = \rho((t, i), (t, j))$  (equation (2.3))  
 $(\hat{Z}, \hat{\rho})$ : extended lockdown space with genealogical distance (p. 7/l. 23)  
 $X_t$  closure of  $\{t\} \times \mathbb{N}$  in  $Z$  (p. 12/l. 14)  
 $X_n(t)$ : relative frequencies in Lemmas 7.1 and 7.2  
 $z(t, i)$ : parent of  $(t, i)$  (p. 7/l. 18)  
 $v_t(i) = \rho((t, i), z(t, i))$  (p. 6/l. -4)  
 $r_t(i, j) = \rho(z(t, i), z(t, j))$  (equation (2.5))  
 $D_t(s, i)$ : lowest level at time  $t$  of a descendant of  $(s, i)$  (p. 9/l. -8)  
 $\tau_{s,i}$ : extinction time of  $(s, i)$  (p. 9/l. -3)  
 $\Theta^{\text{ext}} = \{\tau_{s,i} : s \in \mathbb{R}_+, i \in \mathbb{N}\}$  (equation (2.6))  
 $\Theta_0$ : set of large reproduction times (equation (2.7))  
 $\mu_t^n$ : uniform measure on  $(t, i)$ ,  $i \in [n]$  (p. 11/l. -4)  
 $m_t^n$ : uniform measure on  $(z(t, i), v_t(i))$ ,  $i \in [n]$  (equation (3.2))  
 $\mu_t, m_t$ : weak limits of  $\mu_t^n$  and  $m_t^n$ , respectively (p. 11/l. -3, p. 15/l. 14)  
 $(\Pi_{s,t}, 0 \leq s \leq t)$ : flow of partitions (p. 33/l. 5)  
 $\Pi_t$ : partition of siblings: (p. 41/l. -6)  
 $C_t^{a,\varepsilon,I}$ : subsets of individuals at time  $t$  (p. 33/l. 21)  
 $\Pi_t^{a,\varepsilon}$ : like  $\Pi_t$  but individuals that are in a common  $C_t^{a,\varepsilon,I}$  are also in the same block (p. 33/l. -16)

### Characteristic measures

$\Xi = \Xi_0 + \Xi\{0\}\delta_0$  (p. 10/l. -7)  
 $H_\Xi$ : characteristic measure of  $\eta$  (p. 10/l. -4)  
 $\mathcal{M}_{\text{dust}}, \mathcal{M}_{\text{nd}}, \mathcal{M}_{\text{CDI}}$ : Sets of finite measures on  $\Delta$  with and without dust, and with the coming down from infinity property, respectively (p. 11/l. 4, p. 12/l. -18)

### References

- [1] S. Athreya, W. Löhner, and A. Winter, *The gap between Gromov-vague and Gromov-Hausdorff-vague topology*, Stochastic Process. Appl. **126** (2016), no. 9, 2527–2553. MR-3522292

- [2] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge University Press, Cambridge, 2006. MR-2253162
- [3] J. Bertoin and J.-F. Le Gall, *Stochastic flows associated to coalescent processes*, Probab. Theory Related Fields **126** (2003), no. 2, 261–288. MR-1990057
- [4] M. Birkner, J. Blath, M. Möhle, M. Steinrücken, and J. Tams, *A modified lookdown construction for the Xi-Fleming-Viot process with mutation and populations with recurrent bottlenecks*, ALEA Lat. Am. J. Probab. Math. Stat. **6** (2009), 25–61. MR-2485878
- [5] R. M. Blumenthal, *Excursions of Markov processes*, Probability and its Applications, Birkhäuser Boston, 1992. MR-1138461
- [6] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR-1835418
- [7] I. Dahmer and G. Kersting, *The total external length of the evolving Kingman coalescent*, Probab. Theory Related Fields **167** (2017), 1165–1214. MR-3627436
- [8] I. Dahmer, R. Knobloch, and A. Wakolbinger, *The Kingman tree length process has infinite quadratic variation*, Electron. Commun. Probab. **19** (2014), no. 87. MR-3298276
- [9] A. Depperschmidt, A. Greven, and P. Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Probab. **16** (2011), 174–188. MR-2783338
- [10] A. Depperschmidt, A. Greven, and P. Pfaffelhuber, *Tree-valued Fleming-Viot dynamics with mutation and selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615. MR-3024977
- [11] A. Depperschmidt, A. Greven, and P. Pfaffelhuber, *Path-properties of the tree-valued Fleming-Viot process*, Electron. J. Probab. **18** (2013), no. 84. MR-3109623
- [12] P. Donnelly and T. G. Kurtz, *A countable representation of the Fleming-Viot measure-valued diffusion*, Ann. Probab. **24** (1996), no. 2, 698–742. MR-1404525
- [13] P. Donnelly and T. G. Kurtz, *Particle representations for measure-valued population models*, Ann. Probab. **27** (1999), no. 1, 166–205. MR-1681126
- [14] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley, New York, 1986. MR-838085
- [15] S. N. Evans, *Kingman’s coalescent as a random metric space*, Stochastic models (Ottawa, ON, 1998), CMS Conf. Proc., vol. 26, Amer. Math. Soc., Providence, RI, 2000, pp. 105–114. MR-1765005
- [16] S. N. Evans and A. Winter, *Subtree prune and regraft: a reversible real tree-valued Markov process*, Ann. Probab. **34** (2006), no. 3, 918–961. MR-2243874
- [17] C. Foucart, *Generalized Fleming-Viot processes with immigration via stochastic flows of partitions*, ALEA Lat. Am. J. Probab. Math. Stat. **9** (2012), no. 2, 451–472. MR-3069373
- [18] A. Greven, P. Pfaffelhuber, and A. Winter, *Convergence in distribution of random metric measure spaces ( $\Lambda$ -coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1–2, 285–322. MR-2520129
- [19] A. Greven, P. Pfaffelhuber, and A. Winter, *Tree-valued resampling dynamics Martingale problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3–4, 789–838. MR-3034793
- [20] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser, Boston, MA, 1999. MR-1699320
- [21] S. Gufler, *Invariance principles for tree-valued Cannings chains* (2017). arXiv:1608.08203v2
- [22] S. Gufler, *A representation for exchangeable coalescent trees and generalized tree-valued Fleming-Viot processes*, Electron. J. Probab. **23** (2018).
- [23] O. Hénard, *The fixation line in the  $\Lambda$ -coalescent*, Ann. Appl. Probab. **25** (2015), no. 5, 3007–3032. MR-3375893
- [24] P. Herriger and M. Möhle, *Conditions for exchangeable coalescents to come down from infinity*, ALEA Lat. Am. J. Probab. Math. Stat. **9** (2012), no. 2, 637–665. MR-3069379
- [25] A. S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, New York, 1995. MR-1321597
- [26] G. Kersting, J. Schweinsberg, and A. Wakolbinger, *The evolving beta coalescent*, Electron. J. Probab. **19** (2014), no. 64. MR-3238784

- [27] S. Kliem and W. Löhner, *Existence of mark functions in marked metric measure spaces*, Electron. J. Probab. **20** (2015), no. 73. MR-3371432
- [28] T. G. Kurtz, *Martingale problems for conditional distributions of Markov processes*, Electron. J. Probab. **3** (1998), no. 9, 29 pp. MR-1637085
- [29] C. Labbé, *From flows of  $\Lambda$ -Fleming-Viot processes to lookdown processes via flows of partitions*, Electron. J. Probab. **19** (2014), no. 55. MR-3227064
- [30] W. Löhner, *Equivalence of Gromov-Prohorov- and Gromov's  $\square_\lambda$ -metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17. MR-3037215
- [31] G. Miermont, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 5, 725–781. MR-2571957
- [32] M. Möhle, *Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent*, Stochastic Process. Appl. **120** (2010), no. 11, 2159–2173. MR-2684740
- [33] M. Möhle and S. Sagitov, *A classification of coalescent processes for haploid exchangeable population models*, Ann. Probab. **29** (2001), no. 4, 1547–1562. MR-1880231
- [34] P. Pfaffelhuber and A. Wakolbinger, *The process of most recent common ancestors in an evolving coalescent*, Stochastic Process. Appl. **116** (2006), no. 12, 1836–1859. MR-2307061
- [35] P. Pfaffelhuber, A. Wakolbinger, and H. Weisshaupt, *The tree length of an evolving coalescent*, Probab. Theory Related Fields **151** (2011), no. 3–4, 529–557. MR-2851692
- [36] J. Pitman, *Coalescents with multiple collisions*, Ann. Probab. **27** (1999), no. 4, 1870–1902. MR-1742892
- [37] L. C. G. Rogers and J. W. Pitman, *Markov functions*, Ann. Probab. **9** (1981), no. 4, 573–582. MR-624684
- [38] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge University Press, 2000. MR-1796539
- [39] S. Sagitov, *The general coalescent with asynchronous mergers of ancestral lines*, J. Appl. Probab. **36** (1999), no. 4, 1116–1125. MR-1742154
- [40] J. Schweinsberg, *Coalescents with simultaneous multiple collisions*, Electron. J. Probab. **5** (2000), no. 4. MR-1781024
- [41] A. Véber and A. Wakolbinger, *The spatial Lambda-Fleming-Viot process: an event-based construction and a lookdown representation*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015), no. 2, 570–598. MR-3335017
- [42] A. M. Vershik, *Random and universal metric spaces*, Dynamics and randomness II, Kluwer Acad. Publ., Dordrecht, vol. 10, 2004, pp. 199–228. MR-2153330

**Acknowledgments.** This work is part of the author's PhD thesis. The author thanks Götz Kersting and Anton Wakolbinger for their valuable advice. He also thanks them and the referees for very helpful comments. Partial support from the DFG Priority Programme 1590 "Probabilistic Structures in Evolution" is acknowledged. In 2017/2018, the author is supported by a postdoctoral fellowship of the Minerva Foundation.