ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES
UMR CNRS 7641

91128 PALAISEAU CEDEX (FRANCE). Tél: 01 69 33 41 50. Fax: 01 69 33 30 11

http://www.cmap.polytechnique.fr/

Historical superprocess limits for population models with past dependence and competition

Sylvie Méléard, Chi Viet Tran

R.I. N^0 720 *July 2011*

Historical superprocess limits for population models with past dependence and competition

Sylvie Méléard* and Viet Chi Tran†

July 13, 2011

Abstract

We consider a population structured in trait. When a birth occurs, the trait of the parent is hereditarily transmitted to the offspring unless a mutation occurs. We associate with each individual its lineage consisting of all the traits of his ancestors. The evolution of the population results from the aging, births and deaths of individuals, with a dynamics that may depend on the past history of the lineage and that allows interactions between individuals. We introduce the stochastic process that describes the system and consider its diffusion limit under the assumptions of large populations, individuals of small masses and allometric demographies.

Keywords: Historical superprocess, interacting particle system, limit theorem.
AMS Subject Classification: 60J80, 60J68, 60K35.

1 Introduction

We consider a population of interacting particles evolving with aging and with the occurrence of births and deaths. Each individual is characterized by a vector trait $x \in \mathbb{R}^d$. The trait of an individual remains constant during its life and is transmitted hereditarily unless a mutation occurs. There exist many possible dependencies on the past. Our purpose is to keep tracks of the genealogy of the particles. The dynamics and interaction depend here on the trait past history. For example the birth and death rates of an individual could depend on all the traits of its ancestors or on the time since its trait first appeared in the lineage. With such dependencies on the past, we are inspired mathematically by the works of Dynkin [20] or Perkins [41] and study a historical version of the process where each particle can be represented by its lineage, *i.e.* the path that associates at each time in the past the trait of its ancestor living at that time.

We are interested in the limits of these processes in large populations, when individuals have small masses and allometric demographies (here: short lives and reproduction times). The investigation of these problems for historical processes has been considered by Dynkin [20], Dawson and Perkins [16] in cases where there is no interaction. Kaj and Sagitov [33] considered a similar asymptotic for age-structured processes without interaction and mention the possibility of carrying their study with a historical approach. In previous papers [39, 38], we studied cases where the birth, death and competition rates were functions of the traits and of the physical

*CMAP, Ecole Polytechnique, UMR CNRS 7641, Route de Saclay, 91128 Palaiseau Cédex, France
†Equipe Probabilité Statistique, Laboratoire Paul Painlevé, UMR CNRS 8524, UFR de Mathématiques, Université des Sciences et Technologies Lille 1, Cité Scientifique, 59655 Villeneuve d'Ascq Cédex, France ; CMAP, Ecole Polytechnique

ages of the particles. In Méléard and Tran [38], the convergence is studied, without tackling the genealogies, and averaging phenomena are exhibited.

The evolution of genealogies in population dynamics is a major problem, which motivated an abundant literature and with applications to evolution and population genetics. The historical approach of Dawson and Perkins [14, 16, 41] models evolving genealogies, in the forward physical time and with variable population size (including extinction phenomena). Our model generalizes this approach and emphasizes how the competition between individuals, expressed as nonlinear death-rates, and mutations drive the ecological dynamics. In absence of interaction, it is possible to describe the genealogy of individuals sampled in the population at time t by coalescent processes: Kingman's coalescent [34], Λ-coalescents (see Berestycki [6] and references therein for a survey) or Ξ-coalescents (see *e.g.* Möhle and Sagitov [40], Schweinsberg [45] or Limic [37]). Barton, Etheridge and Véber [4] studied the genealogies of a spatial version of the Λ-Fleming-Viot and obtain in the limit Kingman coalescent, Λ-coalescent or coalescing Brownian motions depending on their various parameters. Depperschmidt, Greven, Pfaffelhuber and Winter [17, 27] proposed to view genealogical trees (for instance Λ-coalescents) as marked ultrametric spaces and described their evolution. All these models allow to incorporate selection and mutation (see also *e.g.* [36, 3, 23]) but not competition between individuals.

In a first section, we construct a historical particle system which dynamics depends on the past. At a given time t, we associate to each particle the lineage that gives for each $s \leq t$ the trait of the ancestor living at this time s, and extended for $s > t$ by the constant function equal to the trait of the individual at t (see Fig. 1). Since each particle keeps a constant trait during its life, these lineages are càdlàg paths constant by parts. We denote by $\mathbb{D}_{\mathbb{R}^d}$ the space of càdlàg paths with values in $\mathbb{R}^d$ embedded with the Skorohod topology (see *e.g.* [7]). The total population is represented by a point measure on $\mathbb{D}_{\mathbb{R}^d}$ where each particle is represented by a Dirac mass. When a particle dies, the corresponding Dirac mass is removed and when a birth occurs, a new Dirac mass is created. As a result, we keep only the genealogical tree of the individuals living at time t, extended to a constant genealogy after t (see Fig. 1 (b)).

(a) (b)

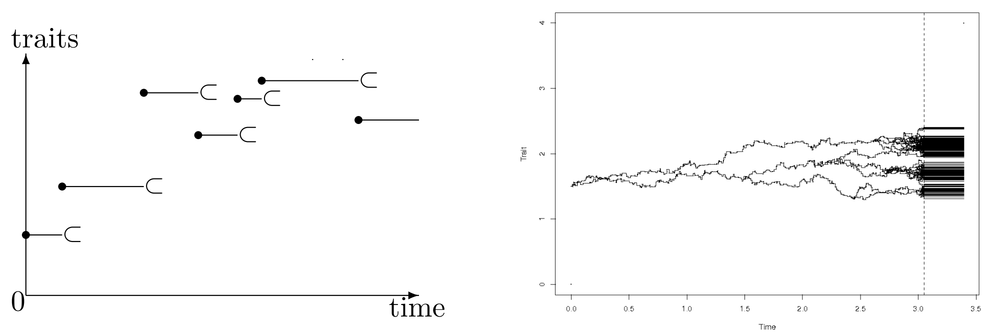

Figure 1: *(a) Example of a path constituted with ancestral traits. (b) Genealogies of the individuals at time $t = 7.8$. We have drawn the support of the time marginals X_t of the historical process. The vertical dotted line corresponds to points with abscissa t. The ancestral paths of the individuals alive at time t provides the genealogical tree of individuals living at time t. Births corresponds to new leaves. Upon death, branches may disappear.*

The diffusive limit in large population is studied in Section 2. We rescale the size of the initial condition proportionally to a parameter $n \in \mathbb{N}^* = \{1, 2, \ldots\}$ that we will let grow to infinity.

The weight of the individuals is renormalized by $1/n$ (to keep the total biomass of the same order when n varies) and the birth and death rates are of order n to reflect allometric demographies, while preserving the demographic balance. To study the convergence of the sequence of stochastic processes indexed by n, we proceed with tightness-uniqueness arguments. The proof of tightness is based on the ideas in Dawson and Perkins [16] for the historical super-Brownian process but requires new arguments to account for interactions. We then identify the limiting values as solution of a nonlinear martingale problem for which uniqueness is stated.
Examples are carried in Section 3. First examples deal with evolution models of adaptive dynamics with local (see [18]) or asymmetric (see [35]) competition. Evolution shows that for a range of parameters, the population separates into groups concentrated around some trait values. We then consider a spatial model (see [1] or [41, Ex. 4.3 p.50]): particles consume ressources where they live and the offspring arriving in previously habited regions are penalized. We will see that this tends to separate the cloud of particles in several distinct families whose common ancestor is very old. Finally, we look at a logistic age and size-structured population (see [39, 38]), where averaging phenomena appear.

Notation: For a given metric space E, we denote by $\mathbb{D}_E = \mathbb{D}(\mathbb{R}_+, E)$ the space of càdlàg functions from $\mathbb{R}_+$ to E. For $E = \mathbb{R}$, we will use the more simple notation $\mathbb{D} = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$. These spaces are embedded with the Skorohod topology associated with the distance on E and meterized by a distance $\mathbf{d}_{\mathrm{Sk}}$ (*e.g.* [7, 31, 32], see also (B.1) in appendix).

For a function $x \in \mathbb{D}_E$ and $t > 0$, we denote by x^t the stopped function defined by $x^t(s) = x(s \wedge t)$ and by x^{t-} the function defined by $x^{t-}(s) = \lim_{r\uparrow t} x^r(s)$. We will also often write $x_t = x(t)$ for the value of the function at time t. For $y, w \in \mathbb{D}_E$ and $t \in \mathbb{R}_+$, we denote by $(y|t|w) \in \mathbb{D}_E$ the following path:

$$(y|t|w) = \begin{cases} y_u & \text{if} \quad u < t \\ w_{u-t} & \text{if} \quad u \geq t. \end{cases} \tag{1.1}$$

When the path w is constant with $\forall u \in \mathbb{R}_+$, $w_u = x$, we will write $(y|t|x)$ with a notational abuse.

We denote by $\mathcal{M}_F(E)$ (resp. $\mathcal{M}_P^n(E)$, $\mathcal{P}(E)$) the set of finite measures on E (resp. of point measures renormalized by $1/n$, of probability measures). These spaces are embedded with the topology of weak convergence.

2 The historical particle system

In this section, we construct the finite interacting historical particle system that we will study. Trait-structured particle systems without dependence on the past have been considered in Fournier and Méléard [26] or Champagnat *et al.* [11]. For populations with age-structure, we refer to Jagers [30, 29] and Méléard and Tran [39, 46] for instance. Here, we are inspired by these works and propose a birth and death particle system where the lineage of each particle, *i.e.* the traits of its ancestors, is encoded into a path of $\mathbb{D}_{\mathbb{R}^d}$.

2.1 Lineage

We consider a discrete population in continuous time where the individuals reproduce asexually and die with rates that depend on a hereditary trait and on their past. To each individual is associated a quantitative trait transmitted from its parent except when a mutation occurs. The rates may express through the traits carried by the ancestors of the individual. One purpose is for example to take into account the accumulation of beneficial and deleterious mutations through generations.

Individuals are characterized by a trait $x \in \mathbb{R}^d$. The lineage or past history of an individual is defined by the succession of ancestral traits with their appearance times and by the succession of ancestral reproduction times (birth of new individuals). To an individual of trait x born at time S_m, having $m-1$ ancestors born after $t=0$ at times $S_1 = 0 < S_2 < \cdots < S_{m-1}$, with $S_{m-1} < S_m$, and of traits $(x_1, x_2, \dots x_{m-1})$, we associate the path

$$y_t = \sum_{j=1}^{m-1} x_j \mathbf{1}_{S_j \le t < S_{j+1}} + x_m \mathbf{1}_{S_m \le t}. \tag{2.1}$$

This path is called the *lineage* of the individual. We denote by $\mathcal{L}$ the set of possible lineages of the form (2.1). Since a path in $\mathcal{L}$ is entirely characterized by the integer m and the sequence $(0, x_1, \dots, s_{m-1}, x_{m-1}, s_m, x)$ of jump times and traits, it is possible to describe each element of $\mathcal{L}$ by an element of $\mathbb{N} \times \bigcup_{m\in\mathbb{N}} (\mathbb{R}_+ \times \mathbb{R}^d)^m$, which we can embed with a natural lexicographical order.

2.2 Population dynamics

Let us introduce a parameter $n \in \mathbb{N}^*$ that will grow to infinity. This parameter can be seen as the order of the carrying capacity, when the total amount of resources is fixed. To keep the total biomass constant, individuals are attributed a weight of $1/n$. The population is represented by a point measure as follows:

$$X_t^n := \frac{1}{n} \sum_{i=1}^{N_t^n} \delta_{y^i_{.\wedge t}} \ \in \ \mathcal{M}_P^n(\mathcal{L}) \subset \mathcal{M}_P^n(\mathbb{D}_{\mathbb{R}^d}), \tag{2.2}$$

where $N_t^n = n \langle X_t^n, 1 \rangle$ is the number of individuals alive at time t.
Following the notation in Fournier and Méléard [26], let us define the map $Y = (Y^i)_{i\in\mathbb{N}^*}$ from $\bigcup_{n\in\mathbb{N}^*} \mathcal{M}_P^n(\mathcal{L})$ in $\mathcal{L}$ defined by: $\forall n$ and N in $\mathbb{N}^*$,

$$Y^j \left(\frac{1}{n} \sum_{i=1}^{N} \delta_{y^i} \right) = \begin{cases} y^j, & \text{if } j \le N \\ 0 & \text{else.} \end{cases}$$

where the individuals are sorted by the lexicographical order. This will be useful to extract a particular individual from the population. When there is no ambiguity, we will write Y^i instead of $Y^i(X)$ for a point measure $X \in \bigcup_{n\in\mathbb{N}^*} \mathcal{M}_P^n(\mathcal{L})$.

The individuals in our population reproduce asexually during their lives, and give birth at random times to new mutant individuals or clones. They also compete and die. We consider allometric demographies where lifetimes and gestation lengths are proportional to the biomass. Birth and death rates are thus of order n, with expressions that characterize the constraint of preservation of the demographic balance. Also, the mutation steps are rescaled by $1/n$.

Let us now define the population dynamics. For $n \in \mathbb{N}^*$, we consider an individual characterized at time t by the lineage $y \in \mathbb{D}_{\mathbb{R}^d}$ in a population $X^n \in \mathbb{D}_{\mathcal{M}_P^n(\mathbb{D}_{\mathbb{R}^d})}$.

Reproduction The birth rate at time t is $b^n(t, y)$, where

$$b^n(t, y) = n\, r(t, y) + b(t, y).$$

The function b is a continuous nonnegative real function on $\mathbb{R}_+ \times \mathbb{D}_{\mathbb{R}^d}$. We are interested in the following forms for instance:

$$b(t,y) = B\Big(\int_0^t y_{t-s}\nu_b(ds)\Big) \tag{2.3}$$

where B is a continuous function bounded by $\bar{B}$ and ν_b is a Radon measure on $\mathbb{R}^+$. The allometric function r is given by:

$$r(t,y) = R\Big(\int_0^t y_{t-s}\nu_r(ds)\Big) \tag{2.4}$$

where R is continuous and bounded below and above by $\underline{R}$ and $\bar{R}$, and where ν_r is a Radon measure on $\mathbb{R}_+$. We also assume that $R^{1/2}$ is Lipschitz continuous.

When an individual with trait y_{t_-} gives birth at time t, the new offspring is either a mutant or a clone. With probability $1-p \in [0,1]$, the new individual is a clone of its parent, with same trait y_{t_-} and same lineage y. With probability $p \in [0,1]$, the offspring is a mutant of trait $y_{t_-} + h$, where h is drawn in the distribution $k^n(h)\,dh$. We associate to this mutant the lineage $(y|t|y_{t_-}+h)$. For the sake of simplicity, we will consider here that the mutation density $k^n(h)$ is a Gaussian density with mean 0 and covariance matrix $\sigma^2\,\mathbf{Id}/n$. However the model could be generalized for instance to mutation densities $k^n(y_{t_-},h)$ with dependence on the parent's trait. We introduce the notation:

$$K^n(dh) = pk^n(h)dh + (1-p)\delta_0(dh). \tag{2.5}$$

Example 2.1. (i) If we choose $\nu_b(ds) = \delta_0(ds)$, then $\int_0^t y_{t-s}\nu_b(ds) = y_t$ is the trait of the individual of the lineage y living at time t.
(ii) If we choose $\nu_b(ds) = e^{-\alpha s}ds$, with $\alpha > 0$, then $\int_0^t y_{t-s}\nu_b(ds) = \int_0^t e^{-\alpha(t-s)}y_s ds$. This means that the traits of recent ancestors have a higher contribution in the birth rate of the individual alive at time t. Such rates may be useful to model social interactions, for instance cooperative breeding where the ancestors contribute to protect and raise their descendants. When ancestors have advantageous traits, they may help their offspring to reproduce in more favorable conditions and increase their birth rates. □

Death To define the death rate, let us consider a bounded continuous interaction kernel $U \in \mathcal{C}_b(\mathbb{R}_+ \times \mathbb{D}^2_{\mathbb{R}^d}, \mathbb{R})$, a bounded continuous function D on $\mathbb{R}_+ \times \mathbb{D}_{\mathbb{R}^d}$ and a Radon measure ν_d weighting the influence of the past population on the present individual y at time t. The death rate is

$$d^n(t,y,X^n) = n\,r(t,y) + d(t,y,X^n),$$

where for a process $X \in \mathbb{D}_{\mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d})}$,

$$d(t,y,X) = D(t,y) + \int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} U(t,y,y')X_{t-s}(dy')\nu_d(ds). \tag{2.6}$$

The first term with function r allows us to preserve the demographic balance. The term $D(t,y)$ is the natural death rate, while $U(t,y,y')$ represents the competition exerted at time t by an individual of lineage y' on our individual of lineage y. We assume that:

$$\begin{aligned}&\exists \bar{D} > 0,\ \forall y \in \mathbb{D},\ \forall t \in \mathbb{R}_+,\ 0 \le D(t,y) < \bar{D},\\ &\exists \underline{U},\ \bar{U} > 0,\ \forall y,y' \in \mathbb{D},\ \forall t \in \mathbb{R}_+,\ 0 < U(t,y,y') < \bar{U}.\end{aligned} \tag{2.7}$$

Example 2.2. The following examples are developed in Section 5.
(i) Kisdi's model with asymmetrical competition:
If we choose $D = 0$, $\nu_d(ds) = \delta_0(ds)$ and the asymmetric competition kernel proposed by Kisdi [35]

$$U(t, y, y') = \frac{2}{K}\Big(1 - \frac{1}{1 + \alpha e^{-\beta(y_t - y'_t)}}\Big), \tag{2.8}$$

then, the death rate becomes:

$$d(t, y, X) = \frac{2}{K}\int_{\mathbb{D}_{\mathbb{R}^d}} \Big(1 - \frac{1}{1 + \alpha e^{-\beta(y_t - y'_t)}}\Big) X_t(dy'). \tag{2.9}$$

(ii) Perkins' variant of Adler's fattened goats [1, 42]:
In this model, trait is space and the genealogy describes the trajectory of the past ancestors. It corresponds to the choice of $D = 0$, of $\nu_d(ds) = e^{-\alpha s} ds$ with $\alpha > 0$ and of

$$U(t, y, y') = K_\varepsilon\big(y_t - y'_t\big).$$

Here, K_ε is a symmetric smooth kernel with maximum at 0, for instance the density function of a centered Gaussian distribution with variance ε.
From the definition of the processes (see (2.1)), if y' belongs to the support of $X_s(dy')$ then almost surely (a.s.) y' is a path stopped at s and for all $t \geq s$, $y'_t = y'_s$. Thus, we obtain in this case

$$d(t, y, X) = \int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} K_\varepsilon\big(y_t - y'_s\big) X_s(dy') e^{-\alpha(t-s)} ds. \tag{2.10}$$

This describes a setting where the goat-like particles consume resources at the point where they are staying, and when they arrive in a region where the population has previously grazed, their death rate is increased. The parameter α describes the speed at which the environment replenishes itself. □

Remark 2.3. A generalization to physical age structure as considered in [38] for instance is possible provided we extend the trait space to add a color to each individual. This consists in associating to each individual an additional path indicating its color. This color is an independent uniform $[0, 1]$-valued r.v. that is drawn at each birth. The lineage of colors of the individual is denoted by $c \in \mathbb{D}(\mathbb{R}_+, [0, 1])$.
The colors allow us to define the birth date of the individual of the lineage alive at time t.

$$\tau_{c,t} \quad = \inf\{s \leq t, \quad c_s = c_t\} \quad = \sup\{s \leq t, \quad c_s \neq c_t\} \tag{2.11}$$

We define by:

$$a(t) := t - \tau_{c,t} \tag{2.12}$$

the age of the latter individual at time t. It is a càdlàg function that is discontinuous at the birth times. □

2.3 Stochastic Differential Equation (SDE)

Following the work of Fournier and Méléard [26], it is possible to propose an SDE driven by a Poisson Point Measure (PPM) to describe in a pathwise manner the evolution of $(X^n_t)_{t\in\mathbb{R}_+}$, for any $n \in \mathbb{N}^*$. The measure representing the population evolves with the occurrences of births and deaths. Since the rates may vary with time, we use acceptance-rejection techniques to obtain these events' occurrences by mean of PPMs. According to the event that happens we add or remove Dirac masses in (2.2). This construction provides an exact simulation algorithm that is extensively used in Section 5.

Definition 2.4. Let us consider on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

1. a random variable $X_0^n \in \mathcal{M}_P^n(\mathbb{D}_{\mathbb{R}^d})$ such that $\mathbb{E}\left(\langle X_0^n, 1\rangle\right) < +\infty$ and such that the support of X_0^n contains a.s. only constant functions,
2. a PPM $Q(ds, di, dh, d\theta)$ on $\mathbb{R}_+ \times E := \mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}^d \times \mathbb{R}_+$ with intensity measure $ds \otimes n(di) \otimes dh \otimes d\theta$ and independent from X_0^n, $n(di)$ being the counting measure on $\mathbb{N}^*$ and ds, dh and $d\theta$ the Lebesgue measures on $\mathbb{R}_+$, $\mathbb{R}^d$ and $\mathbb{R}_+$ respectively.

We denote by $(\mathcal{F}_t^n)_{t\in\mathbb{R}_+}$ the canonical filtration associated to X_0^n and Q, and consider the following SDE with values in $\mathcal{M}_P^n(\mathbb{D}_{\mathbb{R}^d})$:

$$X_t^n = X_0^n + \frac{1}{n}\int_0^t \int_E \mathbf{1}_{\{i\le n\langle X_{s_-}^n,1\rangle\}} \Big[\delta_{(Y^i|s|Y_{s_-}^i+h)}\mathbb{1}_{\theta\le m_1^n(i,s,h)} + \delta_{Y^i}\mathbb{1}_{m_1^n(i,s,h)<\theta\le m_2^n(i,s,h)} - \delta_{Y^i}\mathbb{1}_{m_2^n(i,s,h)<\theta\le m_3^n(i,s,h,X^{n,s_-})}\Big]\, Q(ds, di, dh, d\theta), \tag{2.13}$$

where

$$\begin{aligned} m_1^n(i,s,h) &= p\, b^n(s, Y^i(X_{s_-}^n))k^n(h),\\ m_2^n(i,s,h) &= m_1^n(i,s,h) + (1-p)\, b^n(s, Y^i(X_{s_-}^n))k^n(h)\\ m_3^n(i,s,h,X^{n,s_-}) &= m_2^n(i,s,h) + d^n(s, Y^i(X_{s_-}^n), X^{n,s_-})k^n(h). \end{aligned} \tag{2.14}$$

□

Existence and uniqueness of the solution $(X_t^n)_{t\in\mathbb{R}_+}$ of SDE (2.13), for every $n \in \mathbb{N}^*$, are obtained from a direct adaptation of the proof of Theorem 3.1 in [26], using the following moment estimates and martingale properties.

Proposition 2.5. *Let us assume that the initial populations have sizes proportional to n and that:*

$$\sup_{n\in\mathbb{N}^*} \mathbb{E}\left(\langle X_0^n, 1\rangle^3\right) < +\infty. \tag{2.15}$$

Then:
(i) For all $T > 0$,

$$\sup_{n\in\mathbb{N}^*} \mathbb{E}\Big(\sup_{t\in[0,T]} \langle X_t^n, 1\rangle^p\Big) < +\infty, \quad \text{with } p \in \{2,3\}. \tag{2.16}$$

(ii) For a bounded and measurable function φ,

$$\begin{aligned} \langle X_t^n, \varphi\rangle = \langle X_0^n, \varphi\rangle + M_t^{n,\varphi} + \int_0^t ds \int_{\mathbb{D}_{\mathbb{R}^d}} X_s^n(dy)\Big[& \\ nr(s,y)\Big(\int_{\mathbb{R}^d} \varphi(y|s|y_s+h)K^n(y_s,dh) - \varphi(y)\Big) & \\ + b(s,y)\int_{\mathbb{R}^d} \varphi(y|s|y_s+h)K^n(y_s,dh) - d(s,y,(X^n)^s)\varphi(y)\Big] & \end{aligned} \tag{2.17}$$

where $M^{n,\varphi}$ is a square integrable martingale starting from 0 with quadratic variation:

$$\langle M^{n,\varphi}\rangle_t = \frac{1}{n}\int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} \Big[\big(nr(s,y) + b(s,y)\big)\int_{\mathbb{R}^d} \varphi^2(y|s|y_s+h)K^n(y_s,dh) + \big(nr(s,y) + d(s,y,(X^n)^s)\big)\varphi^2(y)\Big] X_s^n(dy)\, ds. \tag{2.18}$$

Proof. The proofs of (i) and (ii) follow from the proofs of Lemma 5.2 and Theorem 5.6 in [26]. ■

3 The historical superprocess limit

We now investigate a diffusive limit of the sequence of processes X^n defined by (2.13). Let us first introduce a class of test functions that we will use to define the limit process.

Definition 3.1. For a real $\mathcal{C}_b^2$-function g on $\mathbb{R}_+ \times \mathbb{R}^d$ and a real $\mathcal{C}_b^2$-function G on $\mathbb{R}$, we define the continuous function G_g on the path space $\mathbb{D}_{\mathbb{R}^d}$ by

$$G_g(y) = G\Big(\int_0^T g(s, y_s)ds\Big). \tag{3.1}$$

Let us remark that the class generated by finite linear combinations of such functions is stable under addition and separates the points. The latter fact is a consequence of Lemma A.1 in Appendix A.

Notice that if y is a càdlàg path stopped at t then

$$G_g(y) = G\Big(\int_0^t g(s, y_s)ds + \int_t^T g(s, y_t)ds\Big).$$

Also, in the sequel, the following quantity will appear for $t \in [0, T]$ and $y \in \mathbb{D}_{\mathbb{R}^d}$:

$$D^2G_g(t, y) = G'\Big(\int_0^T g(s, y_s)ds\Big)\int_t^T \Delta_x g(s, y_t)ds + G''\Big(\int_0^T g(s, y_s)ds\Big)\sum_{i=1}^d \Big(\int_t^T \partial_{x_i} g(s, y_t)ds\Big)^2. \tag{3.2}$$

This quantity generalizes the Laplacian. For instance, if $G(x) = x$ and if $g(s, x) = g(x)$ does not depend on time, we obtain that $D^2G_g(t, y) = (T - t)\Delta g(y_t)$.

Note that Dawson ([14], p. 203) and Etheridge ([22], p. 24) introduce another class of test functions of the form

$$\varphi(y) = \prod_{j=1}^m g_j(y_{t_j}), \tag{3.3}$$

for $m \in \mathbb{N}^*$, $0 \leq t_1 < \cdots < t_m$ and $\forall j \in \{1, \ldots, m\}$, $g_j \in \mathcal{C}_b^2(\mathbb{R}^d, \mathbb{R})$. This class is not convenient when dealing with càdlàg processes since the functions are not continuous for the Skorohod topology. However, we can use these test functions when it is ensured that we deal with continuous paths y.

If y is a continuous path stopped at t then $\varphi(y) = \prod_{j=1}^m g_j(y_{t_j \wedge t})$. We use the notation introduced in [14] p.203) to generalize the Laplacian to these test functions φ. For a path $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$, a time $t > 0$, we define

$$\widetilde{\Delta}\varphi(t, y) = \sum_{k=0}^{m-1} \mathbb{1}_{[t_k, t_{k+1}[}(t)\Big(\prod_{j=1}^k g_j(y_{t_j})\Delta\Big(\prod_{j=k+1}^m g_j\Big)(y_t)\Big), \tag{3.4}$$

where $t_0 = 0$.

The following lemma which links the test functions (3.1) and (3.3) will be used in the sequel. It is proved in Appendix A

Lemma 3.2. *Let φ be a test function of the form (3.3). Then there exists test functions (φ_q) of the form (3.1) such that for every $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ and $t \in [0, T]$, the sequences $(\varphi_q(y))_{q \in \mathbb{N}^*}$ and $(D^2\varphi_q(t, y))_{q \in \mathbb{N}^*}$ are bounded uniformly in q, t and y and converge respectively to $\varphi(y)$ and $\widetilde{\Delta}\varphi(t, y)$.*

3.1 Main convergence result

Let us assume that the initial conditions converge:

$$\exists X_0 \in \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}), \lim_{n\to+\infty} X_0^n = \bar{X}_0 \text{ for the weak convergence.} \tag{3.5}$$

The main theorem of this section states the convergence of the sequence $(X_t^n)_{n\in\mathbb{N}^*}$:

Theorem 3.3. *Assume* (3.5) *and* (2.15). *Then the sequence* $(X^n)_{n\in\mathbb{N}^*}$ *converges in law in* $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$ *to the superprocess* $\bar{X} \in \mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$ *characterized as follows, for test functions* G_g *of the form (3.1):*

$$M_t^{G_g} = \langle \bar{X}_t, G_g\rangle - \langle \bar{X}_0, G_g\rangle - \int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} \Big(p\, r(s,y)\frac{\sigma^2}{2}\; D^2G_g(s,y) \\ + \gamma(s,y,\bar{X}^s)G_g(y)\Big)\bar{X}_s(dy)\, ds \tag{3.6}$$

is a square integrable martingale with quadratic variation:

$$\langle M^{G_g}\rangle_t = \int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} 2\, r(s,y)\sigma^2 G_g^2(y)\bar{X}_s(dy)\, ds, \tag{3.7}$$

where $D^2G_g(t,y)$ *has been defined in* (3.2) *and where* $\gamma(t,y,\bar{X}^t)$ *defines the growth rate of individuals* y *at time* t *in the population* $\bar{X}$*:*

$$\gamma(t,y,\bar{X}^t) = b(t,y) - d(t,y,\bar{X}^t). \tag{3.8}$$

. □

For the proof of Theorem 3.3, we proceed in a compactness-uniqueness manner. First, we establish the tightness of the sequence $(X^n)_{n\in\mathbb{N}^*}$ (Section 3.2) then use Prohorov's theorem and identify the limiting values as unique solution of the martingale problem (3.6), (3.7).

3.2 Tightness of $(X^n)_{n\in\mathbb{N}^*}$

In this subsection, we shall prove that:

Proposition 3.4. *The sequence* $(\mathcal{L}(X^n))_{n\in\mathbb{N}^*}$ *is tight on* $\mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d})))$.

For this, we use the following criterion adapted from Jakubowski [31] and Dawson and Perkins [16], characterizing the uniform tightness of measure-valued càdlàg processes. The two ingredients are heuristically the compactness of the support of the measures and the uniform tightness of their masses.

Lemma 3.5. $(X^n)_{n\in\mathbb{N}^*}$ *is tight in* $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$ *if:*
(i) $\forall T > 0$, $\forall \varepsilon > 0$, $\exists K \subset \mathbb{D}_{\mathbb{R}^d}$ *compact,*

$$\sup_{n\in\mathbb{N}^*} \mathbb{P}\left(\exists t \in [0,T],\, X_t^n(K_T^c) > \varepsilon\right) \le \varepsilon,$$

where K_T^c *is the complement set of*

$$K_T = \left\{y^t, y^{t-} \;\mid\; y \in K,\, t \in [0,T]\right\} \subset \mathbb{D}_{\mathbb{R}^d}. \tag{3.9}$$

(ii) $\forall G_g$ *of the form (3.1), the family* $((\langle X_t^n, G_g\rangle)_{t\in\mathbb{R}_+})_{n\in\mathbb{N}^*}$ *is uniformly tight in* $\mathbb{D}_{\mathbb{R}_+}$.

Sketch of proof. Recall that $\mathbb{D}_{\mathbb{R}^d}$ embedded with the Skorohod topology is a polish space. By Lemma 7.6 in Dawson and Perkins [16], the set K_T is compact in $\mathbb{D}_{\mathbb{R}^d}$. The set

$$\mathfrak{K}(\varepsilon) = \{\mu \in \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}) \mid \mu(K_T^c) \le \varepsilon\}$$

is then relatively compact in $\mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d})$ by the Prohorov theorem, since it corresponds to a tight family of measures. We can then rewrite Point (i) of Theorem 3.5 in $\forall T > 0$, $\forall \varepsilon > 0$, $\exists \mathfrak{K} \subset \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d})$ relatively compact,

$$\sup_{n\in\mathbb{N}^*} \mathbb{P}\left(\exists t \in [0,T],\ \bar{X}^n_t \notin \mathfrak{K}\right) \le \varepsilon.$$

Moreover the class of functions G_g separates the point and is closed under addition. Thus Points (i) and (ii) of Lemma 3.5 allow us to apply the tightness result of Jakubowski (Theorem 4.6 [31]) and ensure that the sequence of the laws of $(\bar{X}^n)_{n\in\mathbb{N}^*}$ is uniformly tight in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$. ■

Proof of Proposition 3.4. We divide the proof into several steps.

Step 1 Firstly, we consider Point (ii) of Lemma 3.5. Let $T > 0$ and let G_g be of the form (3.1).

Since for every $t \in [0,T]$ and every $A > 0$

$$\mathbb{P}(|\langle X^n_t, G_g\rangle| > A) \le \frac{\|G\|_\infty \sup_{n\in\mathbb{N}^*} \mathbb{E}(\langle X^n_t, 1\rangle)}{A}, \tag{3.10}$$

which tends to 0 when A tends to infinity thanks to (2.16). This proves the tightness of the family of marginals $\langle X^n_t, G_g\rangle$ for $n \in \mathbb{N}^*$.

Then, we use the Aldous and Rebolledo criteria (*e.g.* [32]). For $\varepsilon > 0$ and $\eta > 0$, it is satisfied if there exists $n_0 \in \mathbb{N}^*$ and $\delta > 0$ such that for all $n > n_0$ and for all stopping times $S_n < T_n < (S_n + \delta) \wedge T$:

$$\mathbb{P}(|A^{n,G_g}_{T_n} - A^{n,G_g}_{S_n}| > \eta) \le \varepsilon \quad \text{and} \quad \mathbb{P}(|\langle M^{n,G_g}\rangle_{T_n} - \langle M^{n,G_g}\rangle_{S_n}| > \eta) \le \varepsilon \tag{3.11}$$

where A^{n,G_g} denotes the finite variation process in the r.h.s. of (2.17).

Let us begin with some estimates. We fix $t \in [0,T]$, $h \in \mathbb{R}^d$ and a path $y \in \mathbb{D}_{\mathbb{R}^d}$ stopped at t. Using Taylor-Lagrange formula, there exists $\theta_{y,t,h} \in (0,1)$ such that:

$$\begin{aligned} G_g(y|t|y_{t_-} + h) - G_g(y) &= G\Big(\int_0^T g(s, (y|t|y_{t_-} + h)_s)ds\Big) - G\Big(\int_0^T g(s, y_s)ds\Big) \\ &= G'\Big(\int_0^T g(s, y_s)ds\Big)\Lambda(y,t,h) + \frac{1}{2}G''\Big(\int_0^T g(s,y_s)ds + \theta_{y,t,h}\Lambda(y,t,h)\Big)\Lambda(y,t,h)^2 \end{aligned} \tag{3.12}$$

where

$$\Lambda(y,t,h) = \int_t^T \Big(g(s, y_{t_-} + h) - g(s, y_{t_-})\Big)ds \tag{3.13}$$

converges to zero when h tends to zero. Using Taylor-Lagrange formula again for the integrand in $\Lambda(y,t,h)$, there exists a family $\eta_{y,t,h,s}$ of $(0,1)$ such that:

$$\Lambda(y,t,h) = \int_t^T \Big(h \cdot \nabla_x g(s, y_{t_-}) + \frac{1}{2}\ {}^t h\ [\text{Hess } g(s, y_{t_-} + \eta_{y,t,h,s} h)]\ h\Big)ds. \tag{3.14}$$

Using (3.12) and (3.14), we integrate $G_g(y|t|y_{t_-}+h)-G_g(y)$ with respect to $K^n(dh)$. Since $k^n(h)$ is the density of the Gaussian distribution of mean 0 and covariance $\sigma^2\mathbf{Id}/n$, integrals of odd powers and cross-products of the components of h vanish. Thus:

$$\int_{\mathbb{R}^d}\Big(G_g(y|t|y_{t_-}+h)-G_g(y)\Big)K^n(dh)=G'\Big(\int_0^T g(s,y_s)ds\Big)\frac{\sigma^2 p}{2n}\int_t^T \Delta_x g(s,y_{t_-}+\eta_{t,y,h,s}h)ds$$
$$+\frac{p\sigma^2}{2n}G''\Big(\int_0^T g(s,y_s)ds+\theta_{y,t,h}\Lambda(y,t,h)\Big)\sum_{i=1}^d\Big(\int_0^T \partial_{x_i}g(s,y_{t_-})ds\Big)^2+\frac{C}{n^2}$$

where C is a constant that depends on G, g, σ^2 and p but not on n. Therefore:

$$\lim_{n\to+\infty} n\Big|\int_{\mathbb{R}^d}\Big(G_g(y|u|y_u+h)-G_g(y)\Big)K^n(dh)-\frac{\sigma^2 p}{2n}D^2G_g(u,y)\Big|=0. \tag{3.15}$$

Noting that for G and g in $\mathcal{C}_b^2$, D^2G_g is bounded from the definition (3.1), we obtain the following upper bound:

$$\mathbb{E}(|A^{n,G_g}_{T_n}-A^{n,G_g}_{S_n}|)\leq\delta\,\Big[\Big(\bar{R}\frac{p\sigma^2}{2}(\|D^2G_g\|_\infty+1)+(\bar{B}+\bar{D})\|G\|_\infty\Big)\sup_{n\in\mathbb{N}^*}\mathbb{E}(\sup_{t\in[0,T]}\langle X^n_t,1\rangle)$$
$$+\|G\|_\infty\bar{U}\,\nu_d[0,T]\sup_{n\in\mathbb{N}^*}\mathbb{E}(\sup_{t\in[0,T]}\langle X^n_t,1\rangle^2)\Big]. \tag{3.16}$$

For the quadratic variation process:

$$\mathbb{E}(|\langle M^{n,G_g}\rangle_{T_n}-\langle M^{n,G_g}\rangle_{S_n}|)$$
$$\leq\|G\|^2_\infty\delta\,\Big[\Big(2\bar{r}+\frac{\bar{b}+\bar{d}}{n}\Big)\sup_{n\in\mathbb{N}^*}\mathbb{E}(\sup_{t\in[0,T]}\langle X^n_t,1\rangle)+\frac{\bar{U}\;\nu_d[0,T]}{n}\sup_{n\in\mathbb{N}^*}\mathbb{E}(\sup_{t\in[0,T]}\langle X^n_t,1\rangle^2)\Big]. \tag{3.17}$$

We thus obtain (3.11) by applying the Markov inequality and using the moment estimates of Proposition 2.5.

Step 2 Let us now check that Point (i) of Lemma 3.5 is satisfied. We follow here ideas introduced by Dawson and Perkins [16] who proved the tightness of a system of independent historical branching Brownian particles. Here, we have interacting particles. Let $T\in\mathbb{R}_+$ and $\varepsilon>0$.

Let K be a compact set of $\mathbb{D}_{\mathbb{R}^d}$. We denote by $K^t=\{y^t\,|\,y\in K\}\subset\mathbb{D}_{\mathbb{R}^d}$ the set of the paths of K stopped at time t and we recall that K_T defined in (3.9) is the set of the paths of K stopped at any time before time T and of their left-limited paths stopped at the same time. Let us define the stopping time

$$S^n_\varepsilon=\inf\{t\in\mathbb{R}_+\,|\,X^n_t(K^c_T)>\varepsilon\}. \tag{3.18}$$

From this definition:

$$\mathbb{P}\big(\exists t\in[0,T],\,X^n_t(K^c_T)>\varepsilon\big)=\mathbb{P}(S^n_\varepsilon<T). \tag{3.19}$$

Our purpose is to prove that it is possible to choose K and n_0 such that $\sup_{n\geq n_0}\mathbb{P}(S^n_\varepsilon<T)\leq\varepsilon$. We decompose (3.19) by considering the more tractable $X^n_T((K^T)^c)$ and write:

$$\mathbb{P}(S^n_\varepsilon<T)=\mathbb{P}\Big(S^n_\varepsilon<T,\;X^n_T((K^T)^c)>\frac{\varepsilon}{2}\Big)+\mathbb{P}\Big(S^n_\varepsilon<T,\;X^n_T((K^T)^c)\leq\frac{\varepsilon}{2}\Big)$$
$$\leq\frac{2}{\varepsilon}\mathbb{E}\big(X^n_T((K^T)^c)\big)+\mathbb{P}\Big(S^n_\varepsilon<T,\;X^n_T((K^T)^c)\leq\frac{\varepsilon}{2}\Big) \tag{3.20}$$

by using the Markov inequality. We will show in Steps 3 to 5 that there exists $\eta \in (0,1)$ such that for n large enough,

$$\mathbb{P}\Big(S^n_\varepsilon < T,\ X^n_T((K^T)^c) \leq \frac{\varepsilon}{2}\Big) \leq \mathbb{P}(S^n_\varepsilon < T)(1-\eta). \tag{3.21}$$

Together with (3.20), this entails that:

$$\mathbb{P}(S^n_\varepsilon < T) \leq \frac{2\mathbb{E}\big(X^n_T((K^T)^c)\big)}{\varepsilon\eta}. \tag{3.22}$$

In Step 6, we will also prove that the compact set K can be chosen such that

$$\mathbb{E}\big(X^n_T((K^T)^c)\big) < \frac{\varepsilon^2\eta}{2}. \tag{3.23}$$

This will conclude the proof.

Step 3 Let us prove (3.21). Heuristically, the event $\{S^n_\varepsilon < T,\ X^n_T((K^T)^c) \leq \frac{\varepsilon}{2}\}$ means that most of the trajectories that exited K before S^n_ε have died at time T. On the set $\{S^n_\varepsilon < T\}$, $y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}$ implies $y^T \notin K^T$. Indeed, if the paths stopped at time S^n_ε do not belong to K, then this is also not the case when we stop them at $T > S^n_\varepsilon$. Thus on $\{S^n_\varepsilon < T\}$:

$$X^n_T((K^T)^c) \geq X^n_T(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\}). \tag{3.24}$$

Hence:

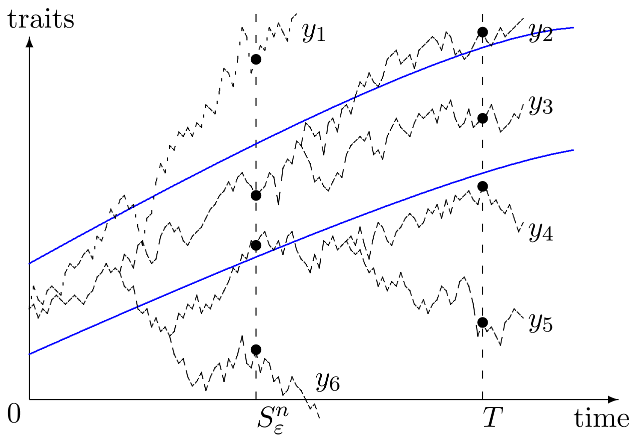

Figure 2: *The compact K corresponds, here, to the region between the two lines. The paths which are drawn correspond to the support of X^n_T. $X^n_T((K^T)^c)$ counts the trajectories that do not belong to K between 0 and T: here we have y_1, y_2, y_4, y_5 and y_6. The quantity $X^n_T(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\})$ counts the trajectories, at time T, that do not belong to K between 0 and S^n_ε: here, we have y_1, y_4, y_5 and y_6 ; although y_2 does not belong to K between time 0 and T, it belongs to K between 0 and S^n_ε. To obtain the trajectories accounting for $X^n_T(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\})$, we can count the descendents of the 3 points at time S^n_ε corresponding to trajectories $y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}$. We can also check that relation* (3.24) *is satisfied.*

$$\begin{aligned}\mathbb{P}\Big(S^n_\varepsilon < T, X^n_T((K^T)^c) \leq \frac{\varepsilon}{2}\Big) \leq & \mathbb{P}\Big(S^n_\varepsilon < T, X^n_T(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\}) \leq \frac{\varepsilon}{2}\Big) \\ = & \mathbb{E}\Big(\mathbb{1}_{S^n_\varepsilon < T}\mathbb{P}\big(X^n_{S^n_\varepsilon + (T-S^n_\varepsilon)}(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\}) \leq \varepsilon/2 \,|\, \mathcal{F}_{S^n_\varepsilon}\big)\Big).\end{aligned} \tag{3.25}$$

Our purpose is to upper bound the probability under the expectation in the r.h.s. of (3.25) by $(1-\eta)$ with $\eta \in (0,1)$. This term is the probability that the population which descends from

particles which at time S^n_ε satisfy $y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}$ has a size less than $\varepsilon/2$. In view of (3.25), we will work on the set $\{S^n_\varepsilon < T\}$ until the end of the proof.
The mass process $\langle X^n_., 1\rangle$ is a semi-martingale with drift and bracket that can be bounded by linear functions of $\langle X^n_., 1\rangle$ with bounded coefficients and independently of n. Hence, using (2.16) and Markov's inequality, it is possible, for any $\eta > 0$, to choose $N > 0$ sufficiently large such that:

$$\mathbb{P}\Big(\sup_{S^n_\varepsilon \le s \le T} \langle X^n_s, 1\rangle > N \,|\, \mathcal{F}_{S^n_\varepsilon}\Big) < \eta. \tag{3.26}$$

Let us introduce the process $(Z^n_t(dy))_{t\in\mathbb{R}_+}$ of independent particles as follows. The particles of Z^n are started at time S^n_ε with the trajectories of $X^n_{S^n_\varepsilon}$ such that $\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\}$ and we define the initial condition:

$$Z^n_{S^n_\varepsilon}(dy) = \mathbb{1}_{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}} X^n_{S^n_\varepsilon}(dy). \tag{3.27}$$

Their birth and death rates are $nr(t,y)$ and $nr(t,y) + \bar{D} + \bar{U}N$. By a coupling argument, we have:

$$\begin{aligned}
&\mathbb{P}\big(X^n_T(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\}) \le \varepsilon/2 \,|\, \mathcal{F}_{S^n_\varepsilon}\big)\\
&\quad \le \mathbb{P}\Big(\langle Z^n_T, 1\rangle \le \frac{\varepsilon}{2}; \sup_{S^n_\varepsilon \le s \le T} \langle X^n_s, 1\rangle \le N \,|\, \mathcal{F}_{S^n_\varepsilon}\Big) + \mathbb{P}\Big(\sup_{S^n_\varepsilon \le s \le T} \langle X^n_s, 1\rangle > N \,|\, \mathcal{F}_{S^n_\varepsilon}\Big)\\
&\quad \le 1 - \mathbb{P}\Big(\inf_{s\in[S^n_\varepsilon, T]} \langle Z^n_s, 1\rangle > \frac{\varepsilon}{2} \,|\, \mathcal{F}_{S^n_\varepsilon}\Big) + \eta
\end{aligned} \tag{3.28}$$

If we can exhibit $\eta > 0$ such that for n large enough

$$\mathbb{E}\Big(\mathbb{1}_{S^n_\varepsilon < T}\ \mathbb{P}\big(\inf_{s\in[S^n_\varepsilon, T]} \langle Z^n_s, 1\rangle > \frac{\varepsilon}{2} \,|\, \mathcal{F}_{S^n_\varepsilon}\big)\Big) > 2\eta \mathbb{P}(S^n_\varepsilon < T), \tag{3.29}$$

then from (3.25) and (3.28), the r.h.s. of (3.25) is strictly smaller than $(1 - 2\eta + \eta)\mathbb{P}(S^n_\varepsilon < T) = (1-\eta)\mathbb{P}(S^n_\varepsilon < T)$, which proves (3.21).

Step 4 Let us prove (3.29). Notice that on the set $\{S^n_\varepsilon < T\}$:

$$\langle Z^n_{S^n_\varepsilon}, 1\rangle = X^n_{S^n_\varepsilon}(\{y^{S^n_\varepsilon} \notin K^{S^n_\varepsilon}\}) \ge X^n_{S^n_\varepsilon}(\{y^{S^n_\varepsilon} \notin K_{S^n_\varepsilon}\}) \ge X^n_{S^n_\varepsilon}(\{y^{S^n_\varepsilon} \notin K_T\}) = X^n_{S^n_\varepsilon}(K^c_T) > \varepsilon.$$

By coupling arguments (using deletions of particles in the initial condition $Z^n_{S^n_\varepsilon}$), and since we are considering a minoration with an infimum in (3.29), we can consider without restriction that $\langle Z^n_{S^n_\varepsilon}, 1\rangle = ([n\varepsilon]+1)/n$, where $[x]$ denotes the integer part of x.

Now, we establish a diffusion approximation of $\langle Z^n_{S^n_\varepsilon+.}, 1\rangle$ when n is large. We know that for all $t \ge 0$:

$$\langle Z^n_{S^n_\varepsilon+t}, 1\rangle = \langle Z^n_{S^n_\varepsilon}, 1\rangle - \big(\bar{D} + \bar{U}N\big)\int_0^t \langle Z^n_{S^n_\varepsilon+s}, 1\rangle ds + M^{n,Z}_t \tag{3.30}$$

where $M^{n,Z}$ is a square integrable martingale such that for all $s \le t$:

$$\begin{aligned}
2\underline{R}\int_s^t \langle Z^n_{S^n_\varepsilon+u}, 1\rangle du \le \langle M^{n,Z}\rangle_t - \langle M^{n,Z}\rangle_s &= \int_s^t \Big\langle Z^n_{S^n_\varepsilon+u}, 2r(S^n_\varepsilon+u, .) + \frac{\bar{D}+\bar{U}N}{n}\Big\rangle du\\
&\le \big(2\bar{R} + \frac{\bar{D}+\bar{U}N}{n}\big)\int_s^t \langle Z^n_{S^n_\varepsilon+u}, 1\rangle du.
\end{aligned} \tag{3.31}$$

Using Proposition 2.5 and adaptations of (3.16) and (3.17), the laws of $(\langle Z^n_{S^n_\varepsilon+.}, 1\rangle, \langle M^{n,Z}\rangle_.)$ are uniformly tight in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^2_+)$. As a consequence, there exists a subsequence, again denoted by $(\langle Z^n_{S^n_\varepsilon+.}, 1\rangle, \langle M^{n,Z}\rangle_.)_{n\in\mathbb{N}^*}$ by abuse of notation, that converges in distribution to a limit, say $(\mathcal{Z}, \mathcal{A})$ where $\mathcal{Z}$ and $\mathcal{A}$ are necessarily continuous. Let us define on the canonical space

$$\mathcal{N}_t = \mathcal{Z}_t - \varepsilon + \int_0^t (\bar{D} + \bar{U}N)\mathcal{Z}_s ds \tag{3.32}$$

Let $0 \leq s_1 \leq \cdots \leq s_k < s < t$ and let $\phi_1, \cdots \phi_k$ be bounded measurable functions on $\mathbb{R}_+$. We define:

$$\Psi(\mathcal{Z}) = \phi_1(\mathcal{Z}_{s_1}) \cdots \phi_k(\mathcal{Z}_{s_k})\Big\{\mathcal{Z}_t - \mathcal{Z}_s + \int_s^t (\bar{D} + \bar{U}N)\mathcal{Z}_u \, du\Big\}. \tag{3.33}$$

From (3.30), $\mathbb{E}(\Psi(\langle Z^n_{S^n_\varepsilon+.}, 1\rangle)) = 0$. Similarly to Prop. 2.5 (ii), we can prove from the SDE (3.30) that $\mathbb{E}(\sup_{t\in[0,T]}\langle Z^n_{S^n_\varepsilon+t}, 1\rangle^3) < +\infty$ for any $T > 0$. Then the sequence $(\Psi(\langle Z^n_{S^n_\varepsilon+.}, 1\rangle))_{n\in\mathbb{N}^*}$ is uniformly integrable and by the continuity of $\mathcal{Z}$, $\lim_{n\to+\infty} \mathbb{E}\big(\Psi(\langle Z^n_{S^n_\varepsilon+.}, 1\rangle\big) = \mathbb{E}(\Psi(\mathcal{Z}))$. Then we deduce that $\mathbb{E}(\Psi(\mathcal{Z})) = 0$, for all Ψ defined by (3.33). Hence, $\mathcal{N}$ is a continuous square integrable martingale, and Theorem 6.1 p. 341 in Jacod and Shiryaev [28] together with Proposition 2.5 implies that its quadratic variation process is $\langle \mathcal{N}\rangle = \mathcal{A}$.

Moreover, using the Skorokhod representation theorem (see *e.g.* [8] Th. 25.6 p.333), there exist a random sequence $(\widetilde{\mathcal{Z}}^n, \widetilde{\mathcal{A}}^n)_{n\in\mathbb{N}^*}$ and a random couple $(\widetilde{\mathcal{Z}}, \widetilde{\mathcal{A}})$ defined on the same probability space, distributed as $(\langle Z^n_{S^n_\varepsilon+.}, 1\rangle, \langle M^{n,Z}\rangle_.)_{n\in\mathbb{N}^*}$ and $(\mathcal{Z}, \mathcal{A})$, and such that

$$\lim_{n\to+\infty} (\widetilde{\mathcal{Z}}^n, \widetilde{\mathcal{A}}^n) = (\widetilde{\mathcal{Z}}, \widetilde{\mathcal{A}}) \qquad \text{a.s.} \tag{3.34}$$

Then, from (3.31), we have a.s. that for all $0 \leq s \leq t$:

$$2\underline{R}\int_s^t \widetilde{\mathcal{Z}}_u \, du \leq \widetilde{\mathcal{A}}_t - \widetilde{\mathcal{A}}_s \leq 2\bar{R}\int_s^t \widetilde{\mathcal{Z}}_u \, du. \tag{3.35}$$

This implies (see e.g. Rudin [44, Chapter 8]) that $\widetilde{\mathcal{A}}$ is a.s. an absolutely continuous function and that there exists a random $\mathcal{F}_t$-measurable function $\rho(u)$ such that $\forall u \in \mathbb{R}_+$, $2\underline{R} \leq \rho(u) \leq 2\bar{R}$ and:

$$\widetilde{\mathcal{A}}_t = \int_0^t \rho(u)\widetilde{\mathcal{Z}}_u \, du \qquad a.s. \tag{3.36}$$

Then there exists a standard real Brownian motion $(B_t)_{t\in\mathbb{R}_+}$ such that almost surely:

$$\widetilde{\mathcal{N}}_t = \widetilde{\mathcal{Z}}_t - \varepsilon + \int_0^t (\bar{D} + \bar{U}N)\widetilde{\mathcal{Z}}_s ds = \int_0^t \sqrt{\rho(u)\widetilde{\mathcal{Z}}_u} dB_u. \tag{3.37}$$

Now that the diffusive limit for $\langle Z^n_{S^n_\varepsilon+.}, 1\rangle$ has been obtained, let us return to (3.29):

$$\begin{aligned}\mathbb{P}\Big(\inf_{s\in[S^n_\varepsilon,T]}\langle Z^n_s, 1\rangle > \frac{\varepsilon}{2} \,|\, \mathcal{F}_{S^n_\varepsilon}\Big)\mathbb{1}_{S^n_\varepsilon<T} =& \mathbb{P}\Big(\inf_{u\in[0,T-S^n_\varepsilon]}\langle Z^n_{S^n_\varepsilon+u}, 1\rangle > \frac{\varepsilon}{2} \,|\, \mathcal{F}_{S^n_\varepsilon}\Big)\mathbb{1}_{S^n_\varepsilon<T}\\ \geq& \mathbb{P}\Big(\inf_{u\in[0,T]}(y|s|\widetilde{\mathcal{Z}}^n)_{s+u} > \frac{\varepsilon}{2}\Big)\Big|_{y=\langle Z^n_{.\wedge S^n_\varepsilon},1\rangle,\ s=S^n_\varepsilon}\mathbb{1}_{S^n_\varepsilon<T}.\end{aligned} \tag{3.38}$$

Notice that for all y and s

$$\lim_{n\to+\infty} \mathbb{P}\Big(\inf_{u\in[0,T]}(y|s|\widetilde{\mathcal{Z}}^n)_{s+u} > \frac{\varepsilon}{2}\Big) = \mathbb{P}\Big(\inf_{u\in[0,T]}(y|s|\widetilde{\mathcal{Z}})_{s+u} \geq \frac{\varepsilon}{2}\Big) = \mathbb{P}_{y,s}\Big(\inf_{u\in[0,T]}\widetilde{\mathcal{Z}}_u \geq \frac{\varepsilon}{2}\Big), \tag{3.39}$$

where the notation $\mathbb{P}_{y,s}$ reminds that the distribution of $\widetilde{\mathcal{Z}}$ depends on ρ which may itself depend on (y, s). If the limit in (3.39) is positive, we are close to (3.29). However, to conclude with

(3.38), we need some uniformity of the convergence in (3.39) with respect to y and s.

For $\zeta > 0$ and $(z,r) \in \mathbb{D} \times [0,T]$, we denote by $B((z,r),\zeta)$ the open ball centered at (z,r) with radius ζ. There exists $\zeta > 0$ small enough such that for all $(y,s) \in \mathcal{B}((z,r),\zeta)$ and $\mathcal{Z} \in \mathbb{D}$, $\mathbf{d}_{\mathrm{Sk}}\big((y|s|\mathcal{Z}),(z|r|\mathcal{Z})\big) < \varepsilon/4$ (see Proposition B.1 in appendix). As a consequence, for this choice of ζ and all $n \in \mathbb{N}^*$,

$$\mathbb{P}\Big(\inf_{u\in[0,T]}(z|r|\widetilde{\mathcal{Z}}^n)_{r+u} > \frac{3\varepsilon}{4}\Big) \leq \mathbb{P}\Big(\inf_{u\in[0,T]}(y|s|\widetilde{\mathcal{Z}}^n)_{s+u} > \frac{\varepsilon}{2}\Big). \tag{3.40}$$

Let $\xi > 0$ be a small positive number. Since the sequence of laws of $\langle Z^n_\cdot, 1\rangle$ is uniformly tight, there exists a compact set K_ξ of $\mathbb{D}_{\mathbb{R}^d}$ such that for sufficiently large n, $\mathbb{P}\big(\langle Z^n_\cdot,1\rangle \notin K_\xi\big) < \xi$. Since $K_\xi \times [0,T]$ is compact, there exists a finite sequence $(z_i,r_i)_{1\leq i\leq M}$ with $M = M(\xi) \in \mathbb{N}^*$ such that

$$K_\xi \times [0,T] \subset \bigcup_{i=1}^{M(\xi)} B\big((z_i,r_i),\zeta\big).$$

With an argument similar to (3.39), there exists, for each $i \in \{1,\dots,M\}$, an integer n_i such that for all $n \geq n_i$,

$$\mathbb{P}\Big(\inf_{u\in[0,T]}(z_i|r_i|\widetilde{\mathcal{Z}}^n)_{r_i+u} > \frac{3\varepsilon}{4}\Big) > \frac{1}{2}\mathbb{P}_{z_i,r_i}\Big(\inf_{u\in[0,T]}\widetilde{\mathcal{Z}}_u \geq \frac{3\varepsilon}{4}\Big). \tag{3.41}$$

Hence, thanks to (3.40), we obtain that for all $(y,s) \in K_\xi \times [0,T]$ and $n \geq \max_{1\leq i\leq M(\xi)} n_i$,

$$\begin{aligned}\mathbb{P}\Big(\inf_{u\in[0,T]}(y|s|\widetilde{\mathcal{Z}}^n)_{s+u} \geq \frac{\varepsilon}{2}\Big) &\geq \min_{1\leq i\leq M}\mathbb{P}\Big(\inf_{u\in[0,T]}(z_i|r_i|\widetilde{\mathcal{Z}}^n)_{r_i+u} > \frac{3\varepsilon}{4}\Big)\\ &> \min_{1\leq i\leq M(\xi)}\frac{1}{2}\mathbb{P}_{z_i,r_i}\Big(\inf_{u\in[0,T]}\widetilde{\mathcal{Z}}_u \geq \frac{3\varepsilon}{4}\Big) =: 2\eta(\xi).\end{aligned} \tag{3.42}$$

Then, from (3.38) and (3.42), the left hand side of (3.29) is lower bounded as follows:

$$\begin{aligned}&\mathbb{E}\Big(\mathbb{1}_{S^n_\varepsilon<T}\ \mathbb{P}\big(\inf_{s\in[S^n_\varepsilon,T]}\langle Z^n_s,1\rangle > \frac{\varepsilon}{2}\,|\,\mathcal{F}_{S^n_\varepsilon}\big)\Big)\\ &\geq \mathbb{E}\Big(\mathbb{P}\Big(\inf_{u\in[0,T]}(y|s|\widetilde{\mathcal{Z}}^n)_{s+u} > \frac{\varepsilon}{2}\Big)\Big|_{y=\langle Z^n_{\cdot\wedge S^n_\varepsilon},1\rangle,\ s=S^n_\varepsilon}\mathbb{1}_{S^n_\varepsilon<T}\mathbb{1}_{\langle Z^n_\cdot,1\rangle\in K_\xi}\Big)\\ &\geq 2\eta(\xi)\ \mathbb{P}\big(S^n_\varepsilon < T,\ \langle Z^n_\cdot,1\rangle \in K_\xi\big).\end{aligned} \tag{3.43}$$

The term $\mathbb{P}\big(S^n_\varepsilon < T,\ \langle Z^n_\cdot,1\rangle \in K_\xi\big)$ in the right hand side converges to $\mathbb{P}(S^n_\varepsilon < T)$ when ξ tends to zero, and there exists $\xi_0 > 0$ sufficiently small such that this term is larger than $\mathbb{P}(S^n_\varepsilon < T)/2$ for every $\xi < \xi_0$ (in case $\mathbb{P}(S^n_\varepsilon < T) = 0$, the proof is done and this is also true). Thus, for $0 < \xi < \xi_0$, the left hand side in (3.43) is lower bounded by $\eta(\xi)\mathbb{P}(S^n_\varepsilon < T)$. This proves (3.29) provided η is positive, what we show in Step 4.

Step 4 Let us prove that η defined in (3.42) is positive. Since it is a minimum over a finite number of terms, let us consider one of the latter. For this, we consider $(z,r) \in \mathbb{D}_{\mathbb{R}^d}$ and our purpose is to prove that

$$\mathbb{P}_{z,r}\Big(\inf_{u\in[0,T]}\widetilde{\mathcal{Z}}_u \geq \frac{3\varepsilon}{4}\Big) > 0.$$

For $M > 0$, let us define the stopping time $\varsigma_M = \inf\{t \geq 0,\ \widetilde{\mathcal{Z}} \geq M\}$ and let us introduce:

$$\tau_{\varepsilon/2} = \inf\big\{t \geq 0,\ \widetilde{\mathcal{Z}}_t \leq \frac{\varepsilon}{2}\big\} \tag{3.44}$$

such that $\mathbb{P}_{z,r}(\inf_{s\in[0,T]} \widetilde{\mathcal{Z}}_s > \varepsilon/2) = \mathbb{P}_{z,r}(\tau_{\varepsilon/2} > T)$. Our purpose is to prove that the latter quantity is positive. Let $\lambda > 0$. From Itô's formula:

$$e^{\lambda \widetilde{\mathcal{Z}}_{t\wedge \varsigma_M}} = e^{\lambda\varepsilon} + \int_0^{t\wedge\varsigma_M} \big(\lambda^2 \bar{R} - \lambda(\bar{D} + \bar{U}N)\big)\widetilde{\mathcal{Z}}_s e^{\widetilde{\mathcal{Z}}_s} ds + \int_0^{t\wedge\varsigma_M} \lambda\sqrt{\rho(s)\widetilde{\mathcal{Z}}_s} e^{\widetilde{\mathcal{Z}}_s} dB_s.$$

Taking the expectation and choosing N sufficiently large ($N > (\lambda\bar{R} - \bar{D})/\bar{U}$), we obtain that $\mathbb{E}\big(\exp(\lambda \widetilde{\mathcal{Z}}_{t\wedge\varsigma_M})\big) \leq \exp(\lambda\varepsilon)$. From (3.37) and since $2\underline{R} \leq \rho(u) \leq 2\bar{R}$, we can classically show that $\mathbb{E}\big(\sup_{t\in[0,T]} \widetilde{\mathcal{Z}}_t^2\big) < +\infty$, from which we deduce that $\lim_{M\to+\infty} \varsigma_M = +\infty$ a.s. Moreover, it follows by Fatou's lemma that for any $t \in [0,T]$, $\mathbb{E}\big(\exp(\lambda\widetilde{\mathcal{Z}}_t)\big) \leq \exp(\lambda\varepsilon)$ and by Jensen's inequality and Fubini's theorem, we have:

$$\mathbb{E}\Big(e^{\int_0^T \frac{(\bar{D}+\bar{U}N)^2}{2\rho(s)}\widetilde{\mathcal{Z}}_s ds}\Big) \leq \frac{1}{T}\int_0^T \mathbb{E}\big(e^{\frac{(\bar{D}+\bar{U}N)^2 T}{2\underline{R}}\widetilde{\mathcal{Z}}_s}\big) ds \leq \mathbb{E}\big(e^{\frac{(\bar{D}+\bar{U}N)^2 T\varepsilon}{2\underline{R}}}\big) < +\infty. \tag{3.45}$$

Novikov's criterion is satisfied, and applying Girsanov's theorem tells us that under the probability $\mathbb{M}$ such that:

$$\frac{d\mathbb{M}}{d\mathbb{P}_{z,r}}\Big|_{\mathcal{F}_t} = \exp\Big(-\int_0^t \frac{(\bar{D}+\bar{U}N)\sqrt{\widetilde{\mathcal{Z}}_u}}{\sqrt{\rho(u)}} dB_u - \frac{1}{2}\int_0^t \frac{(\bar{D}+\bar{U}N)^2 \widetilde{\mathcal{Z}}_u}{\rho(u)} du\Big), \tag{3.46}$$

$\widetilde{\mathcal{Z}}$ is a martingale started at ε. Then we have:

$$\begin{aligned}
&\mathbb{M}(\tau_{\varepsilon/2} \leq T) + \mathbb{M}(\tau_{\varepsilon/2} > T) = 1\\
&\frac{\varepsilon}{2}\mathbb{M}(\tau_{\varepsilon/2} \leq T) + \mathbb{E}^{\mathbb{M}}\big(\widetilde{\mathcal{Z}}_T \mathbb{1}_{\tau_{\varepsilon/2} > T}\big) = \varepsilon.
\end{aligned}$$

If $\mathbb{M}(\tau_{\varepsilon/2} \leq T) = 1$, this yields thus a contradiction since we would obtain $\varepsilon/2 = \varepsilon$ for the second equation. Thus, $\mathbb{M}(\tau_{\varepsilon/2} \leq T) < 1$ and $\mathbb{P}_{z,r}(\tau_{\varepsilon/2} \leq T) < 1$. This shows that $\eta > 0$.

Step 6 It now remains to prove (3.23). We follow Dawson and Perkins [16, Lemma 7.3]. For $n \in \mathbb{N}^*$, we can exhibit, by a coupling argument, a process $\widetilde{X}^n$ constituted of independent particles with birth rate $nr(t,y) + b(t,y)$ and death rate $nr(t,y)$, started at X_0^n and which dominates X^n. In particular, for $T > 0$ and for any compact set $K \subset \mathbb{D}_{\mathbb{R}^d}$, $\mathbb{E}(X_T^n((K^T)^c)) \leq \mathbb{E}(\widetilde{X}_T^n((K^T)^c))$.

The tree underlying $\widetilde{X}^n$ can be obtained by pruning a Yule tree with traits in $\mathbb{R}^d$, where a particle of lineage y at time t gives two offspring at rate $2nr(t,y) + b(t,y)$. One has lineage y and the other has lineage $(y|t|y+h)$ where h is drawn in the distribution $K^n(dh)$. Using Harris-Ulam-Neveu's notation to label the particles (see e.g. Dawson [14]), we denote by $Y^{n,\alpha}$ for $\alpha \in \mathcal{I} = \bigcup_{m=0}^{+\infty}\{0,1\}^{m+1}$ the lineage of the particle with label α. Particles are exchangeable and the common distribution of the process $Y^{n,\alpha}$ is the one of a pure jump process of $\mathbb{R}^d$, where the jumps occur at rate $2nr(t,y) + b(t,y)$ and where the jump sizes are distributed according to the probability measure $\frac{1}{2}\delta_0(ds) + \frac{1}{2}K^n(dh)$. We denote by $\mathbb{P}_x^n$ its distribution starting from $x \in \mathbb{R}^d$. It is standard to prove that the family of laws $(\mathbb{P}_x^n,\ n \in \mathbb{N}^*,\ x \in C)$ is tight as soon as C is compact set of $\mathbb{R}^d$.

At each node of the Yule tree an independent pruning is made: the offspring are kept with probability $(nr(t,y)+b(t,y))/(2nr(t,y)+b(t,y))$ and erased with probability $nr(t,y)/(2nr(t,y)+b(t,y))$. Let us denote by V_t the set of individuals alive at time t and write $\alpha \succ i$ to say that the individual α is a descendent of the individual i:

$$\begin{aligned}\mathbb{E}\big(\widetilde{X}^n_T((K^T)^c)\big) =&\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{N_0^n}\sum_{\alpha\succ i}\mathbb{E}\Big(\mathbb{1}_{Y^{n,\alpha}\notin K^T}\mathbb{1}_{\alpha\in V_T}\Big)\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{N_0^n}\mathbb{P}^n_{X_0^i}\big((K^T)^c\big)\mathbb{E}\Big(\sum_{\alpha\succ i}\mathbb{1}_{\alpha\in V_T}\Big)\right)\\ \leq&\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{N_0^n}\mathbb{P}^n_{X_0^i}\big((K^T)^c\big)e^{\bar{B}T}\right) = e^{\bar{B}T}\int_{\mathbb{D}_{\mathbb{R}^d}} X_0^n(dy)\mathbb{P}^n_{y_0}\big((K^T)^c\big),\end{aligned}$$

where $e^{\bar{B}T}$ is an upper bound of the mean population size at T that descends from a single initial individual, when the growth rate is bounded by $\bar{B}$. For each $\varepsilon>0$ there exists a compact set C of $\mathbb{R}^d$ and a compact set K of $\mathbb{D}_{\mathbb{R}^d}$ such that

$$\sup_{n\in\mathbb{N}^*} X_0^n(C^c)\leq\varepsilon \quad\text{ and }\quad \sup_{n\in\mathbb{N}^*}\sup_{y_0\in C}\mathbb{P}^n_{y_0}\big((K^T)^c\big)\leq\varepsilon,$$

which concludes the proof. ■

3.3 Identification of the limiting values

Let us denote by $\bar{X}\in\mathcal{C}(\mathbb{R}_+,\mathcal{M}_F(\mathbb{D}))$ a limiting process of $(\bar{X}^n)_{n\in\mathbb{N}^*}$. Our purpose here is to characterize the limiting value via the martingale problem that appears in Proposition 3.3. Notice that the limiting process $\bar{X}$ is necessarily almost surely continuous as:

$$\sup_{t\in\mathbb{R}_+}\sup_{\varphi,\,\|\varphi\|_\infty\leq 1}|\langle\bar{X}^n_t,\varphi\rangle-\langle\bar{X}^n_{t_-},\varphi\rangle|\leq\frac{1}{n}. \tag{3.47}$$

For the proof of Proposition 3.3, we will need the following Proposition, which establish the uniqueness of the solution of (3.6)-(3.7). Since the limiting process takes its values in $\mathcal{C}([0,T],\mathcal{M}_F(\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d)))$, we will use the test functions (3.3) instead of (3.1).

Proposition 3.6. *(i) The solutions of the martingale problem* (3.6)-(3.7) *also solve the following martingale problem, where φ is a test function of the form* (3.3) *and $\widetilde{\Delta}$ has been defined in* (3.4)*:*

$$M^\varphi_t=\langle\bar{X}_t,\varphi\rangle-\langle\bar{X}_0,\varphi\rangle-\int_0^t\int_{\mathbb{D}_{\mathbb{R}^d}}\Big(p\,r(s,y)\frac{\sigma^2}{2}\,\widetilde{\Delta}\varphi(s,y)+\gamma(s,y,\bar{X}^s)\varphi(y)\Big)\bar{X}_s(dy)\,ds \tag{3.48}$$

is a square integrable martingale with quadratic variation:

$$\langle M^\varphi\rangle_t=\int_0^t\int_{\mathbb{D}_{\mathbb{R}^d}}2\,r(s,y)\sigma^2\varphi^2(y)\bar{X}_s(dy)\,ds. \tag{3.49}$$

(ii) There exists a unique solution to the martingale problem (3.48)-(3.49)*.*
(iii) There exists a unique solution to the martingale problem (3.6)-(3.7)*.*

In the course of the proof, we will need the following lemma, which proof uses standard arguments with $r(t,y)$ depending on all the trajectory (see (2.4)):

Lemma 3.7. *Let us consider the following SDE on $\mathbb{R}^d$ driven by a standard Brownian motion B:*

$$Y_t=Y_0+\int_0^t\sqrt{\sigma^2p\,r(s,Y^s)}dB_s. \tag{3.50}$$

There exists a unique solution to (3.50)*.*

Proof of Prop. 3.6. It is clear that (iii) follows from (i) and (ii).

Let us begin with the proof of (i). Let us consider a function φ of the form (3.3) and let us assume without restriction that the functions g_j's are positive. First, let us show that (3.48) defines a martingale. Proceeding as for the proof of Th. 5.6 in [26], let us introduce the following function. For $k \in \mathbb{N}^*$, let $0 \leq s_1 \leq \dots s_k < s < t$ and let $\phi_1, \dots \phi_k$ be bounded measurable functions on $\mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d})$. We define for $X \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$:

$$\Psi(X) = \phi_1(X_{s_1}) \dots \phi_k(X_{s_k}) \Big\{ \langle X_t, \varphi \rangle - \langle X_s, \varphi \rangle - \int_s^t \int_{\mathbb{D}_{\mathbb{R}^d}} \big(pr(u,y) \frac{\sigma^2}{2} \widetilde{\Delta}\varphi(y) \\ + \gamma(u, y, \bar{X}^u)) \varphi(y) \Big) \bar{X}_u(dy)\, du \Big\}. \quad (3.51)$$

Let us prove that $\mathbb{E}(\Psi(\bar{X})) = 0$. We consider, for $q \in \mathbb{N}^*$, test functions $\varphi_q(y) = G_{g_q}(y)$ with $G(x) = \exp(x)$ and $g_q(s,x) = \sum_{j=1}^m \log(g_j(x)) k^q(t_j - s)$. From (3.6)-(3.7) and (2.16), we have that

$$M_t^{\varphi_q} = \langle \bar{X}_t, \varphi_q \rangle - \langle \bar{X}_0, \varphi_q \rangle - \int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} \Big(p\, r(s,y) \frac{\sigma^2}{2}\ D^2 \varphi_q(s,y) \\ + \gamma(s, y, \bar{X}^s) \varphi_q(y) \Big) \bar{X}_s(dy)\, ds \quad (3.52)$$

is a square integrable martingale, hence uniformly integrable. The latter property together with Lemma 3.2 implies that $\phi_1(X_{s_1}) \dots \phi_k(X_{s_k})(M_t^{\varphi_q} - M_s^{\varphi_q})$ converges to $\Psi(\bar{X})$ in L^1 and that $\mathbb{E}(\Psi(\bar{X})) = 0$.

Now, let us show that the bracket of M^φ is given by (3.49). We first check that the following process is a martingale:

$$\langle \bar{X}_t, \varphi \rangle^2 - \langle \bar{X}_0, \varphi \rangle^2 - \int_0^t \int_{\mathbb{D}} 2r(s,y) \sigma^2 \varphi^2(y) \bar{X}_s(dy) ds - \int_0^t 2 \langle \bar{X}_s, \varphi \rangle \int_{\mathbb{D}} \Big(pr(s,y) \frac{\sigma^2}{2} \widetilde{\Delta}\varphi(y) \\ + \gamma(s, y, \bar{X}^s) \varphi(y) \Big) \bar{X}_s(dy)\, ds. \quad (3.53)$$

The computation is similar to the way we proved that (3.48) is a martingale in the beginning of this proof, by considering the analogous martingale problem with tests functions φ_q's. Then, using Itô's formula and (3.48),

$$\langle \bar{X}_t, \varphi \rangle^2 - \langle \bar{X}_0, \varphi \rangle^2 - \langle M^\varphi \rangle_t - \int_0^t 2 \langle \bar{X}_s, \varphi \rangle \int_{\mathbb{D}} \Big(pr(s,y) \frac{\sigma^2}{2} \widetilde{\Delta}\varphi(y) \\ + \gamma(s, y, \bar{X}^s) \varphi(y) \Big) \bar{X}_s(dy)\, ds \quad (3.54)$$

is a martingale. Comparing (3.53) and (3.54), we obtain (3.49).

The proof of (ii) is now separated into several steps. Let $\mathbb{P}$ be a solution of the martingale problem (3.6)-(3.7) and let $\bar{X}$ be here the canonical process of $\mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$. We first use Dawson-Girsanov's theorem (see [15, Section 5], [24, Theorem 2.3]) to get rid of the non-linearities. Their result applies because:

$$\mathbb{E}\Big(\int_0^T \int_{\mathbb{D}_{\mathbb{R}^d}} \gamma^2(s, y, \bar{X}^s) \bar{X}_s(dy)\ ds \Big) < +\infty.$$

Indeed, γ is bounded and $\mathbb{E}\Big(\sup_{t\in[0,T]}\langle\bar{X}_t,1\rangle^2\Big) < +\infty$ by taking (2.16) to the limit. Hence, there exists a probability measure $\mathbb{Q}$ on $\mathcal{C}(\mathbb{R}_+,\mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$ equivalent to $\mathbb{P}$ such that under $\mathbb{Q}$, and for every test function φ of the form (3.3), the process

$$\widetilde{M}_t^{\varphi} = \langle\bar{X}_t,\varphi\rangle - \langle\bar{X}_0,\varphi\rangle - \int_0^t\int_{\mathbb{D}_{\mathbb{R}^d}}\frac{pr(s,y)\sigma^2}{2}\widetilde{\Delta}\varphi(s,y)\bar{X}_s(dy)ds, \tag{3.55}$$

is a martingale with quadratic variation (3.7). Thus, if there is uniqueness of the probability measure $\mathbb{Q}$ which solves the martingale problem (3.55)-(3.7) we will deduce the uniqueness of the solution of the martingale problem (3.6)-(3.7).

Let us now prove that the Laplace transform of $\bar{X}$ under $\mathbb{Q}$ is uniquely characterized using SDE Y (3.50). We associate with Y its path-process $W\in\mathcal{C}(\mathbb{R}_+,\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d))$ defined by:

$$W_t = (Y_{t\wedge s})_{s\in\mathbb{R}_+}. \tag{3.56}$$

The path-process W is not homogeneous but it is however a strong Markov process with semi-group defined for all $s\le t$ and all $\varphi\in\mathcal{C}_b(\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d),\mathbb{R})$ by:

$$S_{s,t}\varphi(y) = \mathbb{E}^{\mathbb{Q}}\big(\varphi(W_t)\,|\,W_s = y^s\big). \tag{3.57}$$

Moreover, the infinitesimal generator $\widetilde{A}$ of W at time t is defined for all φ as in (3.3) by:

$$\widetilde{A}\varphi(t,y) = \frac{p\sigma^2}{2}r(t,y)\widetilde{\Delta}\varphi(t,y). \tag{3.58}$$

Then it can be shown that the log-Laplace functional of $\bar{X}_t$ under the probability $\mathbb{Q}$, $L(s,t,y,\varphi) = \mathbb{E}^{\mathbb{Q}}\big(\exp(-\langle\bar{X}_t,\varphi\rangle)\,|\,\bar{X}_s=\delta_{y^s}\big)$ satisfies

$$L(s,t,y,\varphi) = e^{-V_{s,t}\varphi(y)} \tag{3.59}$$

where $V_{s,t}\varphi(y)$ solves:

$$\begin{aligned} V_{s,t}\varphi(y) =& \mathbb{E}\Big(\varphi(W_t) - \int_s^t\frac{p\sigma^2}{2}r(u,W_u)\big(V_{u,t}\varphi(W_u)\big)^2du\,|\,W_s=y^s\Big)\\ =& S_{s,t}\varphi(y) - \int_s^t\frac{p\sigma^2}{2}S_{s,u}\Big(r(u,.)\big(V_{u,t}\varphi(.)\big)^2\Big)(y)du. \end{aligned} \tag{3.60}$$

Adapting Theorem 12.3.1.1 of [14, p.207], there exists a unique solution to (3.60). Indeed, let V^1 and V^2 be two solutions. From (3.60), we see that for $i\in\{1,2\}$:

$$\sup_{s,t,y}|V^i_{s,t}\varphi(y)| \le \sup_y|\varphi(y)| = \|\varphi\|_\infty. \tag{3.61}$$

We have:

$$\begin{aligned} |V^2_{s,t}\varphi(y) - V^1_{s,t}\varphi(y)| =& \Big|\frac{p\sigma^2}{2}\int_s^t S_{s,u}\Big(r(u,.)\big((V^2_{u,t}\varphi(.))^2 - (V^1_{u,t}\varphi(.))^2\big)\Big)(y)du\Big|\\ \le& \frac{p\sigma^2}{2}\int_s^t S_{s,u}\Big(r(u,.)2\|\varphi\|_\infty\big|V^2_{u,t}\varphi(.) - V^1_{u,t}\varphi(.)\big|\Big)(y)du\\ \le& p\sigma^2\|\varphi\|_\infty\bar{R}\int_s^t S_{s,u}\Big(\big|V^2_{u,t}\varphi - V^1_{u,t}\varphi\big|\Big)(y)du. \end{aligned}$$

We conclude with Dynkin's generalized Gronwall inequality (see *e.g.* [14, Lemma 4.3.1]).

In conclusion, the Laplace transform of $\bar{X}_t$ is uniquely characterized for every $t>0$ by $\mathbb{E}_{\bar{X}_0}\big(\exp(-\langle\bar{X}_t,\varphi\rangle)\big) = \exp(-\langle\bar{X}_0,V_{0,t}\varphi\rangle)$. Thus, the one-marginal distributions of the martingale problem (3.55)-(3.7) are uniquely determined and thus, there exists a unique solution to (3.55)-(3.7). ∎

It is now time to turn to the:

Proof of Theorem 3.3. Let $\bar{X} \in \mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$ be a limiting process of the sequence $(\bar{X}^n)_{n\in\mathbb{N}^*}$ and let us denote by $(\bar{X}^n)_{n\in\mathbb{N}^*}$ again the subsequence that converges in law to $\bar{X}$. Since the limiting process is continuous, the convergence holds in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$ embedded with the Skorohod topology, but also with the uniform topology for all $T > 0$ (*e.g.* [7]).

The aim is to identify the martingale problem solved by the limiting value $\bar{X}$. We will see that it satisfies (3.6)-(3.7) which admits a unique solution by Proposition 3.6. This will conclude the proof.

First, we show that (3.6) defines a martingale by proceeding again as for the proof of Th. 5.6 in [26]. For $k \in \mathbb{N}^*$, let $0 \leq s_1 \leq \dots s_k < s < t$ and let $\phi_1, \dots \phi_k$ be bounded measurable functions on $\mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d})$. Let $G \in \mathcal{C}^2_b(\mathbb{R}, \mathbb{R})$, $g \in \mathcal{C}^{0,2}_b(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$ and G_g be functions as in (3.1). We define for $X \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}^d}))$:

$$\begin{aligned}\Phi(X) = \phi_1(X_{s_1}) \dots \phi_k(X_{s_k}) \Big\{ \langle X_t, G_g\rangle - \langle X_s, G_g\rangle - \int_s^t \int_{\mathbb{D}_{\mathbb{R}^d}} \Big(pr(u,y) \frac{\sigma^2}{2} D^2 G_g(u,y) \\ + \gamma(u,y,\bar{X}_u) G_g(y)\Big) \bar{X}_u(dy)\, du \Big\}. \quad (3.62)\end{aligned}$$

Let us prove that $\mathbb{E}(\Phi(\bar{X})) = 0$. From (3.6),

$$\begin{aligned}0 =& \mathbb{E}\Big(\phi_1(\bar{X}^n_{s_1}) \dots \phi_k(\bar{X}^n_{s_k}) \big(M^{n,G_g}_t - M^{n,G_g}_s\big)\Big) \\ =& \mathbb{E}\big(\Phi(\bar{X}^n)\big) + \mathbb{E}\Big(\phi_1(\bar{X}^n_{s_1}) \dots \phi_k(\bar{X}^n_{s_k}) \big(A_n + B_n\big)\Big) \quad (3.63)\end{aligned}$$

where:

$$\begin{aligned}A_n =& \int_s^t \int_{\mathbb{D}_{\mathbb{R}^d}} r(u,y) \Big\{ n \Big(\int_{\mathbb{R}^d} G_g(y|u|y_u + h) K^n(dh) - G_g(y)\Big) \quad (3.64)\\ & - \frac{p\sigma^2}{2} D^2 G_g(u,y)\Big\} \bar{X}^n_u(dy)\, du \\ B_n =& \int_s^t \int_{\mathbb{D}_{\mathbb{R}^d}} b(u,y) \int_{\mathbb{R}^d} \Big(G_g(y|u|y_u + h) - G_g(y)\Big) K^n(dh)\, \bar{X}^n_u(dy)\, du.\end{aligned}$$

As $\bar{X}$ is continuous, Φ is a.s. continuous at $\bar{X}$. Moreover $|\Phi(X)| \leq C\big(\sup_{s\leq t}\langle X_s, 1\rangle + \sup_{s\leq t}\langle X_s, 1\rangle^2\big)$. From this and Prop. 2.5, we deduce that $(\Phi(\bar{X}^n))_{n\in\mathbb{N}^*}$ is a uniformly integrable sequence such that

$$\lim_{n\to+\infty} \mathbb{E}\big(\Phi(\bar{X}^n)\big) = \mathbb{E}\big(\Phi(\bar{X})\big). \quad (3.65)$$

Estimates (2.16) and (3.15) imply that A_n and B_n defined in (3.64) converge in L^1 to zero. Hence (3.63) and (3.65) entail the desired result. The computation of the bracket of the martingale M^{G_g} is classical (see for instance [26]). The proof is complete. ∎

4 Distributions of the genealogies

An important question is to gather information on the lineages of individuals alive in the population.

First, remark that the construction in Steps 2 and 3 of Proposition 3.6 re-establish a representation result proved by Perkins [41, Th. 5.1 p.64]: under $\mathbb{Q}$, $(\bar{X}_t)_{t\in\mathbb{R}_+}$ has the same distribution as the historical superprocess $(K_t)_{t\in\mathbb{R}_+}$ that is the unique solution of:

$$Y_t(y) = Y_0(y) + \int_0^t \sqrt{\sigma^2 pr(s, Y^s(y))}dy_s \tag{4.1}$$

$$K_0 = \bar{X}_0, \qquad \langle K_t, \varphi\rangle = \int_{\mathbb{D}_{\mathbb{R}^d}} \varphi(Y(y)^t)H_t(dy) \tag{4.2}$$

where $(H_t(dy))_{t\in\mathbb{R}_+}$ is under $\mathbb{Q}$ a historical Brownian superprocess (see [16]), and for φ in a sufficiently large class of test functions, of the form (3.3) for instance.

4.1 Lineages drawn at random

For $t > 0$, $\bar{X}_t$ is a random measure on $\mathbb{D}_{\mathbb{R}^d}$ and its restriction to $\mathbb{D}([0,t],\mathbb{R}^d)$ correctly renormalized gives the distribution of the lineage of an individual chosen at random at time t. Let us define $\mu_t(dy) = \bar{X}_t(dy)/\langle \bar{X}_t, 1\rangle$ such that for all measurable test function φ on $\mathbb{D}_{\mathbb{R}^d}$:

$$\langle \mu_t, \varphi\rangle = \frac{\langle \bar{X}_t, \varphi\rangle}{\langle \bar{X}_t, 1\rangle}. \tag{4.3}$$

For instance, choosing $\varphi(y) = \mathbb{1}_A(y)$ for a measurable subset $A \subset \mathbb{D}_{\mathbb{R}^d}$, we obtain the proportion of paths belonging to A under the random probability measure μ_t. Studying such random probability measure remains unfortunately a difficult task and we will also consider its intensity probability measure $\mathbb{E}\mu_t$ defined for all test function φ as: $\langle \mathbb{E}\mu_t, \varphi\rangle = \mathbb{E}(\langle \mu_t, \varphi\rangle)$. This approach has been used for cases where the branching property holds in [2, 21, 25] for instance.

Proposition 4.1. *For $t > 0$, a test function φ as in* (3.3) *and μ_t defined in* (4.3)*:*

$$\begin{aligned}\mathbb{E}\big(\langle \mu_t, \varphi\rangle\big) =& \mathbb{E}\big(\langle \mu_0, \varphi\rangle\big) + \mathbb{E}\Big(\int_0^t \langle \mu_s, pr(s,.)\sigma^2\widetilde{\Delta}\varphi(s,.)\rangle ds\Big)\\ &+\mathbb{E}\Big(\int_0^t \big(\langle \mu_s, \gamma(s,.,\bar{X}_s)\varphi(.)\rangle - \langle \mu_s,\varphi\rangle\langle \mu_s, \gamma(s,.,\bar{X}_s)\rangle\big)ds\Big)\\ &+\mathbb{E}\Big(\int_0^t \frac{2\sigma^2}{\langle \bar{X}_s, 1\rangle}\big(\langle \mu_s,\varphi\rangle\langle \mu_s, r(s,.)\rangle - \langle \mu_s, r(s,.)\varphi(.)\rangle\big)ds\Big)\end{aligned} \tag{4.4}$$

Proof. We consider (3.6)-(3.7) and Itô's formula:

$$\begin{aligned}\mathbb{E}\big(\langle \mu_t, \varphi\rangle\big) =& \mathbb{E}\Big(\frac{\langle \bar{X}_t, \varphi\rangle}{\langle \bar{X}_t, 1\rangle}\Big) = \langle \mu_0, \varphi\rangle + \mathbb{E}\Big(\int_0^t \int_{\mathbb{D}_{\mathbb{R}^d}} \frac{pr(s,y)\sigma^2\widetilde{\Delta}\varphi(s,y) + \gamma(s,y,\bar{X}^s)\varphi(y)}{\langle \bar{X}_s, 1\rangle}\bar{X}_s(dy)\,ds\Big)\\ &-\mathbb{E}\Big(\int_0^t\int_{\mathbb{D}_{\mathbb{R}^d}} \frac{\langle \bar{X}_s, \varphi\rangle}{\langle \bar{X}_s, 1\rangle^2}\gamma(s,y,\bar{X}^s)\bar{X}_s(dy)\,ds\Big)\\ &+\mathbb{E}\Big(\frac{1}{2}\Big[\int_0^t\int_{\mathbb{D}_{\mathbb{R}^d}} \frac{2\langle \bar{X}_s, \varphi\rangle}{\langle \bar{X}_s, 1\rangle^3}2r(s,y)\sigma^2\bar{X}_s(dy)\,ds - 2\int_0^t\int_{\mathbb{D}_{\mathbb{R}^d}} \frac{1}{\langle \bar{X}_s, 1\rangle^2}2r(s,y)\sigma^2\varphi(y)\bar{X}_s(dy)\,ds\Big]\Big)\\ =&\mathbb{E}\big(\langle \mu_0,\varphi\rangle\big) + \mathbb{E}\Big(\int_0^t \langle \mu_s, pr(s,.)\sigma^2\widetilde{\Delta}\varphi(s,.) + \gamma(s,.,\bar{X}_s)\varphi(.)\rangle ds\Big)\\ &+\mathbb{E}\Big(\int_0^t \langle \mu_s,\varphi\rangle\langle \mu_s, \gamma(s,.,\bar{X}_s)\rangle ds\Big) + \mathbb{E}\Big(\int_0^t \frac{\langle \mu_s,\varphi\rangle\langle \mu_s, 2r(s,y)\sigma^2\rangle}{\langle \bar{X}_s, 1\rangle}ds - \int_0^t \frac{\langle \mu_s, 2r(s,.)\sigma^2\varphi\rangle}{\langle \bar{X}_s, 1\rangle}ds\Big)\end{aligned}$$

This ends the proof. ■

In (4.4), we recognize two covariance terms under the measure μ_s:

$$\mathrm{Cov}_{\mu_s}(\gamma(s,.,\bar{X}_s),\varphi) = \langle \mu_s, \gamma(s,.,\bar{X}_s)\varphi(.)\rangle - \langle \mu_s,\varphi\rangle\langle\mu_s,\gamma(s,.,\bar{X}_s)\rangle$$
$$\mathrm{Cov}_{\mu_s}(r(s,.),\varphi) = \langle \mu_s, r(s,.)\varphi(.)\rangle - \langle \mu_s,\varphi\rangle\langle\mu_s, r(s,.)\rangle.$$

These terms can be viewed as generators of jump terms. For example:

$$\mathrm{Cov}_{\mu_s}(\gamma(s,.,\bar{X}_s),\varphi) = \langle \mu_s,\gamma(s,.,\bar{X}_s)\rangle \int_{\mathbb{D}_{\mathbb{R}^d}} \Big(\int_{\mathbb{D}_{\mathbb{R}^d}} \varphi(z)\frac{\gamma(s,z,\bar{X}_s)\mu_s(dz)}{\langle\mu_s,\gamma(s,.,\bar{X}_s)\rangle} - \varphi(y)\Big)\mu_s(dy).$$

Particles are resampled in the distribution μ_s biased by the function $\gamma(s,y,\bar{X}^s)$ which gives more weight to particles having higher growth rate.

Proposition 4.2. *(i)If the function $r(s,y)$ is constant, $r\equiv \bar{R}$, $Cov_{\mu_s}(r(s,.),\varphi)=0$, and similarly if $\gamma(s,y,\bar{X}^s)\equiv\bar{\gamma}$ is constant (including the "neutral" case $\gamma\equiv 0$), $Cov_{\mu_s}(\gamma(s,.,\bar{X}_s),\varphi)=0$.*
(ii) For constant functions r and γ, the lineage distributions under $\mathbb{E}(\mu_t)$ are Brownian motions.

Proof. The proof of (i) is easy. For (ii), we have for any test function φ of the form (3.3):

$$\langle\mathbb{E}(\mu_t),\varphi\rangle = \langle\mathbb{E}(\mu_0),\varphi\rangle + \int_0^t \langle\mathbb{E}(\mu_s), p\bar{R}\sigma^2\widetilde{\Delta}\varphi(s,.)\rangle ds.$$

This proves that under $\mathbb{E}(\mu_t)$, the lineages have the same finite-dimensional distributions as Brownian motions with diffusion coefficient $p\bar{R}\sigma^2$. ■

4.2 Relation with the Fleming-Viot process

For some applications, for instance if we are interested in the time of the most recent common ancestor (MRCA) of two individuals chosen at random in the population, we are interested in quantities of the form:

$$\langle\mu_t\otimes\mu_t,\chi(t,.,.)\rangle = \int_{\mathbb{D}_{\mathbb{R}^d}}\int_{\mathbb{D}_{\mathbb{R}^d}} \chi(t,y,z)\mu_t(dy)\mu_t(dz). \tag{4.5}$$

Proposition 4.3. *(i)For a constant allometry function r, and for a function $\chi(y,z)$ such that $y\mapsto\chi(y,z)$ and $z\mapsto\chi(y,z)$ are of the form (3.3),*

$$\begin{aligned}\mathbb{E}\big(\langle\mu_t\otimes\mu_t,\chi\rangle\big) &= \mathbb{E}\big(\langle\mu_0\otimes\mu_0,\chi\rangle\big) + \frac{pr\sigma^2}{2}\int_0^t \mathbb{E}\big(\langle\mu_s\otimes\mu_s,\widetilde{\Delta}^{(2)}\chi\big)ds \\ &+\mathbb{E}\Big(\int_0^t\int_{\mathbb{D}^2_{\mathbb{R}^d}} \big(\gamma(s,y,\bar{X}^s)+\gamma(s,y,\bar{X}^s) - 2\langle\mu_s,\gamma(s,.,\bar{X}_s)\rangle\big)\chi(y,z)\mu_s\otimes\mu_s(dy,dz)ds\Big) \\ &+\mathbb{E}\Big(2r\sigma^2\int_0^t \frac{1}{\langle\bar{X}_s,1\rangle}\big(\langle\mu_s, y\mapsto\chi(y,y)\rangle - \langle\mu_s\otimes\mu_s,\chi\rangle\big)ds\Big).\end{aligned} \tag{4.6}$$

where $\widetilde{\Delta}^{(2)}\chi(y,z) = \widetilde{\Delta}(y\mapsto\chi(y,z)) + \widetilde{\Delta}(z\mapsto\chi(y,z))$.
(ii) If γ is also constant, we recover the Fleming-Viot process whose generator is defined, for test functions $\phi(\mu)=\langle\mu\otimes\mu,\chi\rangle$ with $\chi(y,z)$ a test function of two variables, by:

$$L^{FV}\chi(X) = \frac{pr\sigma^2}{2}\Big\langle\frac{X}{\langle X,1\rangle}\otimes\frac{X}{\langle X,1\rangle},\widetilde{\Delta}^{(2)}\chi\Big\rangle + \frac{2r\sigma^2}{\langle X,1\rangle}\Big(\int_{\mathbb{D}_{\mathbb{R}^d}}\chi(y,y)\frac{X(dy)}{\langle X,1\rangle} - \langle\frac{X}{\langle X,1\rangle}\otimes\frac{X}{\langle X,1\rangle},\chi\rangle\Big)$$

Proof. Let φ and ψ be two real test functions on $\mathbb{D}_{\mathbb{R}^d}$. Using Itô's formula:

$$\begin{aligned}
&\langle \mu_t, \varphi\rangle\langle \mu_t, \psi\rangle = \langle \mu_0, \varphi\rangle\langle \mu_0, \psi\rangle + M_t^{\varphi,\psi}\\
&+\int_0^t \Big(\frac{pr\sigma^2}{2}\langle \mu_s\otimes\mu_s, \widetilde{\Delta}(\varphi\psi)(s,.)\rangle + \int_{\mathbb{D}_{\mathbb{R}^d}} (\gamma(s,y,\bar{X}^s)+\gamma(s,z,\bar{X}^s))\varphi(y)\psi(z)\mu_s(dy)\mu_s(dz)\Big)ds\\
&-2\int_0^t \Big(\langle \mu_s, \varphi\rangle\langle \mu_s, \psi\rangle\langle \mu_s, \gamma(s,.,\bar{X}_s)\rangle\Big)ds\\
&+2r\sigma^2\int_0^t\int_{\mathbb{R}} \frac{(\varphi(y)-\langle \mu_s,\varphi\rangle)(\psi(y)-\langle \mu_s,\psi\rangle)}{\langle \bar{X}_s, 1\rangle}\bar{X}_s(dy)ds,
\end{aligned} \tag{4.7}$$

where $M^{\varphi,\psi}$ is a square integrable martingale with bracket:

$$\langle M^{\varphi,\psi}\rangle_t = \int_0^t\int_{\mathbb{R}} 2r\sigma^2\frac{(\varphi(y)-\langle \mu_s,\varphi\rangle)(\psi(y)-\langle \mu_s,\psi\rangle)}{\langle \bar{X}_s, 1\rangle}\bar{X}_s(dy)ds. \tag{4.8}$$

For a function $\chi(y,z)=\sum_{k=1}^K \lambda_k\varphi_k(y)\psi_k(z)$ with $K\in\mathbb{N}$ and $\lambda_k\in\mathbb{R}$, we can generalize (4.7) by noting that:

$$\begin{aligned}
\langle \mu_t\otimes\mu_t, \chi\rangle &= \sum_{k=1}^K \lambda_k\langle \mu_t, \varphi_k\rangle\langle \mu_t, \psi_k\rangle\\
\widetilde{\Delta}^{(2)}\chi(y,z) &= \sum_{k=1}^K \lambda_k\widetilde{\Delta}\varphi_k(y)\psi_k(z) + \sum_{k=1}^K \lambda_k\varphi_k(y)\widetilde{\Delta}\psi_k(z).
\end{aligned}$$

Since every bounded measurable function on $\mathbb{D}^2_{\mathbb{R}^d}$ can be approximated by functions of the previous form, for the bounded pointwise topology, the proposition is proved. ■

With Proposition 4.3, we obtain a historical construction, from a branching process, of the Fleming-Viot process with genealogies introduced by [27].

5 Examples

In this section, we give several examples of applications of historical processes in Biology. In the Examples 1 and 2, we investigate two examples of adaptative dynamics. The dynamics does not depend on the past, but we show that the historical processes can bring a new point of view to the problem. In particular, the historical process provides the ancestral paths of living particles. When the superprocess is diffusive, the individual dimension is lost and ancestral paths can not be read from the sole information of the support of the measure that represents the population. Several decompositions of superprocesses have been exposed: see for instance the backbones of [5]. But nothing ensures that the backbones correspond to the real ancestral paths of immortal individuals. In Example 3, we consider an example inspired by Adler Tribe [1] and Perkins [41], where the death rates depend on the past histories of the particles. The last examples deals with the case of an age-structured population.

5.1 A model with competition for resources

We first introduce a model of adaptation with competition for resources that has been considered by Roughgarden [43], Dieckmann and Doebeli [18], Champagnat and Méléard [13]. In this model, the parameters are the following. The trait $x \in [0, x_0]$, with $x_0 = 4$ can be thought as being the size. The birth and death rates are chosen as:

$$r(t,y) = 1, \qquad b(t,y) = \exp\Big(-\frac{(y_t-2)^2}{2\sigma_b^2}\Big), \qquad d(t,y,X) = \int_{\mathbb{D}} \exp\Big(-\frac{(y_t-y'_t)^2}{2\sigma_U^2}\Big) X(dy')$$

The birth rate is indeed maximal at $x^* = 2$ and there is a local competition with neighbors of closed traits. If the competition kernel was flat ($\sigma_U = +\infty$), evolution would favor individuals with maximal growth rate x^*. For $\sigma_U < +\infty$, Champagnat and Méléard [13] proved that, under a time scale that differ from the present one and depending on whether $\sigma_b < \sigma_U$ or $\sigma_b > \sigma_U$, the configuration might be in 0 or constituted of several groups concentrated around different traits values.

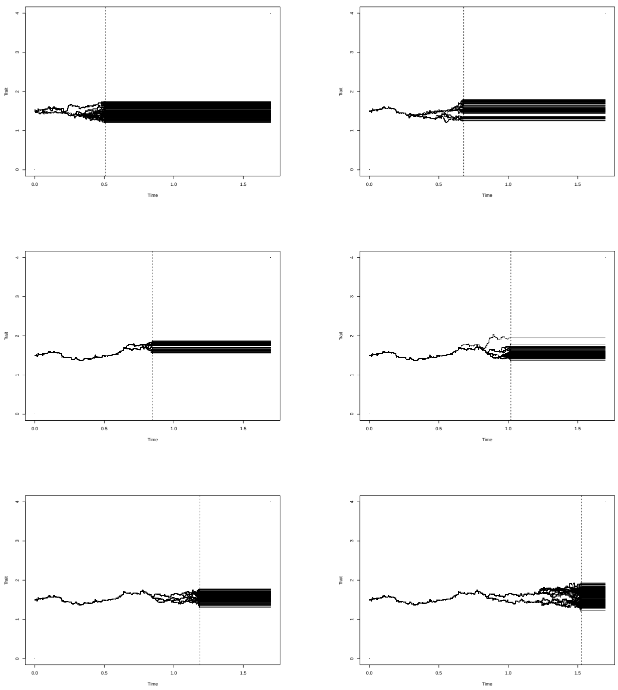

Figure 3: *Dieckmann-Doebeli model.* $\sigma = 0.4$, $\sigma_b = 0.4$, $\sigma_U = 0.3$, $x_0 = 4$, $p = 0.5$, $n = 300$. *The 300 particles are started with the trait* 1.5. *In this example,* $\sigma_b > \sigma_U$ *and we observe the separation of the population into two subgroups concentrated around different trait values.*

5.2 A model with asymmetric competition

This second illustration in adaptive dynamics has been investigated in Champagnat *et al.* [12]. The hereditary trait can be thought here as being the size of an individual, which is assumed to be constant during her life. We use the following birth rate and competition kernel, proposed by Kisdi [35] (see example 2.2):

$$\begin{aligned} b(t,y) &= x_0 - y_t, \\ d(t,y,X) &= \int_{\mathbb{D}} \frac{2}{K}\Big(1 - \frac{1}{1+\alpha\exp(-\beta(y_t - y'_t))}\Big) X(dy') \end{aligned} \tag{5.1}$$

with $x_0 = 4$, $\alpha = 1.2$ and $\beta = 4$.

Kisdi's kernel introduces asymmetry among the individuals. The idea is that larger individuals are more competitive: they exert more competition on the rest of the population and are less sensitive to the pressure for resources created by smaller competitors. The trade-off comes from the fact that these individuals are disadvantaged for births since giving birth to big offspring is more costly. The analysis of such models can be done without a historical approach. However, new questions can be raised. In [12, 39], it is shown that the trait in the population evolves towards values where the mean number of offspring is maximized. Once these states are reached, the population may split in several subgroups characterized by different trait values. Understanding the distribution of surviving lineages (Fig. 4, 5) gives insight on the path taken by evolution and brings information on how the different evolutionary states are reached.

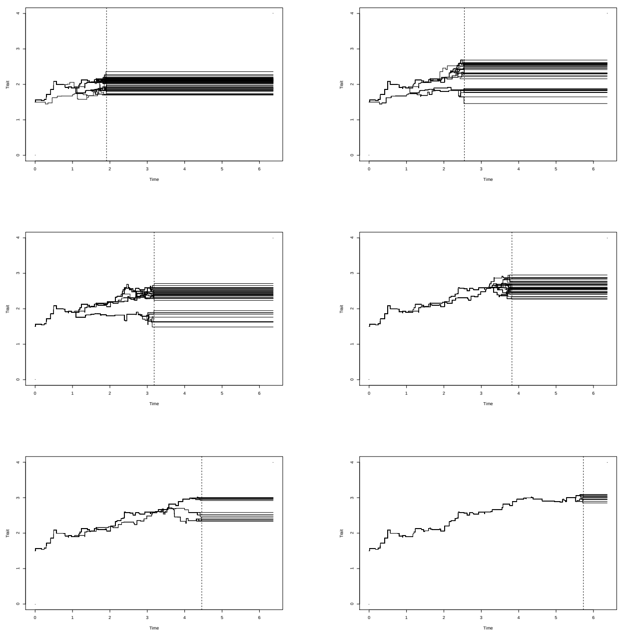

Figure 4: *Kisdi's model.* $\alpha = 1.2$, $\beta = 4$, $x_0 = 4$, $\sigma = 0.2$, $p = 1$, $K = 2$, $n = 10$. *The 50 particles are started with the trait* 1.5.

In Fig. 4, we see that a branching phenomenon appears: at time 3, the different individuals are separated into 3 families with MRCAs around time 1. Because of the selection, the branches are very thin, showing that only a few lineages provide the living population at the current time. Contrarily to Fig. 4, we see in 5 that the MRCAs are distributed relatively regularly along the branches of the tree. This can be a consequence of the higher density of particles and of the weaker selection. Also, the evolution is much slower than in Fig. 4.

5.3 A variant of Adler's fattened goats: a spatial model

In many models, trait or space play a similar role. Spatial models have been extensively studied as toy models for evolution (see Bolker and Pacala [9, 10] or Dieckmann and Law [19]).

Here we consider a spatial model where the competition exerted by past ancestors is softened. This model is a variant of Adler's fattened goats (*e.g.* [1, 41]):

$$r(t, y) = 1, \qquad b(t, y) = b,$$

$$\text{and } d(t, y, X) = \int_0^t \int_{\mathbb{R}^d} \frac{K_\varepsilon\big(y'(s) - y(t)\big)}{K} X_s(dy', dc')\, e^{-\alpha(t-s)} ds \tag{5.2}$$

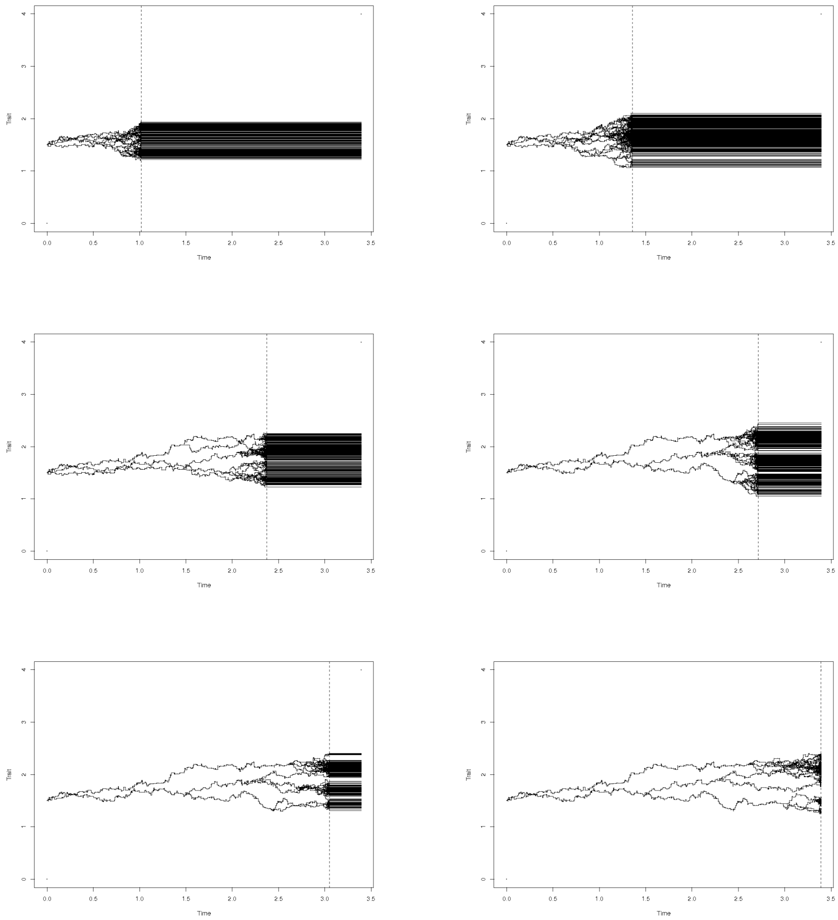

Figure 5: *Kisdi's model.* $\alpha = 1.2$*,* $\beta = 4$*,* $x_0 = 4$*,* $\sigma = 0.2$*,* $p = 1$*,* $K = 100$*,* $n = 100$*. The 100 particles are started with the trait* 1.5*.*

The goat-like particles consume resources at the point where they are staying, and when they arrive in a region where the population has previously grazed, their death rate is increased. The parameter α describes the speed at which the environment replenishes itself. The kernel K_ε can be the density function of a centered Gaussian distribution with variance ε for instance. The parameter K can be seen as a carrying capacity controlling the mass of X_s.

Due to the form of their interaction, we see in Fig. 6 that the goats spread quickly in the whole space and separate into families with very old MRCAs. The families become quickly disjoint and geographically isolated.

5.4 Logistic age and size-structured population

Let us consider the framework of Remark 2.3. Following Méléard and Tran [39], the following example for a population structured by age and size is considered:

$$\begin{aligned} b(t,y,c) &= y_t\big(x_0 - y_t\big)e^{-n(t-\tau_{c,t})}\mathbb{1}_{y_t\in[0,x_0]}, \\ d(t,y,c,X) &= d_0 + \eta\big(x_0 - y_t\big)\mathbb{1}_{y_t\in[0,x_0]}\langle X,1\rangle \end{aligned} \tag{5.3}$$

with $x_0 = 4$, $d_0 = 1/4$ and $\eta = 0.1$. PDE limits, TSS and Canonical Equations are considered in [39]. Here, we choose $r(t,y,c) = 1$. The birth rate is here of the form (2.3). The difficulty of

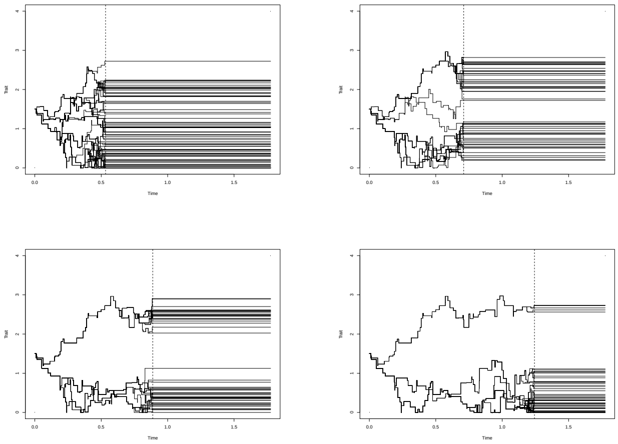

Figure 6: *Simulation for Adler's fattened goats.* $\alpha = 10$, $\varepsilon = 0.8$, $b = 0.75$, $\sigma = 1$, $p = 1$, $K = 50$ *and* $n = 50$. *The 100 initial particles are started at location* 1.5.

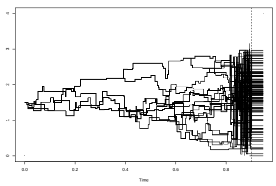

Figure 7: *Simulation for Adler's fattened goats.* $\alpha = 0$, $\varepsilon = 0.8$, $b = 0.75$, $\sigma = 1$, $p = 1$, $K = 50$ *and* $n = 50$. *The 100 initial particles are started at location* 1.5. *In this example, the consumed resources never replenish. Thus individuals die faster.*

these models is that age is a fast component that is hard to track when the process is accelerated. Using time scales separations, as predicted in [38], we can see in Fig. 8 that the age distribution stabilizes to an exponential distribution with rate 1.

In the simulations, we see that the MRCA is less old than in the previous example. This is reminiscent of Fig. 1 (b). In these figures, it is seen that because of the selection, the MRCA is relatively recent. An excursion appears between times 20 and 30. Since the traits then reunite with the upper branch, this branch has escaped selection.

Acknowledgements: This work benefited from the support of the ANR MANEGE (ANR-09-BLAN-0215), from the Chair "Modélisation Mathématique et Biodiversité" of Veolia Environnement-Ecole Polytechnique-Museum National d'Histoire Naturelle-Fondation X. It also benefited from the visit of both authors to the Institute for Mathematical Sciences of the National University of Singapore in 2011.

A Properties of the test functions (3.1) and (3.3)

We begin with a lemma that will be useful to link the classes of test functions (3.1) and (3.3). This lemma also shows that the class of test functions (3.1) separates points.

Lemma A.1. *For* $q \in \mathbb{N}^*$, *recall that we denote by* $k^q(u)$ *the density of the Gaussian distribution with mean 0 and variance* $1/q$. *Let* $g \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R})$. *For* $G(x) = x$ *and* $g_q(s,x) = k^q(t-s)g(x)$,

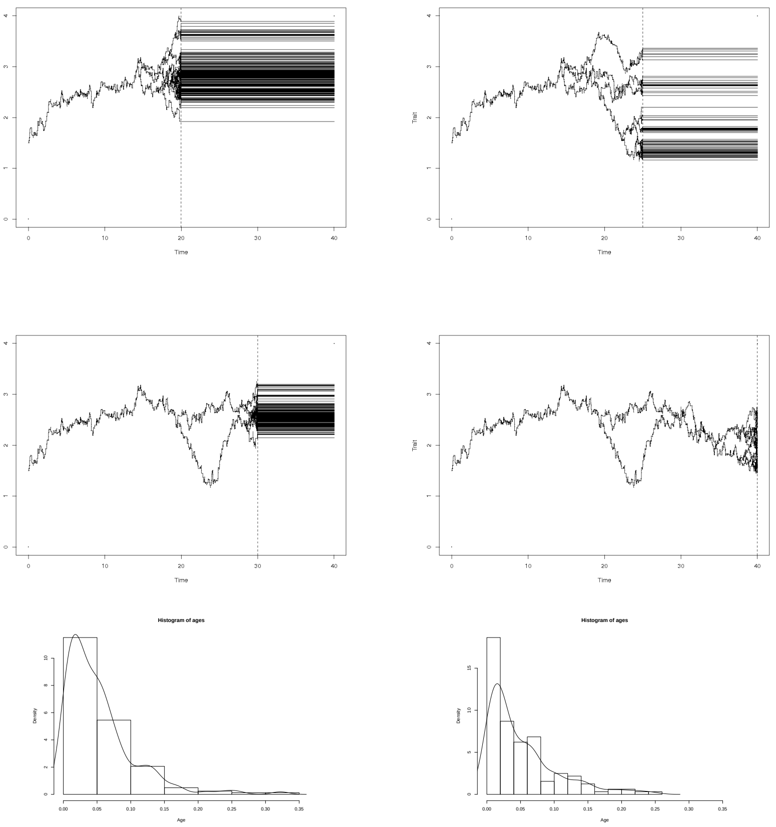

Figure 8: *Simulation for the logistic age and size-structured population. The last line corresponds to the age-distribution at time* $t = 0.5$ *and* $t = 4$*.* $\sigma = 0.2$*,* $p = 1$*,* $x_0 = 4$*,* $d_0 = 0.25$*,* $\eta = 0.1$*,* $n = 20$*. The 100 particles are started with trait* 1.5 *and age 0.*

we have for all $y \in \mathbb{D}_{\mathbb{R}^d}$ *and all* $t \in [0, T]$ *at which* y *is continuous that:*

$$\lim_{q \to +\infty} G_{g_q}(y) = g(y_t).$$

Proof. First notice that all the G_{g_q} are bounded by $\|g\|_\infty$. Let $\varepsilon > 0$. Since y is continuous at t, so is $g \circ y$ and there exists $\alpha > 0$ sufficiently small so that for every $s \in (t - \alpha, t + \alpha)$, $|g(y_s) - g(y_t)| \leq \varepsilon/2$. We can then choose q sufficiently large such that

$$\int_{|t-s|>\alpha} k^q(t-s)ds < \frac{\varepsilon}{4\|g\|_\infty}.$$

Then:

$$\begin{aligned} |G_{g_q}(y) - g(y_t)| =& \Big| \int_0^T k^q(t-s)\big(g(y_s) - g(y_t)\big)ds \Big| \\ \leq& 2\|g\|_\infty \varepsilon \int_{|s-t|\geq\alpha} k^q(t-s)ds + \frac{\varepsilon}{2}\int_{|s-t|<\alpha} k^q(t-s)ds \leq \varepsilon. \end{aligned}$$

■

We are now in position to give the:

Proof of Lemma 3.2. We can assume without restriction that the functions g_j in the definition of φ (3.5) are positive. Let us define for $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$:

$$\varphi_q(y) = \exp\Big(\int_0^T \sum_{j=1}^m \log g_j(y_s) k^q(t_j - s) ds\Big).$$

By Lemma A.1, the term in the integral is bounded uniformly in q and y and converges when q tends to infinity to $\sum_{j=1}^m \log g_j(y_{t_j})$. As a consequence, for every $y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$, the sequence $(\varphi_q(y))_{n\in\mathbb{N}^*}$ is bounded and converges to

$$\exp\Big(\sum_{j=1}^m \log g_j(y_{t_j})\Big) = \prod_{j=1}^m g_j(y_{t_j}) = \varphi(y)$$

when q tends to infinity. Moreover,

$$\begin{aligned} D^2\varphi_q(t,y) =& \exp\Big(\int_0^T \sum_{j=1}^m \log g_j(y_s) k^q(t_j - s) ds\Big) \\ &\times \Big(\int_t^T \sum_{j=1}^m \Delta_x(\log g_j)(y_t) k^q(t_j - s) ds + \sum_{i=1}^d \Big(\int_t^T \sum_{j=1}^m \partial_{x_i}(\log g_j)(y_t) k^q(t_j - s) ds\Big)^2\Big). \end{aligned}$$

When q tends to infinity, we have by Lemma A.1:

$$\begin{aligned} \lim_{q\to+\infty} D^2\varphi_q(t,y) =& \prod_{j=1}^m g_j(y_{t_j})\Big(\sum_{j \mid t_j>t} \frac{\Delta_x g_j(y_t)}{g_j(y_t)} - \sum_{i=1}^d \frac{\big(\partial_{x_i} g_j(y_t)\big)^2}{g_j^2(y_t)} + \sum_{i=1}^d \Big(\sum_{j \mid t_j>t} \frac{\partial_{x_i} g_j(y_t)}{g_j(y_t)}\Big)^2\Big) \\ =& \prod_{j=1}^m g_j(y_{t_j})\Big(\sum_{j \mid t_j>t} \frac{\Delta_x g_j(y_t)}{g_j(y_t)} + 2\sum_{i=1}^d \sum_{\substack{j\neq k \\ t_j,t_k>t}} \frac{\partial_{x_i} g_j(y_t)\partial_{x_i} g_k(y_t)}{g_j(y_t) g_k(y_t)}\Big) \\ =& \prod_{j=1}^m g_j(y_{t_j}) \frac{\Delta_x\big(\prod_{j \mid t_j>t} g_j\big)(y_t)}{\prod_{j \mid t_j>t} g_j(y_t)} = \widetilde{\Delta}\Big(\prod_{j=1}^m g_j\Big)(t,y). \end{aligned}$$

This concludes the proof.

■

B Technical result on concatenated paths

Let $\mathcal{H}$ be the set of increasing bijections from $[0,T]$ to $[0,T]$, where $T > 0$. We recall that the Skorokhod distance is defined for $y, z \in \mathbb{D}$ by:

$$\mathbf{d}_{\mathrm{Sk}}(y,z) = \inf_{\lambda\in\mathcal{H}} \max\left\{\|y\circ\lambda - z\|_\infty, \sup_{t,s<T}\Big|\log\Big(\frac{\lambda(t)-\lambda(s)}{t-s}\Big)\Big|\right\}. \tag{B.1}$$

In the sequel, we consider $y,\ z \in \mathbb{D}$ and $s, r \in [0,T]$. Without loss of generality, we can assume that $s < r$.

Proposition B.1. *If $\mathbf{d}_{Sk}(y,z) < \varepsilon$ and if s and r are sufficiently close so that:*

$$0 \leq \max\Big\{\log\frac{r}{s}, \log\frac{T-s}{T-r}\Big\} \leq \varepsilon. \tag{B.2}$$

Then for all $w \in \mathbb{D}$, $\mathbf{d}_{Sk}((y|s|w),(z|r|w)) < 3\varepsilon$.

In the proof, we will need the following change of time $\lambda_0 \in \mathcal{H}$:

$$\lambda_0(u) = \frac{r}{s} u \mathbb{1}_{u \leq s} + \Big(r + \frac{T-r}{T-s}(u-s) \Big) \mathbb{1}_{u>s}. \tag{B.3}$$

The bijection λ_0 depends on r and s. For u and $v \in [0,T]$, we have:

$$\Big| \log \Big(\frac{\lambda_0(u) - \lambda_0(v)}{u-v} \Big) \Big| \leq \max \Big\{ \log \Big(\frac{r}{s} \Big), \log \Big(\frac{T-s}{T-r} \Big) \Big\}. \tag{B.4}$$

The right hand side converges to 0 when r/s converges to 1, and is upper bounded by ε under the Assumptions (B.2) of Proposition B.1.

Lemma B.2. *For all $w \in \mathbb{D}$. If (B.2) is satisfied, then $\mathbf{d}_{Sk}(w \circ \lambda_0, w) \leq \varepsilon$.*

Proof. The infimum in (B.1) can be upper bounded by choosing $\lambda = \lambda_0^{-1}$, which is the inverse bijection of λ_0:

$$\lambda_0^{-1}(u) = \frac{s}{r} u \mathbb{1}_{u \leq r} + \Big(s + \frac{T-s}{T-r}(u-r) \Big) \mathbb{1}_{u>r}. \tag{B.5}$$

For such choice, we have:

$$\mathbf{d}_{\mathrm{Sk}}(w \circ \lambda_0, w) \leq \max \Big\{ 0, \max \Big(\log \frac{r}{s}, \log \frac{T-s}{T-r} \Big) \Big\} \leq \varepsilon.$$

■

Let us now prove Proposition B.1:

Proof. By the triangular inequality:

$$\begin{aligned} &\mathbf{d}_{\mathrm{Sk}}((y|s|w),(z|r|w)) \leq A + B, \qquad \text{where} \\ &A = d((y|s|w),(y \circ \lambda_0 | r | w \circ \lambda_0)) \\ &B = d((y \circ \lambda_0 | r | w \circ \lambda_0),(z|r|w)). \end{aligned} \tag{B.6}$$

By Lemma B.2, $A \leq \varepsilon$. For the second term, using Lemma B.2 again:

$$B \leq \mathbf{d}_{\mathrm{Sk}}(y \circ \lambda_0, z) + d(w \circ \lambda_0, w) \leq \mathbf{d}_{\mathrm{Sk}}(y \circ \lambda_0, z) + \varepsilon. \tag{B.7}$$

Now, since $\mathbf{d}_{\mathrm{Sk}}(y,z) \leq \varepsilon$, there exists $\lambda \in \mathcal{H}$ such that $\|y \circ \lambda - z\|_\infty \leq 2\varepsilon$ and

$$\sup_{u,v \leq T} \Big| \log \frac{\lambda(u) - \lambda(v)}{u-v} \Big| \leq 2\varepsilon.$$

Then, considering the change of time $\lambda_0^{-1} \circ \lambda$:

$$\begin{aligned} \mathbf{d}_{\mathrm{Sk}}(y \circ \lambda_0, z) &\leq \max \Big\{ \|y \circ \lambda_0 \circ \lambda_0^{-1} \circ \lambda - z\|_\infty, \sup_{u,v \leq T} \Big| \log \frac{\lambda_0^{-1} \circ \lambda(u) - \lambda_0^{-1} \circ \lambda(v)}{u-v} \Big| \Big\} \\ &\leq \max \Big\{ \|y \circ \lambda - z\|_\infty, \sup_{u,v \leq T} \Big| \log \frac{\lambda_0^{-1} \circ \lambda(u) - \lambda_0^{-1} \circ \lambda(v)}{\lambda(u) - \lambda(v)} \Big| + \sup_{u,v \leq T} \Big| \log \frac{\lambda(u) - \lambda(v)}{u-v} \Big| \Big\} \\ &\leq \max \Big(\varepsilon, 3\varepsilon \Big). \end{aligned}$$

■

References

[1] R. Adler and R. Tribe. Uniqueness for a historical SDE with a singular interaction. *Journal of Theoretical Probability*, 11(2):515–533, 1998.

[2] V. Bansaye, J.-F. Delmas, L. Marsalle, and V.C. Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *Annals of Applied Probability*. to appear.

[3] N.H. Barton and A.M. Etheridge. The effect of selection on genealogies. *Genetics*, 166:1115–1131, 2004.

[4] N.H. Barton, A.M. Etheridge, and A. Véber. A new model for evolution in a spatial continuum. *Electronic Journal of Probability*, 15(7):162–216, 2010.

[5] J. Berestycki, A.E. Kyprianou, and A. Murillo-Salas. The prolific backbone for supercritical superprocesses. *Stochastic processes and their applications*, 121:1315–1331, 2011.

[6] N. Berestycki. *Recent progress in coalescent theory*, volume 16. Ensaios Matemáticos, Rio de Janeiro, sociedade brasileira de matemática edition, 2009. 193pp.

[7] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.

[8] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3 edition, 1995.

[9] B. Bolker and S. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, 52:179–197, 1997.

[10] B. Bolker and S. Pacala. Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. *The American Naturalist*, 153:575–602, 1999.

[11] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models via timescale separation. *Theoretical Population Biology*, 69:297–321, 2006.

[12] N. Champagnat, R. Ferrière, and S. Méléard. From individual stochastic processes to macroscopic models in adaptive dynamics. *Stochastic Models*, 24:2–44, 2008.

[13] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*. to appear.

[14] D. A. Dawson. Mesure-valued markov processes. In Springer, editor, *Ecole d'Eté de probabilités de Saint-Flour XXI*, volume 1541 of *Lectures Notes in Math.*, pages 1–260, New York, 1993.

[15] D.A. Dawson. Geostochastic calculus. *Canadian Journal of Statistics*, 6(2):143–168, 1978.

[16] D.A. Dawson and E.A. Perkins. *Historical Processes*, volume 93. American Mathematical Society, Memoirs of the American Mathematical Society edition, 1991.

[17] A. Depperschmidt, A. Greven, and P. Pfaffelhuber. Tree-valued Fleming-Viot dynamics with mutation and selection. 2011. submitted.

[18] U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. *Nature*, 400:354–357, 1999.

[19] U. Dieckmann and R. Law. Relaxation projections and the method of moments. In *The Geometry of Ecological Interactions*, Cambridge University Press, pages 412–455, 2000.

[20] E.B. Dynkin. Branching particle systems and superprocesses. *Annals of Probability*, 19:1157–1194, 1991.

[21] J. Engländer and A. Winter. Law of large numbers for a class of superdiffusions. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(2):171–185, 2006.

[22] A. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, Providence, 2000.

[23] A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. An approximate sampling formula under genetic hitchhiking. *Annals of Applied Probability*, 16(2):685–729, 2006.

[24] S.N. Evans and E.A. Perkins. Measure-valued branching diffusions with singular interactions. *Canadian Journal of Mathematics*, 46:120–168, 1994.

[25] S.N. Evans and D. Steinsaltz. Damage segregation at fissioning may increase growth rates: A superprocess model. *Theoretical Population Biology*, 71:473–490, 2007.

[26] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.

[27] A. Greven, P. Pfaffelhuber, and A. Winter. Tree-valued resampling dynamics: martingale problems and applications. 2010. submitted.

[28] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.

[29] P. Jagers. A general stochastic model for population development. *Skand. Aktuarietidskr*, 52:84–103, 1969.

[30] P. Jagers. *Branching Processes with Biological Applications.* Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, London-New York-Sydney, 1975.

[31] A. Jakubowski. On the Skorokhod topology. *Annales de l'Institut Henri Poincaré*, 22(3):263–285, 1986.

[32] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, 18:20–65, 1986.

[33] I. Kaj and S. Sagitov. Limit processes for age-dependent branching particle systems. *Journal of Theoretical Probability*, 11(1):225–257, 1998.

[34] J.F.C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13:235–248, 1982.

[35] E. Kisdi. Evolutionary branching under asymmetric competition. *J. Theor. Biol.*, 197(2):149–162, 1999.

[36] S. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51:210–237, 1997.

[37] V. Limic. Genealogies of regular exchangeable coalescents with applications to sampling. *Annales de l'Institut Henri Poincaré*, 2011. to appear.

[38] S. Méléard and V.C. Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications.* to appear.

[39] S. Méléard and V.C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of Mathematical Biology*, 58(6):881–921, 2009.

[40] M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Annals of Probability*, 29(4):1547–1562, 2001.

[41] E. Perkins. *On the Martingale Problem for Interactive Measure-Valued Branching Diffusions*, volume 115(549). American Mathematical Society, Memoirs of the American Mathematical Society edition, May 1995.

[42] E. A. Perkins. Dawson-Watanabe superprocesses and mesure-valued diffusions. In Springer, editor, *Ecole d'Eté de probabilités de Saint-Flour*, volume 1781 of *Lectures Notes in Math.*, pages 125–329, New York, 1993.

[43] J.A. Roughgarden. *Theory of population genetics and evolutionary ecology: an introduction.* Macmillan, New York, 1979.

[44] W. Rudin. *Real and Complex Analysis.* McGraw-Hill International Editions, third edition edition, 1987.

[45] J. Schweinsberg. Coalescent with simultaneous multiple collisions. *Electronic Journal of Probability*, 5:1–50, 2000.

[46] V.C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: P&S*, 12:345–386, 2008.