



HAL
open science

Uniform sampling in a structured branching population

Aline Marguet

► **To cite this version:**

| Aline Marguet. Uniform sampling in a structured branching population. 2017. hal-01362366v2

HAL Id: hal-01362366

<https://hal.science/hal-01362366v2>

Preprint submitted on 24 Sep 2017 (v2), last revised 19 Nov 2018 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Uniform sampling in a structured branching population

Aline Marguet*

September 22, 2017

Abstract

We are interested in the dynamic of a structured branching population where the trait of each individual moves according to a Markov process. The rate of division of each individual is a function of its trait and when a branching event occurs, the trait of the descendants at birth depends on the trait of the mother and on the number of descendants. In this article, we explicitly describe the penalized Markov process, named auxiliary process, corresponding to the dynamic of the trait of a "typical" individual by giving its associated infinitesimal generator. We prove a Many-to-One formula and a Many-to-One formula for forks. Furthermore, we prove that this auxiliary process characterizes exactly the process of the trait of a uniformly sampled individual in a large population approximation. We detail three examples of growth-fragmentation models: the linear growth model, the exponential growth model and the parasite infection model.

Keywords: Branching Markov processes, Many-to-One formulas, Size-biased reproduction law.

A.M.S classification: 60J80, 60J85, 60J75, 92D25.

1 Introduction

The characterization of the sampling of individuals in a population is a key issue for branching processes with several motivations in statistics and biology. We refer to the work of Durrett [20] and references therein for the study of the genealogy of a branching Markov process and the study of the degree of relationship between k individuals chosen randomly at time t in the population. In particular, he analyzed the asymptotics of the so-called reduced branching process $N_t(s)$ defined as the number of individuals alive at time s which have offspring alive at time t . An approximation of this process by a pure birth process is given in [44]. The question of finding the coalescing time of individuals in a Galton-Watson tree is addressed in [48] and the coalescent structure of continuous-time Galton-Watson trees is studied in [26]. We refer to [2] and [38] for more results on this question and to [32] for results concerning the Bellman-Harris branching process. The pedigree of a typical individual in a supercritical branching process has also been investigated asymptotically for multi-type branching processes with a finite number of types in [23], with i.i.d life-times in [3] and with an age-structure in [43]. The characterization of the sampling is the key to obtain asymptotic results on the branching process ([37], [5], [15]) and to infer the parameters of the model ([24], [18], [31]).

In this article, we consider a continuous-time structured branching Markov process where the trait of each individual moves according to a Markov process and influences the branching events. The purpose of this article is to characterize the trait of a typical individual uniformly sampled from the population at time t and its associated ancestral lineage. In particular, we exhibit the bias due to the structure of the population and to the sampling. We also describe the traits of a uniformly sampled couple in the current population. Therefore, we provide new applications in a non-neutral framework for cell division (Section 2.2), even for models in a varying environment.

*CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, FRANCE. aline.marguet@polytechnique.edu

We now describe informally the process, while its rigorous construction and characterization as a càdlàg measure-valued process under Assumptions A and B are detailed in Section 2. We assume that individuals behave independently and that for each individual u :

- its trait $(X_t^u)_{t \geq 0}$ evolves as an \mathcal{X} -valued Markov process with infinitesimal generator $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$, where $\mathcal{X} \subset (\mathbb{R}_+)^d$ is a measurable space for some $d \geq 1$,
- it dies at time t at rate $B(X_t^u)$,
- at its death, an individual with trait x is replaced by $k \in \mathbb{N}$ individuals with probability $p_k(x)$ and $m(x) = \sum_{k \geq 1} k p_k(x)$,
- the trait of the j th child among k is distributed as $P_j^{(k)}(x, \cdot)$ for all $1 \leq j \leq k$.

We use the notion of spine, which is a distinguished line of descent in the branching process, and Many-to-One formulas, which have been developed from the notion of size-biased tree, considered by Kallenberg [33], Chauvin and Rouault [13], Chauvin, Rouault and Wakolbinger [14] with a Palm measure approach and Lyons, Peres and Pemantle [40]. For general results on branching processes using these techniques, including the spinal decomposition, we refer to [37] and [1] for discrete-time models and to [23], [25] and [15] for continuous-time branching processes. These previous works ensure in particular that if we denote by V_t the set of individuals alive at time t and by N_t its cardinal, we have the well-known Many-to-One formula:

$$\mathbb{E} \left[\sum_{u \in V_t} f(X_t^u) \right] = \mathbb{E} \left[f(Y_t) e^{\int_0^t B(Y_s)(m(Y_s)-1) ds} \right], \quad (1.1)$$

where f is a non-negative measurable function and $(Y_t)_{t \geq 0}$ follows the dynamic of a tagged-particle i.e. the same dynamic of all the particles between jumps and at a jump, the unique daughter particle is chosen uniformly at random among all the daughter particles. This formula can be seen as a Feynman-Kac formula [17] with a weight on the right-hand side relying on the whole ancestral lineage of current individuals which corresponds to the growth of the population. In this case, under spectral assumptions, the asymptotic behavior of the number of individuals has been well studied in [40], [37], [1], [23] and [10]. We also refer to the work of Bansaye and al. [5] for law of large numbers theorems using Many-to-One formulas.

On the right-hand side of (1.1) appears a Markov process with penalized (or rewarded) trajectories which describes the dynamic of the trait of a typical individual. This corresponds to a time-inhomogeneous Markov process $Y^{(t)}$, indexed by $t \geq 0$, for which we provide the following formula for any non-negative measurable function F on the space of càdlàg processes:

$$\mathbb{E} \left[\sum_{u \in V_t} F(X_s^u, s \leq t) \right] = m(x, 0, t) \mathbb{E} \left[F(Y_s^{(t)}, s \leq t) \right], \quad (1.2)$$

where for $x \in \mathcal{X}$ and $0 \leq s \leq t$,

$$m(x, s, t) := \mathbb{E} [N_t | Z_s = \delta_x], \quad (1.3)$$

and

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

is the empirical measure of the process. We explicit the generator $(\mathcal{A}_s^{(t)})_{s \leq t}$ of this auxiliary process: for all well-chosen functions f , $x \in \mathcal{X}$ and $s < t$, we have:

$$\mathcal{A}_s^{(t)} f(x) = \widehat{\mathcal{G}}_s^{(t)} f(x) + \widehat{B}_s^{(t)}(x) \int_{\mathcal{X}} (f(y) - f(x)) \widehat{P}_s^{(t)}(x, dy),$$

where

$$\widehat{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}(m(\cdot, s, t)f)(x) - f(x)\mathcal{G}(m(\cdot, s, t))(x)}{m(x, s, t)},$$

$$\widehat{B}_s^{(t)}(x) = B(x) \int_{\mathcal{X}} \frac{m(y, s, t)}{m(x, s, t)} m(x, dy),$$

$$\widehat{P}_s^{(t)}(x, dy) = m(y, s, t)m(x, dy) \left(\int_{\mathcal{X}} m(y, s, t)m(x, dy) \right)^{-1},$$

and

$$m(x, A) := \sum_{k \geq 0} p_k(x) \sum_{j=1}^k P_j^{(k)}(x, A),$$

denotes the expected number of children with trait in the Borel set A of an individual of trait x .

Moreover, we give some very simple and interesting examples where we can find the expression of the generator of the auxiliary process: we detail three models for the dynamic of a cell population (see Section 2.2).

The Many-to-One formula (1.2) splits the behavior of the entire population into a term characterizing the growth of the population and a term characterizing the dynamic of the trait. This separation in two terms is the key to the study of the ancestral trait of a uniformly sampled individual. Indeed, we prove in Theorem 4.1, that the auxiliary process describes the ancestral lineage of a sampled individual in a branching population at a fixed time when the initial population is large. More precisely, if we denote by $X^{U(t), \nu}$ the trait of a uniformly sampled individual from a population at time t with initial distribution ν and if $\nu_n = \sum_{i=1}^n \delta_{X_i}$ where X_i are i.i.d. random variables with law ν , under some assumptions, we prove the following convergence in law:

$$X_{[0,t]}^{U(t), \nu_n} \xrightarrow{n \rightarrow +\infty} Y_{[0,t]}^{(t), \pi_t}, \text{ where } \pi_t(dx) = \frac{\mathbb{E}_x(N_t)\nu(dx)}{\int \mathbb{E}_x(N_t)\nu(dx)}, \quad (1.4)$$

and $Y^{(t), \pi_t}$ denotes the auxiliary process with initial condition distributed as π_t . This result shows that the auxiliary process is the appropriate tool for the study of the trait along the ancestral lineage of a sampling. We notice in particular that the dependence on the trait of the average number of individuals in the population plays a crucial part in the creation of a bias.

Finally, we refer the reader to [41] for results on the asymptotic behavior of the process of a sampling. In particular, under some assumptions ensuring the ergodicity of the auxiliary process, a law of large number for the empirical distribution of ancestral trajectories is proven. The asymptotic behavior of the process of a sampling has already been studied in [5] in the case of a constant division rate and in [15] in a spectral framework.

Outline. Section 2 is devoted to the rigorous construction of our process. In Section 2.1, we first describe in detail the model and in Theorem 2.2, we prove the existence and uniqueness of the branching process. Then, in Section 2.2, we introduce our three examples of cell division models: the size-structured model with linear or exponential growth and the parasite infection model. In Section 3, we detail the properties of the Markov process along the spine. In particular, in Theorem 3.1, we prove the Many-to-One formula which describes the dynamic of a typical individual in the population. Finally, we give two other Many-to-One formulas, one for the dynamic of the whole tree in Proposition 3.5 and an other one for the dynamic of a couple of traits in Proposition 3.6. Section 4 concerns the ancestral lineage of a uniform sampling at a fixed time in a large population. More precisely, in Theorem 4.1, we prove the convergence (1.4). In Section 4.2, we give explicitly the dynamic of the auxiliary process for our three examples of cell population models. Finally, in Section 5, we give some useful comments on the model and some other examples.

Notation. We use the classical Ulam-Harris-Neveu notation to identify each individual. Let

$$\mathcal{U} = \bigcup_{n \in \mathbb{N}} (\mathbb{N}^*)^n.$$

The first individual is labeled by \emptyset . When an individual $u \in \mathcal{U}$ dies, its K descendants are labeled $u1, \dots, uK$. If u is an ancestor of v , we write $u \leq v$.

2 Definition and existence of the structured branching process

First, we introduce some useful notations and objects to characterize the branching process. Then, we prove the existence and uniqueness of the measure-valued branching process from scratch in Section 2.1. Henceforth, we work on a probability space denoted by $(\Omega, \mathcal{F}, \mathbb{P})$.

Dynamic of the trait. Let $\mathcal{X} = \mathcal{Y} \times \mathbb{R}_+$ where $\mathcal{Y} \subset (\mathbb{R}_+)^d$ is a measurable space for some $d \geq 1$. It is the state space of the Markov process describing the trait of the individuals. The second component, with values in \mathbb{R}_+ , is a time component. We assume that $(A_t, t \geq 0)$ is a strongly continuous contraction semi-group with associated infinitesimal generator $\mathcal{G} : \mathcal{D}(\mathcal{G}) \subset \mathcal{C}_b(\mathcal{X}) \rightarrow \mathcal{C}_b(\mathcal{X})$, where $\mathcal{C}_b(\mathcal{X})$ denotes the space of continuous bounded function from \mathcal{X} to \mathbb{R} .

Then, according to Theorem 4.4.1 in [21], there is a unique solution to the martingale problem associated with $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$, denoted by $(X_t, t \geq 0)$. It is an \mathcal{X} -valued càdlàg strong Markov process. For all $0 \leq s \leq t$, $x \in \mathcal{X}$, we denote by $\Phi(x, s, t)$ the corresponding stochastic flow i.e. $(\Phi(x, s, t), t \geq s)$ is the unique solution of the martingale problem associated with $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$ satisfying $\Phi(x, s, s) = x$. We have the following properties:

- for all $f \in \mathcal{D}(\mathcal{G})$, $0 \leq s \leq t$ and $x \in \mathcal{X}$:

$$f(\Phi(x, s, t)) - f(x) - \int_s^t \mathcal{G}f(\Phi(x, s, r)) dr, \quad (2.1)$$

is a $\sigma(X_t, t \geq 0)$ -martingale where $\sigma(X_t, t \geq 0)$ is the natural filtration associated with X .

- for each $0 \leq s \leq t$, $\Phi(\cdot, s, t)$ is a measurable map from \mathcal{X} to \mathcal{X} ,
- for each $0 \leq r \leq s \leq t$ and all $x \in \mathcal{X}$, $\Phi(\Phi(x, r, s), s, t) = \Phi(x, r, t)$, almost surely.

We refer the reader to [36] for more properties on stochastic flows.

Remark 2.1. According to the Hille-Yoshida theorem (see [21, Theorem 1.2.6]), $\mathcal{D}(\mathcal{G})$ is dense in $\mathcal{C}_b(\mathcal{X})$ for the topology of uniform convergence.

Division events. An individual with trait x dies at an instantaneous rate $B(x)$, where B is a continuous function from \mathcal{X} to \mathbb{R}_+ . It is replaced by $A_u(x)$ children, where $A_u(x)$ is a \mathbb{N} -valued random variable with distribution $(p_k(x), k \geq 0)$. For convenience, we assume that $p_1(x) \equiv 0$ for all $x \in \mathcal{X}$. The trait at birth of the j th descendant among k is given by the random variable $F_j^{(k)}(x, \theta)$, where $(F_j^{(k)}(\cdot, \cdot), j \leq k, k \in \mathbb{N})$ is a family of measurable functions from $\mathcal{X} \times [0, 1]$ to \mathcal{X} and θ is a uniform random variable on $[0, 1]$. This formalism will prove useful for the use of Poisson point measures. For all $k \in \mathbb{N}$, let $P^{(k)}(x, \cdot)$ be the probability measure on \mathcal{X}^k corresponding to the trait distribution at birth of the k descendants of an individual with trait x . We denote by $P_j^{(k)}(x, \cdot)$ the j th marginal distribution of $P^{(k)}$ for all $k \in \mathbb{N}$ and $j \leq k$ i.e. for all Borel sets $A \subset \mathcal{X}$, we have $P_j^{(k)}(x, A) = P^{(k)}(x, \mathcal{X}^{j-1} \times A \times \mathcal{X}^{k-j})$.

We denote by $\mathcal{M}_P(\mathcal{X})$ the set of point measures on \mathcal{X} . Following Fournier and Méléard [22], we work in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{X}))$, the state of càdlàg measure-valued processes. For any $\bar{Z} \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{U} \times \mathcal{X}))$, we write $\bar{Z}_t = \sum_{u \in V_t} \delta_{(u, X_t^u)}(du, dx)$ and:

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u}, \quad t \geq 0,$$

the marginal measure of $\bar{Z}_t(du, dx)$ on \mathcal{X} , where V_t represents the set of individuals alive at time t . We set $N_t = \#V_t$. Moreover, for any process $\bar{Z} \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{U} \times \mathcal{X}))$, we define recursively the associated sequence of jump times by

$$T_0(\bar{Z}) = 0 \text{ and } T_{k+1}(\bar{Z}) = \inf \{t > T_k(\bar{Z}), N_t \neq N_{T_k(\bar{Z})}\},$$

with the standard convention that $\inf \{\emptyset\} = +\infty$.

In order to ensure the non-explosion in finite time of such a process, we need to consider two sets of hypotheses. The first one controls what happens regarding divisions (in term of rate of division and of mass creation).

Assumption A. *We consider the following assumptions:*

1. *There exist $b_1, b_2 \geq 0$ and $\gamma \geq 1$ such that for all $x \in \mathcal{X}$,*

$$B(x) \leq b_1 |x|^\gamma + b_2.$$

2. *For all $t \geq 0$, there exists $\ell(t) \in \mathbb{R}_+$, increasing in t , such that for all $x = (y, t) \in \mathcal{X}$, $k \in \mathbb{N}$ and $\theta \in [0, 1]$:*

$$\sum_{i=1}^k F_i^{(k)}(x, \theta) \leq x \vee \ell(t), \text{ componentwise.}$$

3. *There exists $\bar{m} \geq 0$ such that for all $x \in \mathcal{X}$,*

$$m(x) = \sum_k k p_k(x) \leq \bar{m}.$$

4. *For all $x \in \mathcal{X}$ and $s \geq 0$, we have:*

$$\lim_{t \rightarrow +\infty} \int_s^t B(\Phi(x, s, r)) dr = +\infty, \text{ almost surely.}$$

The first point controls the life-times of individuals via the division rate. In particular, if $\gamma = 0$, B is bounded and the non-explosion in finite time of the number of individuals in the previously defined process is obvious. In more general framework, we have to consider the other point of Assumption A in order to prove the non-explosion in finite time. The second point of Assumption A means that we consider a fragmentation process with a possibility of mass creation at division when the mass is small enough. In particular, clones are allowed in the case of bounded traits and bounded number of descendants and any finite type branching structured process can be considered. The dependence in t of the threshold ℓ allows us to consider models in a varying environment. The last point of Assumption A ensures that each individual divides after a certain time.

We make a second assumption to control the behavior of traits between divisions.

Assumption B. *There exist $c_1, c_2 \geq 0$ such that for all $x \in \mathcal{X}$:*

$$\mathcal{G}h_\gamma(x) \leq c_1 h_\gamma(x) + c_2,$$

where γ is defined in Assumption A and for $x \in (\mathbb{R}_+)^d$, $h_\gamma(x) = |x|^\gamma = \left(\sum_{i=1}^d x_i\right)^\gamma$.

Assumptions A(1) and B are linked via the parameter γ which controls the balance between the growth of the population and the dynamic of the trait.

2.1 Existence and uniqueness of the structured branching process

We now prove the strong existence and uniqueness of the structured branching process. Let $E = \mathcal{U} \times \mathbb{R}_+ \times [0, 1] \times [0, 1]$ and $M(ds, du, dz, dl, d\theta)$ be a Poisson point measure on $\mathbb{R}_+ \times E$ with intensity $ds \otimes n(du) \otimes dz \otimes dl \otimes d\theta$, where $n(du)$ denotes the counting measure on \mathcal{U} . Let $(\Phi^u)_{u \in \mathcal{U}}$ be a family of independent stochastic flows satisfying (2.1) describing the individual-based dynamics. We assume that M and $(\Phi^u)_{u \in \mathcal{U}}$ are independent. We denote by \mathcal{F}_t the filtration generated by the Poisson point measure M and the family of stochastic flows $(\Phi^u(x, s, t), u \in \mathcal{U}, x \in \mathcal{X}, s \leq t)$ up to time t .

For all $x \in \mathcal{X}$, there exists a function $G(x, \cdot) : [0, 1] \rightarrow \mathbb{N}$ such that:

$$G(x, l) \stackrel{d}{=} (p_k(x), k \in \mathbb{N}),$$

where l is a uniform random variable on $[0, 1]$. This formalism will prove useful in the use of Poisson point measure to describe the jumps in the measure-valued branching process. For convenience, for all $x \in \mathcal{X}$ and θ, l uniform random variables on $[0, 1]$, we write:

$$F_i(x, l, \theta) = F_i^{(G(x, l))}(x, \theta).$$

We denote by

$$\bar{\mathcal{D}}(\mathcal{G}) := \{f : \mathcal{U} \times \mathcal{X} \rightarrow \mathbb{R} \text{ such that } f(u, \cdot) \in \mathcal{D}(\mathcal{G}) \forall u \in \mathcal{U}\}.$$

For all $0 \leq s \leq t$, $f \in \bar{\mathcal{D}}(\mathcal{G})$, $x \in \mathcal{X}$ and $u \in \mathcal{U}$, we consider the \mathcal{F}_t -martingale $(M_{s,t}^{f,u}(x), t \geq s)$ defined by:

$$M_{s,t}^{f,u}(x) := f(u, t, \Phi^u(x, s, t)) - f(u, s, x) - \int_s^t (\mathcal{G}f(u, r, \Phi^u(x, s, r)) + \partial_r f(u, r, \Phi^u(x, s, r))) dr. \quad (2.2)$$

Theorem 2.2. *Under Assumptions A(1-3) and B, there exists a strongly unique \mathcal{F}_t -adapted càdlàg process taking values in $\mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ such that for all $f \in \bar{\mathcal{D}}(\mathcal{G})$ and $t \geq 0$:*

$$\begin{aligned} \langle \bar{Z}_t, f \rangle &= f(\emptyset, 0, x_0) + \int_0^t \int_{\mathcal{U} \times \mathcal{X}} (\mathcal{G}f(u, s, x) + \partial_s f(u, s, x)) \bar{Z}_s(du, dx) ds + M_{0,t}^f(x) \\ &+ \int_0^t \int_E \mathbf{1}_{\{u \in V_{s-}, z \leq B(X_{s-}^u)\}} \left(\sum_{i=1}^{G(X_s^u, l)} f(u, s, F_i(X_s^u, l, \theta)) - f(u, s, X_{s-}^u) \right) M(ds, du, dz, dl, d\theta), \end{aligned} \quad (2.3)$$

where for all $s \geq 0$ and $t \geq s$:

$$M_{s,t}^f(x) = \sum_{k \geq 1} \mathbf{1}_{s \leq T_{k-1}(\bar{Z}) < t} \sum_{u \in V_{T_{k-1}(\bar{Z})}} M_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}^{f,u} \left(X_{T_{k-1}(\bar{Z})}^u \right),$$

is a \mathcal{F}_t -martingale.

The existence and uniqueness of such measure-valued process has first been studied by Fournier and Méléard [22]. We also refer to [46, 6] for different extensions and to [7] for the case of branching processes. Here, we obtain the non-explosion of the branching process in finite time under quite general assumptions (no bounded branching rate, random number of offspring, random transmission of the trait).

The proof of this theorem is split into four lemmas. First, in Lemma 2.3, we prove the existence of a \mathcal{F}_t -adapted càdlàg measure-valued process \bar{Z} solution of (2.3) for all $t \in [0, T_k(\bar{Z})]$ and all $k \in \mathbb{N}$. Then, in Lemma 2.4, we prove that $(M_{s,t}^f, t \geq 0)$ is a \mathcal{F}_t -martingale. Next, in Lemma 2.5, we prove the uniqueness of the increasing sequence $(T_k(\bar{Z}), k \geq 0)$ corresponding to the jump times of a solution \bar{Z} to (2.3) and the uniqueness of a \mathcal{F}_t -adapted càdlàg solution to (2.3) for $t \in [0, T_k(\bar{Z})]$ for all $k \in \mathbb{N}$. Finally, in Lemma 2.6, we prove that the sequence of jump times tends to infinity resulting in the existence and uniqueness of the process on \mathbb{R}_+ .

Lemma 2.3. *There exists a \mathcal{F}_t -adapted càdlàg measure-valued process $(\bar{Z}_t, t \geq 0) \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ which is solution of (2.3) for all $f \in \bar{\mathcal{D}}(\mathcal{G})$ and for all $t \in [0, T_k(\bar{Z}))$, $k \in \mathbb{N}$.*

Proof. See Section A in the appendix. \square

The existence of such processes has already been studied in [7] in the case of a trait following a Feller diffusion. From Lemma 2.3, we deduce the existence of a càdlàg measure-valued process $Z \in \mathcal{M}_P(\mathcal{X})$ solution of (2.3) which is given by the projection of the solution $\bar{Z} \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ on the second coordinate.

Lemma 2.4. *Let $\bar{Z} \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ be a solution of (2.3) whose construction is given in the previous lemma. Let $k \in \mathbb{N}$. For all $0 \leq s \leq t \leq T_k(\bar{Z})$, $x \in \mathcal{X}$ and $f \in \bar{\mathcal{D}}(\mathcal{G})$,*

$$M_{s,t}^f(x) = \sum_{k \geq 1} \mathbf{1}_{\{s \leq T_{k-1}(\bar{Z}) < t\}} \sum_{u \in V_{T_{k-1}(\bar{Z})}} M_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}^{f,u} \left(X_{T_{k-1}(\bar{Z})}^u \right),$$

is an \mathcal{F}_t -martingale.

Proof. Let $k \in \mathbb{N}$ and $0 \leq s \leq t \leq T_k(\bar{Z})$. Let $f \in \bar{\mathcal{D}}(\mathcal{G})$ and $x \in \mathcal{X}$. Then, for all $s \leq r \leq t$, we have:

$$\begin{aligned} & \mathbb{E} \left[M_{s,t}^f(x) | \mathcal{F}_r \right] - M_{s,r}^f(x) \\ = & \mathbb{E} \left[\sum_{k \geq 1} \mathbf{1}_{\{r \leq T_{k-1}(\bar{Z}) < t\}} \sum_{u \in V_{T_{k-1}(\bar{Z})}} M_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}^{f,u} \left(X_{T_{k-1}(\bar{Z})}^u \right) | \mathcal{F}_r \right] \\ = & \mathbb{E} \left[\sum_{k \geq 1} \mathbf{1}_{\{r \leq T_{k-1}(\bar{Z}) < t\}} \sum_{u \in V_{T_{k-1}(\bar{Z})}} \mathbb{E} \left[M_{T_{k-1}(\bar{Z}), T_k(\bar{Z}) \wedge t}^{f,u} \left(X_{T_{k-1}(\bar{Z})}^u \right) | \mathcal{F}_{T_{k-1}(\bar{Z})} \right] | \mathcal{F}_r \right] = 0, \end{aligned}$$

because $(M_{s,t}^{f,u}(x), t \geq s)$ is a \mathcal{F}_t -martingale. \square

Next, we prove the uniqueness of the sequence of jump times $(T_k(\bar{Z}), k \geq 0)$ associated with a solution $\bar{Z} \in \mathcal{M}_P(\mathcal{U} \times \mathcal{X})$ to (2.3) and the uniqueness of the solution on $[0, T_k(\bar{Z}))$, for all $k \in \mathbb{N}$. We refer to [46] for similar results.

Lemma 2.5. *The increasing sequence $(T_k(\bar{Z}), k \geq 0)$ corresponding to the jump times of a solution \bar{Z} to (2.3) is strongly unique. Moreover, the strong uniqueness of a \mathcal{F}_t -adapted càdlàg measure-valued solution to (2.3) holds, for $t \in [0, T_k(\bar{Z}))$ and for all $k \in \mathbb{N}$.*

Proof. See Section B in the appendix. \square

Lemma 2.6. *Under Assumptions A(1-3) and B, the strongly unique sequence of jump times of a solution \bar{Z} to (2.3) tends to infinity as k tends to infinity, almost surely.*

Proof. Let $T > 0$. To shorten notation, we write T_k instead of $T_k(\bar{Z})$. We prove that almost surely there is no accumulation of jumps on $[0, T]$ of the solution of (2.3) previously constructed on $[0, T_k[$, for all $k \in \mathbb{N}$. Let $k \in \mathbb{N}$ and $(\bar{Z}_t, t \leq T_k)$ be the solution of (2.3) up to the k th division time. Using equation (2.3) applied to the constant function equal to 1, we have for all $t \leq T_k \wedge T$:

$$\begin{aligned} \mathbb{E}_{\delta_x}(N_t) &= 1 + \int_0^t \mathbb{E}_{\delta_x} \left(\sum_{u \in V_s} B(X_s^u) (m(X_s^u) - 1) \right) ds \\ &\leq 1 + \bar{m}b_1 \int_0^t \mathbb{E}_{\delta_x} \left(\sum_{u \in V_s} |X_s^u|^\gamma \right) ds + \bar{m}b_2 \int_0^t \mathbb{E}_{\delta_x}(N_s) ds, \end{aligned} \tag{2.4}$$

where the inequality comes from Assumption A(1) and A(3). Recalling that $h_\gamma(x) = \left(\sum_{i=1}^d x_i\right)^\gamma$, for $x \in (\mathbb{R}_+)^d$, we have using (2.3):

$$\begin{aligned} \mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} |X_t^u|^\gamma \right] &= |x|^\gamma + \int_0^t \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} \mathcal{G}h_\gamma(X_s^u) \right] ds \\ &\quad + \int_0^t \int_{[0,1]} \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} B(X_s^u) \sum_{k \geq 0} p_k(X_s^u) \left(\sum_{j=1}^k |F_j^{(k)}(X_s^u, \theta)|^\gamma - |X_s^u|^\gamma \right) \right] d\theta ds. \end{aligned}$$

Next, using Assumption A(2) and B, we get that:

$$\mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} |X_t^u|^\gamma \right] \leq |x|^\gamma + \int_0^t \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} (c_1 |X_s^u|^\gamma + c_2) \right] + \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} B(X_s^u) \ell(s) \mathbf{1}_{\{|X_s^u|^\gamma \leq \ell(s)\}} \right] ds.$$

Finally, using Assumption A(1) and the fact that $t \mapsto \ell(t)$ is increasing, we get:

$$\mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} |X_t^u|^\gamma \right] \leq |x|^\gamma + c_1 \int_0^t \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} |X_s^u|^\gamma \right] ds + (c_2 + (b_1 \ell(t) + b_2) \ell(t)) \int_0^t \mathbb{E}_{\delta_x} [N_s] ds.$$

Adding this inequality to (2.4) we obtain for all $t \leq T_k \wedge T$:

$$\mathbb{E}_{\delta_x} [N_t] + \mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} |X_t^u|^\gamma \right] \leq 1 + |x|^\gamma + A(T) \int_0^t \left(\mathbb{E}_{\delta_x} [N_s] + \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} |X_s^u|^\gamma \right] \right) ds,$$

where $A(T) = c_1 + c_2 + b_1 \ell(T)^2 + b_2 \ell(T) + (b_1 + b_2) \bar{m}$. According to Grönwall Lemma, we obtain for all $t \leq T_k \wedge T$:

$$\mathbb{E}_{\delta_x} [N_t] + \mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} |X_t^u|^\gamma \right] \leq (1 + |x|^\gamma) e^{A(T)t} < \infty.$$

Finally, the average number of individuals in the population at time t is bounded for t in compact sets and there is no explosion of the population in finite time. \square

Before moving to the next section, we introduce the first-moment semi-group $(R_{s,t}, t \geq s)$ associated with the branching process: for all $s \geq 0, t \geq s$ and $x \in \mathcal{X}$, let

$$R_{s,t} f(x) = \mathbb{E} \left[\sum_{u \in V_t} f(X_t^u) \mid Z_s = \delta_x \right], \quad (2.5)$$

where f is a measurable function. Applying equation (2.3) to $f \equiv 1$, we obtain by taking the expectation :

$$R_{s,t} \mathbf{1}(x) = m(x, s, t) = 1 + \int_s^t \mathbb{E} \left[\sum_{u \in V_r} B(X_r^u) (m(X_r^u) - 1) \mid Z_s = \delta_x \right] dr. \quad (2.6)$$

In particular, if $B \equiv b$ and $m(x) = m$ for all $x \in \mathcal{X}$, we obtain $m(x, s, t) = e^{b(m-1)(t-s)}$.

2.2 Some growth-fragmentation models for cell population dynamics

In this section, we consider growth-fragmentation processes: at division, the trait of the ancestor is shared between the children and the number of individuals in the population increases. Moreover, we focus on models where the trait moves according to a diffusion given by its associated generator of following form:

$$\mathcal{G}f(x) = r(x)f'(x) + \sigma^2(x)f''(x),$$

where r and σ are measurable functions. This class covers several dynamics for the trait. Here, we present three of them. In particular, we give an explicit formula for the average number of individuals in the population at time t . We first give a useful equation concerning models with such a dynamic. For all $s \geq 0$, $t \geq s$ and $x \in \mathcal{X}$, applying (2.3) to $f(x) = x$ and taking the expectation, we obtain:

$$R_{s,t}\mathbf{Id}(x) = x + \int_s^t R_{s,u}r(x)du, \quad (2.7)$$

where $(R_{s,t})_{t \geq s}$ is defined in (2.5).

2.2.1 Linear growth model

We consider here a size-structured model. More precisely, the size of each cell grows linearly at a rate $a > 0$ supposed to be identical for each cell and divisions occur at rate $B(x) = \alpha x$, $\alpha > 0$. At fission, the cell splits into two daughter cells of size $\frac{x}{2}$, when x denotes the size of the mother at splitting. Deciding whether the cells' growth follows a linear or an exponential dynamic has fueled a large debate in the literature (see [16] and references therein). The linear growth model has been considered for example in [19] for the calibration of a deterministic growth-fragmentation model from experimental data and in [30] for the estimation of the division rate.

Using the previous notation, the process $(X_t, t \geq 0)$ describing the size of a cell starting from x_0 is given by:

$$X_t = x_0 + at,$$

and the associated generator is given for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ by:

$$\mathcal{G}f(x) = af'(x).$$

Then, the branching process $(Z_t, t \geq 0)$ is solution of the following equation, for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned} \langle Z_t, f \rangle = & \langle Z_0, f \rangle + \int_0^t \int_{\mathbb{R}_+} af'(x)Z_s(dx)ds \\ & + \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+} \mathbf{1}_{\{u \in V_{s^-}, z \leq \alpha X_{s^-}^u\}} \left(2f\left(\frac{X_s^u}{2}\right) - f(X_{s^-}^u) \right) M(ds, du, dz), \end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+$ with intensity $ds \otimes n(du) \otimes dz$. The first integral corresponds to the dynamic of the population between two divisions. The integral with respect to the Poisson point measure represents to the jump part of the process and the indicator function corresponds to the fact that an individual u jumps at time s if it is in the population at time s^- and if the division rate at $X_{s^-}^u$ is large enough. In this case u , it is removed from the population and two descendants with trait $X_{s^-}^u/2$ appear.

The validity of Assumptions A and B is trivial for this model with $\gamma = 1$. Let us compute the average number of individuals in the population at time t . For all $s \leq t$ and $x \in \mathbb{R}$, we have using (2.6):

$$m(x, s, t) = 1 + \alpha \int_s^t \mathbb{E} \left(\sum_{u \in V_r} X_r^u \mid Z_s = \delta_x \right) dr. \quad (2.8)$$

Combining (2.7) and (2.8), we obtain:

$$m(x, s, t) = 1 + \alpha \int_s^t \left(x + a \int_s^r m(x, s, \tau) d\tau \right) dr,$$

and for all $x \in \mathcal{X}$ and $s \geq 0$, $m(x, s, \cdot)$ is the solution of the following Cauchy problem with unknown f :

$$\begin{cases} f''(t) = a\alpha f(t), \\ f(s) = 1, \quad f'(s) = \alpha x. \end{cases}$$

with explicit solution given by:

$$m(x, s, t) = \frac{1}{2} \left(e^{\bar{a}(t-s)} + e^{-\bar{a}(t-s)} \right) + \frac{x}{2} \sqrt{\frac{\alpha}{a}} \left(e^{\bar{a}(t-s)} - e^{-\bar{a}(t-s)} \right),$$

where $\bar{a} = \sqrt{a\alpha}$. The population size is exponential in time as in the neutral case.

2.2.2 Exponential growth model in a varying environment

We assume here that the growth of the cells is exponential at rate a . This exponential growth model has been studied in [18] in the case of a specific growth rate for each individual in order to infer the division rate of the population. Here, we assume that the division rate is a function of time, mimicking a varying environment. More precisely, we set $B(x, t) = \alpha(t)x$, with α a positive function. The generator for the dynamic of the size is given for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ by:

$$\mathcal{G}f(x) = axf'(x).$$

We still assume that the branching is binary and that the size of the descendants at birth are both $x/2$ if x is the size of the mother at splitting. Then, the branching process $(Z_t, t \geq 0)$ is solution of the following equation, for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned} \langle Z_t, f \rangle &= \langle Z_0, f \rangle + \int_0^t \int_{\mathbb{R}_+} axf'(x)Z_s(dx)ds \\ &\quad + \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+} \mathbf{1}_{\{u \in V_{s^-}, z \leq \alpha(s)X_{s^-}^u\}} \left(2f\left(\frac{X_{s^-}^u}{2}\right) - f(X_{s^-}^u) \right) M(ds, du, dz), \end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+$ with intensity $ds \otimes n(du) \otimes dz$. Moreover, we have using (2.7) with $r(x) = ax$:

$$\mathbb{E} \left(\sum_{u \in V_t} X_t^u \mid Z_s = \delta_x \right) = xe^{a(t-s)}.$$

Combining this with equation (2.6), we obtain:

$$m(x, s, t) = 1 + x \int_s^t \alpha(r) e^{a(r-s)} dr.$$

In particular, if $\alpha(r) \equiv \alpha$ with α a positive constant, we obtain:

$$m(x, s, t) = 1 + \frac{\alpha x}{a} \left(e^{a(t-s)} - 1 \right).$$

The growth is again exponentially fast in time.

2.2.3 Parasite infection model

This model is a continuous version of Kimmel's multilevel model for plasmids [34] which has already been studied in the case of a constant or monotone division rate by Bansaye and Tran in [7]. It models the proliferation of a parasite infection in a cell population. More precisely, we assume here that the trait $(X_t, t \geq 0)$ is a Markov process describing the quantity of parasites in each cell which evolves as a Feller diffusion process:

$$X_t = X_0 + \int_0^t gX_s ds + \int_0^t \sqrt{2\sigma^2 X_s} dB_s,$$

where $(B_s)_{s \geq 0}$ is standard Brownian motion and $g, \sigma > 0$ are some fixed parameters. The generator for the dynamic of the quantity of parasites is given for any function $f \in \mathcal{C}^2(\mathbb{R}_+)$ by:

$$\mathcal{G}f(x) = gx f'(x) + \sigma^2 x f''(x).$$

We assume here that a cell with a quantity x of parasites will potentially divide at a rate $B(x) = \alpha x + \beta$, $\alpha, \beta > 0$ into two daughter cells with a quantity δx and $(1 - \delta)x$ of parasites respectively, where δ is a random variable with uniform distribution on $[0, 1]$. We need β to be strictly positive so that even cells without any parasites divide after some time. The branching process $(Z_t, t \geq 0)$ is then solution of the following equation, for any function $f \in \mathcal{C}^2(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned} \langle Z_t, f \rangle &= \langle Z_0, f \rangle + \int_0^t \int_{\mathbb{R}_+} (gx f'(x) + \sigma^2 x f''(x)) Z_s(dx) ds + \int_0^t \sum_{u \in V_s} \sqrt{2\sigma^2 X_s^u} f'(X_s^u) dB_s^u \\ &+ \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+ \times [0,1]} \mathbf{1}_{\{u \in V_{s^-}, z \leq \alpha X_{s^-}^u + \beta\}} (f(\delta X_{s^-}^u) + f((1 - \delta)X_{s^-}^u) - f(X_{s^-}^u)) M(ds, du, dz, d\delta), \end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+ \times [0, 1]$ with intensity $ds \otimes n(du) \otimes dz \otimes d\delta$ and $(B_s^u, s \geq 0)_{u \in \mathcal{U}}$ is a family of standard Brownian motions. In particular, the generator corresponding to first moment semi-group is given for any function $f \in \mathcal{C}^2(\mathbb{R})$ and $x \in \mathcal{X}$ by:

$$\mathcal{F}_{\text{inf}} f(x) = gx f'(x) + \sigma^2 x f''(x) + (\alpha x + \beta) \left(\int_0^1 [f(\delta x) + f((1 - \delta)x)] d\delta - f(x) \right).$$

Therefore, we notice that if (V, λ) are eigenelements of \mathcal{F}_{inf} , we have $\mathcal{F}_{\text{inf}} V(0) = \beta V(0)$ so that $V(0) = 0$ if $\lambda \neq \beta$ and we cannot apply usual techniques using eigenelements.

Let us compute the average number of individuals in the population after time t . We have using (2.6):

$$m(x, s, t) = 1 + \alpha \int_s^t \mathbb{E} \left[\sum_{u \in V_r} X_r^u | Z_s = \delta_x \right] dr + \beta \int_s^t m(x, s, r) dr.$$

Again, using (2.7), we obtain:

$$\mathbb{E} \left[\sum_{u \in V_r} X_r^u | Z_s = \delta_x \right] = x e^{g(r-s)}.$$

Then, combining the two previous equations, we get after differentiation:

$$\partial_t m(x, s, t) = \alpha x e^{g(t-s)} + \beta m(x, s, t),$$

and finally:

$$m(x, s, t) = \frac{\alpha x}{g - \beta} e^{g(t-s)} + \left(1 - \frac{\alpha x}{g - \beta} \right) e^{\beta(t-s)},$$

if $g \neq \beta$ and:

$$m(x, s, t) = (1 + \alpha(t - s)) x e^{\beta(t-s)},$$

if $g = \beta$. In the three examples above, the mean number of individuals in the population is an affine function of the trait of the initial individual. However, this is not the rule. For example, Cloez developed in [15](Corollary 6.1.) the case of a dynamic of the trait following an Ornstein-Uhlenbeck process where the dependence in x is not affine.

For other examples and comments, including a link with the integro-differential model, we refer to Section 5.

3 The trait of sampled individuals at a fixed time : Many-to-One formulas

In order to characterize the trait of a uniformly sampled individual, the spinal approach ([13],[40]), consists in following a "typical" individual in the population whose behavior summarizes the behavior of the entire population. Biggins [9] used this approach for the study of branching random walks extending Kingman results [35]. The spinal approach has then been extended to various frameworks ([28],[37],[25]). In particular, Georgii and Baake [23] used spine techniques in a spectral framework to describe the asymptotic distribution of the trait of a uniformly sampled individual in the population and its ancestral lineage in the case of a finite set of possible trait.

In this section, we specify the generator of the process describing the trait along the spine. The existence of our auxiliary process does not rely on the existence of spectral elements for the mean operator of the branching process.

With a slight abuse of notation, for all $u \in V_t$ and $s < t$, we denote by X_s^u the trait of the unique ancestor living at time s of u .

3.1 The auxiliary process

Let us define

$$\mathcal{D}(\mathcal{A}) = \{f \in \mathcal{D}(\mathcal{G}) \text{ s.t. } m(\cdot, s, t)f \in \mathcal{D}(\mathcal{G}) \forall t \geq 0, \forall s \leq t\}.$$

From now on, we assume that for all $x \in \mathcal{X}$, $t \geq 0$ and $s \leq t$, $m(x, s, t) \neq 0$.

We now recall the operator and functions needed for the definition of the auxiliary process, and introduce additional notations. For all $f \in \mathcal{D}(\mathcal{A})$, $x \in \mathcal{X}$ and $s < t$, we write:

$$\widehat{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}(m(\cdot, s, t)f)(x) - f(x)\mathcal{G}(m(\cdot, s, t))(x)}{m(x, s, t)}, \quad (3.1)$$

$$\widehat{B}_s^{(t)}(x) = B(x)\Lambda(x, s, t), \quad (3.2)$$

$$\widehat{P}_s^{(t)}(x, dy) = \Lambda^{-1}(x, s, t) \frac{m(y, s, t)}{m(x, s, t)} m(x, dy), \quad (3.3)$$

where:

$$\Lambda(x, s, t) = \int_{\mathcal{X}} \frac{m(y, s, t)}{m(x, s, t)} m(x, dy).$$

In order to prove a Many-to-One formula, we need to consider the following assumption:

Assumption C. *There exists a function C such that for all $j \leq k$, $j, k \in \mathbb{N}$ and $0 \leq s \leq t$, we have:*

$$\sup_{x \in \mathcal{X}} \sup_{s \in [0, t]} \int_{\mathcal{X}} \frac{m(y, s, t)}{m(x, s, t)} P_j^{(k)}(x, dy) \leq C(t), \quad \forall t \geq 0.$$

Assumption D. For all $t \geq 0$ we have:

- for all $x \in \mathcal{X}$, $s \mapsto m(x, s, t)$ is differentiable on $[0, t]$ and its derivative is continuous on $[0, t]$,
- for all $x \in \mathcal{X}$, $f \in \mathcal{D}(\mathcal{A})$, $s \mapsto \mathcal{G}(m(\cdot, s, t)f)(x)$ is continuous,
- $\mathcal{D}(\mathcal{A})$ is dense in $\mathcal{C}(\mathcal{X})$ for the topology of uniform convergence.

The last item of this assumption allows us to extend our formulas to all measurable functions with respect to the Skorokod topology using a monotone class argument. Moreover, combining Lemma 3.4 and Remark 2.1, this assumption is in particular satisfied if $\mathcal{D}(\mathcal{G})$ is stable by product.

Theorem 3.1. Under Assumptions A(1-3), B, C and D, for all $t \geq 0$, for all $x_0 \in \mathcal{X}$ and for all non-negative measurable functions $F : \mathbb{D}([0, t], \mathcal{X}) \rightarrow \mathbb{R}_+$ we have:

$$\mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_t} F(X_s^u, s \leq t) \right] = m(x_0, 0, t) \mathbb{E}_{x_0} \left[F(Y_s^{(t)}, s \leq t) \right], \quad (3.4)$$

where $(Y_s^{(t)}, s \leq t)$ is a time-inhomogeneous Markov process whose law is characterized by its associated infinitesimal generators $(\mathcal{A}_s^{(t)})_{s \leq t}$ given for $f \in \mathcal{D}(\mathcal{A})$ and $x \in \mathcal{X}$ by:

$$\mathcal{A}_s^{(t)} f(x) = \widehat{\mathcal{G}}_s^{(t)} f(x) + \widehat{B}_s^{(t)}(x) \int_{\mathcal{X}} (f(y) - f(x)) \widehat{P}_s^{(t)}(x, dy). \quad (3.5)$$

Formula (3.4) has a natural interpretation in terms of semi-groups. If f is a non-negative measurable function, we set for any $0 \leq r \leq s \leq t$ and any $x \in \mathcal{X}$:

$$P_{r,s}^{(t)} f(x) := \frac{\mathbb{E} \left[\sum_{u \in V_t} f(X_s^u) \mid Z_r = \delta_x \right]}{m(x, r, t)} = \mathbb{E} \left[f(Y_s^{(t)}) \mid Y_r^{(t)} = x \right]. \quad (3.6)$$

In other words, $(P_{r,s}^{(t)})_{r \leq s \leq t}$ is a conservative (non-homogeneous) semi-group and the auxiliary process $Y^{(t)}$ is its time-inhomogeneous associated Markov process corresponding to the right-hand side of (1.1). We can exhibit this process using a change of probability measure. Indeed, by Feynman-Kac's formula, we have

$$P_{r,s}^{(t)} f(x) = m(x, r, t)^{-1} \mathbb{E} \left[e^{\int_r^s B(X_v)(m(X_v)-1)dv} m(X_s, s, t) f(X_s) \mid X_r = x \right],$$

where the Markov process $(X_s, r \leq s \leq t)$ corresponds to dynamic of the tagged-particle which infinitesimal generator \mathcal{M} is given by

$$\mathcal{M}f(x) = \mathcal{G}f(x) + B(x)m(x) \sum_{k \geq 0} \frac{k p_k(x)}{m(x)} \frac{1}{k} \sum_{i=1}^k \int_{\mathcal{X}} (f(y) - f(x)) P_i^{(k)}(x, dy).$$

Then, the change of probability measure given by the $\sigma(X_l, l \leq s)$ -martingale

$$M_s^{(t)} := \frac{e^{\int_r^s B(X_s)(m(X_s)-1)ds} m(X_s, s, t)}{m(x, r, t)}, \quad \text{for } r \leq s \leq t$$

exhibits the probability measure corresponding to the auxiliary process.

Before proving Theorem 3.1, we give some links between our approach and previous works on this subject. In the neutral case, i.e. B and $(p_k)_{k \in \mathbb{N}}$ constants, the auxiliary process coincides with the

one in [5] i.e. for all $f \in \mathcal{D}(\mathcal{G})$ and $x \in \mathcal{X}$, the infinitesimal generator of the auxiliary process is given by:

$$\mathcal{A}f(x) = \mathcal{G}f(x) + Bm \sum_{k \geq 0} \widehat{p}_k \left(\frac{1}{k} \sum_{j=1}^k \int_{\mathcal{X}} (f(y) - f(x)) P_j^{(k)}(x, dy) \right),$$

where $\widehat{p}_k = kp_k m^{-1}$ denote the biased reproduction law. The dynamic of this auxiliary process heavily depends on the comparison between $m(x, s, t)$ and $m(y, s, t)$, for $x, y \in \mathcal{X}$. It emphasizes several bias due to growth of the population. First, the auxiliary process jumps more than the original process, if jumping is beneficial in terms of number of descendants. This phenomenon of time-acceleration also appears for examples in [13], [40] or [25]. Moreover, the reproduction law favors the creation of a large number of descendant as in [5] and the non-neutrality favors individuals with an "efficient" trait at birth in terms of number of descendants. Finally, a new bias appears on the dynamic of the trait because of the combination of the random evolution of the trait and non-neutrality. Indeed, if the dynamic of the trait is deterministic, we have $\widehat{\mathcal{G}}_s^{(t)} f(x) = \mathcal{G}f(x)$.

The auxiliary process has been guessed through a discretization of the model using the expression of the auxiliary process in [4]. More precisely, we considered the discrete time auxiliary process with a probability of branching at each step depending on the division rate B . Then, we looked at the limit of the generator of this process as the discretization step goes to zero in order to obtain a formula for the continuous-time version of the auxiliary process. However, the proof of Theorem 3.1 does not rely on a discretization argument but on the uniqueness of the solution to the integro-differential equation (3.8). The proof is decomposed in four parts: first, in Lemma 3.2, we prove that the integro-differential equation (3.8) admits a unique solution which corresponds to the semi-group of the auxiliary process defined in (3.6). Afterwards, in Lemma 3.3, we prove that the infinitesimal generator of this auxiliary process verifies (3.5). Then, we prove Theorem 3.1 for any function such that $F(x) = f_1(x_{t_1}) \dots f_k(x_{t_k})$, $x \in \mathbb{D}([0, t], \mathcal{X})$, by induction on $k \in \mathbb{N}$. Finally, we extend the set of functions for which (3.4) is satisfied using a monotone class argument.

Let $t \geq 0$. We define the following family of semi-groups for $f \in \mathcal{D}(\mathcal{A})$:

$$Q_{s,r}^{(t)} f(x) = \frac{A_{r-s}(m(\cdot, r, t)f)(x)}{m(x, s, t)}, \quad s \leq r \leq t.$$

We also define:

$$\widetilde{\mathcal{G}}_s^{(t)} f(x) = \frac{\mathcal{G}(m(\cdot, s, t)f)(x) + f(x)\partial_s m(x, s, t)}{m(x, s, t)}. \quad (3.7)$$

Lemma 3.2. *Let $t \geq 0$. Under Assumptions A(1-3), B, C and D, for all $x_0 \in \mathcal{X}$ and $t_0 \leq t$, the family of probability measures $(P_{t_0,s}^{(t)}(x_0, \cdot), t_0 \leq s \leq t)$ is the unique solution of the following equation with unknown $(\mu_{t_0,s}(x_0, \cdot), t_0 \leq s \leq t)$:*

$$\begin{aligned} \mu_{t_0,s}(x_0, f) = & f(t_0, x_0) + \int_{t_0}^s \int_{\mathcal{X}} \left(\widetilde{\mathcal{G}}_r^{(t)} f(r, x) + \partial_r f(r, x) \right) \mu_{t_0,r}(x_0, dx) dr \\ & + \int_{t_0}^s \int_{\mathcal{X}} \left[\widehat{B}_r^{(t)}(x) \int_{\mathcal{X}} f(r, y) \widehat{P}_r^{(t)}(x, dy) - B(x)f(r, x) \right] \mu_{t_0,r}(x_0, dx) dr, \end{aligned} \quad (3.8)$$

for all function f such that, $f(s, \cdot) \in \mathcal{D}(\mathcal{A})$ for all $s \geq 0$ and $s \mapsto f(s, x)$ is continuously differentiable for all $x \in \mathcal{X}$.

Proof. Let $t \geq 0$ and let f be as in the statement of the lemma. The proof falls naturally into two parts. We first prove that $(P_{t_0,s}^{(t)}(x_0, \cdot), t_0 \leq s \leq t)$ is a solution of (3.8). We show that for all $t_0 \leq s \leq t$, $x_0 \in \mathcal{X}$,

$$m(x_0, t_0, t)P_{t_0,s}^{(t)} f(t_0, x_0) = \mathbb{E}(\langle Z_s, f(s, \cdot)m(\cdot, s, t) \rangle | Z_{t_0} = \delta_{x_0}).$$

Indeed, from (3.6), the left-hand side of the above equation is equal to:

$$\begin{aligned}
m(x_0, t_0, t)P_{t_0, s}^{(t)}f(t_0, x_0) &= \mathbb{E} \left[\sum_{v \in V_s} \sum_{\substack{u \in V_t \\ u \geq v}} f(s, X_s^v) \mid Z_{t_0} = \delta_{x_0} \right] \\
&= \mathbb{E} \left[\sum_{v \in V_s} f(s, X_s^v) \mathbb{E}_{\delta_{x_0}} \left(\sum_{\substack{u \in V_t \\ u \geq v}} 1 \mid \mathcal{F}_s \right) \mid Z_{t_0} = \delta_{x_0} \right] \\
&= \mathbb{E} \left[\sum_{v \in V_s} f(s, X_s^v) \mathbb{E} \left(\sum_{u \in V_t} 1 \mid Z_s = \delta_{X_s^u} \right) \right] \\
&= \mathbb{E} \left[\sum_{v \in V_s} f(s, X_s^v) m(X_s^v, s, t) \mid Z_{t_0} = \delta_{x_0} \right].
\end{aligned}$$

Then, applying (2.3) to the function $f(s, \cdot)m(\cdot, s, t)$ and taking the expectation, we obtain:

$$\begin{aligned}
&\mathbb{E} \left[\sum_{v \in V_s} f(s, X_s^v) m(X_s^v, s, t) \mid Z_{t_0} = \delta_{x_0} \right] = m(x_0, t_0, t)f(t_0, x_0) \\
&+ \int_{t_0}^s \int_{\mathcal{X}} (\mathcal{G}(f(r, \cdot)m(\cdot, r, t)))(x) + f(r, x)\partial_r m(x, r, t) + \partial_r f(r, x)m(x, r, t) R_{t_0, r}(x_0, dx) dr \\
&+ \int_{t_0}^s \int_{\mathcal{X}} B(x) \left(\sum_{k \geq 0} p_k(x) \sum_{j=1}^k \int_{\mathcal{X}} f(r, y) m(y, r, t) P_j^{(k)}(x, dy) - f(r, x) m(x, r, t) \right) R_{t_0, r}(x_0, dx) dr.
\end{aligned} \tag{3.9}$$

Finally, dividing by $m(x_0, t_0, t)$, we obtain that $(P_{t_0, s}^{(t)}(x_0, \cdot), t_0 \leq s \leq t)$ is a solution of (3.8).

We now prove the uniqueness of a solution to (3.8). Without loss of generality, we assume that $t_0 = 0$. This part of the proof is adapted from [7]. Let $(\gamma_{s, t}^1, s \leq t)$ and $(\gamma_{s, t}^2, s \leq t)$ be two solutions of equation (3.8). Let us recall that the total variation norm is given for all measures γ^1, γ^2 on \mathcal{X} with finite mass by:

$$\|\gamma^1 - \gamma^2\|_{TV} = \sup_{\phi \in \mathcal{C}_b(\mathcal{X}, \mathbb{R}), \|\phi\|_{\infty} \leq 1} |\gamma^1(\phi) - \gamma^2(\phi)|,$$

where $\mathcal{C}_b(\mathcal{X}, \mathbb{R})$ denotes the set of continuous bounded functions from \mathcal{X} to \mathbb{R} . The idea is to find a function which cancels the first integral in (3.8). Let $x \in \mathcal{X}$, $t \geq 0$ and $r \leq t$. We begin by computing the differential of $(Q_{s, r}^{(t)}f(x), s \leq r \leq t)$ with respect to s . First, $s \mapsto A_{r-s}(m(\cdot, r, t)f)(x)$ is differentiable because $x \mapsto m(x, r, t)f(x) \in \mathcal{D}(\mathcal{G})$ and according to the backward equation, its derivative is $s \mapsto \mathcal{G}(A_{r-s}(m(\cdot, r, t)f))(x) = A_{r-s}(\mathcal{G}(m(\cdot, r, t)f))(x)$. Furthermore, $s \mapsto m(x, s, t)^{-1}$ is differentiable because $s \mapsto m(x, s, t)$ is differentiable according to the first point of Assumption D and because $m(x, s, t) \neq 0$ for all $x \in \mathcal{X}$, $t \geq 0$ and $s \leq t$. Then, we have for all $s \geq 0$ and $r \geq s$:

$$\begin{aligned}
\partial_s Q_{s, r}^{(t)}f(x) &= \frac{\partial_s A_{r-s}(m(\cdot, r, t)f)(x)}{m(x, s, t)} - \frac{\partial_s m(x, s, t)}{m(x, s, t)^2} A_{r-s}(m(\cdot, r, t)f)(x) \\
&= -\frac{\mathcal{G}(A_{r-s}(m(\cdot, r, t)f))(x)}{m(x, s, t)} - \frac{\partial_s m(x, s, t)}{m(x, s, t)} \frac{A_{r-s}(m(\cdot, r, t)f)(x)}{m(x, s, t)} \\
&= -\left(\frac{\mathcal{G}(m(\cdot, r, t)Q_{s, r}^{(t)}f)(x)}{m(x, s, t)} + \frac{\partial_s m(x, s, t)}{m(x, s, t)} Q_{s, r}^{(t)}f(x) \right).
\end{aligned}$$

Therefore, for all $s \leq t$, $f \in \mathcal{D}(\mathcal{A})$, we have:

$$\partial_s Q_{s,r}^{(t)} f(x) = -\tilde{\mathcal{G}}_s^{(t)} Q_{s,r}^{(t)} f(x). \quad (3.10)$$

Let $f \in \mathcal{D}(\mathcal{A})$ be such that $\|f\|_\infty \leq 1$. Let $\tau_n(x) = \inf \{t \geq 0, X_t \notin \mathcal{B}(x, n)\}$ where $\mathcal{B}(x, n)$ is the \mathcal{X} -ball of radius n and centered in x . Let us define for all $x \in \mathcal{X}$, $s \leq r \leq t$ and $n \in \mathbb{N}$:

$$Q_{s,r}^{(t),n} f(x) = \frac{\mathbb{E}_x [m(X_{r \wedge \tau_n(x)-s}, r \wedge \tau_n(x), t) f(X_{r \wedge \tau_n(x)-s})]}{m(x, s, t)}.$$

We still have $\partial_s Q_{s,r}^{(t),n} f(x) = -\tilde{\mathcal{G}}_s^{(t)} Q_{s,r}^{(t),n} f(x)$. Moreover, for all $s \leq r \leq t$ and all $x \in \mathcal{X}$, we have:

$$\left| Q_{s,r}^{(t),n} f(x) \right| \leq \frac{\mathbb{E}_x [m(X_{r \wedge \tau_n(x)-s}, r \wedge \tau_n(x), t)]}{m(x, s, t)} \leq \frac{\mathbb{E}_x [m(X_{r \wedge \tau_n(x)-s}, r \wedge \tau_n(x), t)]}{\mathbb{E} \left[\mathbf{1}_{\Omega_r \wedge \tau_n(x)} m(X_{r \wedge \tau_n(x)}, r \wedge \tau_n(x), t) \mid Z_s = \delta_x \right]},$$

where $\Omega_r = \{T_1(\bar{Z}) > r\}$. Conditioning with respect to $\sigma(X_s, s \leq r \wedge \tau_n(x))$ on the denominator we obtain:

$$\begin{aligned} \left| Q_{s,r}^{(t),n} f(x) \right| &\leq \frac{\mathbb{E}_x [m(X_{r \wedge \tau_n(x)-s}, r \wedge \tau_n(x), t)]}{\mathbb{E} \left[\exp \left(- \int_0^{r \wedge \tau_n(x)} B(X_u) du \right) m(X_{r \wedge \tau_n(x)}, r \wedge \tau_n(x), t) \mid X_s^\emptyset = x \right]} \\ &\leq \frac{\mathbb{E} [m(X_{r \wedge \tau_n(x)}, r \wedge \tau_n(x), t) \mid X_s = x]}{\exp(-r \bar{B}_n(x)) \mathbb{E} [m(X_{r \wedge \tau_n(x)}, r \wedge \tau_n(x), t) \mid X_s^\emptyset = x]} \leq e^{r \bar{B}_n(x)}, \end{aligned} \quad (3.11)$$

where $\bar{B}_n(x) = \sup_{y \in \mathcal{B}(x, n)} B(y)$.

Let $T_n = \inf \{s \leq t, \gamma_{s,t}^1(x_0, \mathcal{B}(x_0, n)^c) + \gamma_{s,t}^2(x_0, \mathcal{B}(x_0, n)^c) > 0\}$ where $\mathcal{B}(x_0, n)^c$ is the complementary of the \mathcal{X} -ball of radius n and centered in x_0 with the convention that $\inf \emptyset = +\infty$. Then, using that $(\gamma_{s,t}^i, s \leq t)$, for $i = 1, 2$, are solutions of (3.8), we have for all $s \leq r \leq t$:

$$\begin{aligned} \langle \gamma_{s \wedge T_n, t}^i(x_0, \cdot), Q_{s \wedge T_n, r}^{(t),n} f \rangle &= Q_{0,r}^{(t),n} f(x_0) \\ &+ \int_0^{s \wedge T_n} \int_{\mathcal{X}} \left[\hat{B}_u^{(t)}(x) \int_{\mathcal{X}} Q_{u,r}^{(t),n} f(y) \hat{P}_u^{(t)}(x, dy) - B(x) Q_{u,r}^{(t),n} f(x) \right] \gamma_{u,t}^i(x_0, dx) du. \end{aligned}$$

Using (3.11), we get: we have:

$$\begin{aligned} &\left| \gamma_{s \wedge T_n, t}^1(x_0, Q_{s \wedge T_n, r}^{(t),n} f) - \gamma_{s \wedge T_n, t}^2(x_0, Q_{s \wedge T_n, r}^{(t),n} f) \right| \\ &= \left| \int_0^{s \wedge T_n} \int_{\mathcal{X}} \left[\hat{B}_u^{(t)}(x) \int_{\mathcal{X}} Q_{u,r}^{(t),n} f(y) \hat{P}_u^{(t)}(x, dy) - B(x) Q_{u,r}^{(t),n} f(x) \right] (\gamma_{u,t}^1 - \gamma_{u,t}^2)(x_0, dx) du \right| \\ &= \left| \int_0^{s \wedge T_n} \int_{\mathcal{X}} B(x) \left[\int_{\mathcal{X}} Q_{u,r}^{(t),n} f(y) \frac{m(y, u, t)}{m(x, u, t)} m(x, dy) - Q_{u,r}^{(t),n} f(x) \right] (\gamma_{u,t}^1 - \gamma_{u,t}^2)(x_0, dx) du \right| \\ &\leq (C(t) \bar{m} + 1) e^{r \bar{B}_{r(n, x_0)}(x_0)} \bar{B}_n(x_0) \int_0^{s \wedge T_n} \|\gamma_{u,t}^1 - \gamma_{u,t}^2\|_{TV} du \\ &\leq (C(t) \bar{m} + 1) e^{r \bar{B}_{r(n, x_0)}(x_0)} \bar{B}_n(x_0) \int_0^s \|\gamma_{u \wedge T_n, t}^1 - \gamma_{u \wedge T_n, t}^2\|_{TV} du, \end{aligned}$$

where $r(n, x_0) = 2n + |x_0| + \ell(s)^{1/\gamma}$ and $C(t)$ is defined in Assumption C. Then Grönwall's lemma implies that $\|\gamma_{s \wedge T_n, t}^1 - \gamma_{s \wedge T_n, t}^2\|_{TV} = 0$. Taking the limit as n tends to $+\infty$, we obtain $\|\gamma_{s,t}^1 - \gamma_{s,t}^2\|_{TV} = 0$ and the uniqueness of the solution to (3.8). \square

Lemma 3.3. Let $t \geq 0$. Under Assumption D, the generator of the semi-group $(P_{r,s}^{(t)}, r \leq s \leq t)$ defined in (3.6) is $(\mathcal{A}_s^{(t)}, s \leq t)$ defined on $\mathcal{D}(\mathcal{A})$.

For the proof of this Lemma, we need a preliminary result which proof is given in Section C in the appendix.

Lemma 3.4. For all $t \geq 0$ and $s \leq t$, $m(\cdot, s, t) \in \mathcal{D}(\mathcal{G})$.

We can now prove Lemma 3.3.

Proof. Let $t \geq 0$ and $f \in \mathcal{D}(\mathcal{A})$. If we take the expectation of (2.3) and differentiate with respect to t , we get that for all function g such that, $g(s, \cdot) \in \mathcal{D}(\mathcal{A})$:

$$\partial_t R_{s,t} g(x, s) = \mathcal{G}g(x, s) + \partial_s g(x, s) + B(x) \left(\int_{\mathcal{X}} g(y, s) m(x, dy) - g(x, s) \right) := \mathcal{R}g(x, s),$$

for all $x \in \mathcal{X}$ and $s \leq t$, because $t \mapsto \mathbb{E}[\langle Z_t, f \rangle]$ is continuous whenever f is continuous. Next, according to Assumption D, we have the following first order Taylor expansion: for all $x \in \mathcal{X}$, $r < t$ and $h > 0$,

$$\begin{aligned} P_{r,r+h}^{(t)} f(x) &= \frac{R_{r,r+h}(m(\cdot, r+h, t)f)(x)}{m(x, r, t)} \\ &= f(x) + \frac{\mathcal{R}(m(\cdot, r, t)f)(x)}{m(x, r, t)} h + \frac{\partial_r m(x, r, t)f(x)}{m(x, r, t)} h + o(h). \end{aligned}$$

Then

$$\lim_{h \rightarrow 0} \frac{P_{r,r+h}^{(t)} f(x) - f(x)}{h} = \frac{\mathcal{R}(m(\cdot, r, t)f)(x)}{m(x, r, t)} + \frac{\partial_r m(x, r, t)f(x)}{m(x, r, t)},$$

and we obtain:

$$\mathcal{A}_r^{(t)} f(x) = \tilde{\mathcal{G}}_r^{(t)} f(x) + \hat{B}_r^{(t)}(x) \left[\int_{\mathcal{X}} f(y) \hat{P}_r^{(t)}(x, dy) - B(x)f(x) \right], \quad r \leq t. \quad (3.12)$$

However, taking $f \equiv 1$ in (3.9), which is possible according to Lemma 3.4, and differentiating with respect to s yields:

$$\begin{aligned} &\mathbb{E} \left[\sum_{u \in V_s} (\mathcal{G}(m(\cdot, s, t))(X_s^u) + \partial_s m(x, s, t)) \mid Z_{t_0} = \delta_{x_0} \right] \\ &= -\mathbb{E} \left[B(X_s^u) \left(\sum_{k \geq 0} p_k(X_s^u) \sum_{j=1}^k \int_{\mathcal{X}} m(y, s, t) P_j^{(k)}(X_s^u, dy) - m(X_s^u, s, t) \right) \mid Z_{t_0} = \delta_{x_0} \right]. \end{aligned}$$

Then, for $s = t_0$ we get :

$$\partial_s m(x, t_0, t) = -\mathcal{G}(m(\cdot, t_0, t))(x) + B(x)m(x, t_0, t) - B(x) \sum_{k \geq 0} p_k(x) \sum_{j=1}^k \int_{\mathcal{X}} m(y, t_0, t) P_j^{(k)}(x, dy), \quad (3.13)$$

for all $t_0 \geq 0$. Combining (3.12), (3.13), we obtain formula (3.5) for the generator of the auxiliary process. \square

Proof of Theorem 3.1. We prove the result by induction on $k \in \mathbb{N}$ for any separable function $F = f_1 \dots f_k$ with $f_i \in \mathcal{D}(\mathcal{A})$ for all $i = 1 \dots k$. We consider the following proposition denoted by \mathcal{H}_k : for all $0 < s_1 \leq s_2 \leq \dots \leq s_k \leq t$, for all $x_0 \in \mathcal{X}$ and $f_1, \dots, f_k \in \mathcal{D}(\mathcal{A})$:

$$\mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_t} f_1(X_{s_1}^u) \dots f_k(X_{s_k}^u) \right] = m(x_0, 0, t) \mathbb{E}_{x_0} \left[f_1(Y_{s_1}^{(t)}) \dots f_k(Y_{s_k}^{(t)}) \right].$$

First, \mathcal{H}_1 holds by (3.6). Assuming that \mathcal{H}_{k-1} is true for some $k > 1$, we now prove \mathcal{H}_k . Let $0 < s_1 \leq s_2 \leq \dots \leq s_k \leq t$ and f_1, \dots, f_k be measurable non-negative functions such that $f_i \in \mathcal{D}(\mathcal{A})$ for all $1 \leq i \leq k$. We have using the Markov property:

$$\begin{aligned} & \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_t} f_1(X_{s_1}^u) \dots f_k(X_{s_k}^u) \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_{s_{k-1}}} f_1(X_{s_1}^u) \dots f_{k-1}(X_{s_{k-1}}^u) \mathbb{E} \left[\sum_{\substack{v \in V_t \\ v \geq u}} f_k(X_{s_k}^v) \mid \mathcal{F}_{s_{k-1}} \right] \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_{s_{k-1}}} f_1(X_{s_1}^u) \dots f_{k-1}(X_{s_{k-1}}^u) \mathbb{E} \left[\sum_{v \in V_t} f_k(X_{s_k}^v) \mid Z_{s_{k-1}} = \delta_{X_{s_{k-1}}^u} \right] \right]. \end{aligned}$$

We can now use the result proved in the case $k = 1$ and the last term on the right hand side is equal to:

$$\begin{aligned} & \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_{s_{k-1}}} f_1(X_{s_1}^u) \dots f_{k-1}(X_{s_{k-1}}^u) m(X_{s_{k-1}}^u, s_{k-1}, t) \mathbb{E} \left[f_k(Y_{s_k}^{(t)}) \mid Y_{s_{k-1}}^{(t)} = X_{s_{k-1}}^u \right] \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_t} f_1(X_{s_1}^u) \dots f_{k-1}(X_{s_{k-1}}^u) \mathbb{E} \left[f_k(Y_{s_k}^{(t)}) \mid Y_{s_{k-1}}^{(t)} = X_{s_{k-1}}^u \right] \right] \\ &= m(x_0, 0, t) \mathbb{E}_{x_0} \left[f_1(Y_{s_1}^{(t)}) \dots f_{k-1}(Y_{s_{k-1}}^{(t)}) \mathbb{E} \left[f_k(Y_{s_k}^{(t)}) \mid Y_{s_{k-1}}^{(t)} \right] \right], \end{aligned}$$

where the last equality is obtained using the induction hypothesis.

Finally, using Assumption D and a monotone-class argument, we extend the result to all measurable function with respect to the Skorokod topology (see details in Appendix D). \square

We now develop two other Many-to-One formulas: one to characterize the trait of the individuals over the whole tree and the other to characterize the trait of a couple of individuals.

3.2 A Many-to-One formula for the whole tree

We denote by:

$$\mathcal{T} = \bigcup_{s \geq 0} V_s \subset \mathcal{U},$$

the set of all individuals in the population. Let us recall that for $u \in \mathcal{T}$, $\alpha(u)$ and $\beta(u)$ are random variables representing respectively the time of birth and death of u .

Proposition 3.5. *Under Assumptions A, B, C and D, for all $x_0 \in \mathcal{X}$ and for any non-negative measurable function $F : \mathbb{D}(\mathbb{R}_+, \mathcal{X}) \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$, we have:*

$$\mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in \mathcal{T}} F(X_{[\alpha(u), \beta(u)]}^u, \beta(u)) \right] = \int_0^{+\infty} m(x_0, 0, s) \mathbb{E}_{x_0} \left[F(Y_{[0, s]}^{(s)}, s) B(Y_s^{(s)}) \right] ds. \quad (3.14)$$

The left-hand side of (3.14) describes the dynamic of the trait of all individuals that were in the population. The right-hand side is the equivalent in terms of auxiliary process. Then, according to this result, the sum of the contributions of all individuals in the population is equal to the average of the auxiliary process with respect to the mean number of individuals in the population. The weight B in the right-hand side comes from the density of the lifetimes. The terms might be infinite.

Proof. We follow [15] (Lemma 3.8) and provide a proof for the whole trajectories. First, we recall that for any $u \in \mathcal{T}$ and any Borel set $A \subset \mathbb{R}_+$:

$$\mathbb{P}\left(\beta(u) \in A \mid (X_s^u)_{s \geq 0}, \alpha(u)\right) = \int_A B(X_t^u) \exp\left(-\int_{\alpha(u)}^t B(X_s^u) ds\right) dt.$$

Then, for all non-negative measurable functions $f : \mathbb{D}(\mathbb{R}_+, \mathcal{X}) \rightarrow \mathbb{R}_+$, we have:

$$\begin{aligned} & \mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_{\alpha(u)}^{\beta(u)} F(X_{[0,s]}^u, s) B(X_s^u) ds \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \left(\int_{\alpha(u)}^{\tau} F(X_{[0,s]}^u, s) B(X_s^u) ds \right) B(X_\tau^u) \exp\left(-\int_{\alpha(u)}^{\tau} B(X_r^u) dr\right) d\tau \right]. \end{aligned}$$

Next, using Fubini's Theorem, we obtain that the right-hand side of the above equation is:

$$\begin{aligned} & \mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \left(\int_s^{+\infty} B(X_r^u) \exp\left(-\int_{\alpha(u)}^r B(X_r^u) dr\right) d\tau \right) F(X_{[0,s]}^u, s) B(X_s^u) ds \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \exp\left(-\int_{\alpha(u)}^s B(X_r^u) dr\right) F(X_{[0,s]}^u, s) B(X_s^u) ds \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} F(X_{[0,\beta(u)]}^u, \beta(u)) \right], \end{aligned} \tag{3.15}$$

where the first equality is comes from of Assumption A(4). But $\{\alpha(u) \leq s < \beta(u), u \in \mathcal{T}\} = \{u \in V_s\}$, then:

$$\mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_{\alpha(u)}^{\beta(u)} F(X_{[0,s]}^u, s) B(X_s^u) ds \right] = \mathbb{E}_{\delta_{x_0}} \left[\int_0^{+\infty} \mathbf{1}_{\{u \in V_s\}} F(X_{[0,s]}^u, s) B(X_s^u) ds \right]. \tag{3.16}$$

Finally combining (3.15) and (3.16) we get:

$$\begin{aligned} \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in \mathcal{T}} f(X_{[0,\beta(u)]}^u, \beta(u)) \right] &= \sum_{u \in \mathcal{U}} \mathbb{E}_{\delta_{x_0}} \left[\mathbf{1}_{\{u \in \mathcal{T}\}} F(X_{[0,\beta(u)]}^u, \beta(u)) \right] \\ &= \sum_{u \in \mathcal{U}} \mathbb{E}_{\delta_{x_0}} \left[\int_0^{+\infty} \mathbf{1}_{\{u \in V_s\}} F(X_{[0,s]}^u, s) B(X_s^u) ds \right] \\ &= \int_0^{+\infty} \mathbb{E}_{\delta_{x_0}} \left[\sum_{u \in V_s} F(X_{[0,s]}^u, s) B(X_s^u) \right] ds \\ &= \int_0^{+\infty} m(x_0, 0, s) \mathbb{E}_{x_0} \left[F(Y_{[0,s]}^{(s)}, s) B(Y_s^{(s)}) \right] ds, \end{aligned}$$

where the last equality comes from the Many-to-One formula (3.4). \square

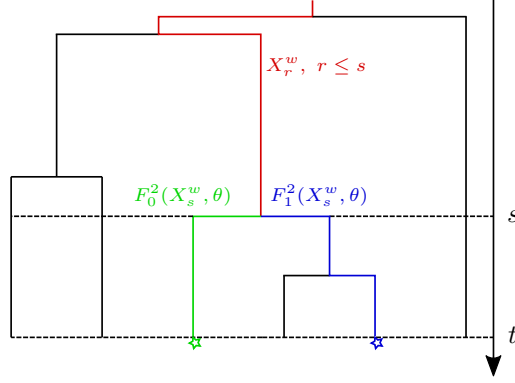


Figure 3.1: Forks.

3.3 Many-to-One formulas for forks

In this section, we characterize the law of a couple of lineage coming from two individuals alive at time t . For former results on the subject, we refer to [5] for such formulas in the neutral case and to [27] for many-to-few-formulas on weighted k -fold sums over particles in the case of local branching. We aim at characterizing the dynamic of the trait of a couple of individual along the spine using our auxiliary process. Those formulas have already proved useful to control the variance of estimators [31].

For any two functions f, g , defined respectively on two intervals I_f, I_g , for any $[a, b] \subset I_f, [c, d] \subset I_g$, we define the concatenation $[f_{[a,b]}, g_{[c,d]}]$ by:

$$[f_{[a,b]}, g_{[c,d]}](t) = \begin{cases} f(t) & , \text{ if } t \in [a, b), \\ g(t + c - b) & , \text{ if } t \in [b, b + (d - c)). \end{cases}$$

Proposition 3.6. *Under Assumptions A, B, C and D, for any $t \geq 0, x_0 \in \mathcal{X}$ and for any non-negative measurable function $F : \mathbb{D}([0, t], \mathcal{X})^2 \rightarrow \mathbb{R}_+$:*

$$\mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{u, v \in V_t \\ u \neq v}} F(X_{[0,t]}^u, X_{[0,t]}^v) \right] = \int_0^t m(x_0, 0, s) \mathbb{E}_{x_0} \left[B(Y_s^{(s)}) J_{s,t} F(Y_{[0,s]}^{(s)}) \right] ds, \quad (3.17)$$

where for $(x_r, r \leq s) \in \mathbb{D}([0, s], \mathcal{X})$:

$$J_{s,t} F(x) = \sum_{a \neq b \in \mathbb{N}} \sum_{k \geq \max(a,b)} p_k(x_s) \int_0^1 m(F_a^{(k)}(x_s, \theta), s, t) m(F_b^{(k)}(x_s, \theta), s, t) H_{s,t} F(x, F_a^{(k)}(x_s, \theta), F_b^{(k)}(x_s, \theta)) d\theta,$$

and for all $s \leq t, (x_s, s \leq t) \in \mathbb{D}([0, t], \mathcal{X})$ and $y_1, y_2 \in \mathcal{X}$:

$$H_{s,t} F(x, y_1, y_2) = \mathbb{E} \left[F \left([x_{[0,s]}; Y_{[s,t]}^{(t),1}], [x_{[0,s]}; Y_{[s,t]}^{(t),2}] \right) \mid \left(Y_s^{(t),1}, Y_s^{(t),2} \right) = (y_1, y_2) \right],$$

and $(Y_s^{(t),1}, s \leq t), (Y_s^{(t),2}, s \leq t)$ are two independent copies of $(Y_s^{(t)}, s \leq t)$.

According to this proposition, the sum of the contributions of each couple in the population at time t corresponds to an integral of a product of a count term and a term characterizing the dynamic of the trait the couple. The integral is over all the possible death time $s \in [0, t]$ for the most recent

common ancestor w of u and v , where $u \neq v \in V_t$. For example, in the case of Figure 3.1, if we pick the green star and the blue star, the lineage of their most recent common ancestor w is in red. For the count term, $m(x_0, 0, s)$ corresponds to the choice w among the individuals in the population at time s and $m\left(F_a^{(k)}(x_s, \theta), s, t\right) m\left(F_b^{(k)}(x_s, \theta), s, t\right)$ corresponds to the choice of u and v among the descendants of w . In the example, with our choice of w , there is only one choice for u and two for v . Before s , the traits along the ancestral lineage of u and v are identical. After the death of w , the dynamic of the trait of the ancestor of u and the ancestor of v become independent conditionally to the trait of w at death. This explains the term $H_{s,t}F\left(x, F_a^{(k)}(x_s, \theta), F_b^{(k)}(x_s, \theta)\right)$ above where x , $(Y_s^{(t),1}, s \leq t)$ and $(Y_s^{(t),2}, s \leq t)$ describe the dynamic of the trait along the red, green and blue path respectively.

This formula is similar to the Many-to-Two formula proved in [27] but as in the Many-to-One formula (3.4), the count terms are separated from the terms corresponding to the dynamic of the trait of a "typical" individual contrary to the formula in [27]. This decomposition is useful for the study of the asymptotic behavior of the branching process. We refer the reader the [41] for an example of use of this formula to prove a law of large numbers.

Proof. Let $t \geq 0$ and $x_0 \in \mathcal{X}$. First we prove (3.17) for $F(x, y) = f_1(x)f_2(y)$, where $f_i : \mathbb{D}([0, t], \mathcal{X}) \rightarrow \mathbb{R}_+$ are non-negative measurable functions for $i = 1, 2$. Let us denote by A the left-hand side of (3.17). We explicit the most recent common ancestor w of two individuals u, v living at time t and we obtain:

$$\begin{aligned} A &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{w \in \mathcal{U}} \sum_{a_1 \neq a_2 \in \mathbb{N}} \sum_{\tilde{u}_1, \tilde{u}_2 \in \mathcal{T}} \mathbf{1}_{\{t \geq \beta(w)\}} \mathbf{1}_{\{wa_1 \tilde{u}_1 \in V_t, wa_2 \tilde{u}_2 \in V_t\}} \prod_{i=1,2} f_i \left(\left[X_{[0, \beta(w)]}^w; X_{[\beta(w), t]}^{wa_i \tilde{u}_i} \right] \right) \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{wa_1 \neq wa_2 \in \mathcal{T} \\ a_1, a_2 \in \mathbb{N}}} \mathbf{1}_{\{t \geq \beta(w)\}} \mathbb{E} \left[\prod_{i=1,2} \sum_{\substack{u_i \in V_t \\ u_i \geq wa_i}} f_i \left(\left[X_{[0, \beta(w)]}^w; X_{[\beta(w), t]}^{u_i} \right] \right) \middle| \mathcal{F}_{\beta(w)} \right] \right]. \end{aligned}$$

Then, applying successively the branching property and the Markov property, we have:

$$\begin{aligned} A &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{wa_1 \neq wa_2 \in \mathcal{T} \\ a_1, a_2 \in \mathbb{N}}} \mathbf{1}_{\{t \geq \beta(w)\}} \prod_{i=1,2} \mathbb{E} \left[\sum_{\substack{u_i \in V_t \\ u_i \geq wa_i}} f_i \left(\left[X_{[0, \beta(w)]}^w; X_{[\beta(w), t]}^{u_i} \right] \right) \middle| X_{[0, \beta(w)]}^{wa_i} \right] \right] \\ &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{wa_1 \neq wa_2 \in \mathcal{T} \\ a_1, a_2 \in \mathbb{N}}} \mathbf{1}_{\{t \geq \beta(w)\}} \prod_{i=1,2} \mathbb{E} \left[\sum_{\substack{u_i \in V_t \\ u_i \geq wa_i}} f_i \left(\left[\tilde{x}; X_{[\beta(w), t]}^{u_i} \right] \right) \middle| X_{\beta(w)}^{wa_i} \right] \right]_{\tilde{x} = X_{[0, \beta(w)]}^w}. \end{aligned}$$

Next, we use the Many-to-One formula (3.4) and we explicit the distribution of the trait at birth of wa and wb :

$$\begin{aligned}
A &= \mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{wa_1 \neq wa_2 \in \mathcal{T} \\ a_1, a_2 \in \mathbb{N}}} \mathbf{1}_{\{t \geq \beta(w)\}} \prod_{i=1,2} m \left(X_{\beta(w)}^{wa_i}, \beta(w), t \right) \mathbb{E} \left[f_i \left([\tilde{x}; Y_{[\beta(w), t]}^{(t)}] \right) \middle| Y_{\beta(w)}^{(t)} = X_{\beta(w)}^{wa_i} \right]_{\tilde{x} = X_{[0, \beta(w)]}^w} \right] \\
&= \mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{wa_1 \neq wa_2 \in \mathcal{T} \\ a_1, a_2 \in \mathbb{N}}} \mathbf{1}_{\{t \geq \beta(w)\}} \sum_{k \geq \max(a_1, a_2)} p_k \left(X_{\beta(w)}^w \right) \int_0^1 \prod_{i=1,2} m \left(F_{a_i}^{(k)} \left(X_{\beta(w)}^w, \theta \right), \beta(w), t \right) \right. \\
&\quad \left. \times \mathbb{E} \left[f_i \left([\tilde{x}; Y_{[\beta(w), t]}^{(t)}] \right) \middle| Y_{\beta(w)}^{(t)} = F_{a_i}^{(k)} \left(X_{\beta(w)-}^w, \theta \right) \right]_{\tilde{x} = X_{[0, \beta(w)]}^w} d\theta \right].
\end{aligned}$$

Applying the Many-to-One formula over the whole tree (3.14) yields:

$$\begin{aligned}
A &= \int_0^t m(x_0, 0, s) \mathbb{E}_x \left[B \left(Y_s^{(s)} \right) \sum_{a_1 \neq a_2 \in \mathbb{N}} \sum_{k \geq \max(a_1, a_2)} p_k \left(Y_s^{(s)} \right) \right. \\
&\quad \left. \int_0^1 \prod_{i=1,2} m \left(F_{a_i}^{(k)} \left(Y_s^{(s)}, \theta \right), s, t \right) \mathbb{E} \left[f_i \left([\tilde{x}; Y_{[s, t]}^{(t)}] \right) \middle| Y_s^{(t)} = F_{a_i}^{(k)} \left(Y_s^{(s)}, \theta \right) \right]_{\tilde{x} = Y_{[0, s]}^{(s)}} d\theta \right] ds \\
&= \int_0^t m(x_0, 0, s) \mathbb{E}_{x_0} \left[B \left(Y_s^{(s)} \right) J_{s, t} \left(f_1 \otimes f_2 \right) \left(Y_{[0, s]}^{(s)} \right) \right] ds,
\end{aligned}$$

where $f_1 \otimes f_2(x) = f_1(x)f_2(x)$. Finally, we obtain (3.17) using a monotone class argument. \square

Let us explicit a particular case of formula (3.17). We define:

$$J_2(f, g)(x) = \sum_{a \neq b} \sum_{k \geq \max(a, b)} p_k(x) \int_0^1 f \left(F_a^{(k)}(x, \theta) \right) g \left(F_b^{(k)}(x, \theta) \right) d\theta. \quad (3.18)$$

J_2 represents the average trait at birth of two uniformly chosen children from an individual of type x . For simplicity of notation, we write $J_2 f(x)$ instead of $J_2(f, f)(x)$. Let us recall that:

$$P_{r, s}^{(t)} f(x) = \mathbb{E} \left[f \left(Y_s^{(t)} \right) \middle| Y_r^{(t)} = x \right].$$

Corollary 3.7. *Under Assumptions A, B, C and D, for any non-negative measurable functions f_t, g_t from $\mathcal{X} \times \mathbb{R}^+$ to \mathbb{R} and any $x_0 \in \mathcal{X}$ we have for $s \leq t$:*

$$\begin{aligned}
\mathbb{E}_{\delta_{x_0}} \left[\sum_{\substack{u, v \in V_t \\ u \neq v}} f_t \left(X_s^u \right) g_t \left(X_s^v \right) \right] &= \int_s^t m(x_0, 0, r) \mathbb{E}_{x_0} \left[f_t \otimes g_t \left(Y_s^{(r)} \right) B \otimes J_2 m(\cdot, r, t) \left(Y_r^{(r)} \right) \right] dr \\
&\quad + \int_0^s m(x_0, 0, r) \mathbb{E}_{x_0} \left[B \otimes J_2 \left(m(\cdot, r, t) P_{r, s}^{(t)} f_t, m(\cdot, r, t) P_{r, s}^{(t)} g_t \right) \left(Y_r^{(r)} \right) \right] dr. \quad (3.19)
\end{aligned}$$

The first integral corresponds to the couple of individuals alive at time t whose most recent common ancestor died after time s . It is for example the case on Figure 3.2 if you pick two red stars on the tree on the left-hand side. The product $m(x_0, 0, r) J_2(m(\cdot, r, t)(y))$, with $y \in \mathcal{X}$ corresponds to the average number of such couple at time t whose most recent common ancestor died at time r with $s \leq r \leq t$. The second integral corresponds to couples $(u, v) \in V_t$ of individuals whose most recent common ancestor w died before s . It is the case on Figure 3.2 if you pick one blue star and one green star on the tree on the right-hand side. In this case, unlike in the previous one, the value of the trait

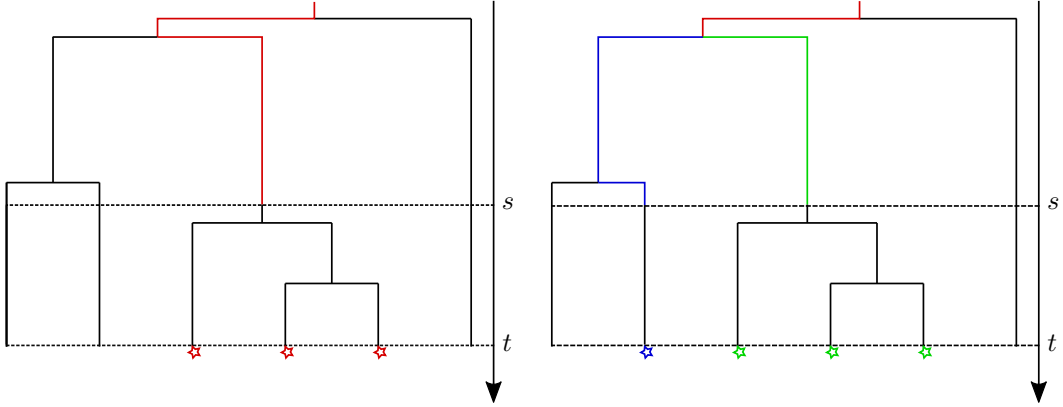


Figure 3.2: Trees and forks.

of the individuals at time s is not the same. The dynamic of the trait on the blue lineage and on the green lineage are independent conditionally to the trait of their common ancestor at death. This explains the terms $P_{r,s}^{(t)} f_t$ and $P_{r,s}^{(t)} g_t$ that appear in the second integral. As before, the remaining terms depending on the average number of individuals in the population is a count term.

4 Ancestral lineage of a uniform sampling at a fixed time with a large initial population

The Many-to-One formula (3.4) gives us the law of the trait of a uniformly sampled individual in an "average" population. But the characterization of the law of the trait of a uniformly sampled individual in the effective population is more complex because the number of individuals alive at time t is stochastic and depends on the dynamic of the trait of individuals. As the auxiliary process takes into account the bias in the population due to the number of individuals, it characterizes the law of a uniformly sampled individual only when the bias are in place i.e. when there is a certain amount of individuals. Indeed, the dynamic of the first individual in the population is not biased. That is why we now look at the ancestral lineage of a uniform sampling in a large population.

4.1 Convergence of the sampling process on a fixed time interval

It only makes sense to speak of a uniformly sampled individual at time t if the population does not become extinct before time t . For all $t \geq 0$, let $\Omega_t = \{N_t > 0\}$ denote the event of survival of the population. Let $\nu \in \mathcal{M}_P(\mathcal{X})$ be such that:

$$\mathbb{P}_\nu(\Omega_t) > 0. \quad (4.1)$$

We set

$$\nu_n := \sum_{i=1}^n \delta_{X_i}, \quad (4.2)$$

where X_i are i.i.d. random variables with distribution ν . For $t \geq 0$, we denote by $U(t)$ the random variable with uniform distribution on V_t conditionally on Ω_t and by $(X_s^{U(t)}, s \leq t)$ the process describing the trait of a sampling along its ancestral lineage. If X is a stochastic process, we denote by X^ν the process with initial distribution $\nu \in \mathcal{M}_P(\mathcal{X})$. In particular, for all $t \geq 0$, $Y^{(t),\nu}$ corresponds

to the auxiliary process with $Y_0^{(t)} \sim \nu$. For all $0 \leq s \leq t$,

$$m(\nu, s, t) = \mathbb{E}(N_t | Z_s = \nu),$$

denote the average number of individuals in the population after time t starting from a population distributed as ν at time s .

Theorem 4.1. *Under Assumptions A(1-3), B, C and D, for any $t \geq 0$, the sequence $(X_{[0,t]}^{U(t), \nu_n})_{n \geq 0}$ converges in law in $\mathbb{D}([0, t], \mathcal{X})$ to $Y_{[0,t]}^{(t), \pi_t}$ where:*

$$\pi_t(dx) = \frac{m(x, 0, t)\nu(dx)}{m(\nu, 0, t)}.$$

Proof. Let $t \geq 0$. Let $(X_i)_{1 \leq i \leq n}$ be i.i.d random variables with distribution ν and $\nu_n = \sum_{i=1}^n \delta_{X_i}$. Let $F : \mathbb{D}([0, t], \mathcal{X}) \rightarrow \mathbb{R}_+$ be a bounded measurable function. First, we notice that:

$$\frac{1}{n} N_t^{\nu_n} = \frac{1}{n} \sum_{i=1}^n N_t^{(i)}, \quad (4.3)$$

where $N_t^{(i)}$ are independent copies of N_t with initial distribution δ_{X_i} . According to the law of large numbers, (4.3) converges almost surely as n tends to infinity to $m(\nu, 0, t) = \int_{\mathcal{X}} m(x, 0, t)\nu(dx)$. Next, let $\Omega_t(\nu_n) = \{N_t^{\nu_n} > 0\}$. $(\Omega_t(\nu_n))_{n \geq 0}$ is an increasing sequence. According to (4.1), there exists $0 < \varepsilon(t) \leq 1$ such that:

$$\mathbb{P}(\Omega_t(\nu_n)^C) \leq (1 - \varepsilon(t))^n \xrightarrow{n \rightarrow +\infty} 0,$$

so that:

$$\mathbf{1}_{\{\Omega_t(\nu_n)^C\}} \xrightarrow{n \rightarrow +\infty} 0, \text{ almost surely.}$$

We have:

$$\mathbb{E} \left[F \left(X_{[0,t]}^{U(t), \nu_n} \right) \right] = \mathbb{E} \left[\mathbf{1}_{\{\Omega_t(\nu_n)\}} \frac{1}{N_t^{\nu_n}} \sum_{u \in V_t^{\nu_n}} F \left(X_{[0,t]}^u \right) \right] \mathbb{P}(\Omega_t(\nu_n))^{-1}.$$

Let $V_t^{(i)}, i = 1 \dots n$ be independent identically distributed populations at time t coming from an individual with trait $X_i \sim \nu$ at 0. Then:

$$\mathbb{E} \left[\mathbf{1}_{\{\Omega_t(\nu_n)\}} \frac{1}{N_t^{\nu_n}} \sum_{u \in V_t^{\nu_n}} F \left(X_{[0,t]}^u \right) \right] = \mathbb{E} \left[\mathbf{1}_{\{\Omega_t(\nu_n)\}} \frac{n}{N_t^{\nu_n}} \frac{1}{n} \sum_{i=1}^n \sum_{u \in V_t^{(i)}} F \left(X_{[0,t]}^u \right) \right]. \quad (4.4)$$

According to the strong law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n \sum_{u \in V_t^{(i)}} F \left(X_{[0,t]}^u \right) \xrightarrow{n \rightarrow +\infty} \mathbb{E}_\nu \left[\sum_{u \in V_t} F \left(X_{[0,t]}^u \right) \right], \text{ almost surely.}$$

Taking the limit in (4.4) as n tends to infinity, we have by dominated convergence because F is bounded:

$$\mathbb{E} \left[F \left(X_{[0,t]}^{U(t), \nu_n} \right) \right] \xrightarrow{n \rightarrow +\infty} \frac{1}{m(\nu, 0, t)} \int_{\mathcal{X}} \mathbb{E}_x \left[\sum_{u \in V_t} F \left(X_{[0,t]}^u \right) \right] \nu(dx),$$

because $\mathbb{P}(\Omega_t(\nu_n)) \rightarrow 1$ as n tends to infinity. Finally, applying the Many-to-one formula (3.4), we obtain:

$$\mathbb{E} \left[F \left(X_{[0,t]}^{U(t), \nu_n} \right) \right] \xrightarrow{n \rightarrow +\infty} \frac{\int_{\mathcal{X}} m(x, 0, t) \mathbb{E}_x \left[F \left(Y_{[0,t]}^{(t)} \right) \right] \nu(dx)}{m(\nu, 0, t)}.$$

□

Remark 4.2. If we start with n individuals with the same trait x , we obtain:

$$\mathbb{E} \left[F \left(X_{[0,t]}^{U(t), \nu_n} \right) \right] \xrightarrow{n \rightarrow +\infty} \mathbb{E}_x \left[F \left(Y_{[0,t]}^{(t)} \right) \right].$$

Therefore, the auxiliary process describes exactly the dynamic of the trait of a uniformly sampled individual in the large population limit, if all the starting individuals have the same trait. If the initial individuals have different traits at the beginning, the large population approximation of a uniformly sampled individual is a linear combination of the auxiliary process.

Remark 4.3. One can easily generalize this results to a k -tuple of individuals uniformly picked at time t . But if you start with a population of size n and you pick k individuals uniformly at random at time t , when n tends to infinity, the probability that those k individuals comes from the same initial individual is zero. Then, the trajectories of their traits are independent and we get for example in the case $k = 2$, for any $f, g : \mathbb{D}([0, t], \mathcal{X}) \rightarrow \mathbb{R}_+$ bounded measurable functions:

$$\mathbb{E} \left[f \left(X_{[0,t]}^{U_1(t), \nu_n} \right) g \left(X_{[0,t]}^{U_2(t), \nu_n} \right) \right] \xrightarrow{n \rightarrow +\infty} \mathbb{E}_x \left[f \left(Y_{[0,t]}^{(t), 1} \right) g \left(Y_{[0,t]}^{(t), 2} \right) \right],$$

where $U_1(t), U_2(t)$ are independent random variables with uniform distribution on V_t and the processes $(Y_s^{(t), 1}, s \leq t), (Y_s^{(t), 2}, s \leq t)$ are i.i.d. distributed as $(Y_s^{(t)}, s \leq t)$.

Remark 4.4. An other way of characterizing the trait of a uniformly sampled individual via the auxiliary process is to look at the long time behavior of the process instead of looking at the large population behavior. This has been done in [5] in the case of a constant division rate and in [41] in a general framework using the ergodicity of ancestral lineages.

4.2 The trait of a uniformly sampled individual for growth-fragmentation models

The auxiliary process is a good way of getting simulated random variables corresponding to the trait of a uniformly sampled individual with large initial population or to the trait of a uniformly sampled individual for large times (see [41]). Indeed, it is much more quicker to simulate one trajectory of the auxiliary process rather than the dynamic of an entire population. In this section, we detail the auxiliary process for our three examples introduced in Section 2.2.

4.2.1 Linear growth model

For the linear growth model with binary division (Section 2.2.1), Assumption C is satisfied for $C \equiv 1$ and the large population limit of the ancestral process of a sampling grows linearly between two jumps and jumps at time s at rate

$$\widehat{B}_s^{(t)}(x) = \alpha x \left(1 + \frac{1 + e^{2\sqrt{\alpha}(t-s)}}{1 - x\sqrt{\frac{\alpha}{a}} + e^{2\sqrt{\alpha}(t-s)}(1 + x\sqrt{\frac{\alpha}{a}})} \right).$$

At a jump, there is a unique descendant with trait $\frac{x}{2}$ if x is the trait of its parent at the splitting time. In particular, the rate of division of the limiting process is bigger than the rate of division in a cell line for the original process. It means that in the large population limit, a typical individual has overcome more division than any individual.

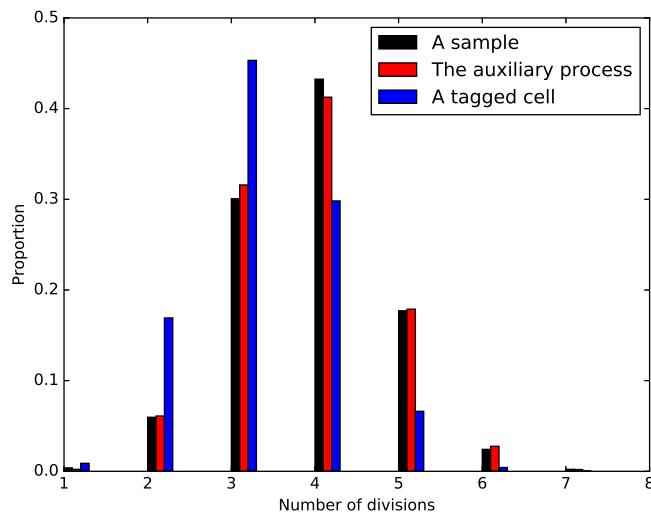


Figure 4.1: Distribution of the number of divisions in the lineage of a uniformly sampled individual (black bars), of the auxiliary process (red bars) and of a tagged cell (blue bars). For each case, we used 5000 realizations of each process until time $t = 50$ with parameters $a = 0.1$ and $x_0 = 1$. The distribution of the number of divisions almost coincides for the auxiliary process and a sampled individual. However, the distribution of the number of divisions for a tagged cell is different from the two previous ones. Indeed, it is more likely to sample an individual whose ancestors divided many times, that is why, the distributions of the number of divisions for the auxiliary process and for a uniformly sampled individual are centered on bigger values than the distribution of the number of divisions for a tagged cell.

4.2.2 Exponential growth model in a varying environment

For the exponential growth model in a varying environment with binary division (Section 2.2.2), Assumption C is satisfied for $C \equiv 1$ and the associated auxiliary process grows exponentially between two jumps and jumps at time s at rate

$$\widehat{B}_s^{(t)}(x) = (\alpha(s)x + \beta) \left(1 + \frac{1}{1 + x \int_s^t \alpha(r) e^{(a-\beta)(r-s)} dr} \right).$$

The rate of division of the limiting process is again bigger than the division rate of any individual. At a jump, there is a unique descendant with trait $\frac{x}{2}$ if x is the trait of its parent at the splitting time.

This example is a good illustration of the fact that the large population limit of the size of a uniformly sampled individual does not correspond to the size of a tagged cell, i.e. the size along a lineage where at each division, you choose randomly one daughter cell. In fact, as the division rate of the auxiliary process is larger than B , the number of divisions along the lineage of a uniformly sampled individual is bigger than the number of divisions along the lineage of tagged cell, resulting in a difference on the size of the individuals. However, the distribution of the number of divisions along the lineage of a uniformly sampled individual coincides with the one for the auxiliary process. On Figure 4.1, we can see those distributions: the two first distributions, corresponding to the distribution of the number of divisions along the lineage of a uniformly sampled individual and of the auxiliary process, are centered on a bigger number of divisions than the third distribution corresponding to a tagged cell.

4.2.3 Parasite infection model

For this cell division model with parasite infection, Assumption C is satisfied for $C \equiv 1$ and the auxiliary process evolves as a Feller diffusion with infinitesimal generator:

$$\mathcal{F}_s^{(t)} f(x) = \left(gx + 2\sigma^2 \frac{\alpha x (e^{g(t-s)} - e^{\beta(t-s)})}{\alpha x (e^{g(t-s)} - e^{\beta(t-s)}) + (g - \beta)e^{\beta(t-s)}} \right) f'(x) + \sigma^2 x f''(x),$$

so that the drift of the limit of the process of the ancestral trait of a sampling is bigger than the original drift in the population. Then, the limiting process jumps at time s at rate

$$\widehat{B}_s^{(t)}(x) = (\alpha x + \beta) \left(1 + \frac{1}{1 + \frac{\alpha x}{g - \beta} (e^{(g-\beta)(t-s)} - 1)} \right),$$

so that the division rate of the limiting process is also bigger than the rate of division in a cell line for the original process.

The trait of the newborn cell is distributed according to the following probability law:

$$\widehat{P}_s^{(t)}(x, dy) = \mathbf{1}_{\{0 \leq y \leq x\}} \frac{2(g - \beta) + 2\alpha y (e^{(g-\beta)(t-s)} - 1)}{2(g - \beta) + \alpha x (e^{(g-\beta)(t-s)} - 1)} \frac{dy}{x}.$$

In fact, because cells divide faster when they have more parasites inside them, it is a good strategy, in order to have a lot of descendants in a long time scale, to choose to give a lot of parasites to your daughter cell. Moreover, the evolution of the trait is biased: the drift in the Feller diffusion is more important for the auxiliary process because a cell with more parasites divides faster so that it produces more descendants.

5 Further comments and examples

We can apply the results of this work to various models and we choose to detail in this article only three of them based on biological and computational considerations. However, we review in this section some other interesting models.

5.1 The age-structured population model

In this model, the quantity of interest is the age of each individual which grows linearly. The life-time of each individual is a random variable with cumulative distribution function G . Such models have been first introduced by Bellman and Harris in [8] and have recently been studied in order to infer the division rate [31]. Let $B : \mathbb{R}_+ \rightarrow \mathbb{R}$ be the rate of division of each cell defined via:

$$G(t) = 1 - \exp \left(- \int_0^t B(s) ds \right).$$

The branching process $(Z_t)_{t \geq 0}$ is solution of the following equation, for any function $f \in \mathcal{C}^1(\mathbb{R}_+)$ and any $x \in \mathcal{X}$:

$$\begin{aligned} \langle Z_t, f \rangle &= \langle Z_0, f \rangle + \int_0^t \int_{\mathbb{R}_+} f'(x) Z_s(dx) ds \\ &\quad + \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+ \times \mathbb{N}} \mathbf{1}_{\{u \in V_{s-}, \theta \leq B(X_{s-}^u)\}} (kf(0) - f(X_{s-}^u)) M(ds, du, d\theta, dk), \end{aligned}$$

where M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+ \times \mathbb{N}$ with intensity $ds \otimes n(du) \otimes d\theta \otimes p(dk)$, where p denotes the distribution of the number of descendants.

In order to get information on the average number of individuals in the population at time t , we follow Harris in [29] (Chap. 6) and we obtain:

$$m(0, 0, t) = 1 - G(t) + m \int_0^t m(0, 0, t - u) dG(u),$$

where m is the average number of descendants at division. From this expression, we can derive a renewal equation for $m(x, 0, t)$, for $x \geq 0$. We cannot find an explicit solution to this renewal equation except in the case of an exponentially distributed life-time but we know the asymptotic behavior of a solution (see [29]). In particular, if G is non-lattice and $m > 1$,

$$m(0, 0, t) \underset{t \rightarrow +\infty}{\sim} c(\alpha, m) n_1 e^{\alpha t}, \quad \text{where } \int_0^\infty e^{-\alpha t} dG(t) = 1/m,$$

and $c(\alpha, m)$ and n_1 are explicitly given in [29] Theorem 17.1. The rate of division of the auxiliary process is thus given for large t by:

$$\widehat{B}_s^{(t)} \sim B(x) \frac{e^{-\alpha x} (1 - G(x))}{1 - m \int_0^x e^{-\alpha u} dG(u)}.$$

5.2 Multi-type branching process and switching

An example of phenomenon that we would like to understand using a model on a finite state space is the phenotypic switching, i.e. the capacity to achieve multiple internal states in response to a single set of external inputs. Examples of studies of switching can be found in [45] or [39]. For an asymptotic characterization of the ancestral lineage of a typical individual for models with a trait on a finite state space, we refer to [23]. We assume here that an individual can be in state 0 or 1 which is constant during its lifetime. An individual in state $x = 0, 1$ divide at rate $B(x) = b_x$ and at division, it is replaced by 2 individuals. We denote by p the probability of switching at birth. We assume that this probability does not depend on the trait. Therefore, the trait only affects the lifetime of individuals. We obtain for the generator of the first moment semi-group for any function f taking values in $\{0, 1\}$ and any $x \in \{0, 1\}$:

$$\mathcal{F}_{\text{switch}} f(x) = B(x) (2f(x)(1 - p) + 2f(\bar{x})p - f(x)),$$

where $\bar{x} = 1 - x$. After some computations, we obtain:

$$\mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} X_t^u \right] = x + (b_1(1 - 2p) - 2pb_0) \int_0^t \mathbb{E}_{\delta_x} \left[\sum_{u \in V_s} X_s^u \right] ds + 2pb_0 \int_0^t \mathbb{E}_{\delta_x} [N_s] ds.$$

Then if we write:

$$\mu(t) = \mathbb{E}_{\delta_x} [N_t], \quad \nu(t) = \mathbb{E}_{\delta_x} \left[\sum_{u \in V_t} X_t^u \right], \quad \forall t \geq 0,$$

we obtain:

$$\begin{pmatrix} \partial_t \mu \\ \partial_t \nu \end{pmatrix} = \begin{pmatrix} b_0 & b_1 - b_0 \\ 2pb_0 & b_1(1 - 2p) - 2pb_0 \end{pmatrix} \begin{pmatrix} \mu \\ \nu \end{pmatrix}.$$

For example, for $p = 1/2$, writing $\gamma = \frac{b_0}{b_1}$, we have:

$$m(1, s, t) = m(0, s, t) + \left[e^{\sqrt{b_0 b_1}(t-s)} - e^{-\sqrt{b_0 b_1}(t-s)} \right] \frac{1}{2\sqrt{\gamma}} (1 - \gamma).$$

In particular, the transition kernel of the auxiliary process is given by:

$$\widehat{P}_s^{(t)}(x, dy) = \frac{m(x, s, t) \delta_x(dy) + m(\bar{x}, s, t) \delta_{\bar{x}}(dy)}{m(x, s, t) + m(\bar{x}, s, t)},$$

so that if $\gamma > 1$, i.e. $b_0 > b_1$, the auxiliary process switches more from 1 to 0 at a jump because $m(0, s, t) > m(1, s, t)$.

5.3 Markovian jump processes for the dynamic of the trait

The dynamic of some characteristics of a cell are non-continuous and thus cannot be described by a diffusion type process. For example, this is the case for the dynamic of populations inside a cell such as plasmids or extra-chromosomal DNA. Then, an other generalization of Kimmel's multilevel model for plasmids [34] is the following: we assume that the trait of each individual evolves as a birth and death process with birth rate $\lambda > 0$ and death rate $\mu > 0$. We assume here that $\lambda - \mu > 0$. The generator of the process corresponding to the dynamic of the trait is then given for any measurable function $f : \mathbb{N} \rightarrow \mathbb{R}_+$ and any $x \in \mathbb{N}$ by:

$$\mathcal{G}f(x) = \lambda(f(x+1) - f(x)) + \mu(f(x-1) - f(x)).$$

We assume that a cell with x plasmids divides at a rate $B(x)$ and that at division, the plasmids are randomly allocated to one of the two daughter cell. The branching process $(Z_t, t \geq 0)$ is solution of the following equation, for any measurable function $f : \mathbb{N} \rightarrow \mathbb{R}_+$ and any $x \in \mathcal{X}$:

$$\begin{aligned} \langle Z_t, f \rangle &= \langle Z_0, f \rangle + \int_0^t \int_{\mathbb{R}_+} \sum_{u \in V_s} [\mathbf{1}_{\{\theta \leq \lambda X_s^u\}} (f(X_s^u + 1) - f(X_s^u)) \\ &\quad + \mathbf{1}_{\{\lambda X_s^u \leq \theta \leq (\lambda + \mu) X_s^u\}} (f(X_s^u - 1) - f(X_s^u))] Q^u(ds, d\theta) \\ &\quad + \int_0^t \int_{\mathcal{U} \times \mathbb{R}_+ \times [0,1]} \mathbf{1}_{\{u \in V_{s-}, z \leq B(X_{s-}^u)\}} (f(\delta X_{s-}^u) + f((1 - \delta) X_{s-}^u) - f(X_{s-}^u)) M(ds, du, dz, d\delta), \end{aligned}$$

where $(Q^u)_{u \in \mathcal{U}}$ is a family of Poisson point measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $ds \otimes d\theta$ and M is a Poisson point measure on $\mathbb{R}_+ \times \mathcal{U} \times \mathbb{R}_+ \times [0, 1]$ with intensity $ds \otimes n(du) \otimes dz \otimes d\delta$.

For example, for the division rate $B(x) = x$, we obtain for the average number of individuals in the population after a time t :

$$m(x, s, t) = 1 + \frac{x}{\lambda - \mu} \left(e^{(\lambda - \mu)t} - 1 \right).$$

In particular, the motion of the auxiliary process between jumps is given by the following generator:

$$\begin{aligned} \widehat{\mathcal{G}}_s^{(t)} f(x) &= \lambda \left[1 + \frac{e^{(\lambda - \mu)(t-s)} - 1}{\lambda - \mu + x (e^{(\lambda - \mu)(t-s)} - 1)} \right] (f(x+1) - f(x)) \\ &\quad + \mu \left[1 - \frac{e^{(\lambda - \mu)(t-s)} - 1}{\lambda - \mu + x (e^{(\lambda - \mu)(t-s)} - 1)} \right] (f(x-1) - f(x)). \end{aligned}$$

The birth rate of the plasmid population for the auxiliary process is bigger than λ and the death rate is smaller than μ . This can be explained again by the fact that cells with a lot of plasmids divides more so that they are more represented at sampling.

5.4 Link with the integro-differential model

The study of the average process associated with the measure-valued branching process Z is interesting in the sense that it characterizes the macroscopic evolution of the population. For a more detailed study of this link see for example [12]. The following result is a corollary of Theorem 2.2 of Section 2. We recall that for all $s \geq 0$, $t \geq s$ and $x \in \mathcal{X}$,

$$R_{s,t} f(x) = \mathbb{E} \left[\sum_{u \in V_t} f(X_t^u) \mid Z_s = \delta_x \right],$$

where f is a measurable function.

Corollary 5.1. *Let $f \in D(\mathcal{G})$, $s \geq 0$ and $x_0 \in \mathcal{X}$. Under Assumptions A(1-3) and B, the measure $(R_{s,t}(x_0, \cdot))_{t \geq 0}$ is the unique solution to the following integro-differential equation:*

$$\begin{aligned} R_{s,t}f(s, x_0) &= f(s, x_0) + \int_s^t \int_{\mathcal{X}} (\mathcal{G}f(r, x) + \partial_r f(r, x)) R_{s,r}(x_0, dx) ds \\ &\quad + \int_s^t \int_{\mathcal{X}} B(x) \left[\sum_{k \geq 0} p_k(x) \sum_{j=1}^k \int_{\mathcal{X}} f(r, y) P_j^{(k)}(x, dy) - f(r, x) \right] R_{s,r}(x_0, dx) ds, \end{aligned} \quad (5.1)$$

where $(R_{s,t})_{t \geq s}$ is defined in (2.5).

One can prove this result taking the expectation in (2.3) and using the same arguments as in the proof of Corollary 2.4 in [15].

Let $n(t, \cdot) := R_{0,t}(x_0, \cdot)$. Equation (5.1) can be rewritten as:

$$\begin{cases} \partial_t n(t, x) = \mathcal{G}^T n(t, x) + \sum_{k \geq 0} \sum_{j=1}^k K_j^{(k)}(B p_k n(t, \cdot)) - B(x) n(t, x), \\ n(0, x) dx = \delta_{x_0}(dx). \end{cases}$$

where \mathcal{G}^T is the adjoint operator of \mathcal{G} and $K_j^{(k)}$ is the adjoint operator of $f \mapsto \int_{\mathcal{X}} f(y) P_j^{(k)}(x, dy)$ as in [15].

For example, in the case of the cell division model with exponential growth introduced in Section 2.2.2, we obtain in a weak sense:

$$\partial_t n(t, x) + \partial_x (axn(t, x)) = 4B(2x)n(t, 2x) - B(x)n(t, x).$$

This is a classical growth-fragmentation equation as the one studied in [42] or [12]. The solutions of the associated eigenvalue problem permit in particular to quantify the asymptotic global growth rate of the population.

Acknowledgements. I would like to thank Vincent Bansaye for his guidance during this work. I also thank Marc Hoffmann and Bertrand Cloez for many helpful discussions on the subject of this paper. I also thank an anonymous referee for its useful comments. I acknowledge partial support by the Chaire Modélisation Mathématique et Biodiversité of Veolia Environnement - École Polytechnique - Museum National Histoire Naturelle - F.X. This work is supported by the "IDI 2014" project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

A Proof of Lemma 2.3

We give a recursive construction of the solution to (2.3). For all $u \in \mathcal{U}$, we denote the birth time and the death time of u respectively by $\alpha(u)$ and $\beta(u)$. Let $x_0 \in \mathcal{X}$ be given. We construct a structured population $Y^k = (\bar{Z}^k, (X_s^u, s \geq T_k(\bar{Z}^k), u \in V_{T_k(\bar{Z}^k)}))$, where $\bar{Z}^k \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{U} \times \mathcal{X}))$ is such that $T_{k+1} = +\infty$. We set $\alpha(\emptyset) = 0$, $X_0^\emptyset = x_0$, $V_0 = \{\emptyset\}$ and $\bar{Z}_t^0 \equiv \delta_{(\emptyset, x_0)}$ for all $t \geq 0$, so that:

$$Y^0 = (\bar{Z}^0, (\Phi^\emptyset(x_0, 0, t), t \geq 0)).$$

Let $k \geq 1$. We now construct Y^{k+1} . For all $u \in V_{T_k(\bar{Z}^k)}$ such that $\alpha(u) = T_k(\bar{Z}^k)$, we set $X_t^u = \Phi^u(X_{\alpha(u)}^u, \alpha(u), t)$, for all $t \geq \alpha(u)$. For all $u \in V_{T_k(\bar{Z}^k)}$, let:

$$\beta(u) = \inf \left\{ t > \alpha(u), \int_{\alpha(u)}^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq B(X_{s^-}^u)\}} M(ds, \{u\}, dz, [0, 1], [0, 1]) > 0 \right\}.$$

Let $T = \inf \{ \beta(u), u \in V_{T_k(\bar{Z}^k)} \}$. Let $(T, U_{k+1}, \theta_{k+1}, L_{k+1}, A_{k+1})$ be the unique quintuplet of random variables such that $M(\{T\}, \{U_{k+1}\}, \{\theta_{k+1}\}, \{L_{k+1}\}, \{A_{k+1}\}) = 1$. Let:

$$V_T = V_{T^-} \setminus \{U_{k+1}\} \cup \{U_{k+1}1, \dots, U_{k+1}G(U_{k+1}, T, L_{k+1})\},$$

and for all $i = 1, \dots, G(X_T^{U_{k+1}}, L_{k+1})$, we set $\alpha(U_{k+1}i) = T$ and:

$$X_{\alpha(U_{k+1}i)}^{U_{k+1}i} = F_i(X_T^{U_{k+1}}, L_{k+1}, A_{k+1}).$$

We set

$$\begin{aligned} \bar{Z}_t^{k+1} &= \bar{Z}_t^k, \text{ for all } t \in [0, T_k(\bar{Z}^k)], \\ \bar{Z}_t^{k+1} &= \sum_{u \in V_{T_k(\bar{Z}^k)}} \delta_{(u, X_t^u)}, \text{ for all } t \in [T_k(\bar{Z}^k), T[, \\ \bar{Z}_t^{k+1} &= \sum_{u \in V_T} \delta_{(u, X_t^u)}, \text{ for all } t \geq T. \end{aligned}$$

Finally, we set $Y^{k+1} = (\bar{Z}^{k+1}, (X_s^u, s \geq T, u \in V_T))$ so that $T_{k+1}(\bar{Z}^{k+1}) = T$.

Let \bar{Z} be the measure-valued branching process on \mathbb{R}_+ satisfying for all $k \in \mathbb{N}$ and all $t \geq 0$:

$$\bar{Z}_{t \wedge T_k(\bar{Z}^k)} = \bar{Z}_t^k.$$

Therefore, $T_k(\bar{Z}) = T_k(\bar{Z}^k)$ for all $k \in \mathbb{N}$. To shorten notation, we write T_k instead of $T_k(\bar{Z})$ until the end of the proof.

Let $f \in \mathcal{D}(\mathcal{G})$. We now prove by induction the following property:

$$\mathcal{H}_k : \{ \forall t \in [T_k, T_{k+1}), \langle \bar{Z}_t, f \rangle \text{ is a solution to (2.3)} \}.$$

First, \mathcal{H}_0 is obviously true. Assume that \mathcal{H}_{k-1} is true. Let $t \in [T_k, T_{k+1})$. We recall that U_k denotes the individual who dies at time T_k . We denote by:

$$V_{t,1} = V_{T_{k-1}} \setminus \{U_k\}, \quad V_{t,2} = \{u \in V_t | \alpha(u) = T_k\},$$

the set of all individuals born strictly before T_k except U_k and the descendants of U_k , respectively. We have:

$$\sum_{u \in V_t} f(u, t, X_t^u) = \sum_{u \in V_{t,1}} f(u, t, X_t^u) + \sum_{u \in V_{t,2}} f(u, t, X_t^u),$$

and

$$f(u, t, X_t^u) = f(u, t, \Phi^u(X_{T_{k-1}}, T_{k-1}, t)).$$

As none of the individuals in $V_{T_{k-1}} \setminus \{U_k\}$ divides on $[T_{k-1}, t]$, we obtain using (2.2):

$$f(u, t, X_t^u) = f(u, T_{k-1}, X_{T_{k-1}}^u) + \int_{T_{k-1}}^t (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_{k-1}, t}^{f, u}(X_{T_{k-1}}^u).$$

Then, we split both the integral term and the martingale in two terms to separate the behavior of the population before T_k and after T_k . We add and subtract the contribution corresponding to U_k to get

a sum over all individuals alive at time T_{k-1} :

$$\begin{aligned} \sum_{u \in V_{t,1}} f(u, t, X_t^u) &= \sum_{u \in V_{T_{k-1}} \setminus \{U_k\}} \left[\int_{T_k}^t (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_k, t}^{f, u}(X_{T_k}^u) \right] \\ &\quad - f\left(U_k, T_k^-, X_{T_k^-}^{U_k}\right) + \sum_{u \in V_{T_{k-1}}} \left[f\left(u, T_{k-1}, X_{T_{k-1}}^u\right) \right. \\ &\quad \left. + \int_{T_{k-1}}^{T_k} (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_{k-1}, T_k}^{f, u}(X_{T_{k-1}}^u) \right]. \end{aligned} \quad (\text{A.1})$$

Using the induction hypothesis, we have:

$$\begin{aligned} \sum_{u \in V_{T_{k-1}}} f\left(u, T_{k-1}, X_{T_{k-1}}^u\right) &= f(u, 0, x_0) \\ &+ \int_0^{T_{k-1}} \int_{U \times \mathcal{X}} (\mathcal{G}f(u, s, x) + \partial_s f(u, s, x)) \bar{Z}_s(dudx) ds + M_{0, T_{k-1}}^f(x_0) \\ &+ \int_0^{T_{k-1}} \int_E \mathbf{1}_{\{u \in V_{s^-}, z \leq B(X_{s^-}^u)\}} \left(\sum_{i=1}^{G(X_s^u, l)} f(u, s, F_i(X_s^u, l, \theta)) - f(u, s, X_{s^-}^u) \right) M(ds, du, dz, dl, d\theta). \end{aligned} \quad (\text{A.2})$$

Moreover, for all $s \in [T_{k-1}, T_k[$, $V_s = V_{T_{k-1}}$, so that we have:

$$\sum_{u \in V_{T_{k-1}}} \int_{T_{k-1}}^{T_k} (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds = \int_{T_{k-1}}^{T_k} \sum_{u \in V_s} (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds. \quad (\text{A.3})$$

Finally, we obtain combining (A.1), (A.2) and (A.3):

$$\begin{aligned} \sum_{u \in V_{t,1}} f(u, t, X_t^u) &= f(\emptyset, 0, x_0) + \int_0^{T_k} \int_{U \times \mathcal{X}} (\mathcal{G}f(u, s, x) + \partial_s f(u, s, x)) \bar{Z}_s(dudx) ds \\ &+ M_{0, T_{k-1}}^f(x_0) + \sum_{u \in V_{T_{k-1}}} M_{T_{k-1}, T_k}^{f, u}(X_{T_{k-1}}^u) \\ &+ \int_0^{T_{k-1}} \int_E \mathbf{1}_{\{u \in V_{s^-}, z \leq B(X_{s^-}^u)\}} \left(\sum_{i=1}^{G(X_s^u, l)} f(u, s, F_i(X_s^u, l, \theta)) - f(u, s, X_{s^-}^u) \right) M(ds, du, dz, dl, d\theta) \\ &- f\left(U_k, T_k^-, X_{T_k^-}^{U_k}\right) + \sum_{u \in V_{T_{k-1}} \setminus \{U_k\}} \left[\int_{T_k}^t (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_k, t}^{f, u}(X_{T_k}^u) \right]. \end{aligned} \quad (\text{A.4})$$

Next, using again (2.2), we have:

$$\sum_{u \in V_{t,2}} f(u, t, X_t^u) = \sum_{u \in V_{t,2}} \left[f(u, T_k, X_{T_k}^u) + \int_{T_k}^t (\mathcal{G}f(u, s, X_s^u) + \partial_s f(u, s, X_s^u)) ds + M_{T_k, t}^{f, u}(X_{T_k}^u) \right]. \quad (\text{A.5})$$

Moreover, by definition of $V_{t,2}$:

$$\sum_{u \in V_{t,2}} f(u, T_k, X_{T_k}^u) = \int_{T_{k-1}}^t \int_E \mathbf{1}_{\{u \in V_{s^-}, z \leq B(X_{s^-}^u)\}} \sum_{i=1}^{G(X_s^u, l)} f(u, s, F_i(X_s^u, l, \theta)) M(ds, du, dz, dl, d\theta). \quad (\text{A.6})$$

Adding the martingale terms of (A.4) and (A.5), we obtain:

$$\begin{aligned} M_{0,T_{k-1}}^f(x_0) + \sum_{u \in V_{T_{k-1}}} M_{T_{k-1},T_k}^{f,u} \left(X_{T_{k-1}}^u \right) + \sum_{u \in V_{T_{k-1}} \setminus \{U_k\}} M_{T_{k-1},t}^{f,u} \left(X_{T_{k-1}}^u \right) \\ + \sum_{u \in V_{T_k}, \alpha(u)=T_k} M_{T_k,t}^{f,u} \left(X_{T_k}^u \right) = M_{0,T_k}^f(x_0). \end{aligned} \quad (\text{A.7})$$

Finally, we obtain the result combining (A.5),(A.6) and (A.7).

B Proof of Lemma 2.5

Let $\bar{Z}^{(1)}$ and $\bar{Z}^{(2)}$ be two solutions of (2.3) associated with the previously defined family of flows and Poisson point measure. For all $k \in \mathbb{N}$, we write $T_k^{(i)} = T_k(\bar{Z}^{(i)})$, $i = 1, 2$. We assume that $\bar{Z}_0^{(1)} = \bar{Z}_0^{(2)} = \delta_x$, for some $x \in \mathcal{X}$. We have $T_0^{(1)} = T_0^{(2)} = 0$. We prove by induction on $k \in \mathbb{N}$ the following proposition:

$$\mathcal{H}_k : T_{k+1}^{(1)} = T_{k+1}^{(2)} \text{ and } \forall t \in [T_k^{(1)}, T_{k+1}^{(1)}), \forall f \in \bar{\mathcal{D}}(\mathcal{G}), \langle \bar{Z}_t^{(1)}, f \rangle = \langle \bar{Z}_t^{(2)}, f \rangle.$$

First, \mathcal{H}_0 is true because:

$$T_1^{(1)} = T_1^{(2)} = \inf \left\{ t > 0, \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq B(\Phi^\theta(x,0,s))\}} M(ds, \{\emptyset\}, dz, [0, 1], [0, 1]) > 0 \right\}$$

and for all $t \in [0, T_1)$, $\bar{Z}_t^{(1)} = \bar{Z}_t^{(2)} = \delta_{\Phi^\theta(x,0,t)}$. Let us assume that \mathcal{H}_{k-1} is true. We first prove the second point of \mathcal{H}_k . By (2.3), we have for $i = 1, 2$:

$$\langle \bar{Z}_{T_k}^{(i)}, f \rangle = \langle \bar{Z}_{T_{k-1}}^{(i)}, f \rangle + \int_{T_{k-1}}^{T_k} \int_{\mathcal{U} \times \mathcal{X}} (\mathcal{G}f(u, s, x) + \partial_s f(u, s, x)) \bar{Z}_s^{(i)}(du, dx) ds + M_{T_{k-1}, T_k}^{f, (i)}(x) \quad (\text{B.1})$$

$$\begin{aligned} + \int_{T_{k-1}}^{T_k} \int_E \mathbf{1}_{\{u \in V_{s^-}^{(i)}, z \leq B(X_{s^-}^{u, (i)})\}} \left(\sum_{i=1}^{G(X_s^u, l)} f(u, s, F_i(X_s^u, l, \theta)) - f(u, s, X_{s^-}^{u, (i)}) \right) \\ M(ds, du, dz, dl, d\theta). \end{aligned} \quad (\text{B.2})$$

As the jump integral (B.2) depends only on the process strictly before $T_k^{(1)}$, we obtain, using the induction hypothesis, that $\langle \bar{Z}_{T_k}^{(1)}, f \rangle = \langle \bar{Z}_{T_k}^{(2)}, f \rangle$. The evolution of the trait for $t \in [T_k^{(1)}, T_{k+1}^{(1)} \wedge T_{k+1}^{(2)})$ only depends on the family of flows given at the beginning and which are the same for both solutions. Hence, it remains to prove that $T_{k+1}^{(1)} = T_{k+1}^{(2)}$. And it is the case because this jump time only depend on the state of the population at $T_k^{(1)}$, on the flows and on the Poisson point measure M . Finally, for all $t \in [T_k^{(1)}, T_{k+1}^{(1)})$, we have: $\langle \bar{Z}_t^{(1)}, f \rangle = \langle \bar{Z}_t^{(2)}, f \rangle$.

Moreover, the measure-valued process is entirely characterized by $\{\langle \bar{Z}_t, f \rangle, f \in \bar{\mathcal{D}}(\mathcal{G})\}$ according to Remark 2.1. Therefore, there is a unique càdlàg measure-valued strong solution to (2.3) up to the k th jump time for all $k \in \mathbb{N}$.

C Proof of Lemma 3.4

Let $r, s < t$ and $x \in \mathcal{X}$. We consider the event $\Omega_r = \{\text{no division before } r\}$. Then:

$$\frac{1}{r} (\mathbb{E}(m(X_r, s, t) | X_0 = x) - m(x, s, t)) = A(x, r, s, t) + B(x, r, s, t) + C(x, r, s, t),$$

where

$$\begin{aligned} A(x, r, s, t) &= \frac{1}{r} \left(m(X_r, s, t) | X_0 = x - \mathbb{E}(\langle Z_r, m(\cdot, s, t) f \rangle \mathbf{1}_{\Omega_r} | Z_0 = \delta_x) \right), \\ B(x, r, s, t) &= \frac{1}{r} \mathbb{E}(\langle Z_r, m(\cdot, s, t) \rangle (\mathbf{1}_{\Omega_r} - 1) | Z_0 = \delta_x), \\ C(x, r, s, t) &= \frac{1}{r} \left(\mathbb{E}(\langle Z_r, m(\cdot, s, t) \rangle | Z_0 = \delta_x) - m(x, s, t) \right). \end{aligned}$$

First,

$$A(x, r, s, t) = \frac{1}{r} \mathbb{E}_x \left(m(X_r, s, t) \mathbf{1}_{\Omega_r^c} \right).$$

Conditioning with respect to $\sigma(X_u, u \leq r - s)$ we obtain:

$$A(x, r, s, t) = \frac{1}{r} \mathbb{E}_x \left(m(X_r, s, t) \left(1 - e^{-\int_0^r B(X_u) du} \right) \right) \xrightarrow{r \rightarrow 0} m(x, s, t) B(x).$$

Next, we have:

$$B(x, r, s, t) = -\frac{1}{r} \mathbb{E}(\langle Z_r, m(\cdot, s, t) \rangle \mathbf{1}_{\Omega_r^c} | Z_0 = \delta_x).$$

Then, let us denote T_1 the random variable corresponding to the lifetime of the first individual. We have using the Markov property and the branching property:

$$\begin{aligned} B(x, r, s, t) &= -\frac{1}{r} \mathbb{E} \left(\sum_{u \in V_r} m(X_r^u, s, t) \mathbf{1}_{T_1 < r} | Z_0 = \delta_x \right) \\ &= -\frac{1}{r} \mathbb{E} \left(\mathbf{1}_{T_1 < r} \sum_{k \geq 0} p_k \left(X_{T_1}^\emptyset \right) \right. \\ &\quad \left. \sum_{j=0}^k \int_0^1 \mathbb{E} \left(\sum_{u \in V_r} m(X_r^u, s, t) | X_{T_1}^u = F_j^{(k)}(X_{T_1}^\emptyset, \theta) \right) P_j^{(k)}(X_{T_1}^\emptyset, d\theta) | Z_0 = \delta_x \right). \end{aligned}$$

Next, exhibiting the distribution of T_1 we obtain:

$$\begin{aligned} B(x, r, s, t) &= -\frac{1}{r} \int_0^r \mathbb{E} \left(B \left(X_v^\emptyset \right) e^{-\int_0^v B(X_l^\emptyset) dl} \sum_{k \geq 0} p_k \left(X_v^\emptyset \right) \right. \\ &\quad \left. \sum_{j=0}^k \int_0^1 \mathbb{E} \left(\sum_{u \in V_r} m(X_r^u, s, t) | X_v^u = F_j^{(k)}(X_v^\emptyset, \theta) \right) P_j^{(k)}(X_v^\emptyset, d\theta) | Z_0 = \delta_x \right). \end{aligned}$$

Finally:

$$B(x, r, s, t) \xrightarrow{r \rightarrow 0} -B(x) \sum_{k \geq 0} p_k(x) \sum_{j=1}^k \int_0^1 m \left(F_j^{(k)}(x, \theta), s, t \right) P_j^{(k)}(x, d\theta).$$

For the last term we have:

$$\begin{aligned}
C(x, r, s, t) &= \frac{1}{r} \left(\mathbb{E}(\langle Z_r, m(\cdot, s, t) \rangle \mid Z_0 = \delta_x) - m(x, s, t) \right) \\
&= \frac{1}{r} \left(\mathbb{E} \left(\sum_{u \in V_r} m(X_r^u, s, t) \mid Z_0 = \delta_x \right) - m(x, s, t) \right) \\
&= \frac{1}{r} \left(\mathbb{E} \left(\sum_{u \in V_r} \mathbb{E} \left(\sum_{u \in V_{t-s}} 1 \mid Z_0 = \delta_{X_r^u} \right) \mid Z_0 = \delta_x \right) - m(x, s, t) \right) \\
&= \frac{1}{r} (m(x, s, t+r) - m(x, s, t)) \xrightarrow[r \rightarrow 0]{} -\partial_t m(x, s, t)
\end{aligned}$$

Finally,

$$\lim_{r \rightarrow 0} \frac{1}{r} (A_r(m(\cdot, s, t))(x) - m(x, s, t)),$$

exists and $m(\cdot, s, t) \in \mathcal{D}(\mathcal{G})$.

D Details of the proof of Theorem 3.1

We detail here the use of the monotone-class theorem in the proof of Theorem 3.1.

Using Remark 2.1, (3.4) is satisfied for any function of the form $F = \mathbf{1}_{B_1} \dots \mathbf{1}_{B_n}$, where B_i are Borel sets, for $i = 1 \dots n$, for all $n \in \mathbb{N}$. Let us define:

$$\mathcal{H} = \{F : \mathbb{D}([0, t], \mathcal{X}) \rightarrow \mathbb{R}_+ \text{ bounded and measurable such that (3.4) is satisfied for } F\},$$

and:

$$I = \left\{ \bigcap_{i=1}^n \{x \in \mathbb{D}([0, t], \mathcal{X}), x(s_i) \in B_i\}, n \in \mathbb{N}, s_1, \dots, s_n \in \mathbb{R}_+, B_1, \dots, B_n \text{ Borel sets} \right\}.$$

First, I is a π -system and $\sigma(I) = \mathcal{D}$ where \mathcal{D} is the Borel σ -field associated with the Skorokod topology on $\mathbb{D}([0, t], \mathcal{X})$ ([11] Theorem 12.5). Then, applying the monotone-class theorem ([47], Theorem 3.14.), we obtain that \mathcal{H} contains all measurable functions with respect to the Skorokhod topology.

References

- [1] K. B. Athreya. Change of measures for Markov chains and the $L \log L$ theorem for branching processes. *Bernoulli*, pages 323–338, 2000.
- [2] K. B. Athreya. Coalescence in critical and subcritical Galton-Watson branching processes. *Journal of Applied Probability*, 49(3):627–638, 2012.
- [3] K. B. Athreya, S. R. Athreya, and S. K. Iyer. Supercritical age-dependent branching Markov processes and their scaling limits. *Bernoulli*, 17(1):138–154, 2011.
- [4] V. Bansaye. Ancestral lineages and limit theorems for branching Markov chains. *To appear in Journal of Theoretical Probability*, 2015. Available at <http://www.cmapx.polytechnique.fr/~bansaye/recherche.html>.
- [5] V. Bansaye, J.-F. Delmas, L. Marsalle, and V. C. Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *The Annals of Applied Probability*, pages 2263–2314, 2011.

- [6] V. Bansaye and S. Méléard. Stochastic models for structured populations. *Mathematical Biosciences Institute Lecture Series. Springer edition*, 2015.
- [7] V. Bansaye and V. C. Tran. Branching Feller diffusion for cell division with parasite infection. *Aléa*, 8(81-127):241–242, 2011.
- [8] R. Bellman and T. Harris. On age-dependent binary branching processes. *Annals of Mathematics*, pages 280–295, 1952.
- [9] J. D. Biggins. Martingale convergence in the branching random walk. *Journal of Applied Probability*, pages 25–37, 1977.
- [10] J. D. Biggins and A. E. Kyprianou. Measure change in multitype branching. *Advances in Applied Probability*, pages 544–581, 2004.
- [11] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [12] F. Campillo, N. Champagnat, and C. Fritsch. Links between deterministic and stochastic approaches for invasion in growth-fragmentation-death models. *Journal of mathematical biology*, pages 1–41, 2016.
- [13] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probability Theory and Related Fields*, 80(2):299–314, 1988.
- [14] B. Chauvin, A. Rouault, and A. Wakolbinger. Growing conditioned trees. *Stochastic Processes and their Applications*, 39(1):117–130, 1991.
- [15] B. Cloez. Limit theorems for some branching measure-valued processes. *Advances in Applied Probability*, 49(2):549–580, 2017.
- [16] S. Cooper. Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical Biology and Medical Modelling*, 3(1):10, 2006.
- [17] P. Del Moral. *Feynman-Kac Formulae*. Springer, 2004.
- [18] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3):1760–1799, 2015.
- [19] M. Doumic-Jauffret, P. Maia, and J. Zubelli. On the calibration of a size-structured population model from experimental data. *Acta Biotheoretica*, 58(4):405–413, 2010.
- [20] R. Durrett. The genealogy of critical branching processes. *Stochastic Processes and their Applications*, 8(1):101–116, 1978.
- [21] S. N. Ethier and T. G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.
- [22] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, pages 1880–1919, 2004.
- [23] H.-O. Georgii and E. Baake. Supercritical multitype branching processes: the ancestral types of typical individuals. *Advances in Applied Probability*, pages 1090–1110, 2003.
- [24] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *The Annals of Applied Probability*, 17(5/6):1538–1569, 2007.

- [25] R. Hardy and S. C. Harris. A spine approach to branching diffusions with applications to L_p -convergence of martingales. In *Séminaire de Probabilités XLII*, pages 281–330. Springer, 2009.
- [26] S. Harris, S. Johnston, and M. Roberts. The coalescent structure of continuous-time Galton-Watson trees. *arXiv:1707.07993*, 2017.
- [27] S. C. Harris and M. I. Roberts. The many-to-few lemma and multiple spines. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):226–242, 2017.
- [28] S. C. Harris and D. Williams. Large deviations and martingales for a typed branching diffusion, 1. *Astérisque*, 236:133–154, 1996.
- [29] T. E. Harris. *The theory of branching processes*. Courier Corporation, 2002.
- [30] V. H. Hoang. Estimating the division kernel of a size-structured population. *arXiv:1509.02872*, 2015.
- [31] M. Hoffmann and A. Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Processes and their Applications*, 2015.
- [32] J. J. Hong. Coalescence in Bellman-Harris and multi-type branching processes. *Graduate Theses and Dissertations*, 2011.
- [33] O. Kallenberg. Stability of critical cluster fields. *Mathematische Nachrichten*, 77(1):7–43, 1977.
- [34] M. Kimmel. Quasistationarity in a branching model of division-within-division. In *Classical and Modern Branching Processes*, pages 157–164. Springer, 1997.
- [35] J. F. C. Kingman. The first birth problem for an age-dependent branching process. *The Annals of Probability*, pages 790–801, 1975.
- [36] H. Kunita. *Stochastic flows and stochastic differential equations*, volume 24. Cambridge university press, 1997.
- [37] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. In *Classical and modern branching processes*, pages 181–185. Springer, 1997.
- [38] A. Lambert and L. Popovic. The coalescent point process of branching trees. *The Annals of Applied Probability*, 23(1):99–144, 2013.
- [39] S. Leibler and E. Kussell. Individual histories and selection in heterogeneous populations. *Proceedings of the National Academy of Sciences*, 107(29):13183–13188, 2010.
- [40] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *The Annals of Probability*, 23(3):1125–1138, 1995.
- [41] A. Marguet. A law of large numbers for branching markov processes by the ergodicity of ancestral lineages. *arXiv:1707.07993*, 2017.
- [42] S. Mischler and J. Scher. Spectral analysis of semigroups and growth-fragmentation equations. In *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis*. Elsevier, 2015.
- [43] O. Nerman and P. Jagers. The stable doubly infinite pedigree process of supercritical branching populations. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65(3):445–460, 1984.
- [44] N. O'Connell. The genealogy of branching processes and the age of our most recent common ancestor. *Advances in Applied Probability*, 27(2):418–442, 1995.

- [45] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, 427(6976):737–740, 2004.
- [46] V. C. Tran. *Stochastic particle models for problems of adaptive evolution and for the approximations of statistical solutions*. Thesis, Université de Nanterre - Paris X, 2006.
- [47] D. Williams. *Probability with martingales*. Cambridge university press, 1991.
- [48] A. M. Zubkov. Limiting distributions of the distance to the closest common ancestor. *Theory of Probability and Its Applications*, 20(3):602–612, 1976.